Computational Sensorimotor Learning (Spring'21)

Pulkit Agrawal

Feb 16 2021

Course Logistics

Website
http://pulkitag.github.io/6.884

Piazza

Gradescope

First Half of the Course

Weekly Assignments in CoLab

Second Half of the Course

Course Project

Grading Policy

50% Assignments

40% Project

10% Class Participation

Course TAs

Tao Chen

Joshua Gruenstein

OH on website

Email
csl-staff@mit.edu

Lectures

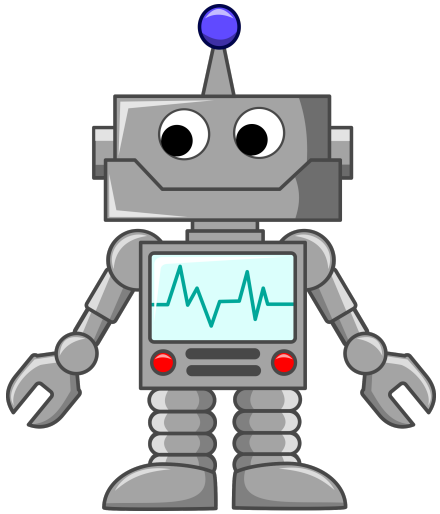
On Zoom

Encourage everyone to have their videos on

Recorded and will be posted on YouTube later


I want a bottle of coke …


I want a bottle of coke …



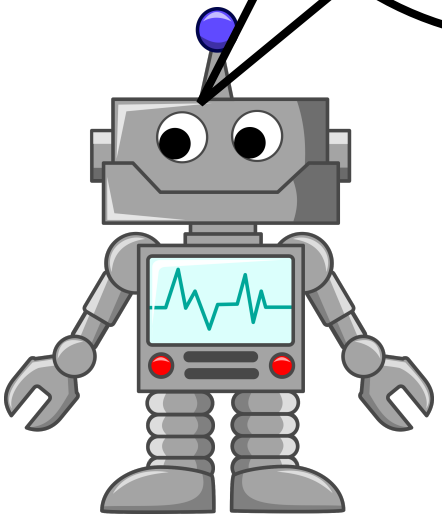
I want a bottle of coke …

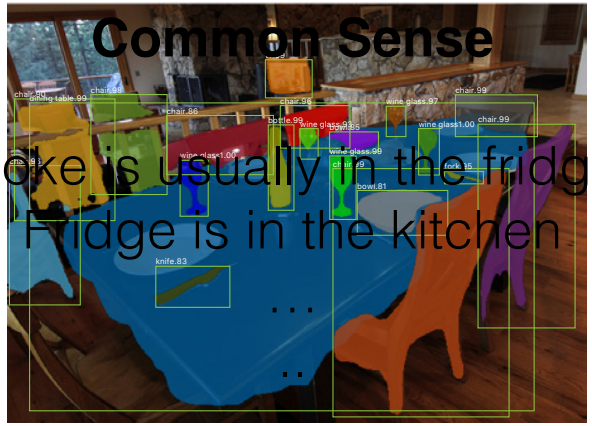
Want a bottle of coke

NLP

What is coke?

Where is coke?





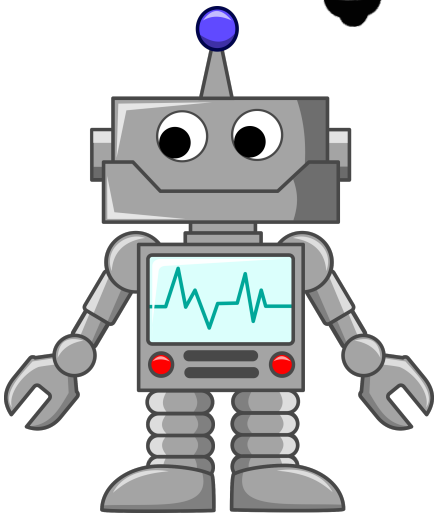
Common Sense

Coke is usually in the fridge

Fridge is in the kitchen

…

..

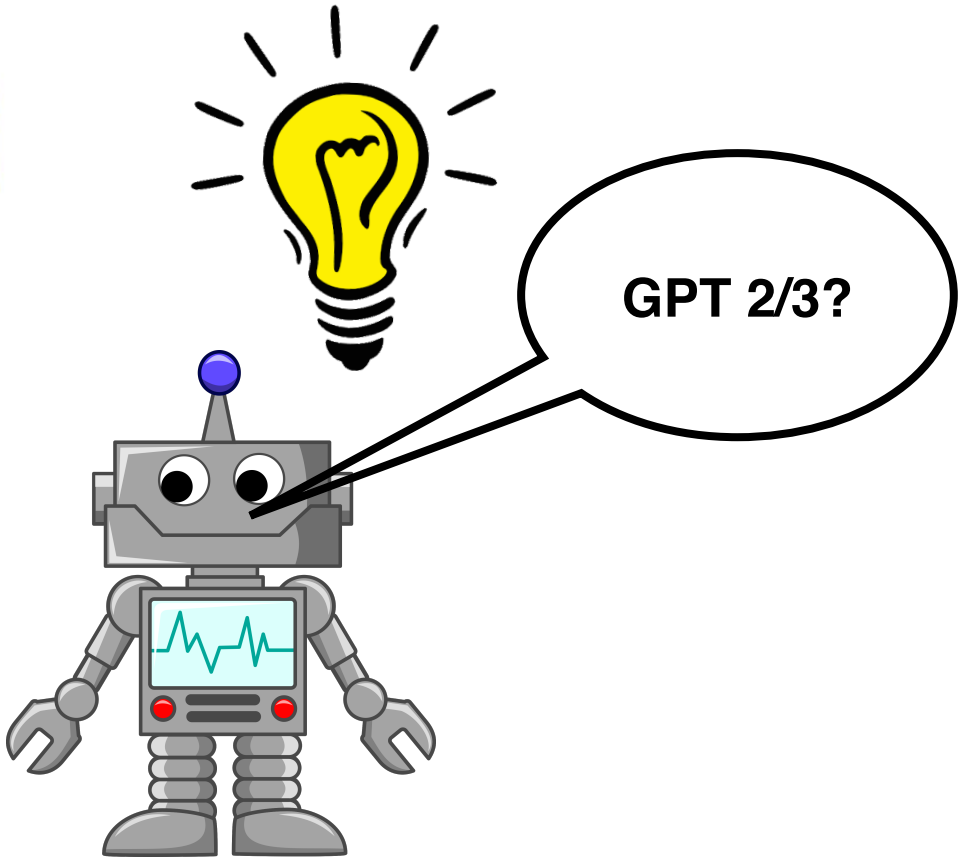


Common Sense

Coke is usually in the fridge
Fridge is in the kitchen
…
..

Common Sense

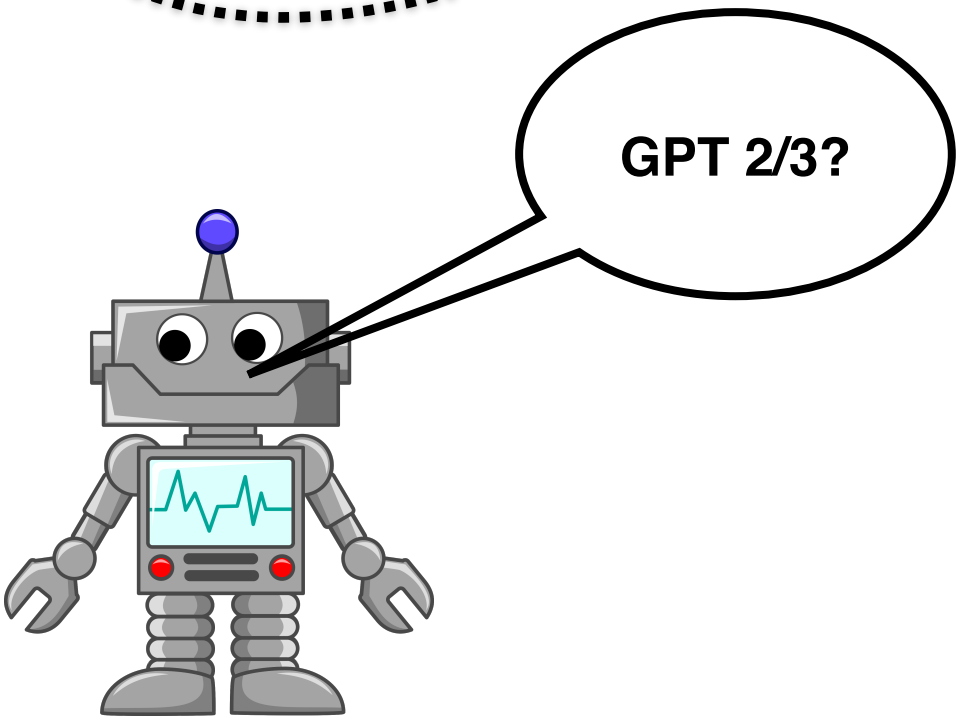
Coke is usually in the fridge
Fridge is in the kitchen
…
..



the coke is in ???

Condense information
from lots of text data

Language Models



(examples from GPT2)

the coke is in the bottom of a small, empty bucket," according to

buy a coke at the gas station down the road from mine.

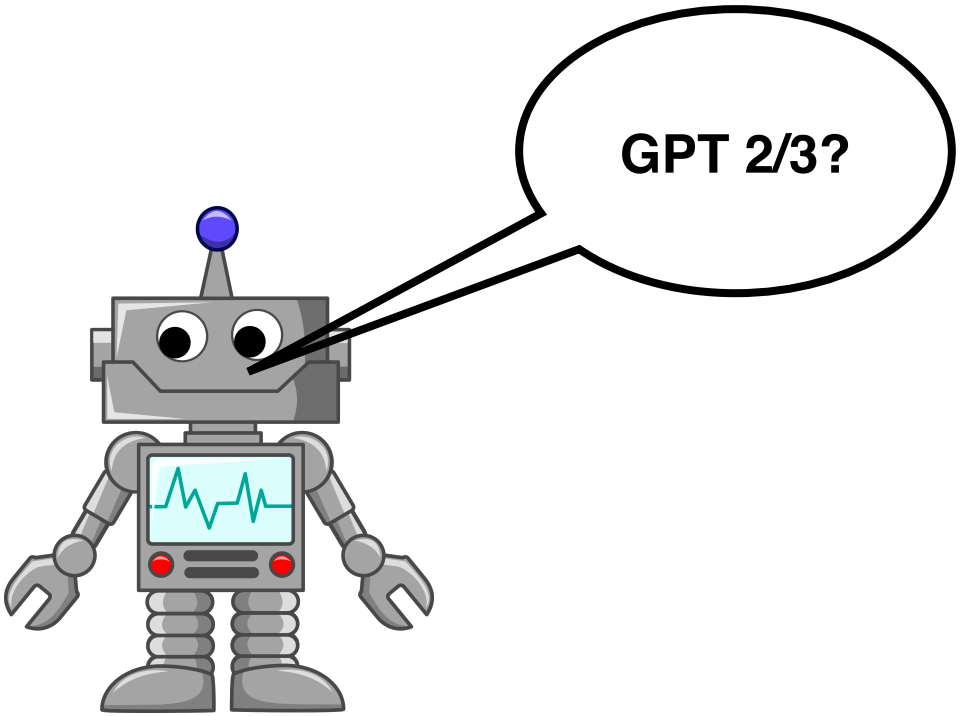
The coke is in the fridge with the rest of the food."

the toaster is in the microwave oven, the toaster is in the microwave

To get out you need to go through the door marked "the tunnel", and

Sometimes it may work

.. lets for a moment
assume this …



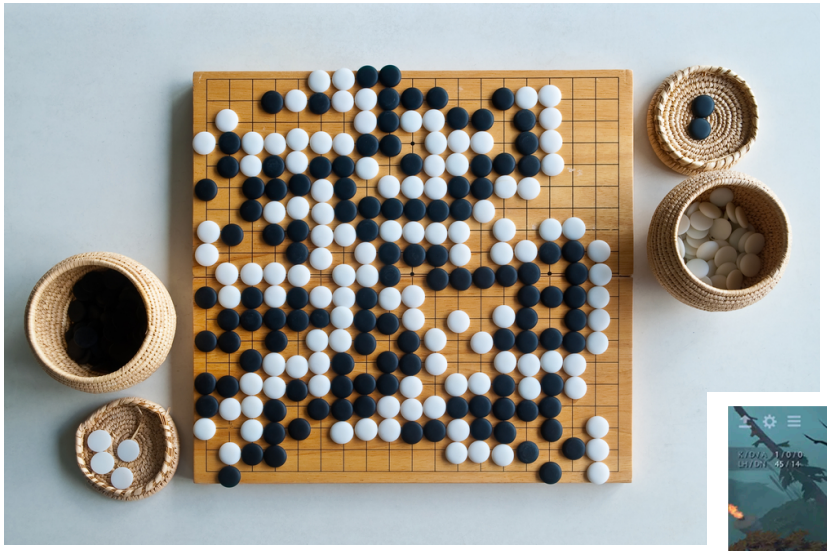
(examples from GPT2)

Lets say we get how to make pancakes …

(short circuit)

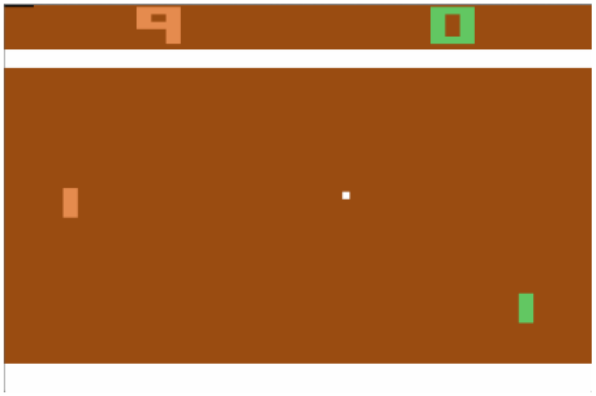
How do we learn such skills?

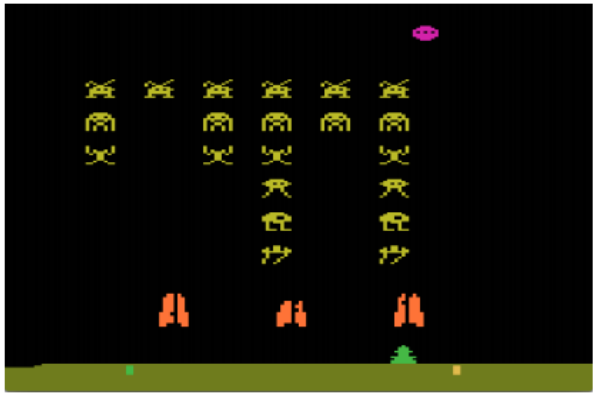
AlphaGo



Open AI Five playing DOTA

Success Stories of Skill Learning





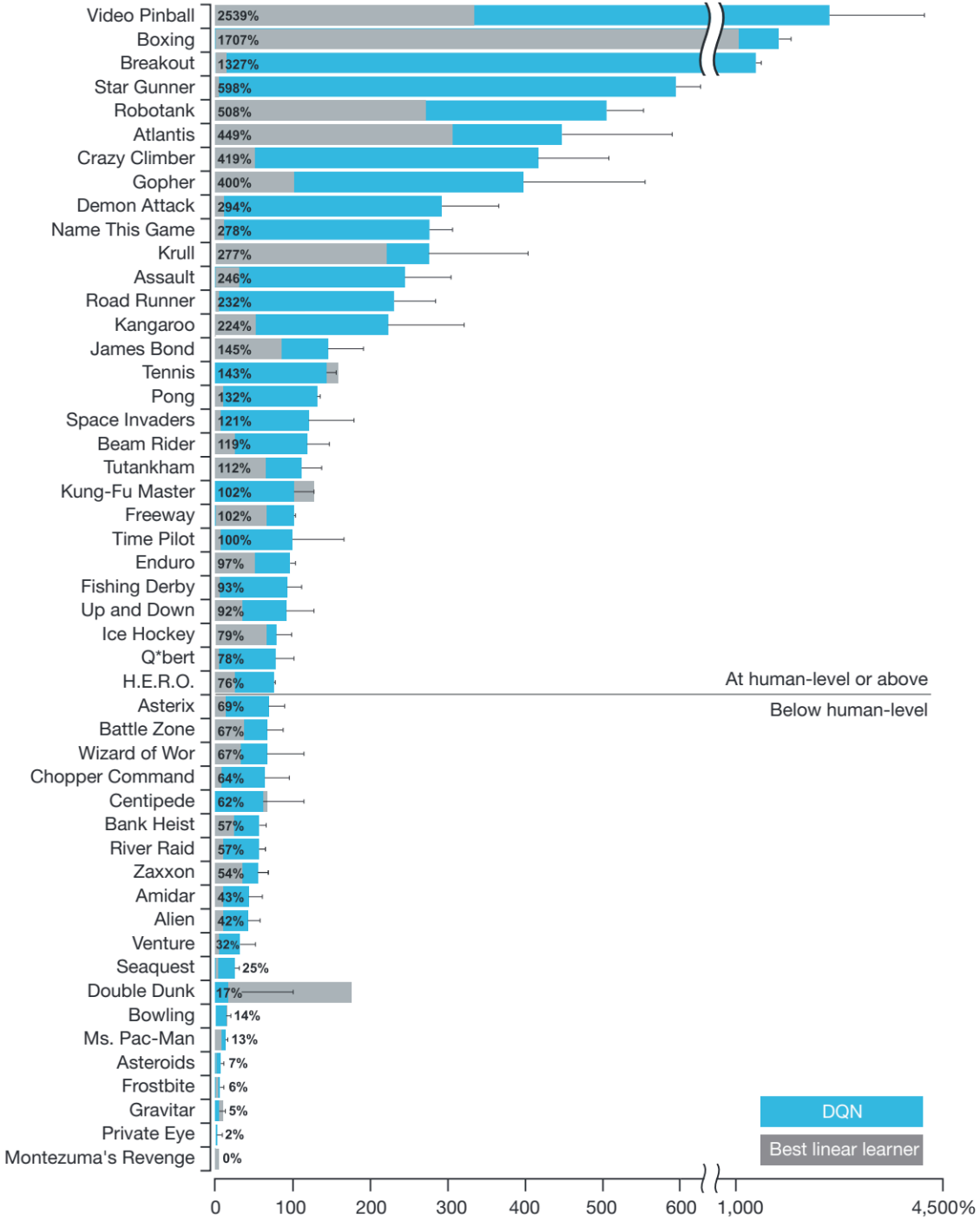
ATARI Games

Playing Atari with Deep Reinforcement Learning, Mnih et al., 2016

Starting out - 10 minutes of training

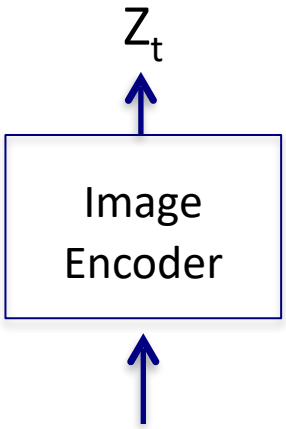
The algorithm tries to hit the ball back, but
it is yet too clumsy to manage.

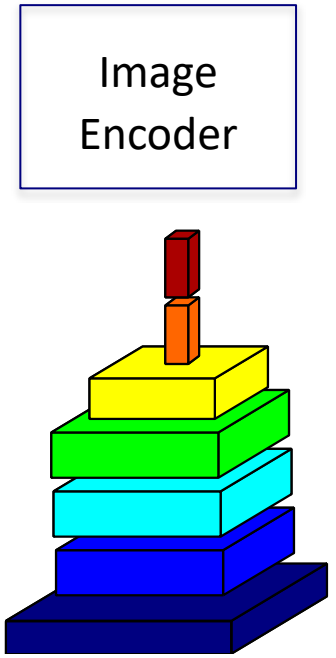
Super-Human Performance



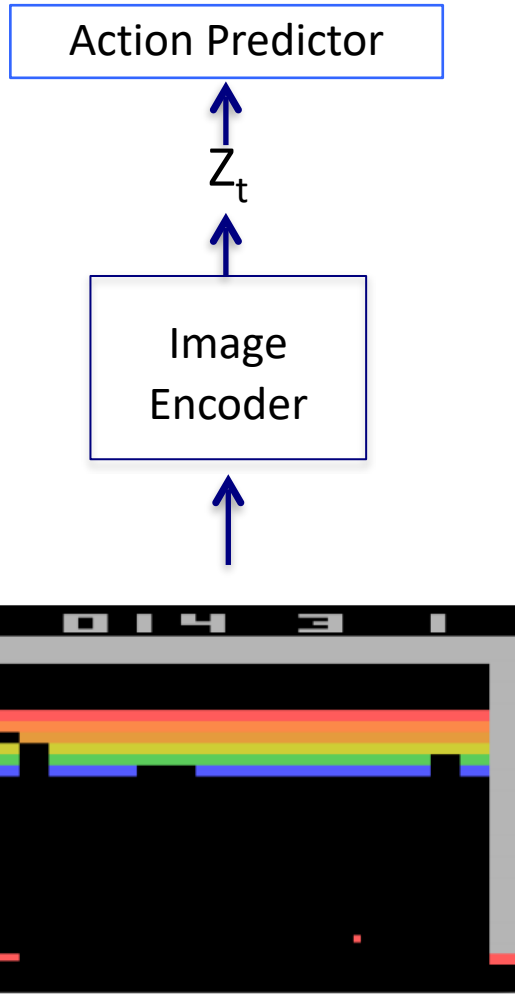
Playing ATARI Games

Image Encoder
is a
Multi layer Neural
Network



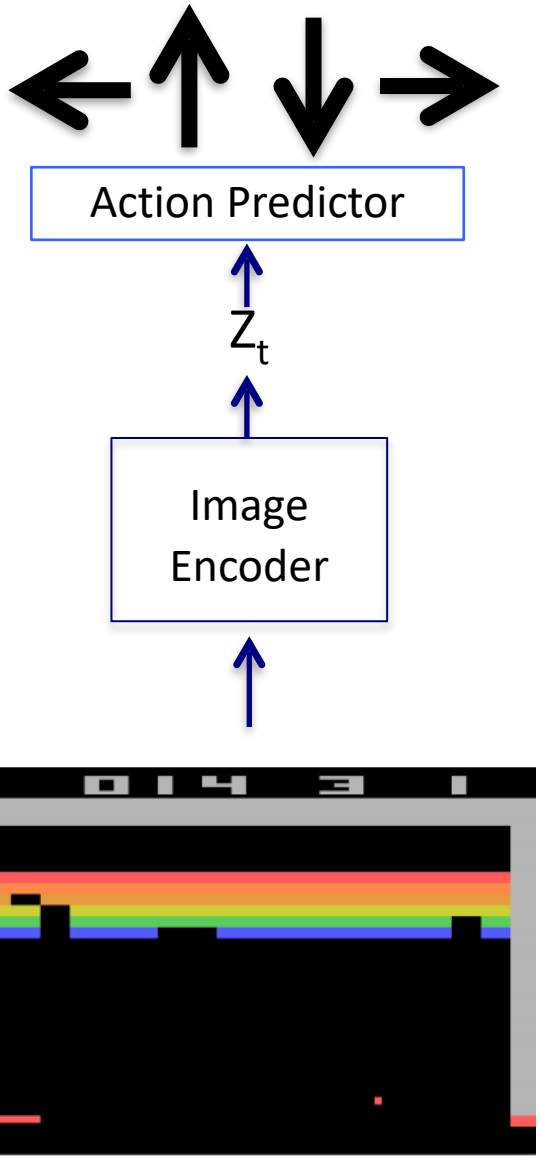


Playing ATARI Games

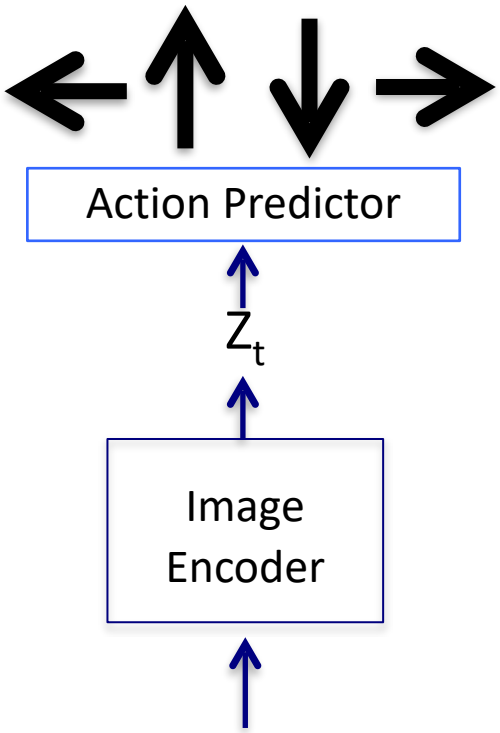


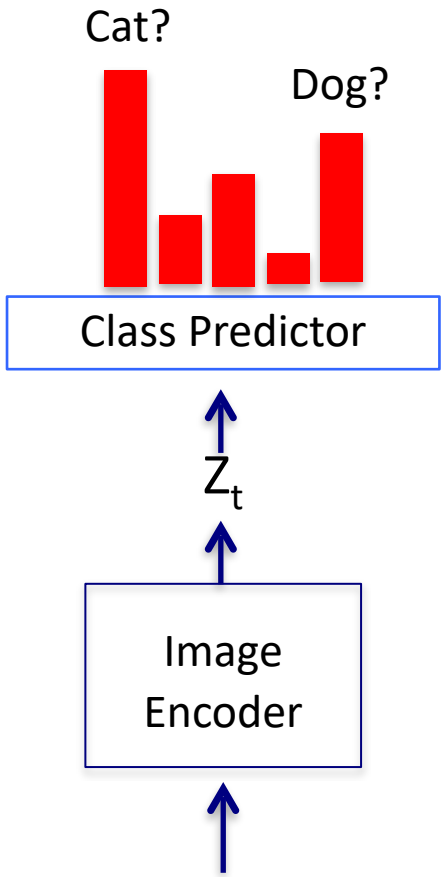
Action Predictor
is a
Multi layer Neural
Network

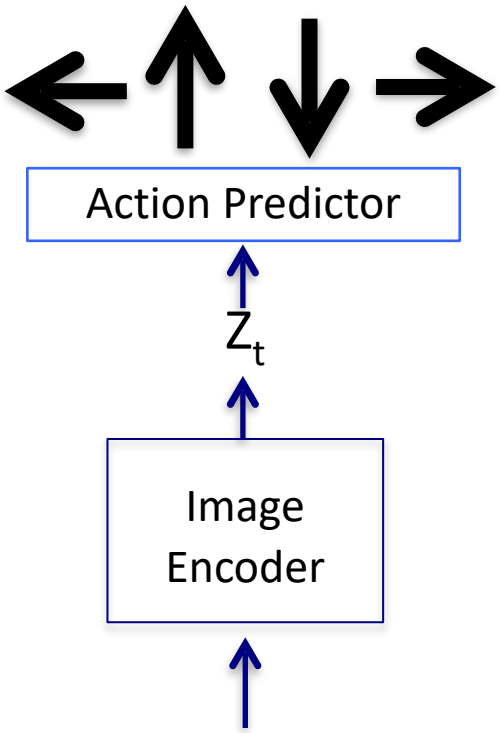
Playing ATARI Games

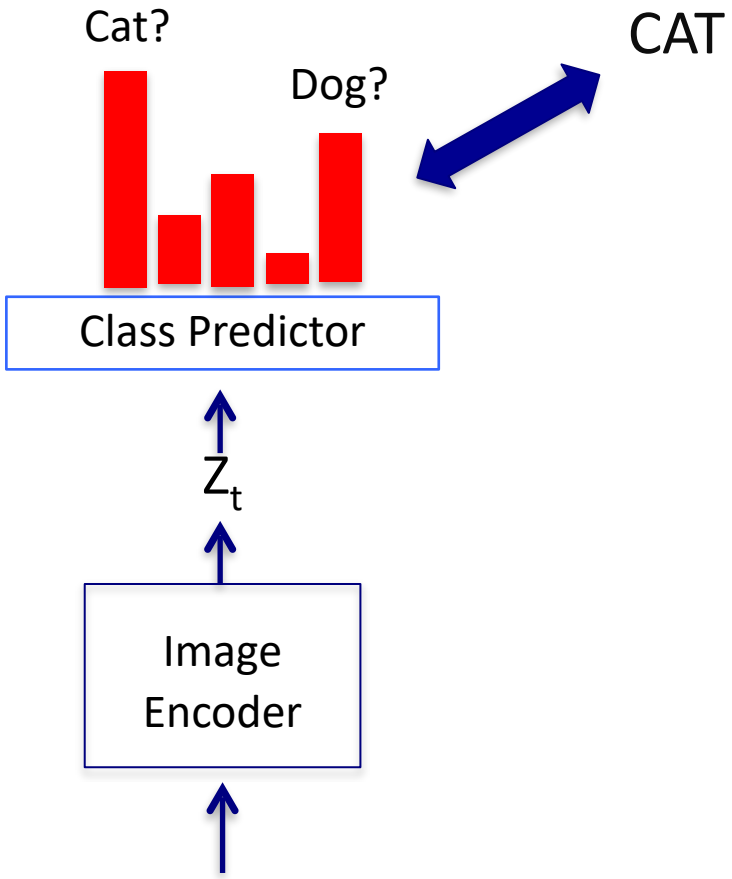


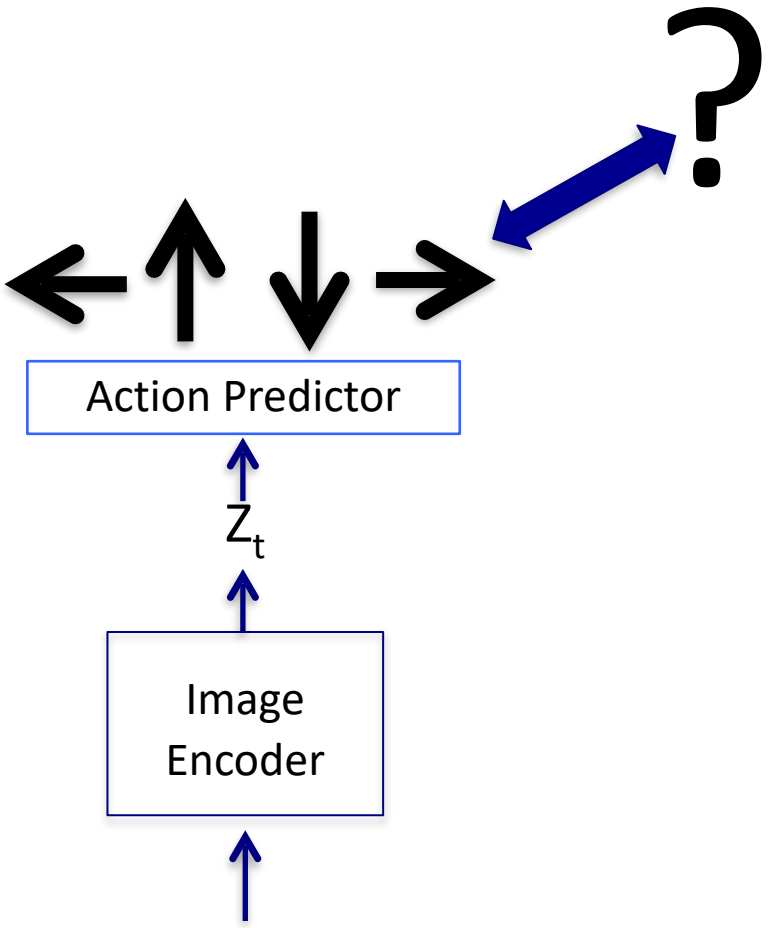
Similarity to Image Classification

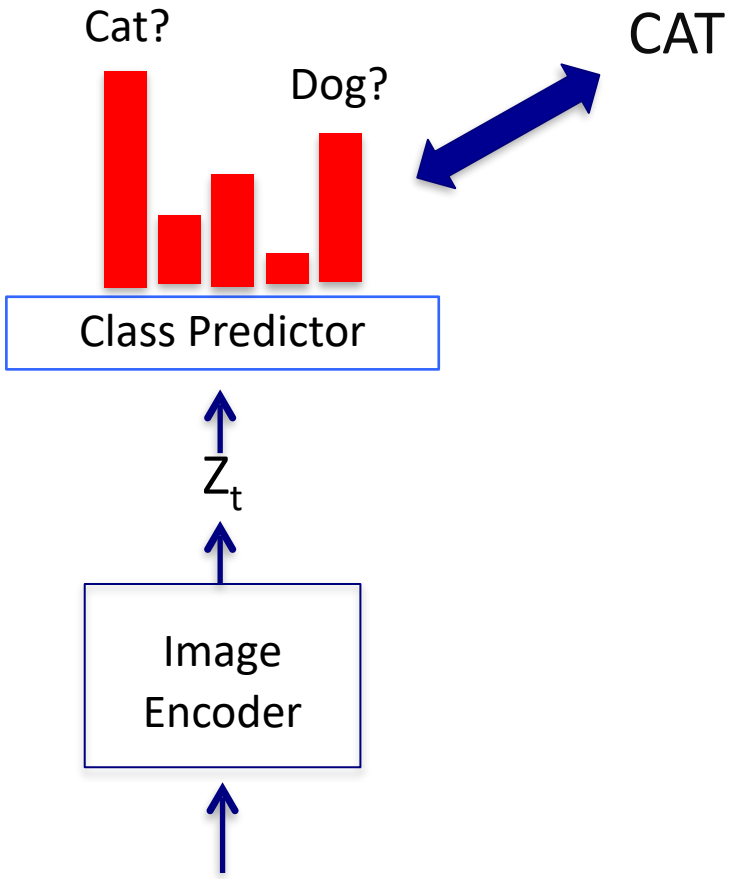


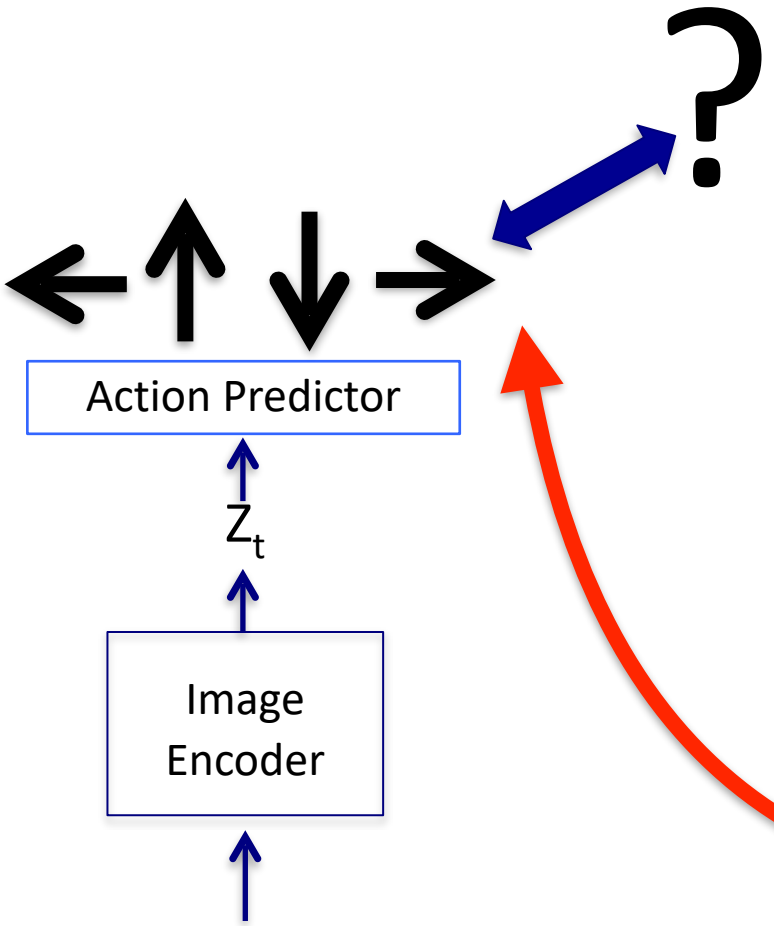



Action Predictor
Z_t
Image Encoder


Cat?
Dog?
CAT
Class Predictor
Z_t
Image Encoder


?
Action Predictor
Z_t
Image Encoder


Cat?
Dog?
CAT
Class Predictor
Z_t
Image Encoder



How to get the
"action" label?

Human Knowledge!

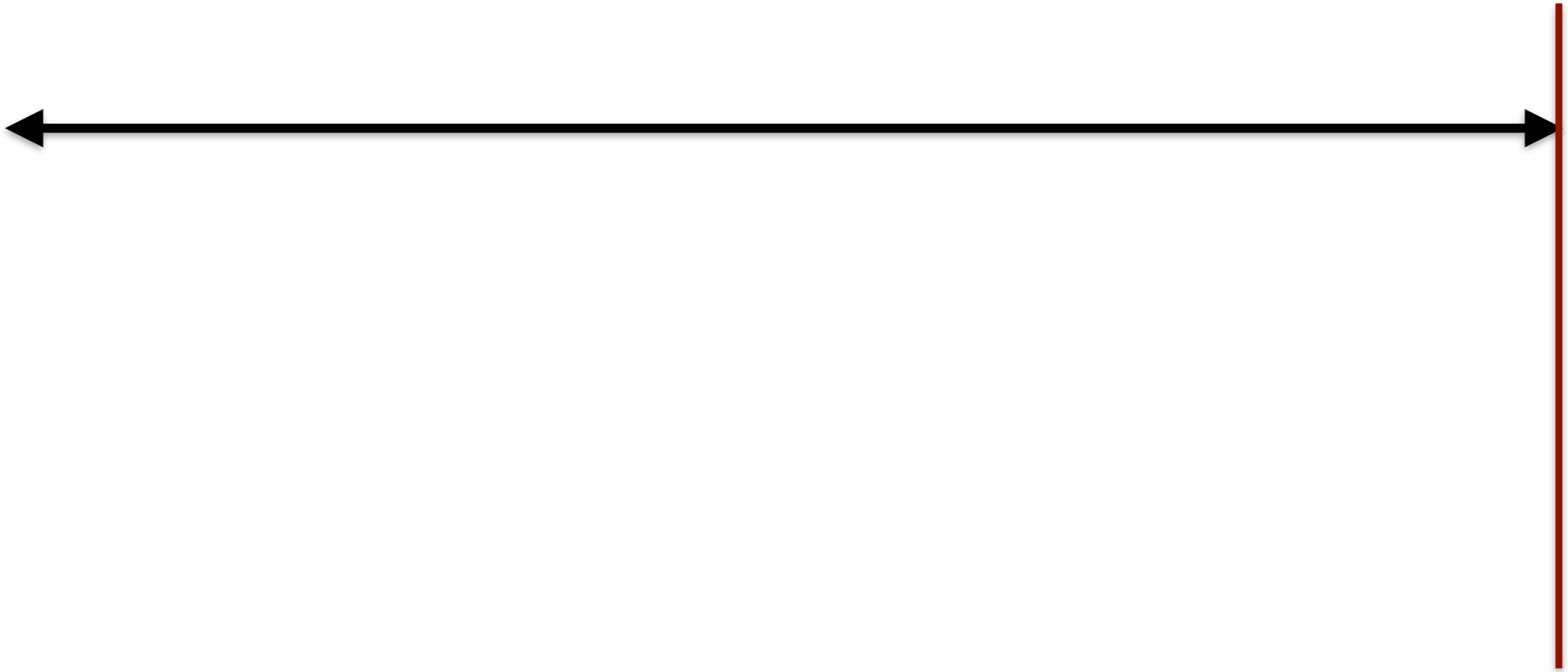
Unimate: First Industrial Robot (1954)

Used by General Motors for Welding

Unimate: Repetition after Memorization

Self-Learnt Behavior

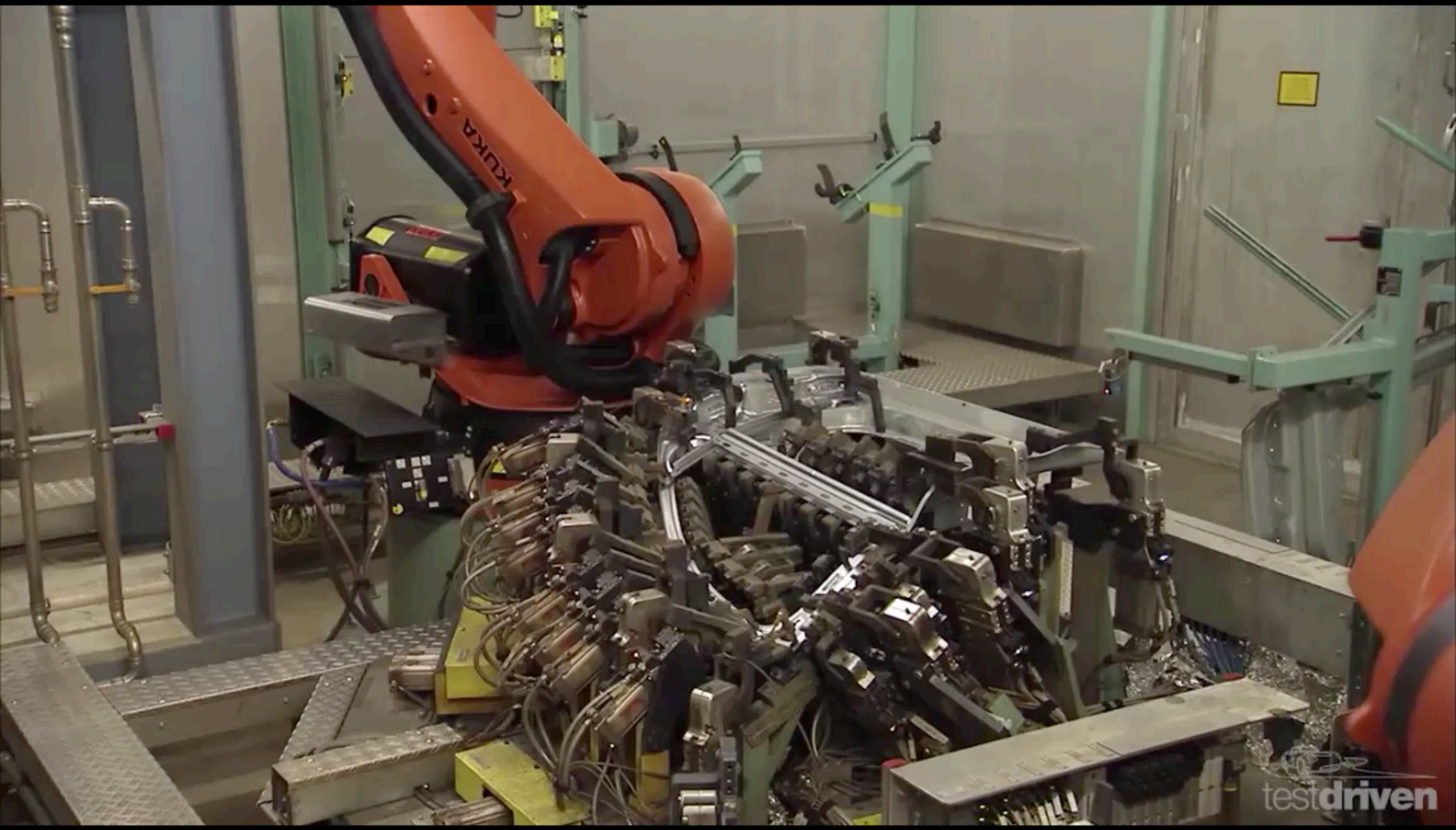
Hard-Coded Behavior



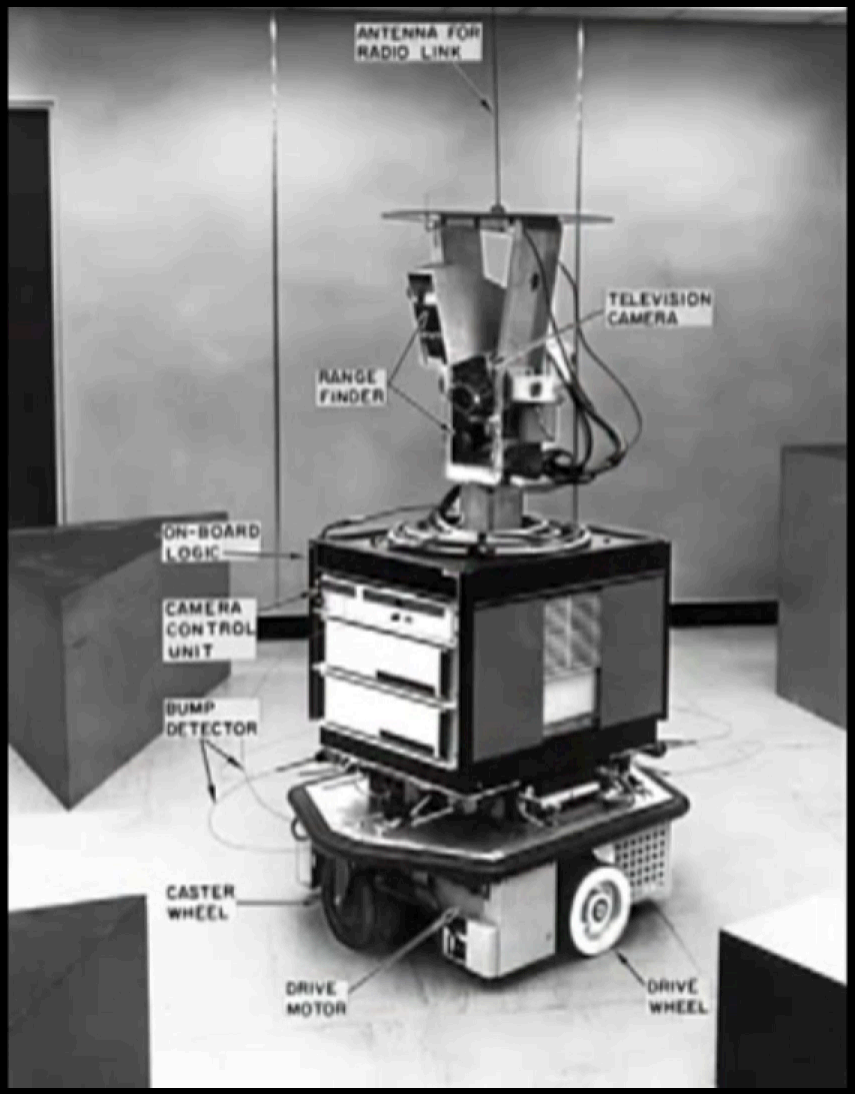
Unimate

(Memorized Behavior Cloning)

Even today, most industrial robots work like this!

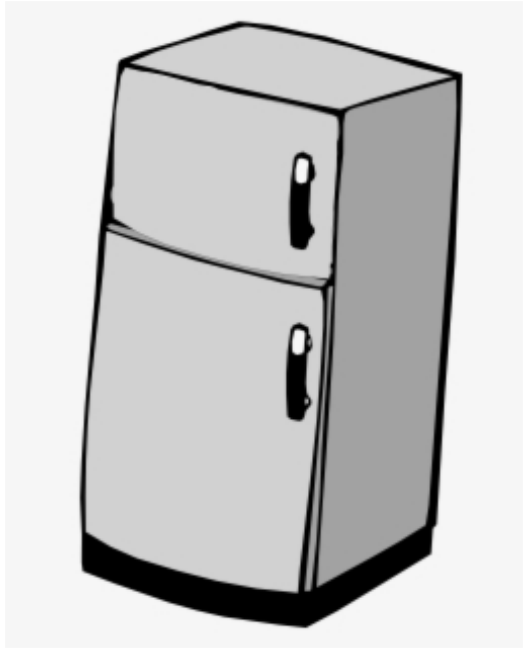


Shakey (1966-1972)

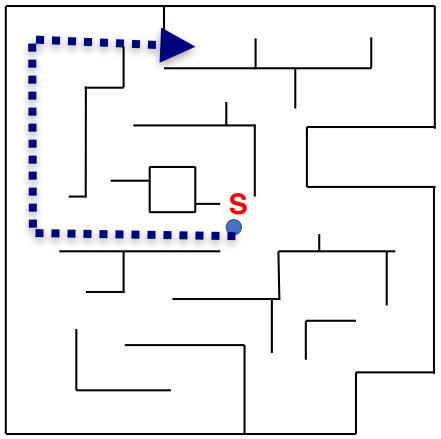


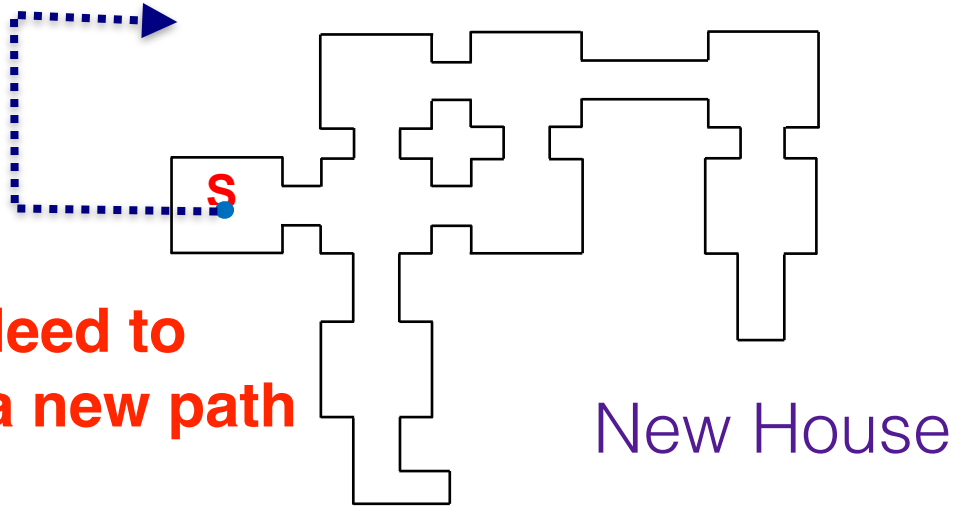
Consider the following problem

Current Observation



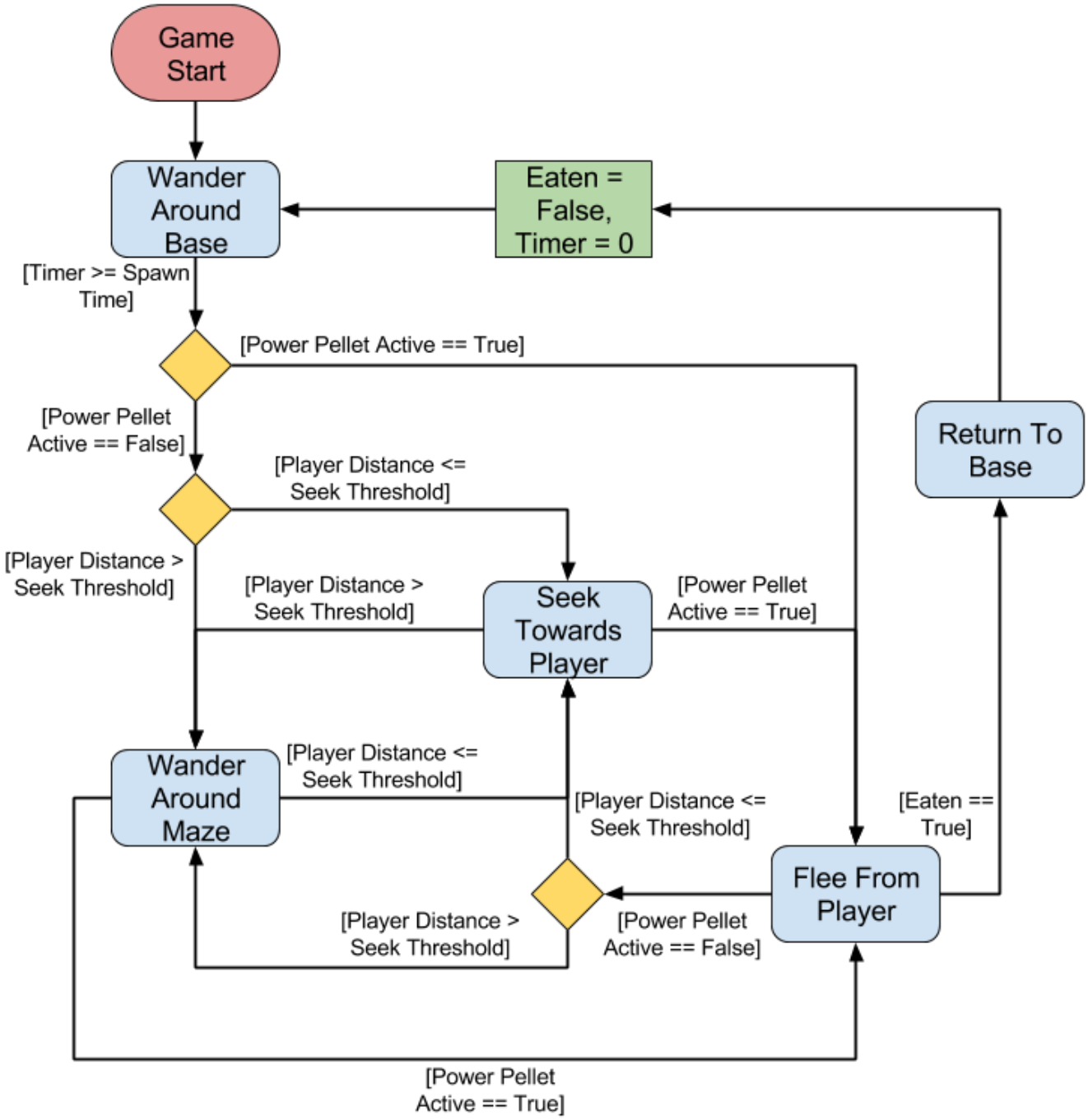
Goal





Need to find a new path

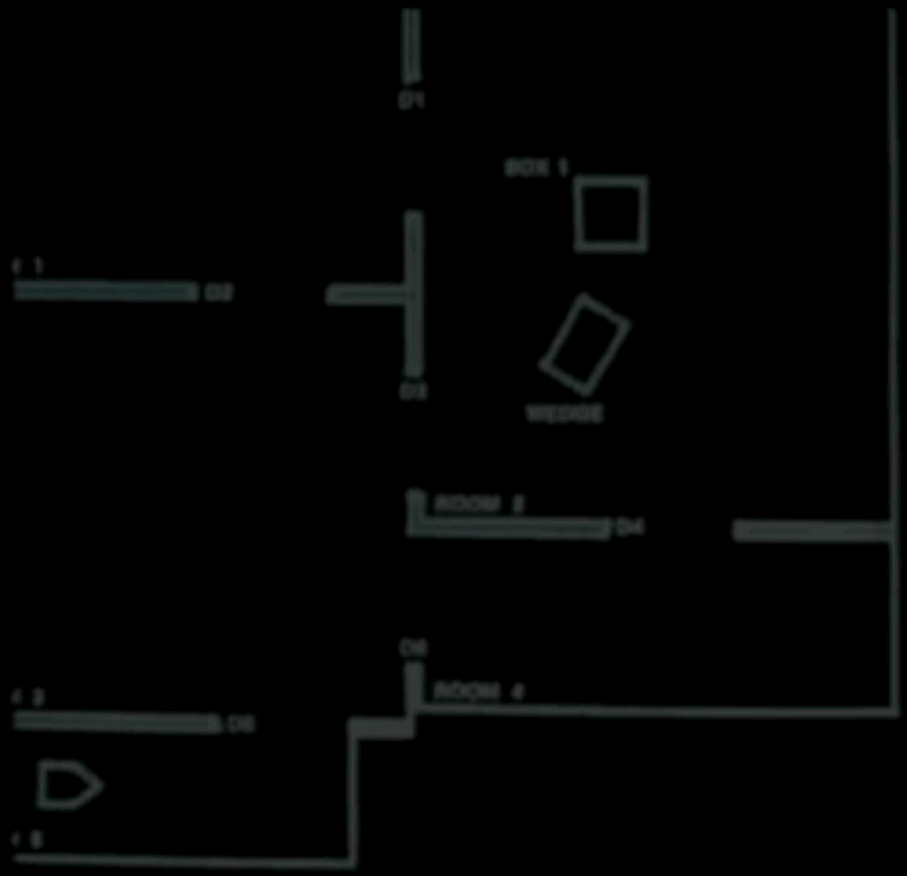
New House



Variable Name	Variable Definition
Spawn Time	The amount of time a ghost must wait before exiting the base
Timer	A timer on the ghost that is compared to Spawn Time
Power Pellet Active	Whether or not the player has just eaten a Power Pellet
Player Distance	The distance between the player and the current ghost
Seek Threshold	The maximum distance from the player the ghost can be to seek the player
Eaten	Whether or not a ghost has been eaten

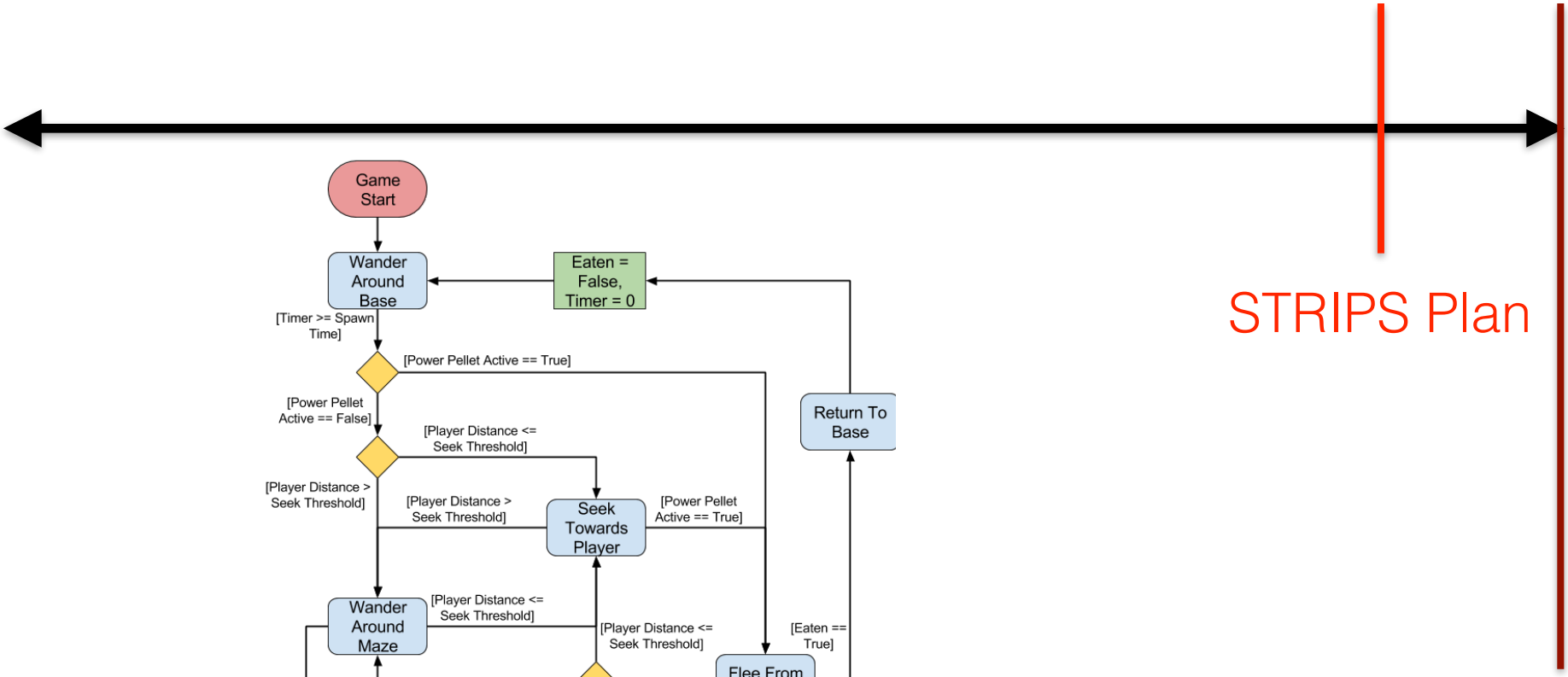
Symbol	Definition
	Terminal
	State
	Decision
	Process
	Transition
[asdfx]	Condition

Shakey (1966-1972)



Self-Learnt Behavior

Hard-Coded Behavior



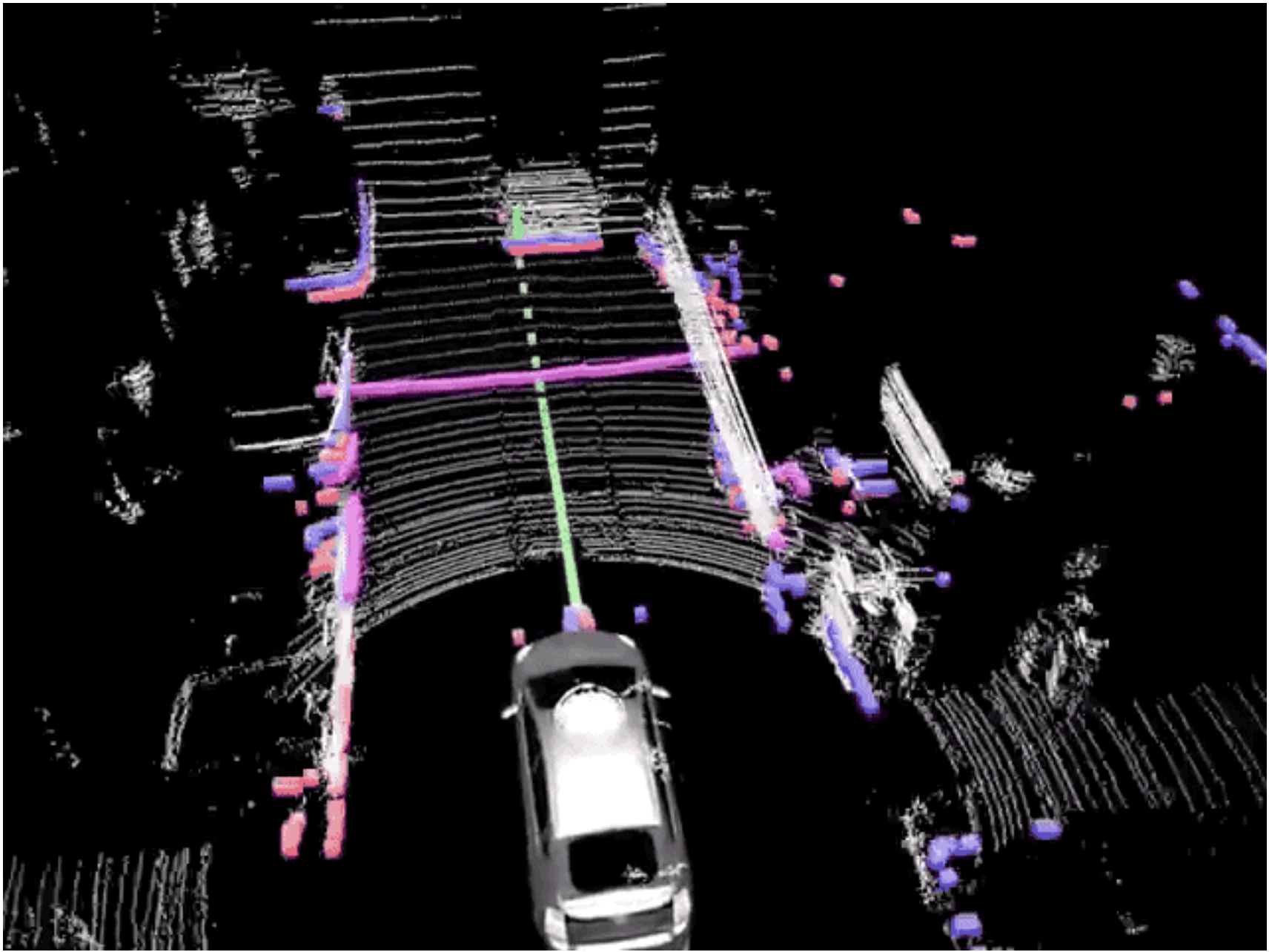
STRIPS Plan

Unimate

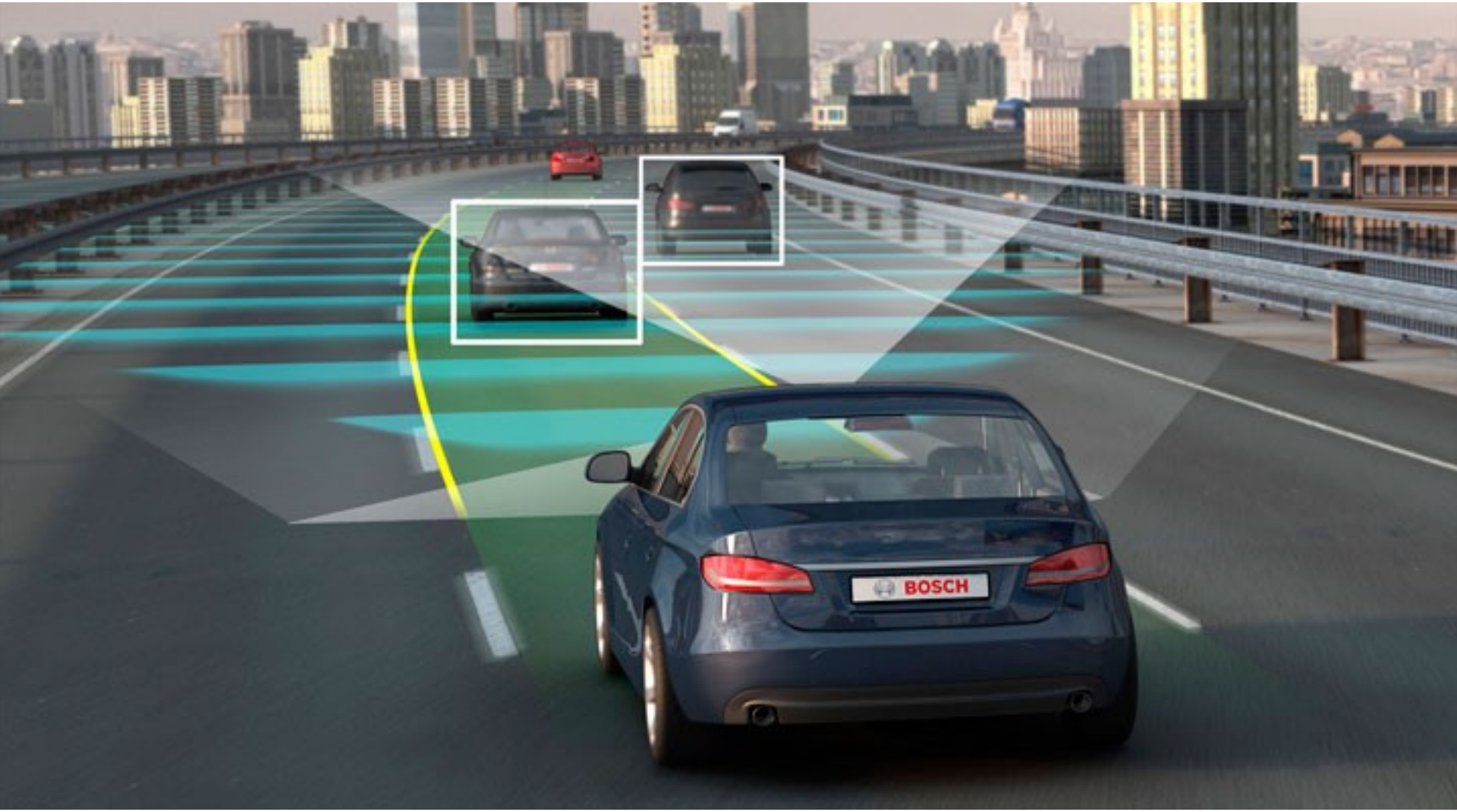
(Memorized Behavior Cloning)

Consider Self-Driving Cars

Map of the environment



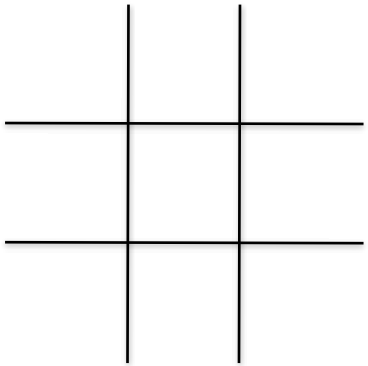
Adding Constraints such as don’t collide

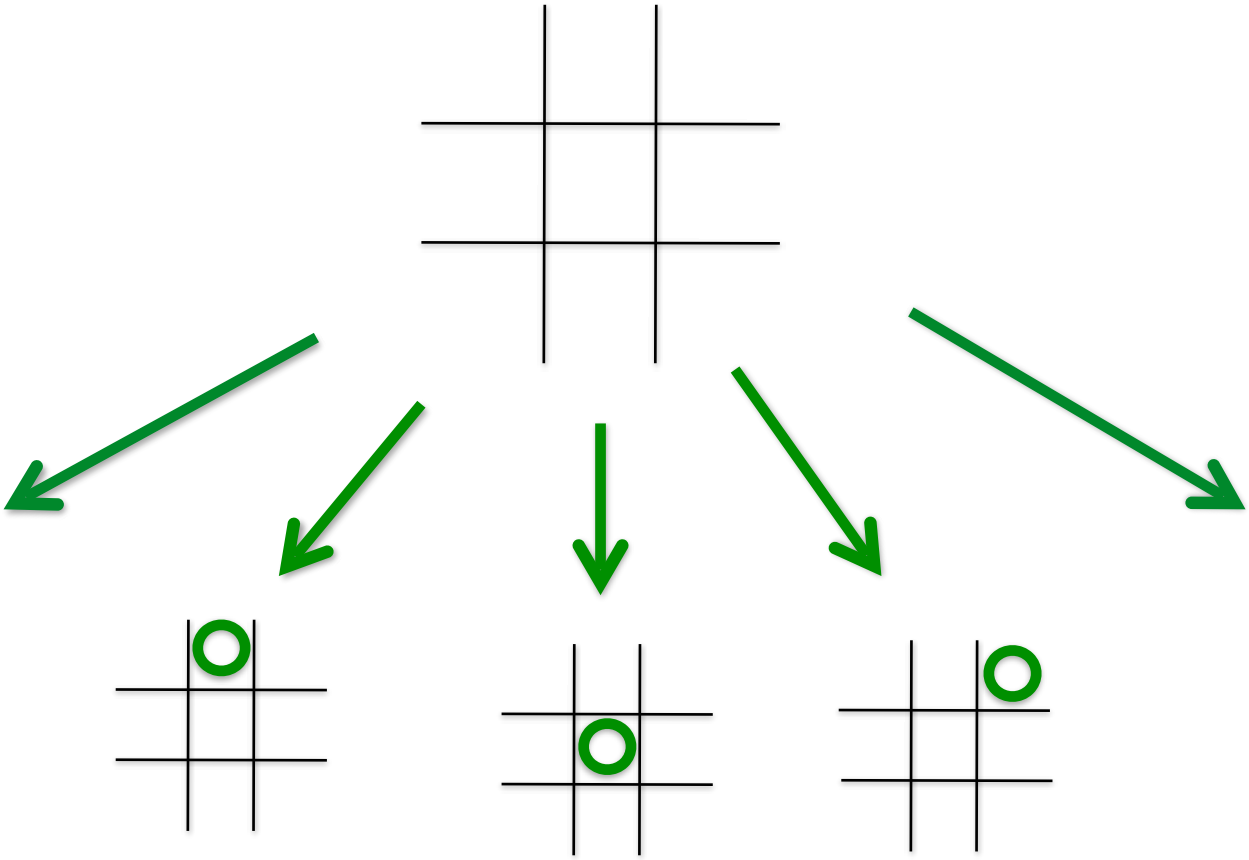


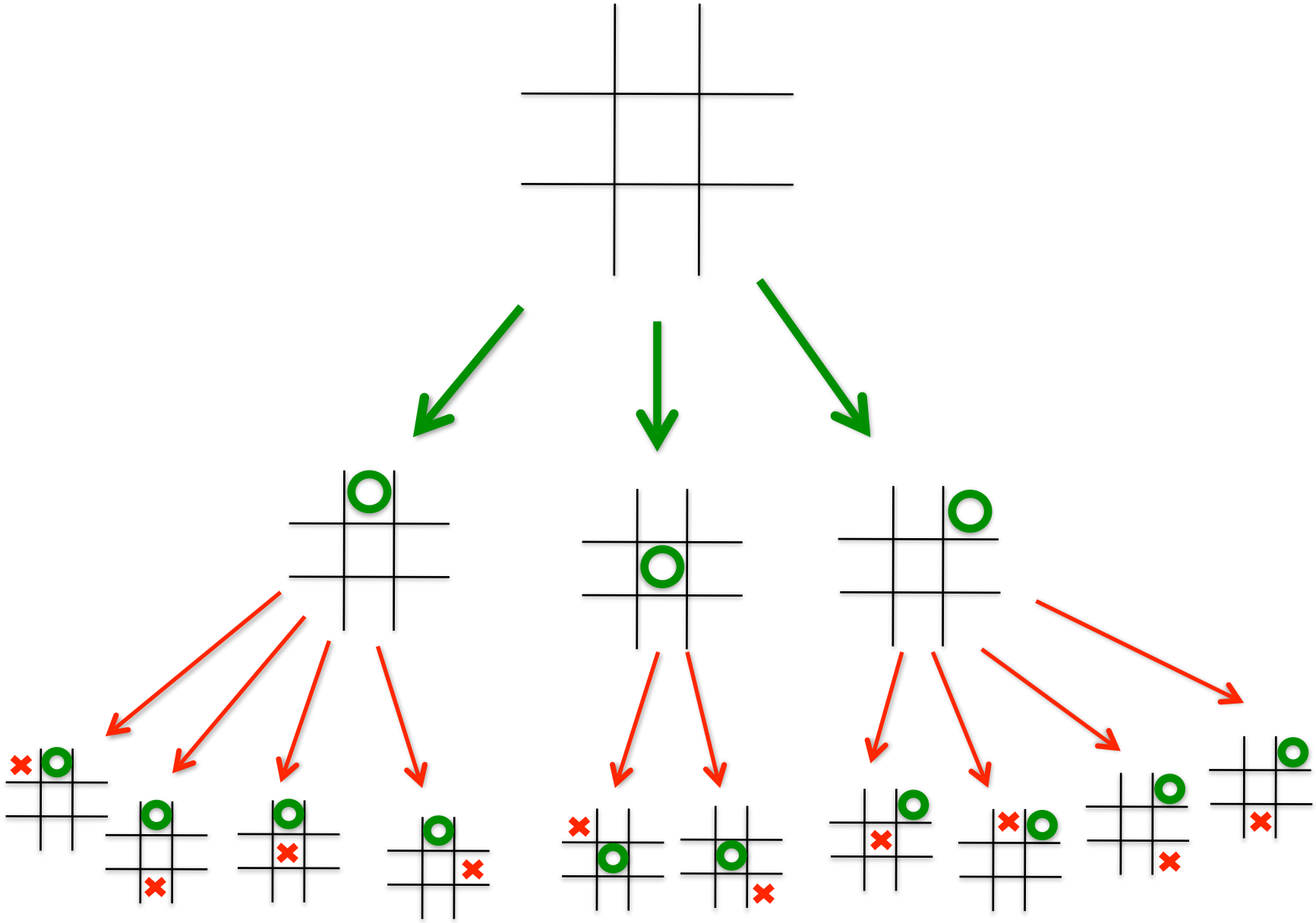
How to plan car’s actions?

Planning in a simpler domain: Games!

Consider
Tic-Tac-Toe





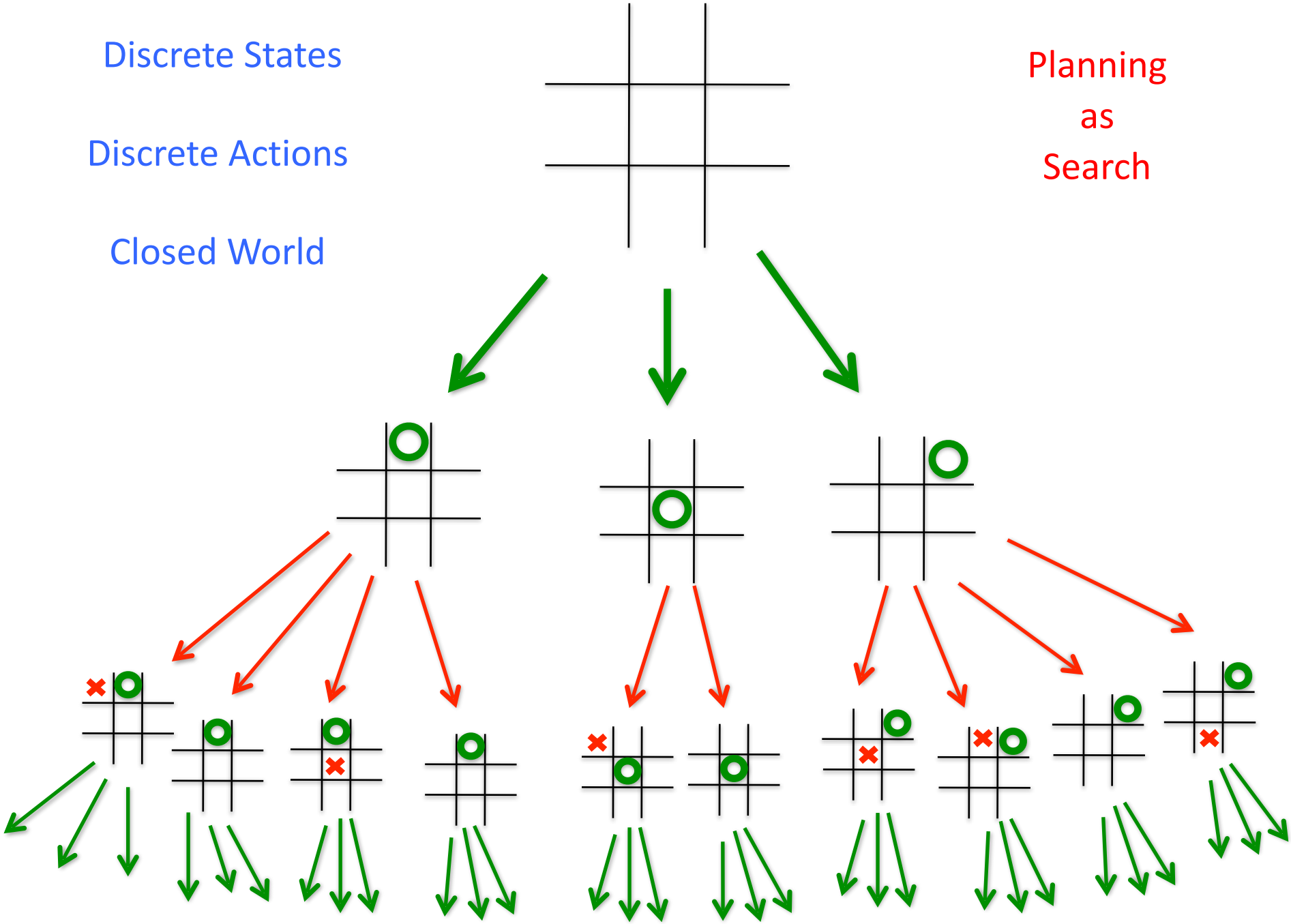


Discrete States

Discrete Actions

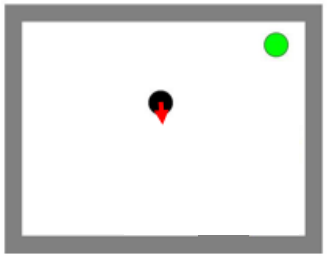
Closed World

Planning
as
Search

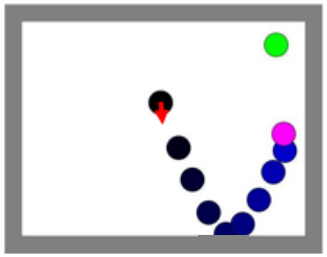


Intuition behind planning (in continuous domain)

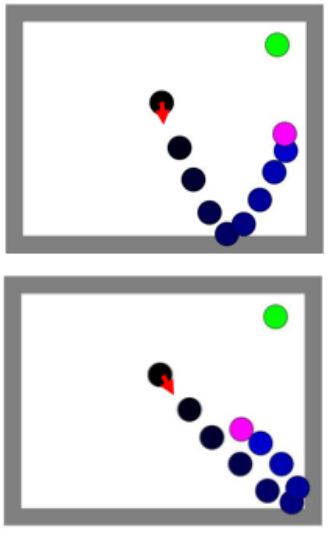
Task: hit the green ball



“Imagine” the effect of applied force

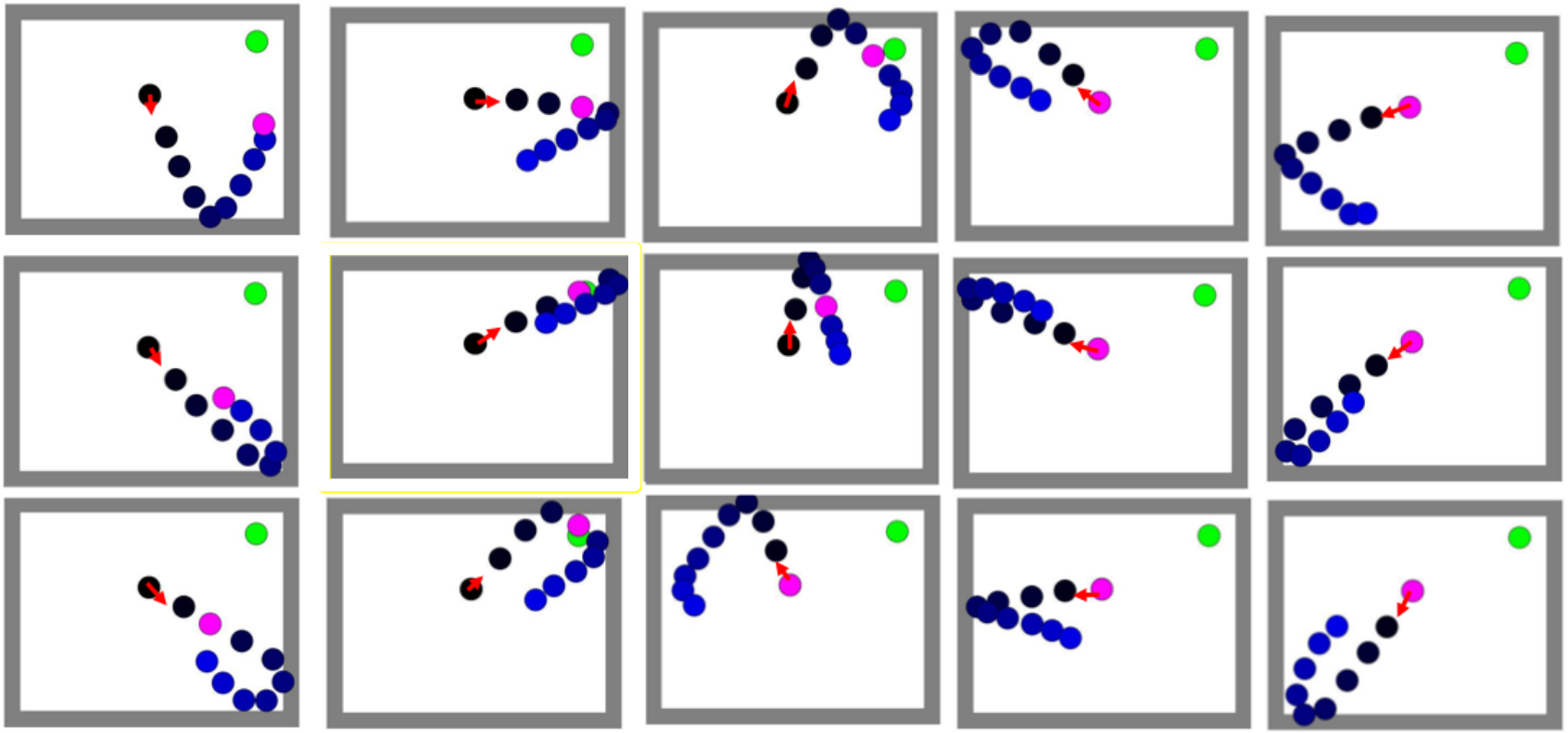


“Imagine” the effect of applied force



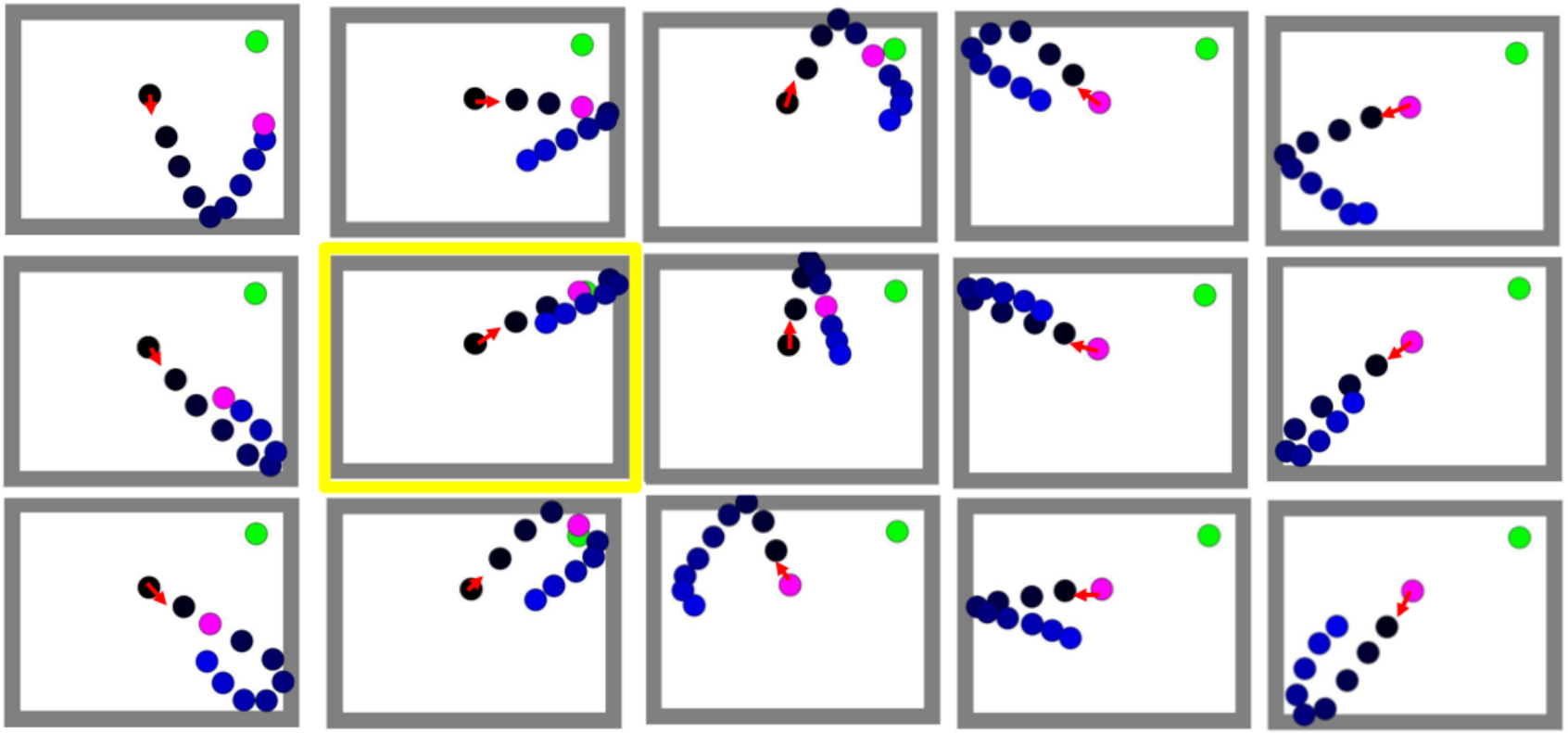
Planning Actions

Visually imagine the effect of different forces



Planning Actions

Visually imagine the effect of different forces



Choose the optimal force which "in simulation" leads to best result

Simulation using "models of how the world works" (e.g. physics)

Using Physics

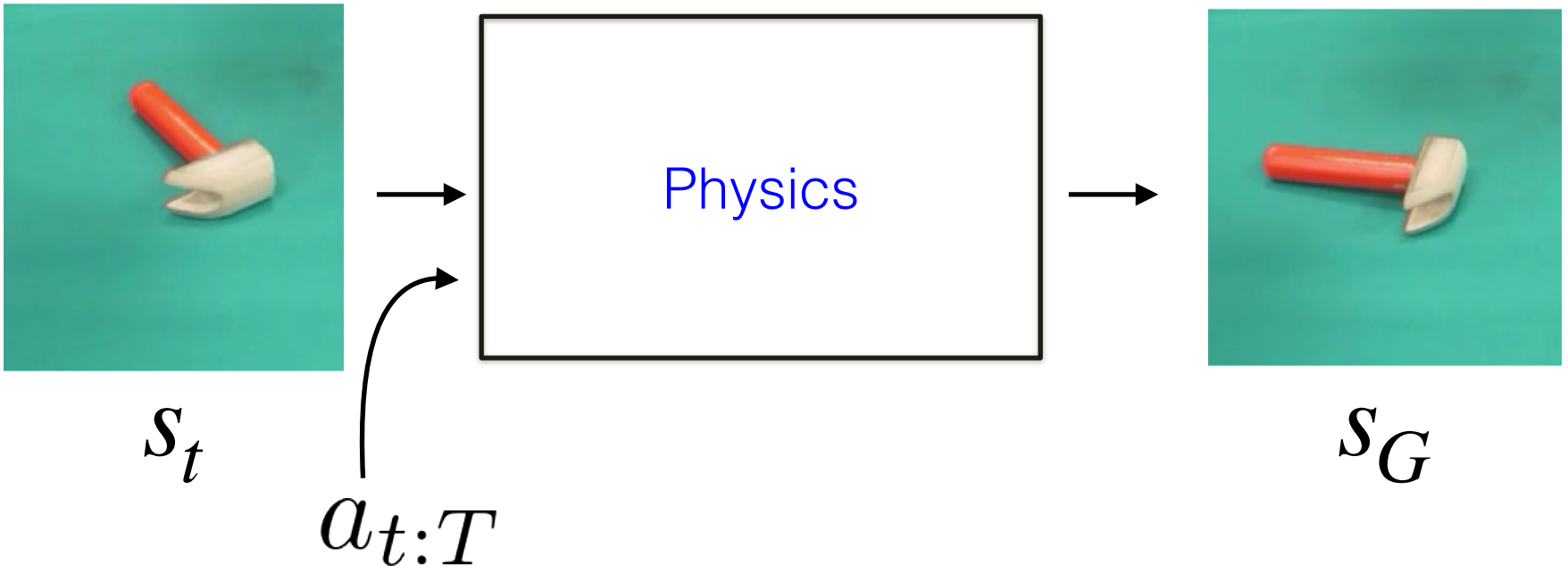
actions?

s_t

s_G

(position, pose of the hammer)

Using Physics



Using Physics

s_t

$a_{t:T}$

Physics

$s_{t+1} = f(s_t, a_t)$

s_G

Using Physics

s_t

$a_{t:T}$

Physics

$$s_{t+1} = f(s_t, a_t)$$

s_G

$$\min_{a_{t:T}} \|s_T - s_G\|^2$$

cost function

Using Physics

s_t

$a_{t:T}$

Physics

$$s_{t+1} = f(s_t, a_t)$$

s_G

$$\min_{a_{t:T}} \|s_T - s_G\|^2$$

subject to $s_{t+1} = f(s_t, a_t) \quad \forall t \in [1, T]$

Model Based Control

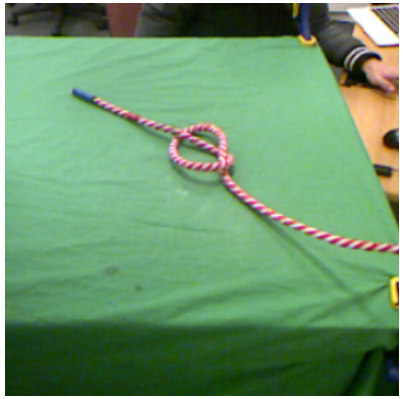
Using Physics

s_t

$a_{t:T}$

Physics

$$s_{t+1} = f(s_t, a_t)$$



s_G

$$\min_{a_{t:T}} \|s_T - s_G\|^2$$

subject to $s_{t+1} = f(s_t, a_t) \;\; \forall t \in [1, T]$

Using Physics

s_t

$a_{t:T}$

Physics

$$s_{t+1} = f(s_t, a_t)$$

s_G

$$\min_{a_{t:T}} \|s_T - s_G\|^2$$

subject to $s_{t+1} = f(s_t, a_t) \quad \forall t \in [1, T]$

Using Physics

s_t

$a_{t:T}$

Physics

$s_{t+1} = f(s_t, a_t)$

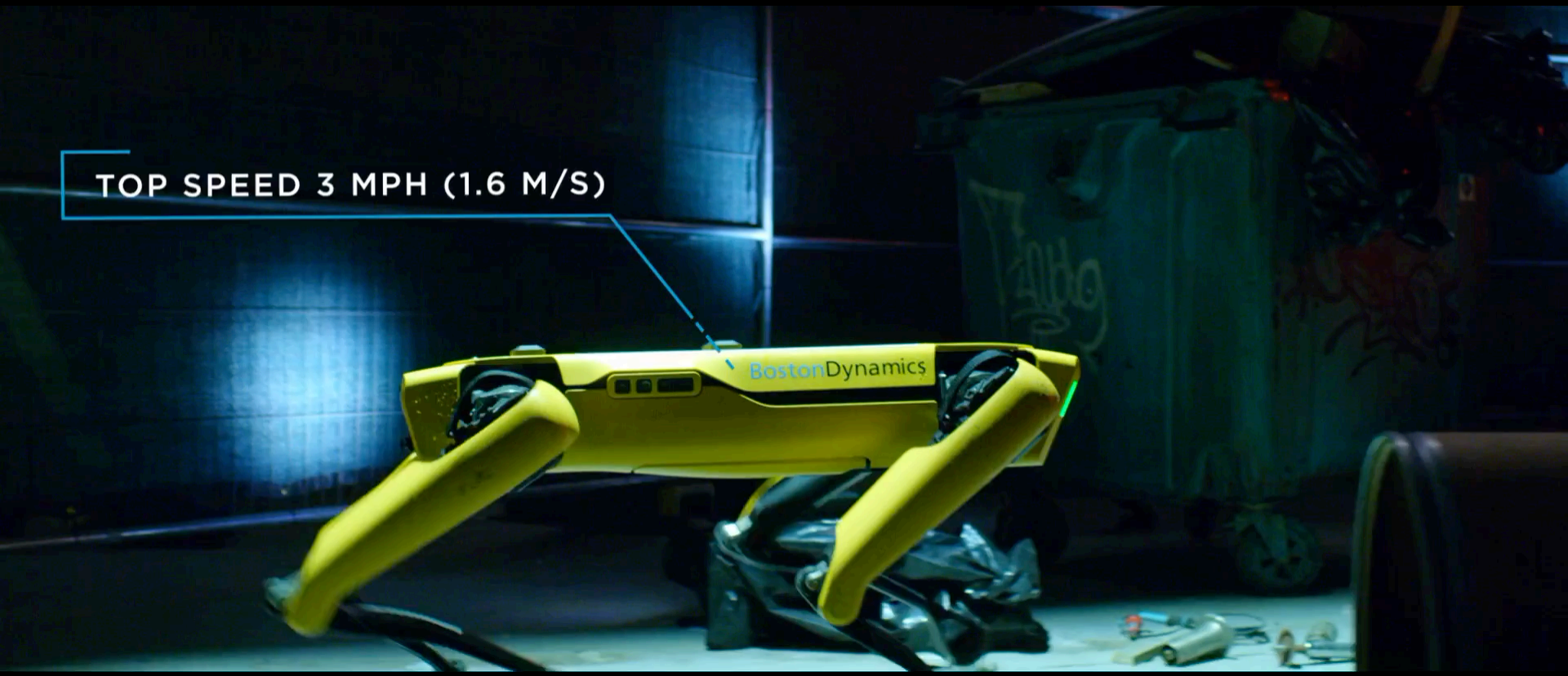
s_G

$$\min_{a_{t:T}} \|s_T - s_G\|^2$$

subject to $s_{t+1} = f(s_t, a_t) \quad \forall t \in [1, T]$

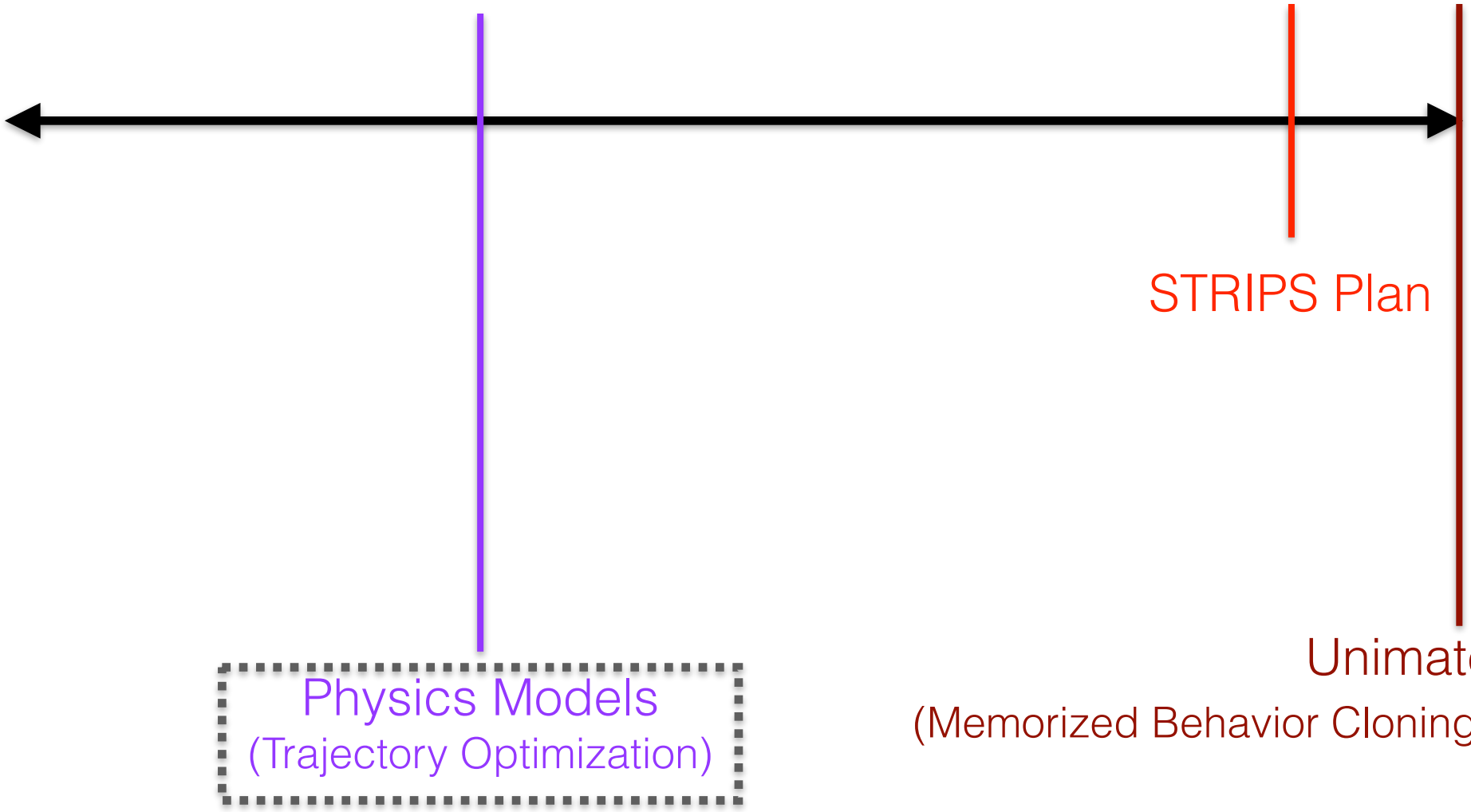
Impressive Specialists

Impressive Specialists

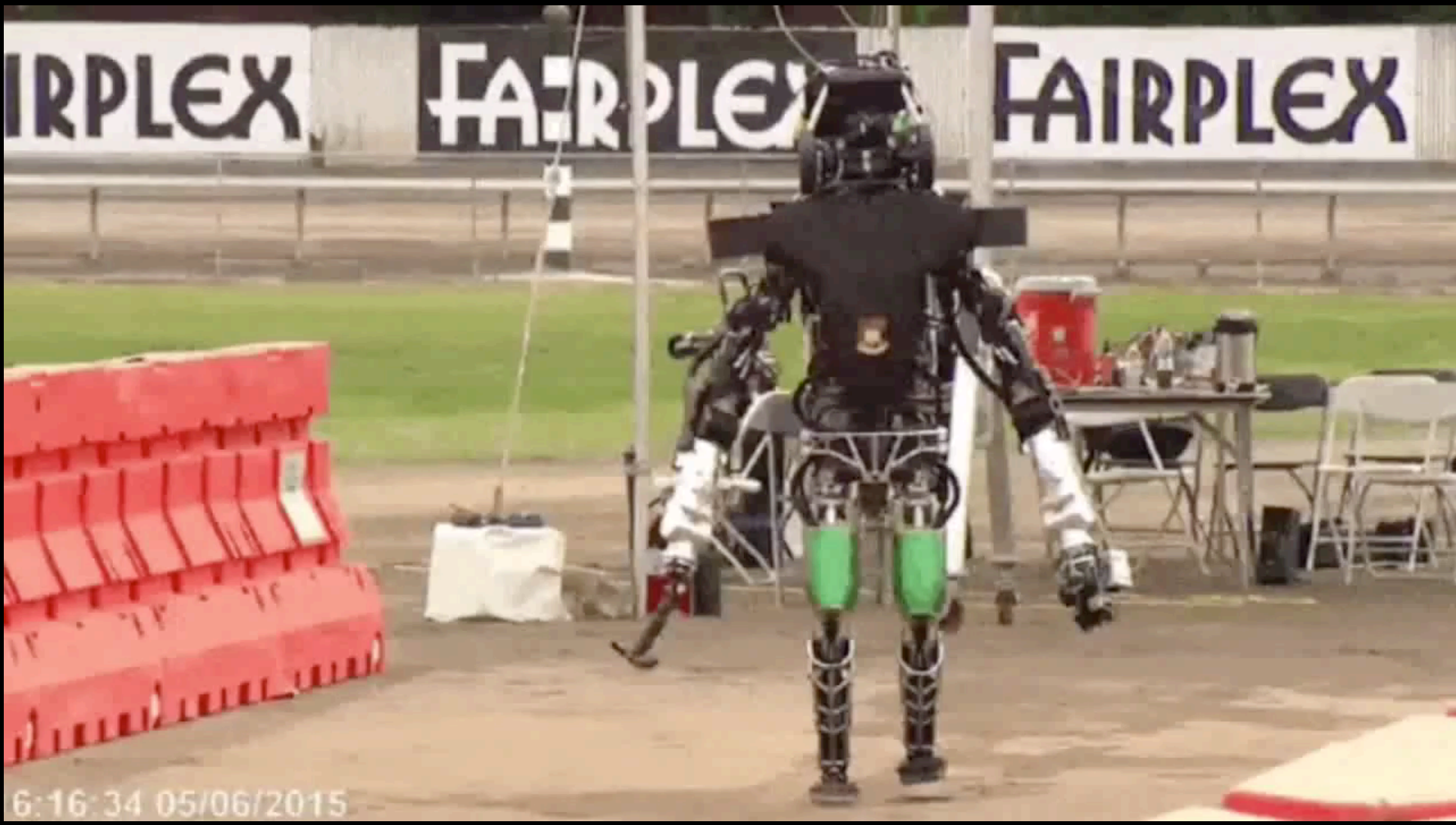


Self-Learnt Behavior

Hard-Coded Behavior

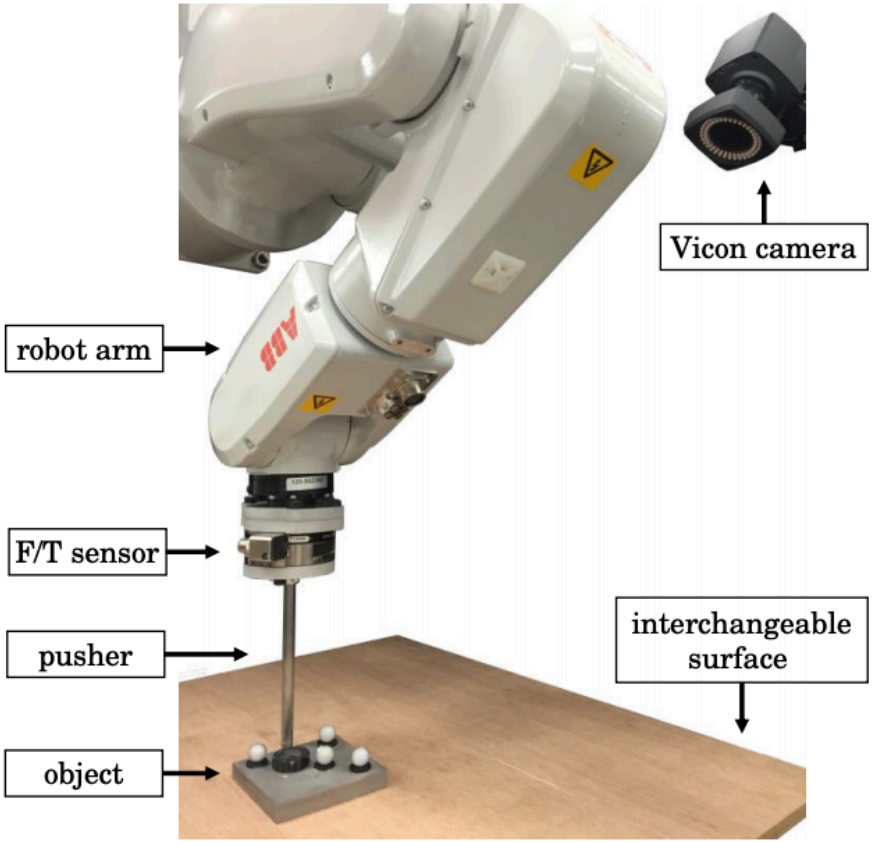


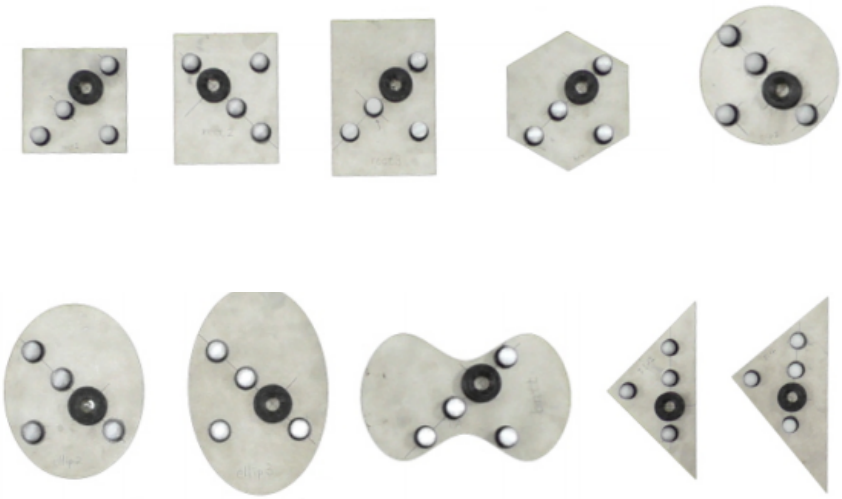
Limitations of these systems



Credits: DARPA Challenge 2015

Studying Object Pushing





More than a Million Ways to be Pushed, Yu et al., 2016

However, **predicting object motion** is not easy.

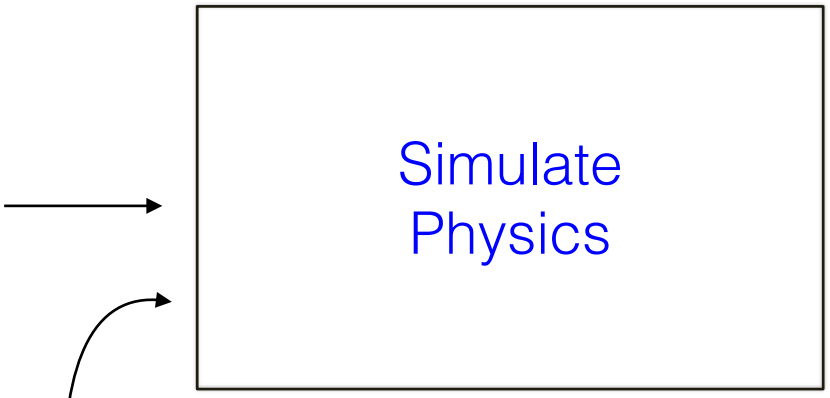
More than a Million Ways to be Pushed, Yu et al., 2016

Take Away

Model Based methods are good as the specified models!

Some Considerations!

s_t



$a_{t:T}$

s_G

One can observe the image
Features like position, pose must be inferred

Can't feed images into a physics engine!

Some Considerations!

Mass

Friction

Shape

•

•

Simulate Physics

System Identification

Issues in Classical Model Based Control

Mass

Friction

Shape

•

•

Simulate Physics

Mass

Friction

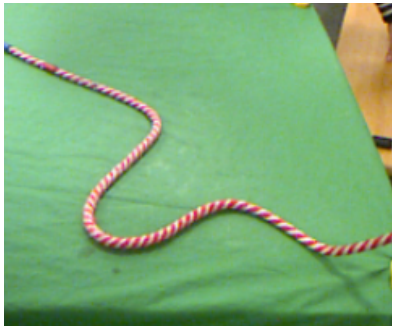
Shape

Stiffness

•

•

Simulate Physics



For different task: tedious system identification

Issues in Classical Model Based Control

Mass

Friction

Shape

Simulate Physics

How do we estimate these parameters?

Issues in Classical Model Based Control

Mass

Friction

Shape

Simulate Physics

Computer Vision

Issues in Classical Model Based Control

Mass

Friction

Shape

Simulate Physics

Computer Vision

can be very hard to predict from pixels!

Issues in Classical Model Based Control

Mass

Friction

Shape

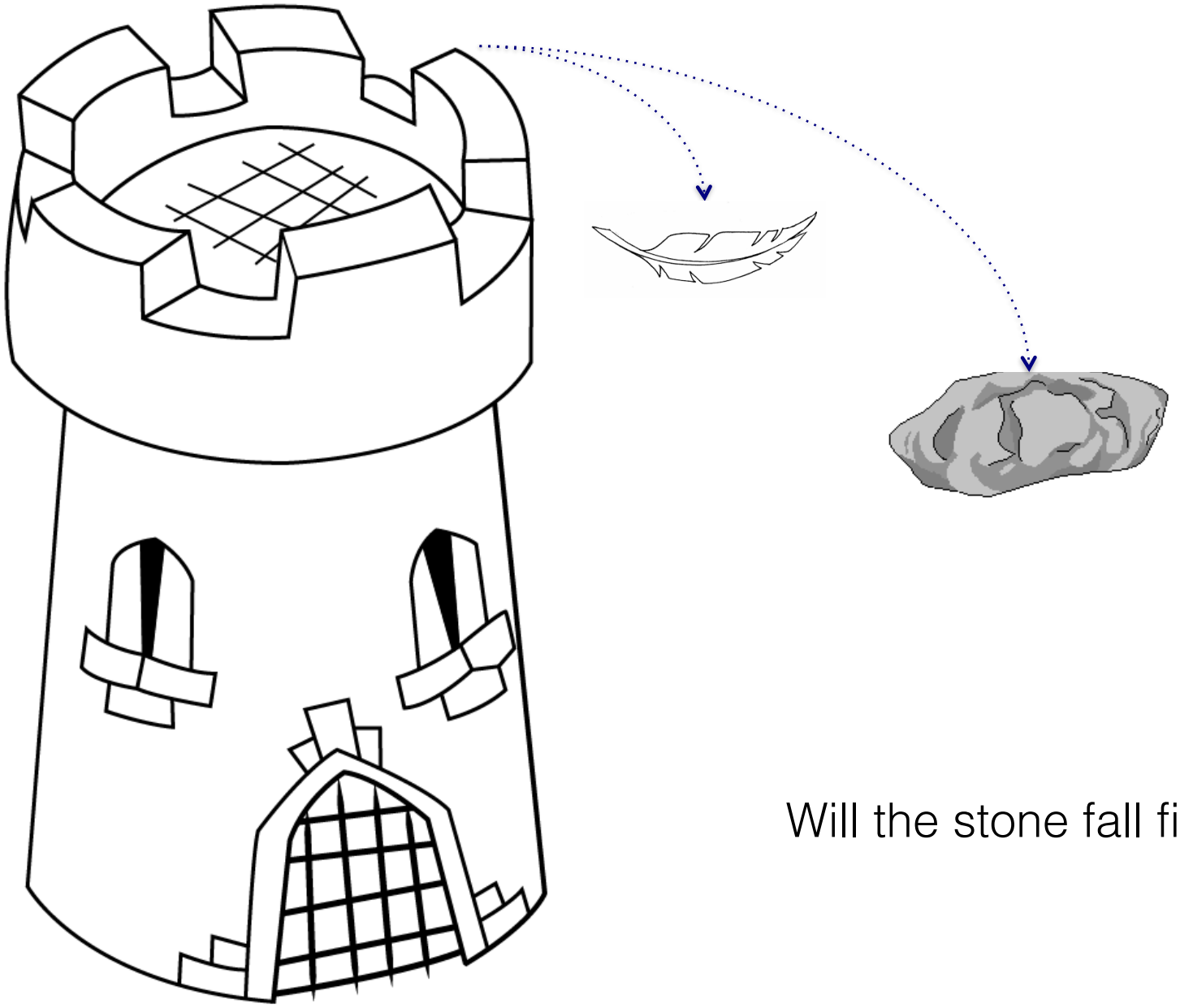
Simulate Physics

Computer Vision

can be very hard to predict from pixels!

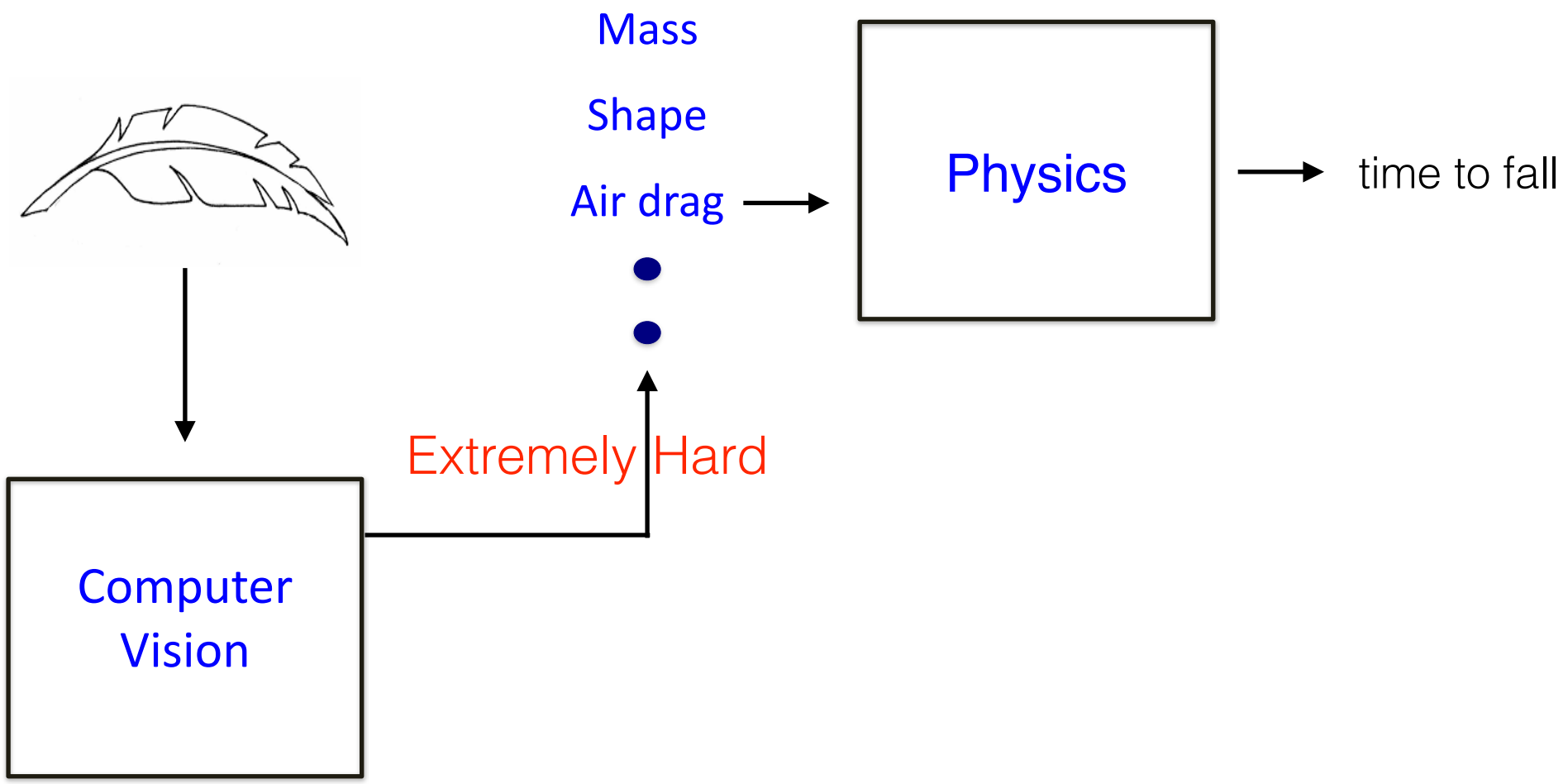
Errors accumulate!

In general the model is unknown in the observation space!

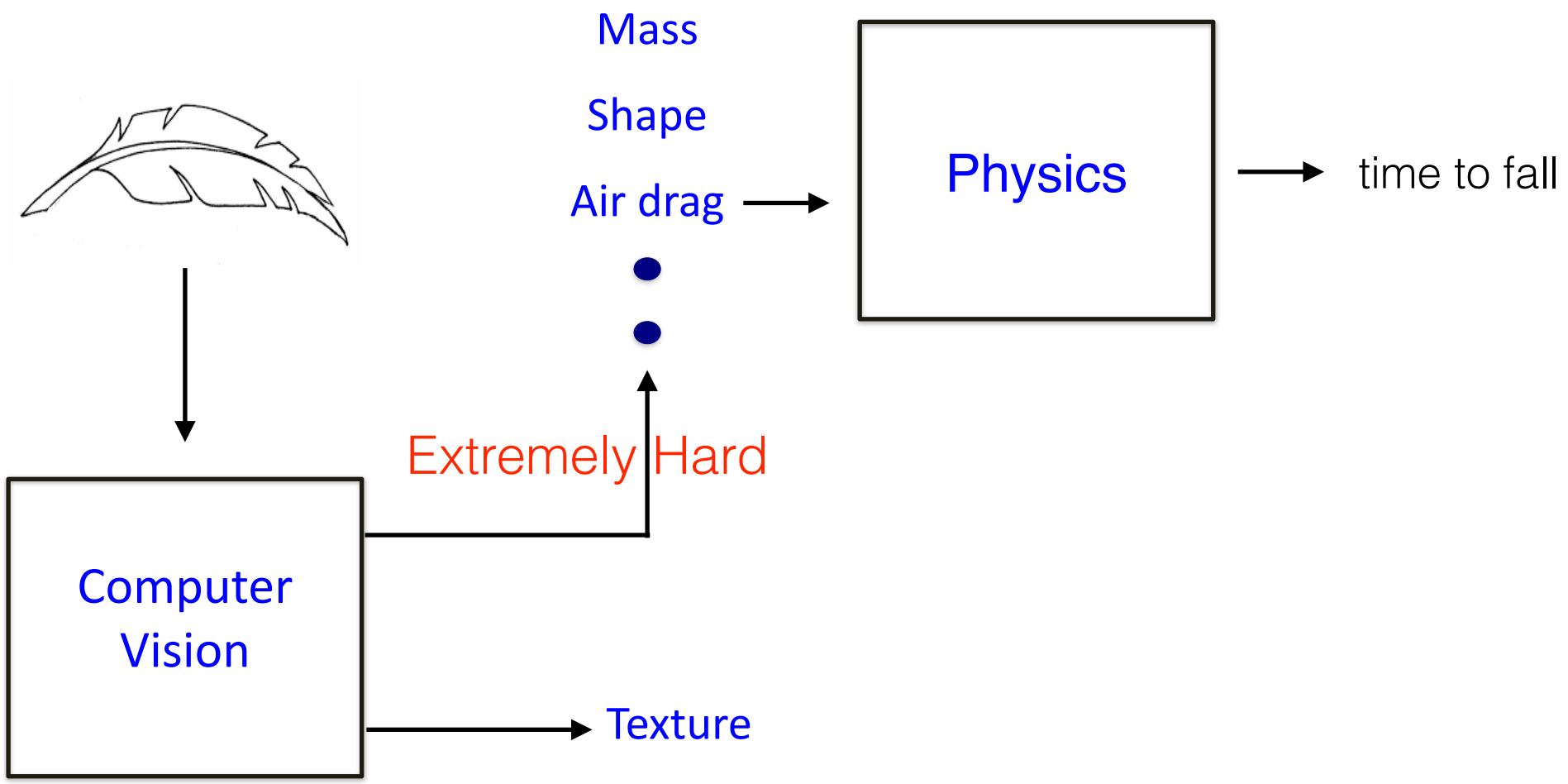


Will the stone fall first?

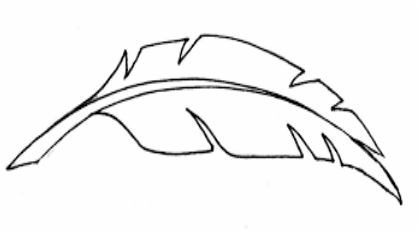
Second Approach: Classical Model Based Control



Second Approach: Classical Model Based Control



Second Approach: Classical Model Based Control



Mass

Shape

Air drag →

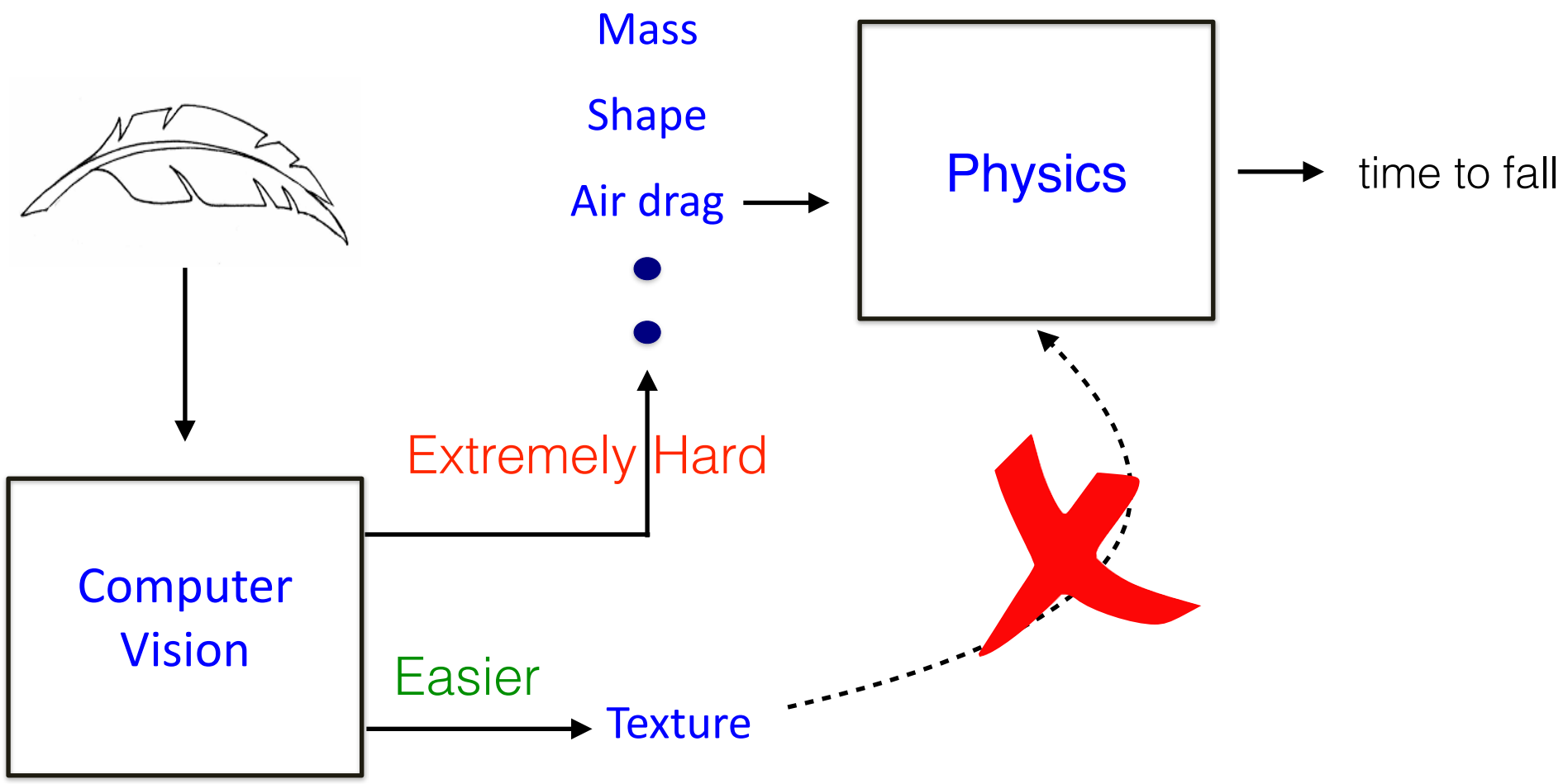
Physics → time to fall

Extremely Hard

Computer Vision

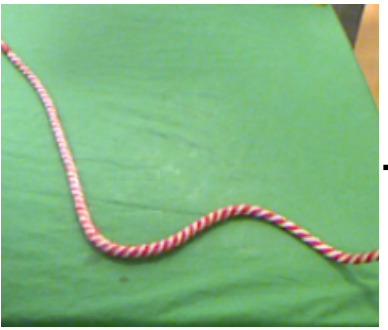
Easier → Texture

Second Approach: Classical Model Based Control



Some representations are easier for vision, others for physics!

Issue with classical model based control

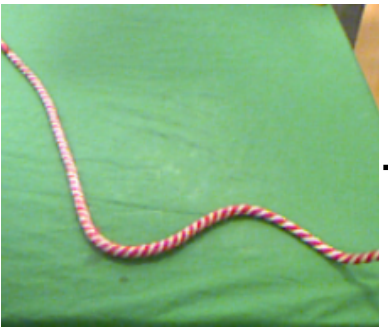


Computer Vision

Simulate Physics

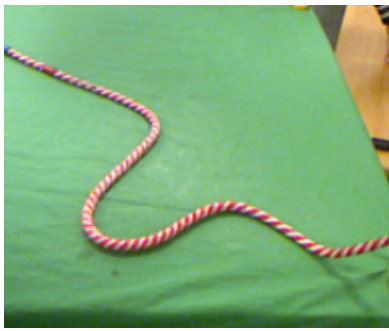
What is the appropriate representation?

Issue with classical model based control



Computer Vision

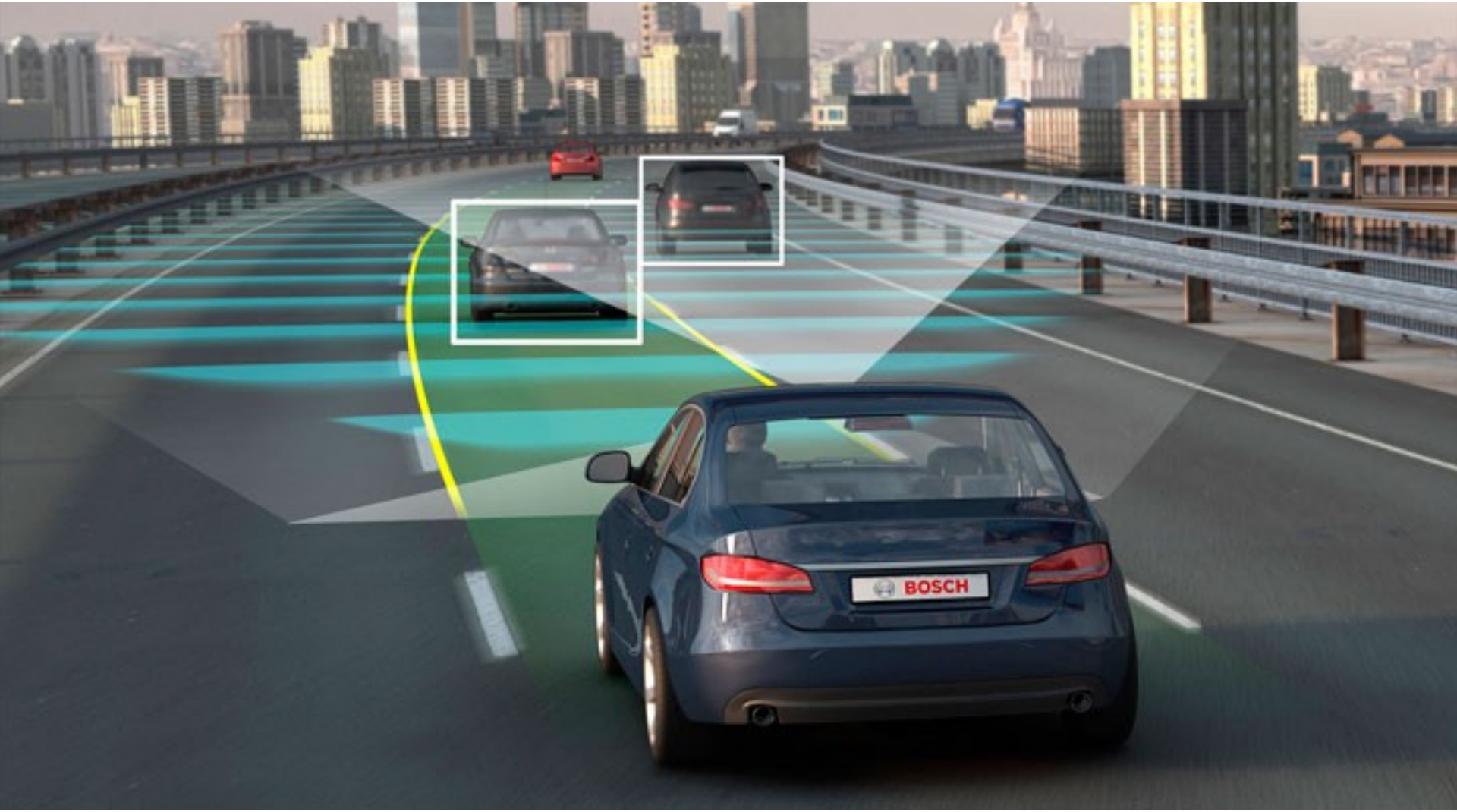
Simulate Physics



Is this sufficient?

Reconsider Self-Driving Cars

Not only drive on highways ..



How to encode these in a model?

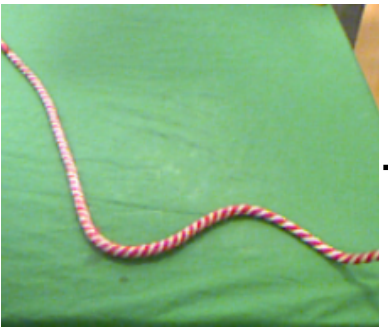
How to encode these in a model?

How to encode these in a model?

How to encode these in a model?

Enumerating all possible conditions is very hard!

Issue with classical model based control

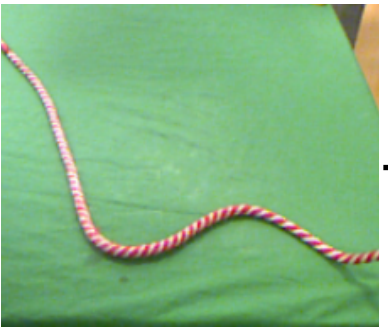


Computer Vision

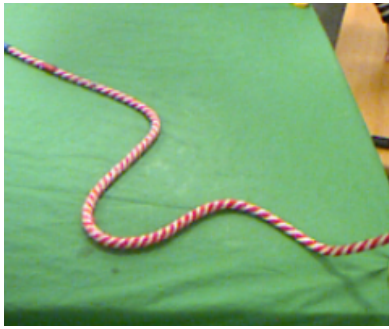
Simulate Physics

What is the appropriate representation?

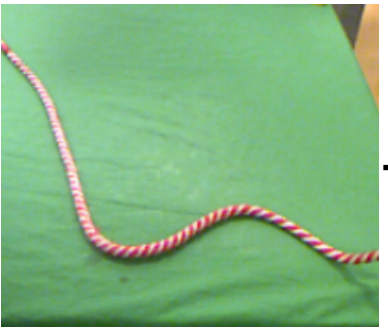
Directly predict what happens next



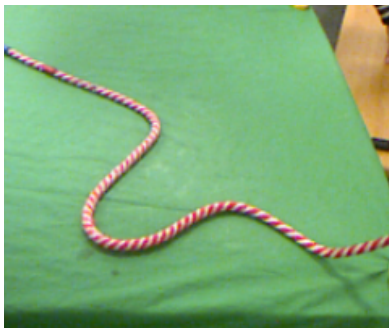
Computer Vision + Simulate Physics



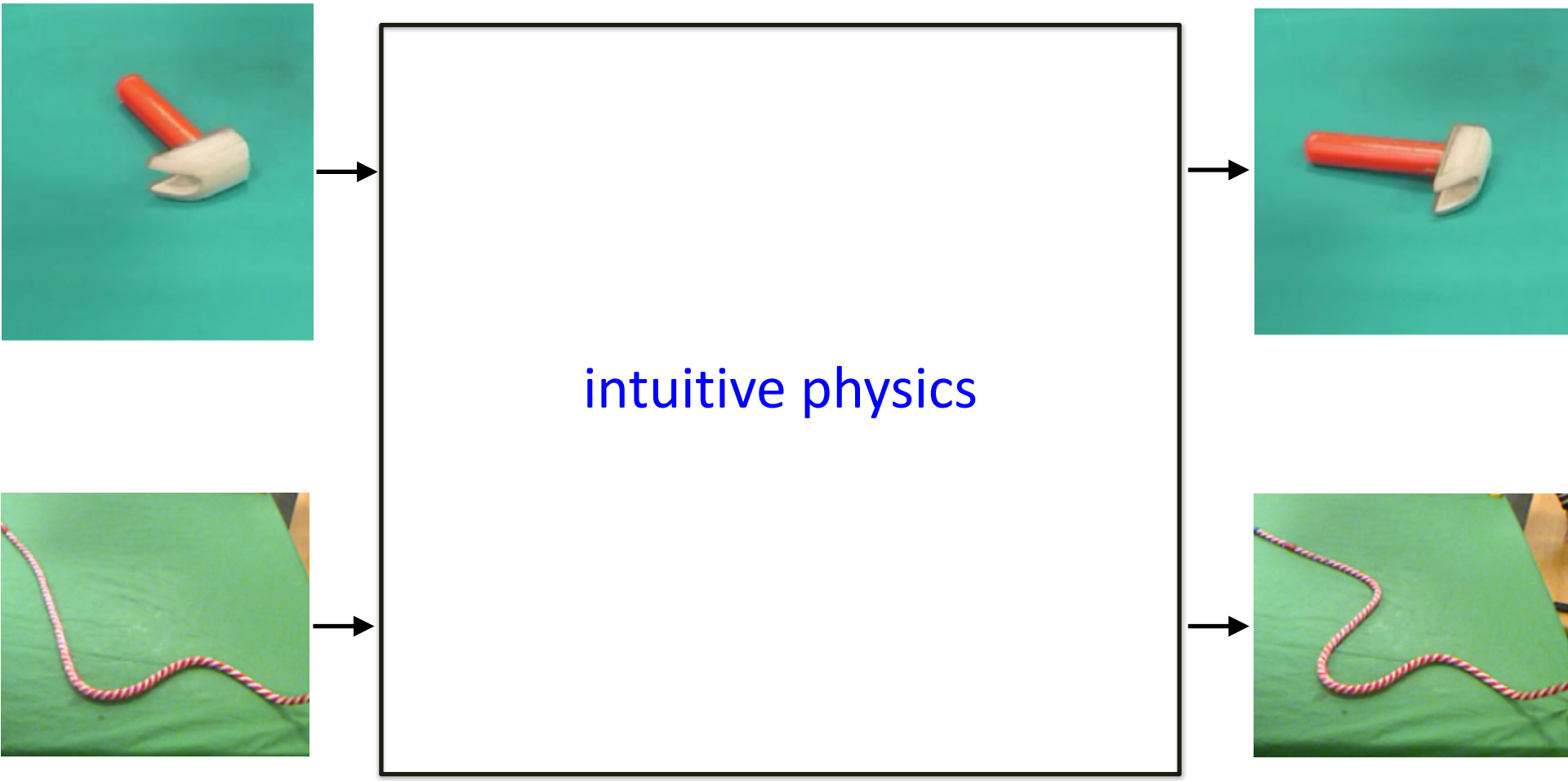
Models that can actually work from pixel inputs



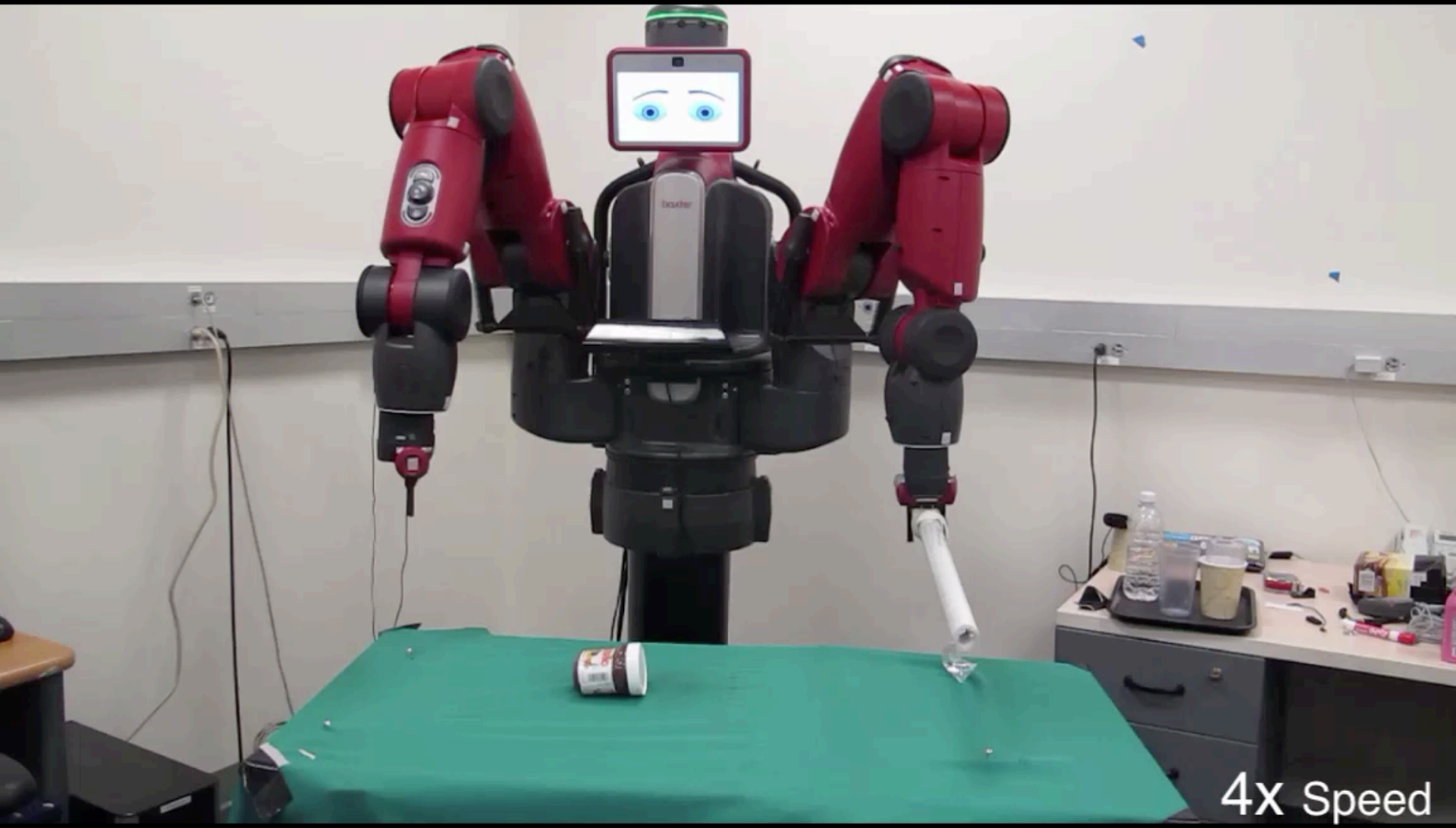
intuitive physics



How to learn such models?



The robot performs experiments by poking objects



Related concurrent efforts in large scale robotic data collection

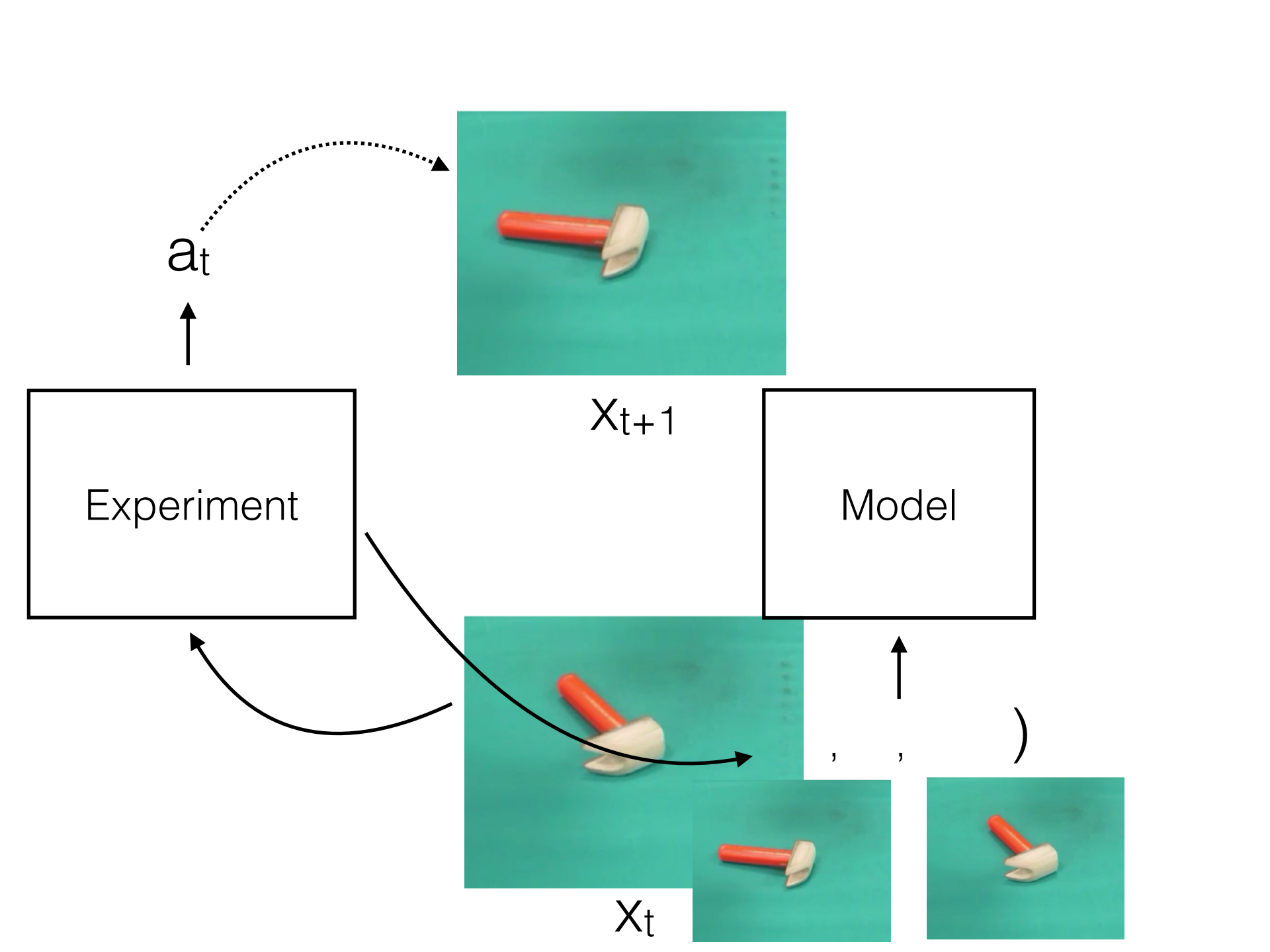
Supersizing Self-Supervision, Pinto et al. 2016 The curious robot, Pinto et al. 2016 Learning hand-eye coordination, Levine et al., 2016

a_t

x_{t+1}

Experiment

x_t


a_t
x_{t+1}
Experiment
Model
x_t
,
,
)

Useful Model: Predict what will happen next

a_t

x_t

Useful Model: Predict what will happen next

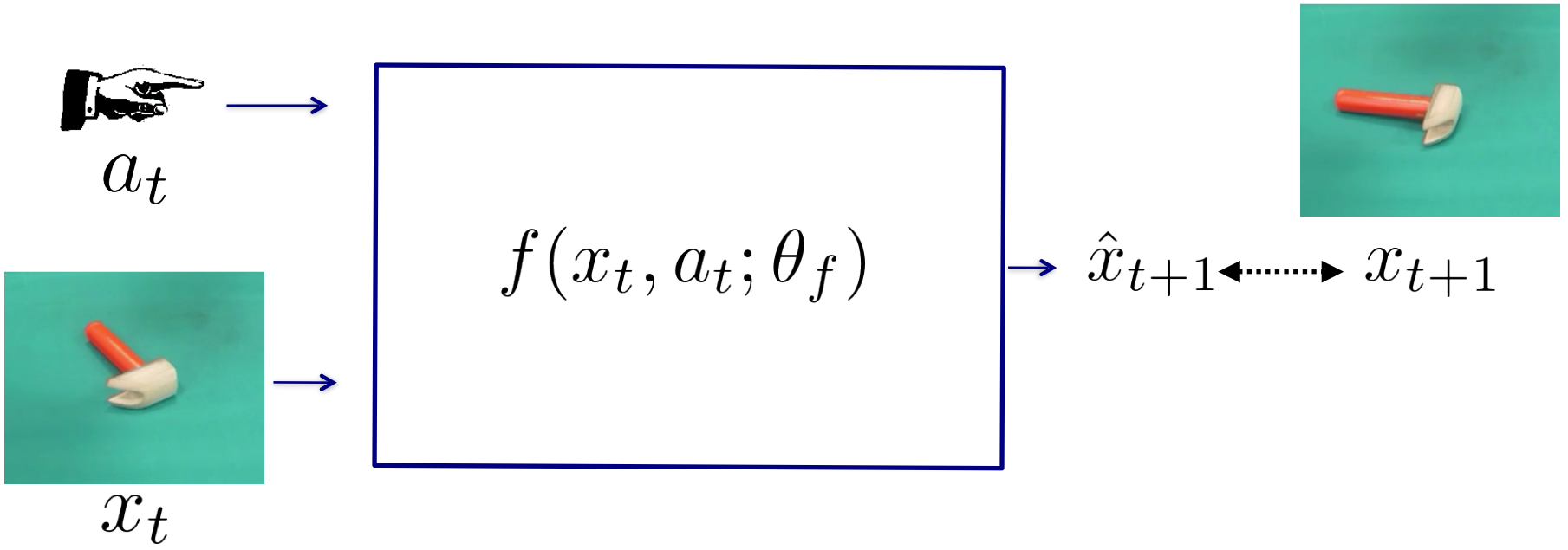
a_t

x_t

$$f(x_t, a_t; \theta_f)$$

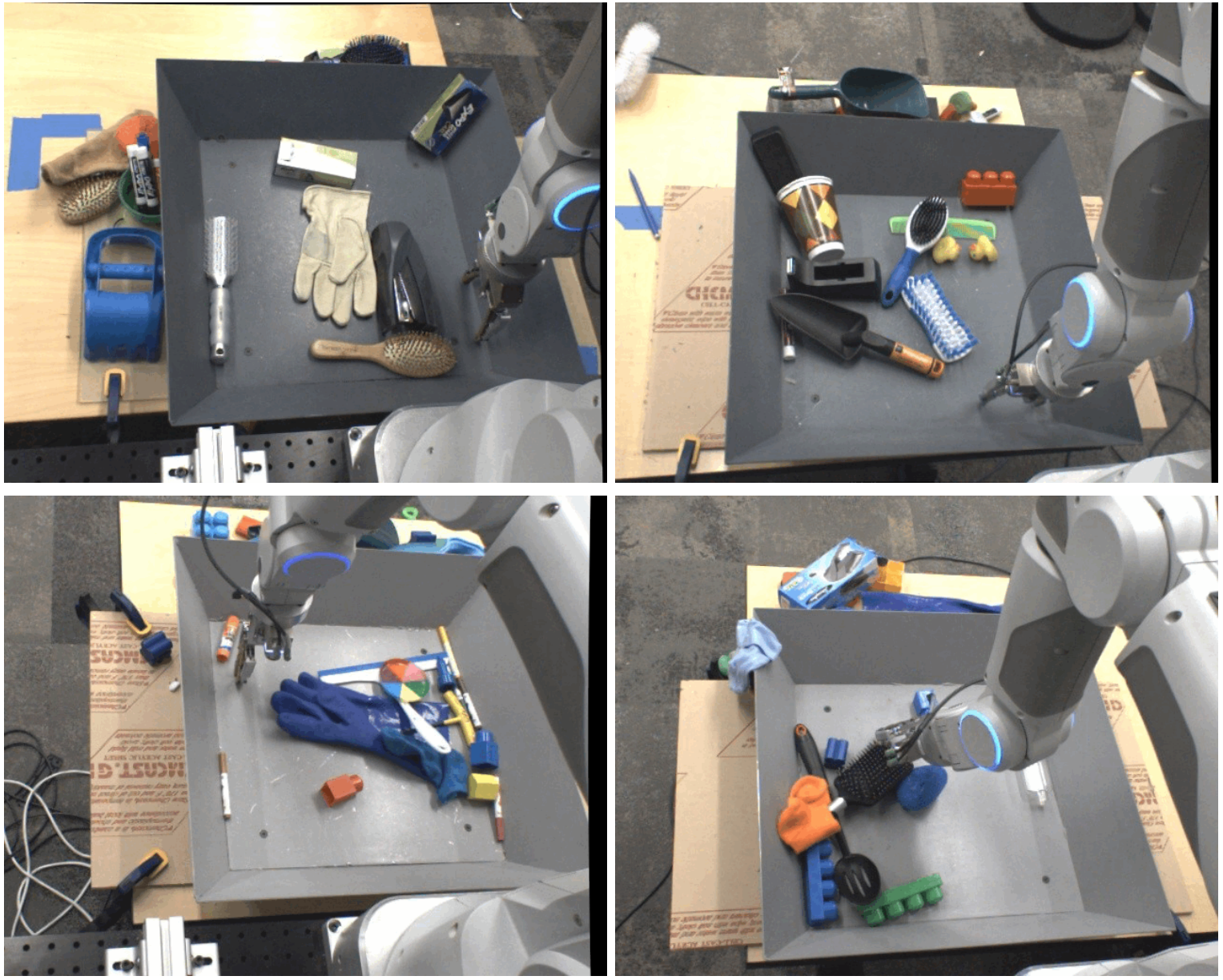
$\hat{x}_{t+1}$ ◀·····▶ x_{t+1}

Useful Model: Predict what will happen next



Forward model in pixel space

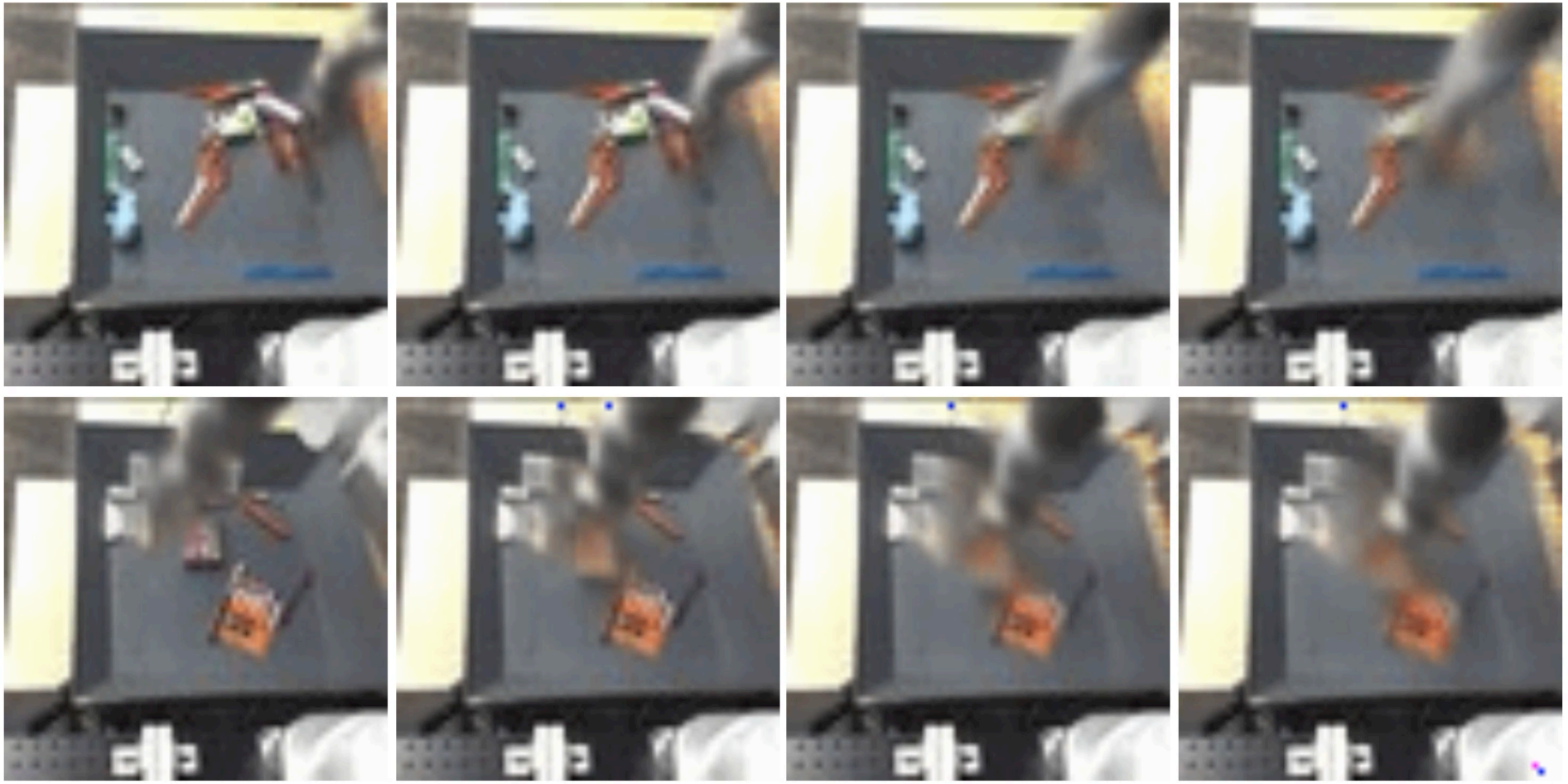
Petrovic et al., 2006	Oh et al., 2015	Xue et al., 2016
Goodfellow et al., 2014	Mathieu et al., 2015	Vondrick et al., 2016
Ranzato et al., 2014	Vondrick et al., 2015	Finn et al., 2017



Unsupervised Learning for Physical Interaction through Video Prediction, Finn et al., 2016

Model Predictions

0x action // 0.5x action // 1x action // 1.5x action



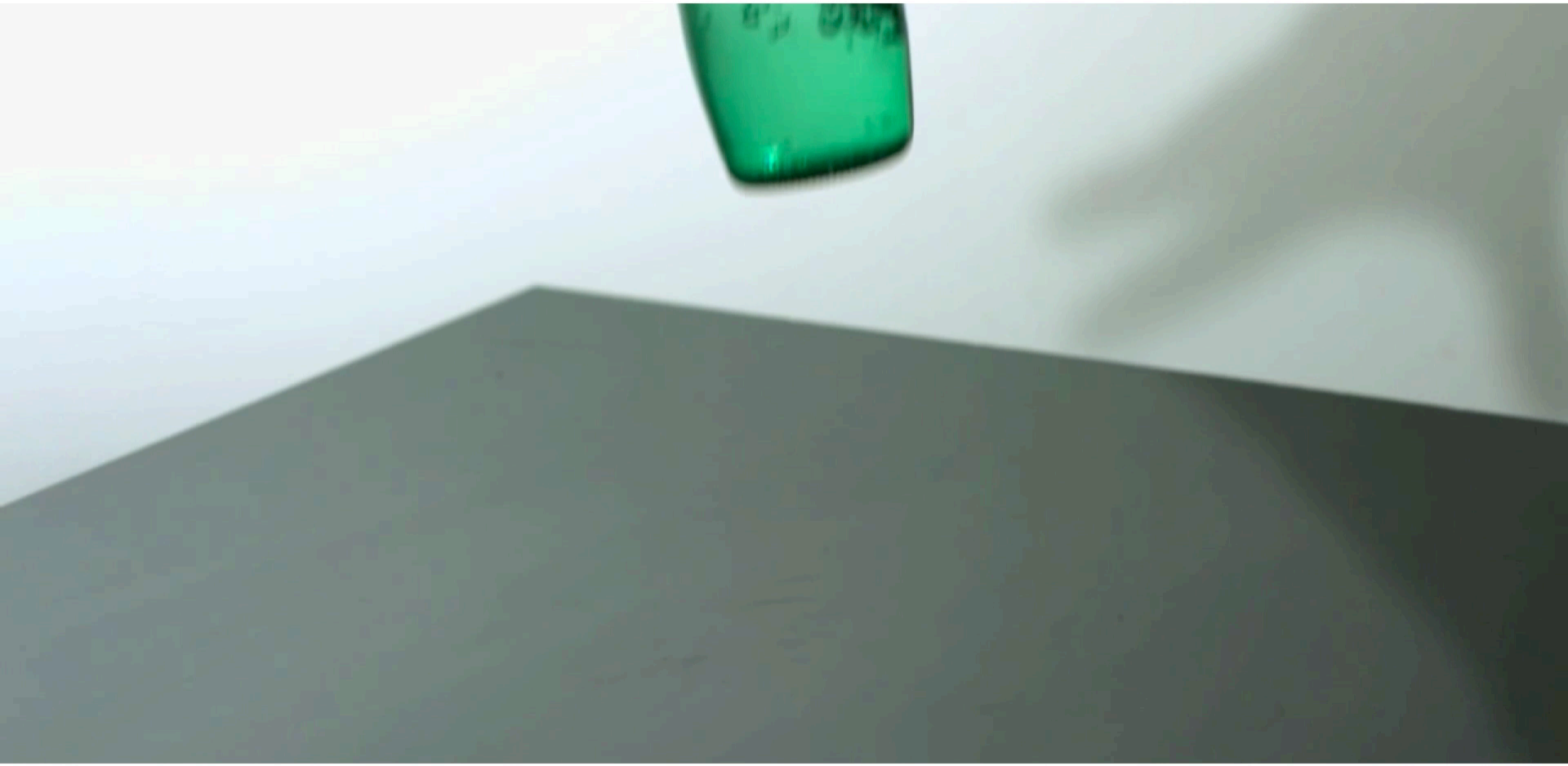
Unsupervised Learning for Physical Interaction through Video Prediction, Finn et al., 2016

Not only hard,

but is this the right model to build ???

Consider a glass bottle

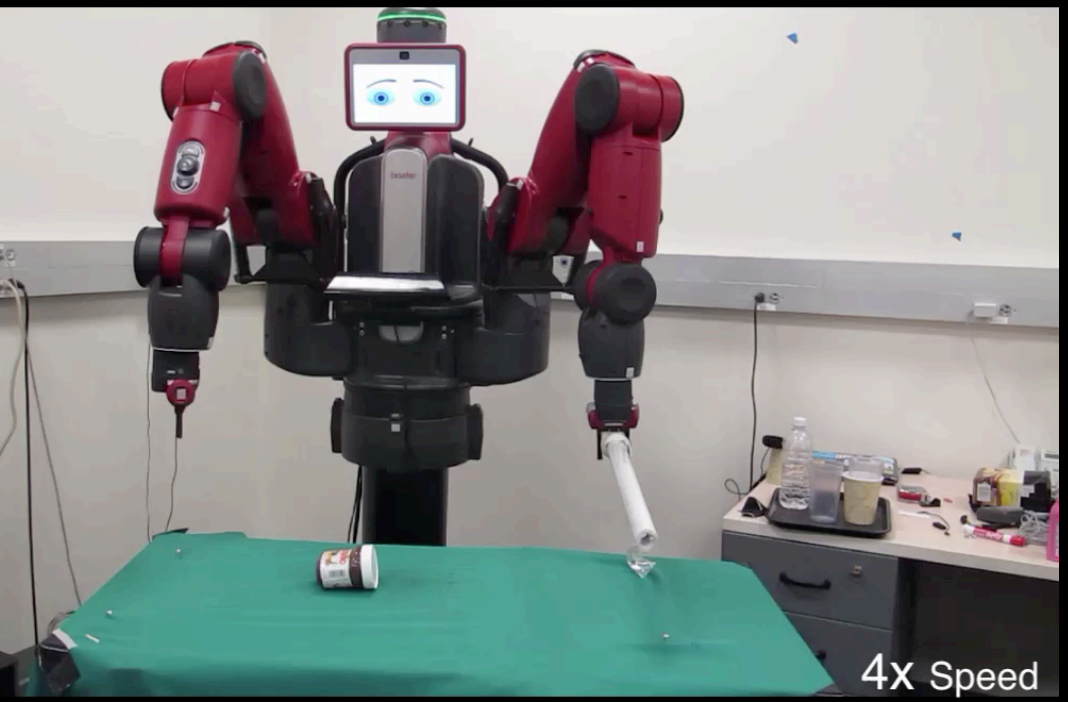
What will happen on dropping the bottle?



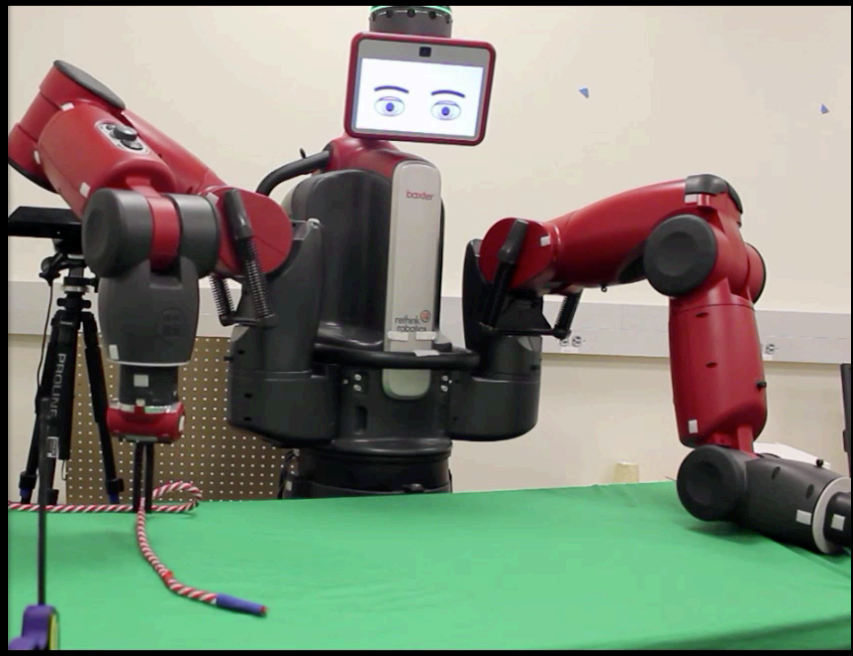
Different Feature Abstractions afford Different Predictions

Easy to predict: bottle breaks
but
Hard to predict: exact location of glass pieces

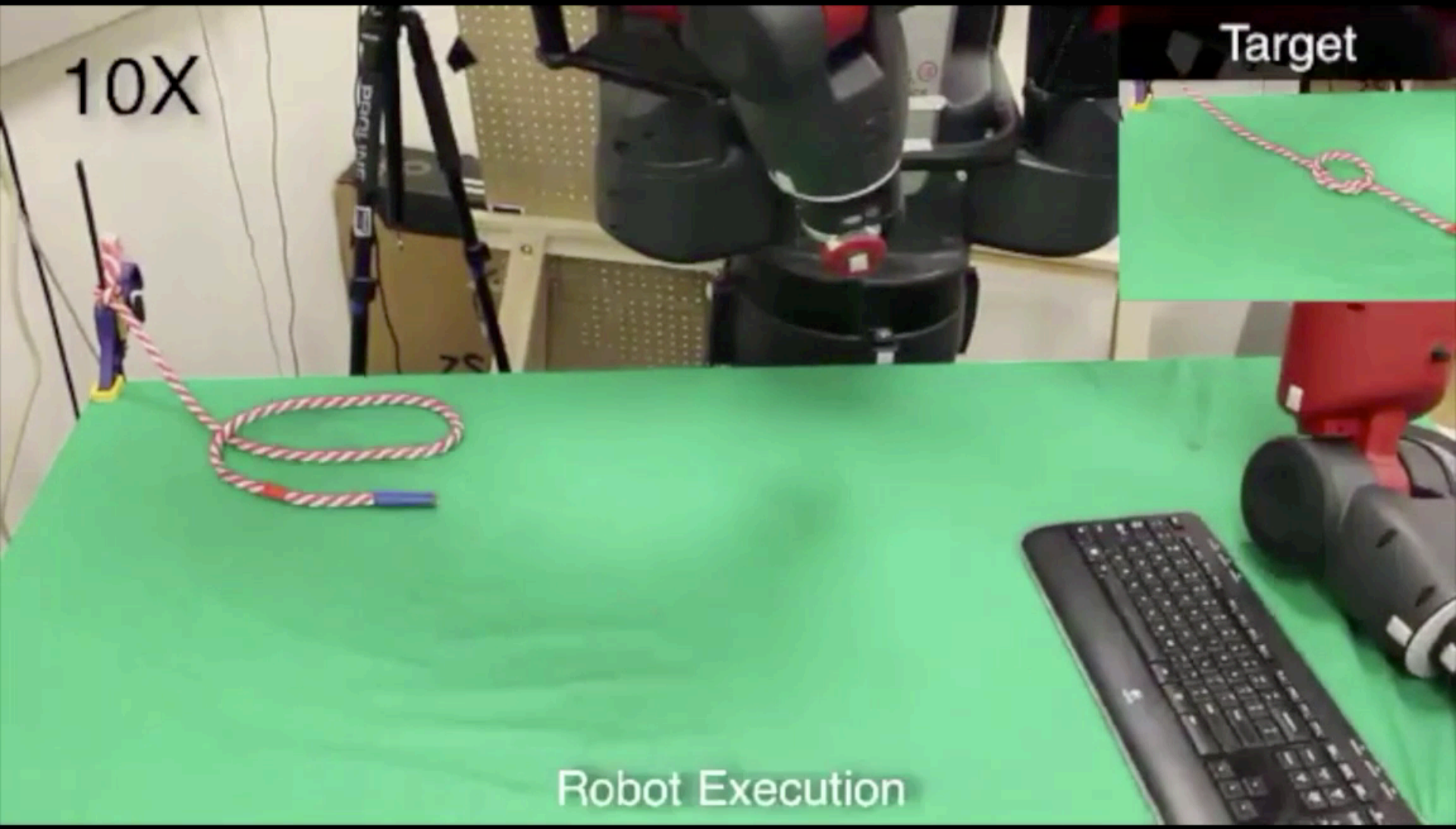
How to learn the appropriate feature space?



Learning About the World



Rope Manipulation

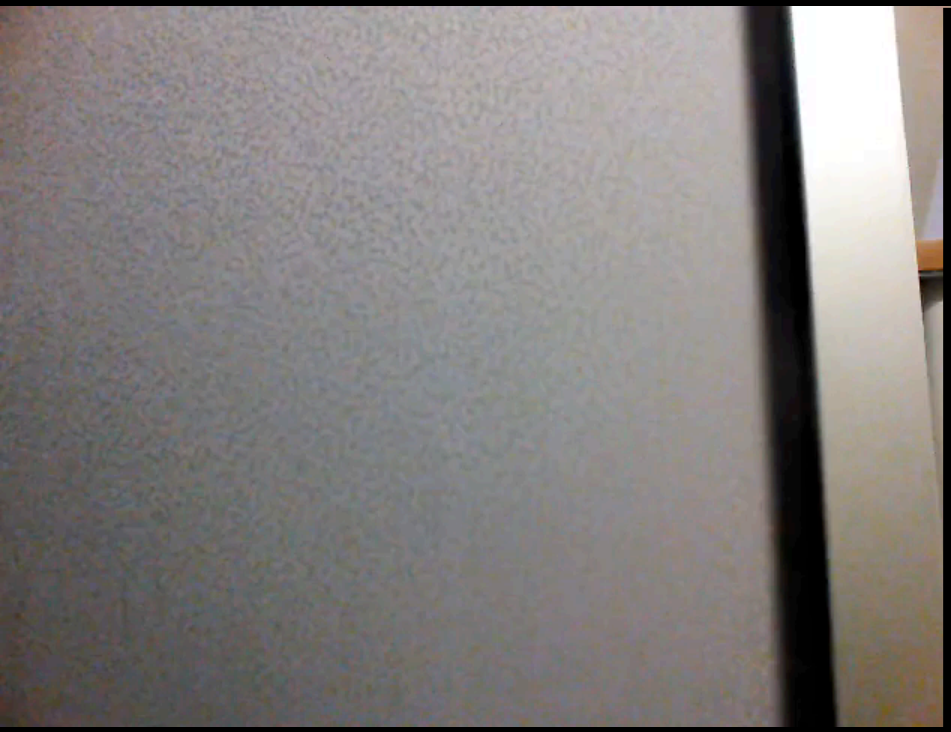


robot gets only RGB images as input!

Combining Self-Supervision and Imitation for Vision Based Rope Manipulation, Ashvin Nair*, Dian Chen*, **Pulkit Agrawal***, Phillip Isola, Pieter Abbeel , Jitendra Malik, Sergey Levine, ICRA 2017 (*equal contribution)

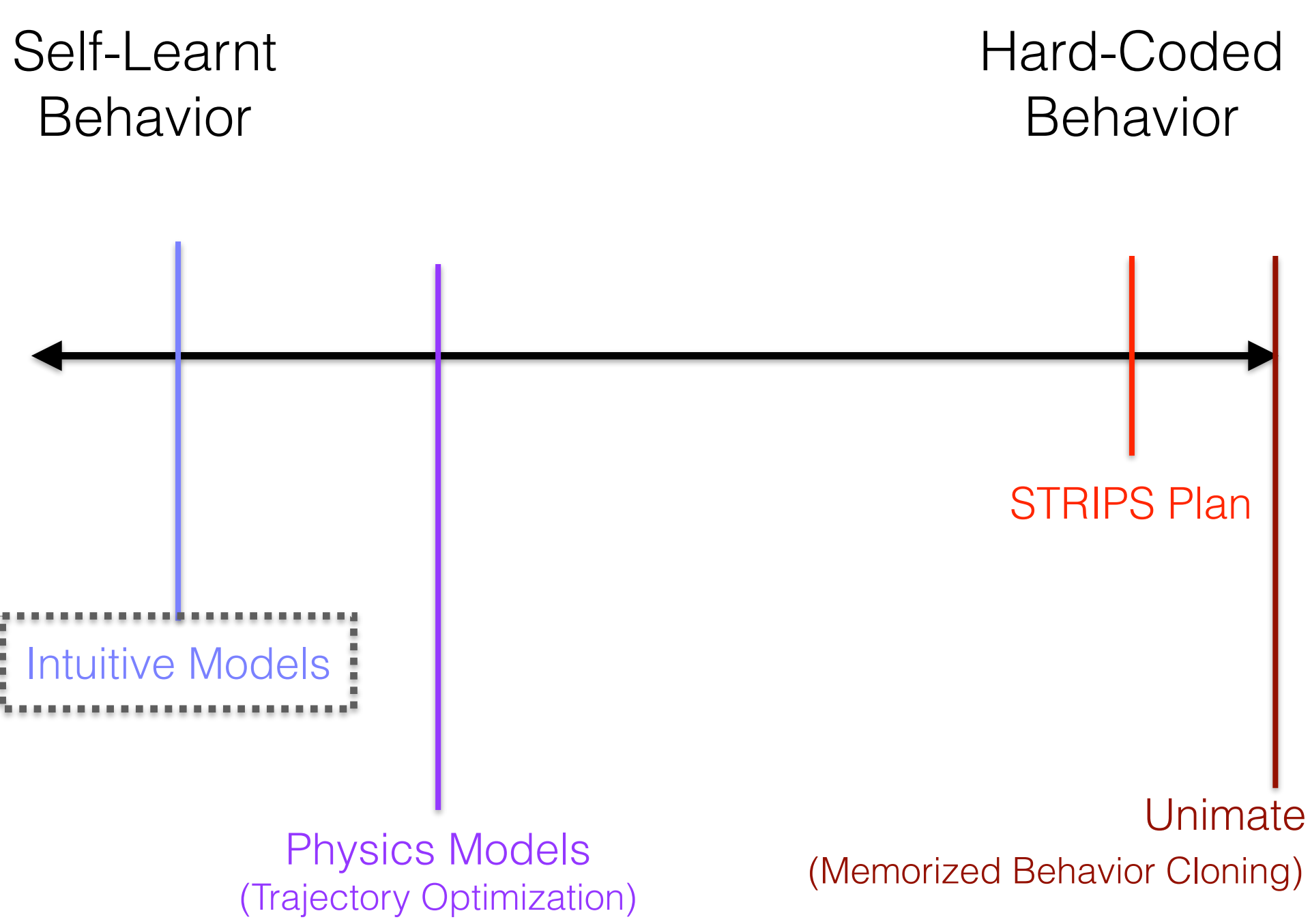
Robot's Emergent Behavior

Current Image

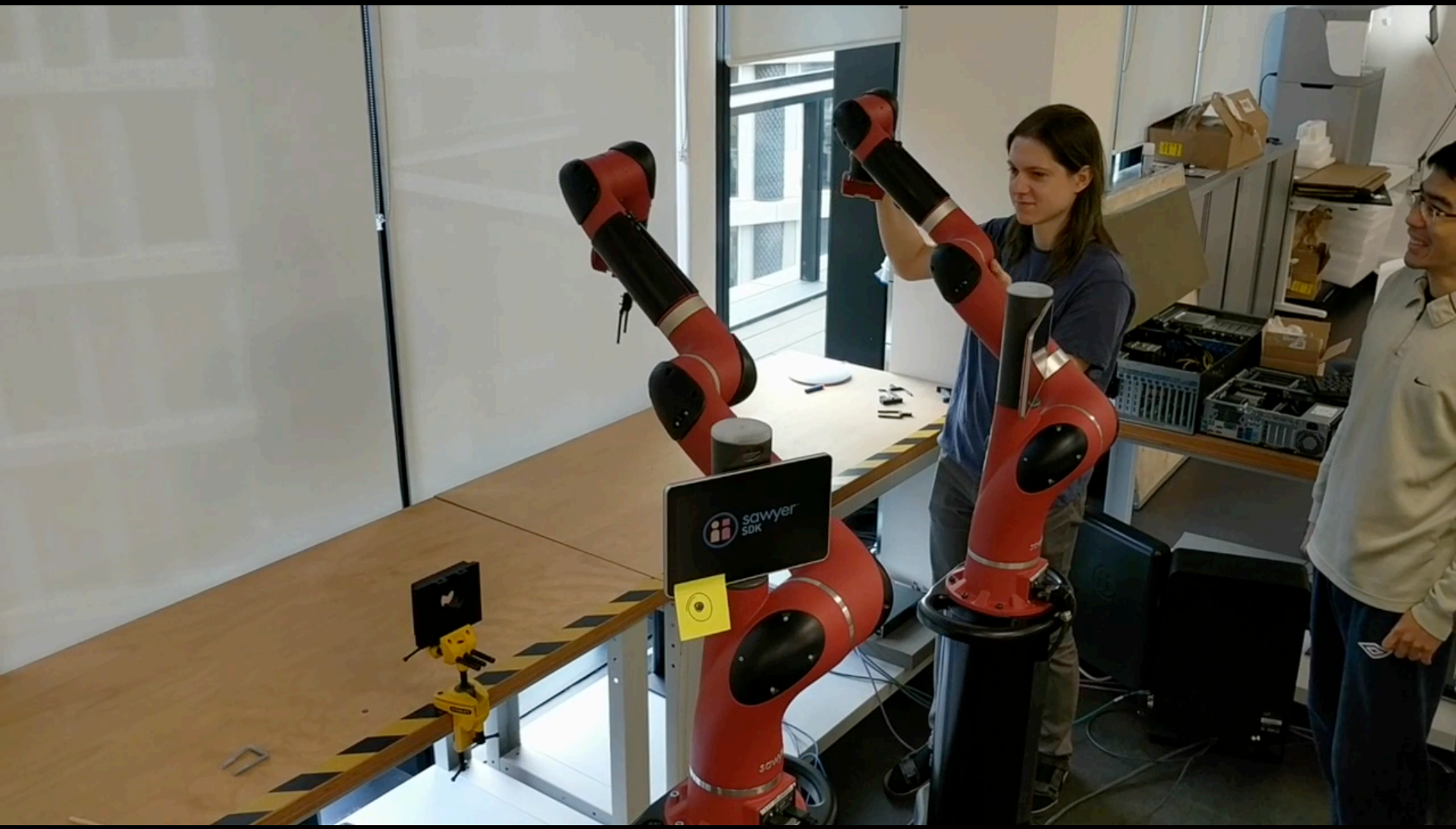


Goal Image

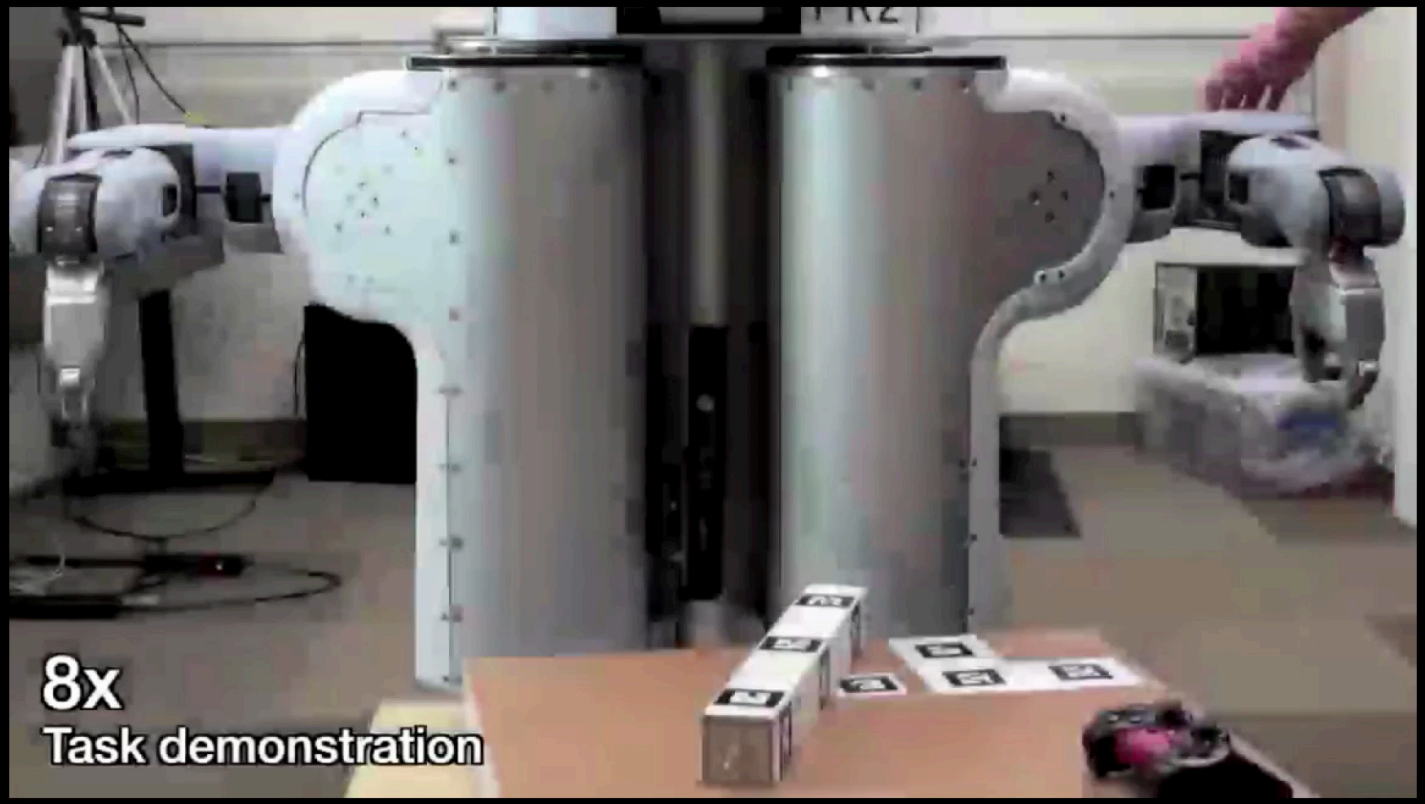
Zero Shot Visual Imitation, Pathak D.*, Mahmoudieh P*., Luo M.*, **Agrawal P. *,** Shentu Y., Chen D., Shelhamer E., Malik J., Darrell, T. (ICLR 2018, *equal contribution)


Self-Learnt Behavior
Hard-Coded Behavior
STRIPS Plan
Intuitive Models
Physics Models
(Trajectory Optimization)
Unimate
(Memorized Behavior Cloning)

DDPG with Demonstrations (DDPGfD)



Leveraging Demonstrations for Deep Reinforcement Learning on Robotics Problems with Sparse Rewards, 2018

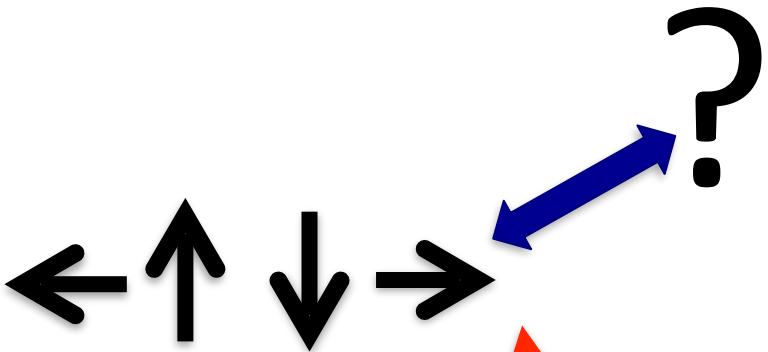


$$x_t, a_t, x_{t+1}, a_{t+1}, \ldots, a_{T-1}, x_T$$

Visual Observation

Robot's Action

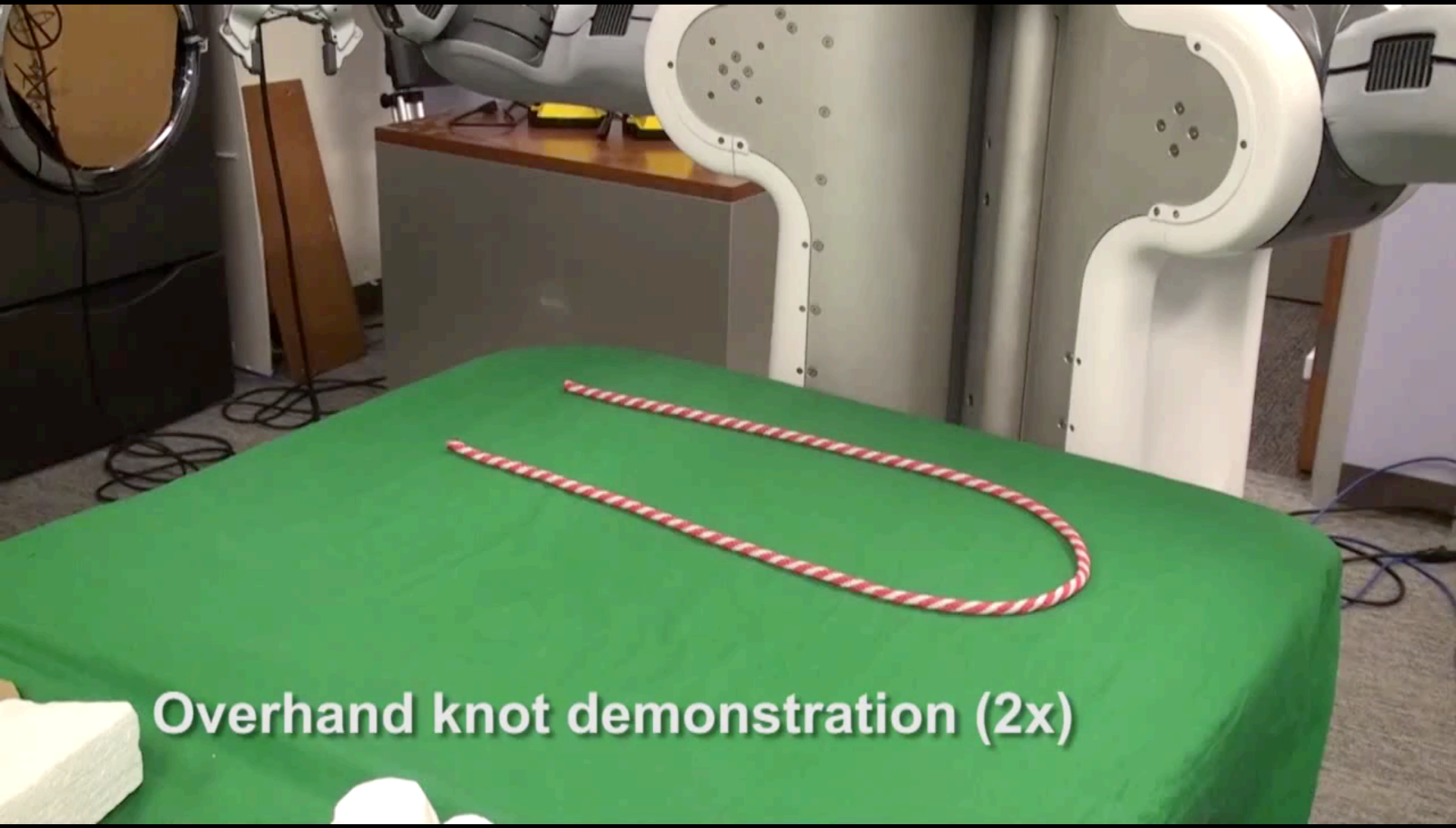
Video Credits: Scott Niekum, Amherst



How to get the
"action" label?

Behavior Cloning

Towards Generalization



Learning from Demonstrations Through the Use of Non-Rigid Registration, Schulman et al., ISRR 2013

Old Task

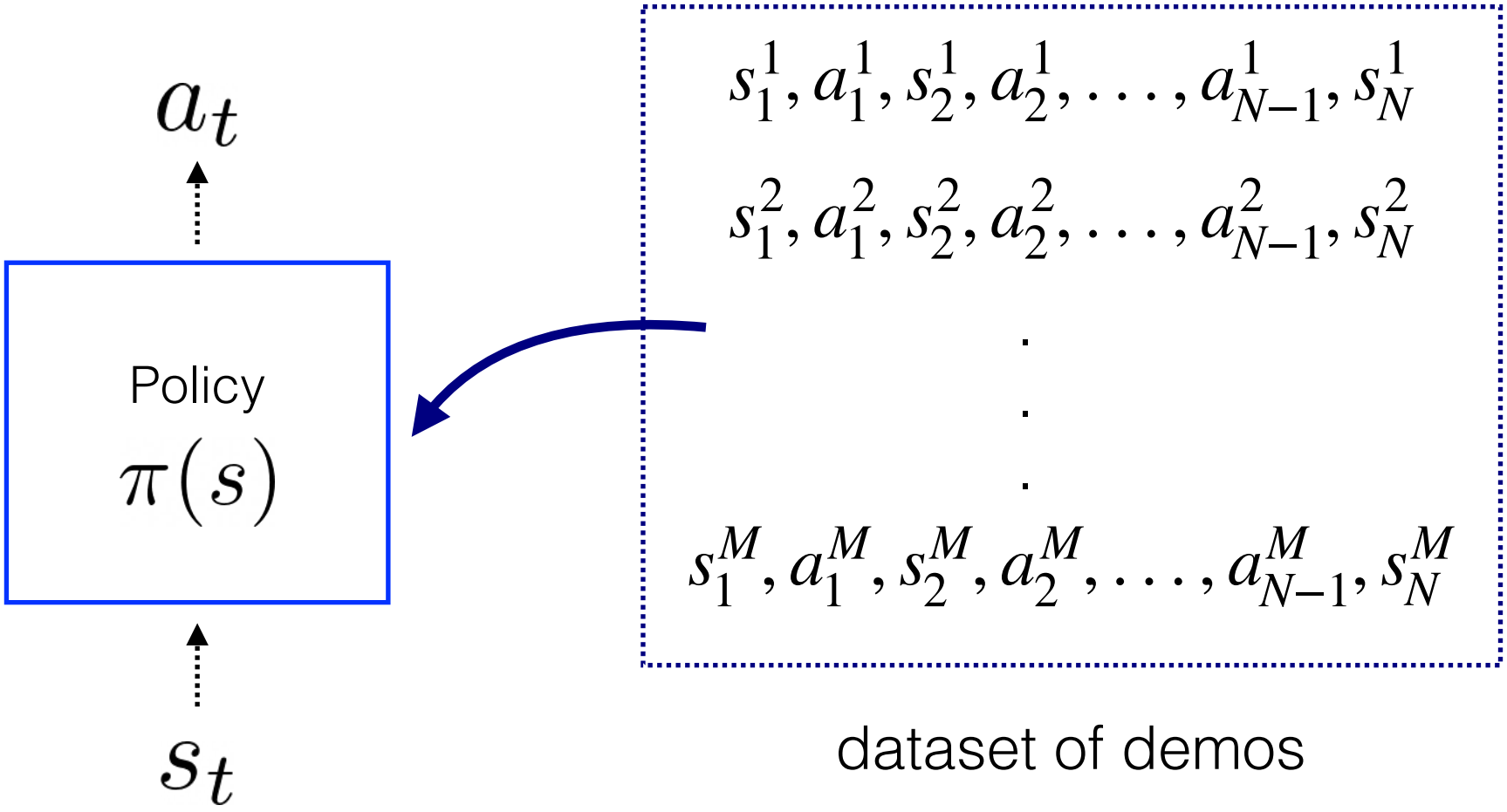
New Task

from demo

need to infer actions

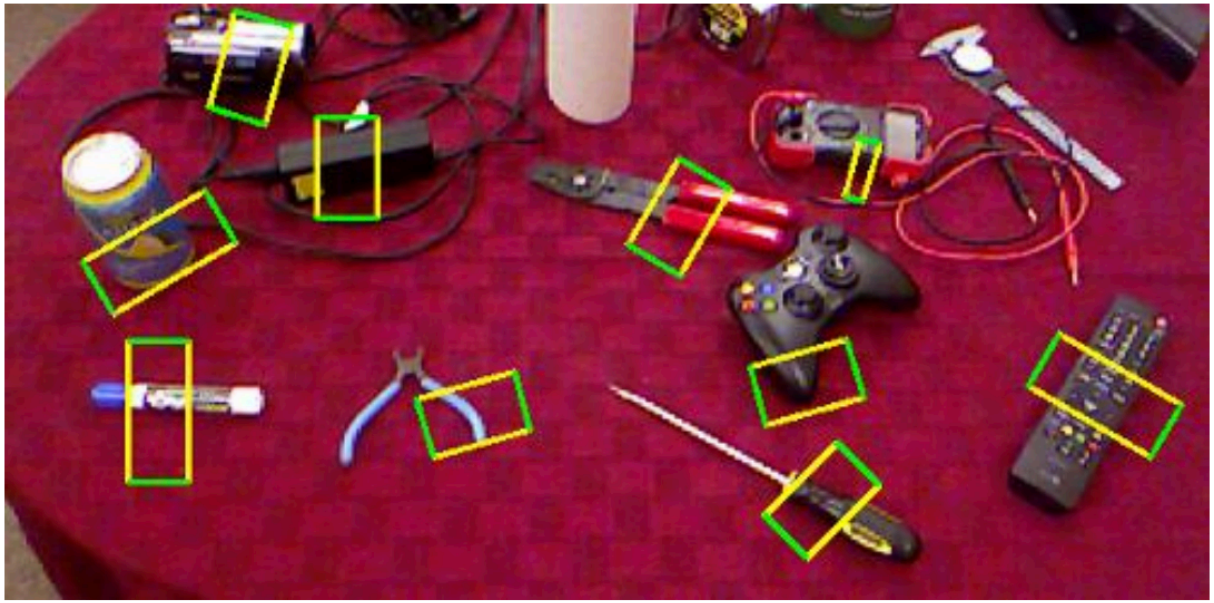
Learning from Demonstrations Through the Use of Non-Rigid Registration, Schulman et al., ISRR 2013

Performing such Alignment can be Non-Trivial



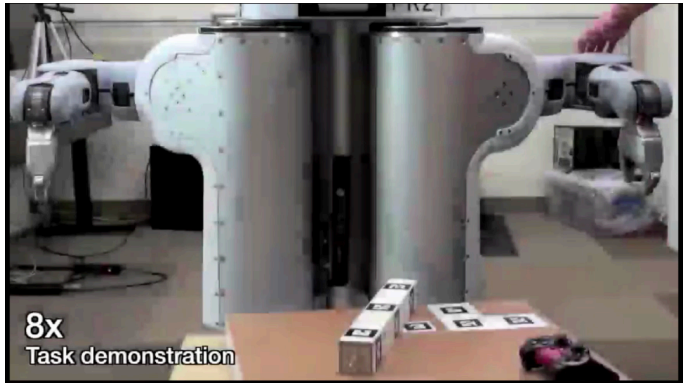
$s_t^{M+1} \rightarrow a_t$???

Grasping by Collecting Human Data



Self-Supervised Learning automates this process (more later)

Self-Learnt Behavior



Hard-Coded Behavior

Intuitive Models

Physics Models
(Trajectory Optimization)

Behavior Cloning
(Learning from Demonstrations)

STRIPS Plan

Unimate
(Memorized Behavior Cloning)

Behavior cloning can be tedious

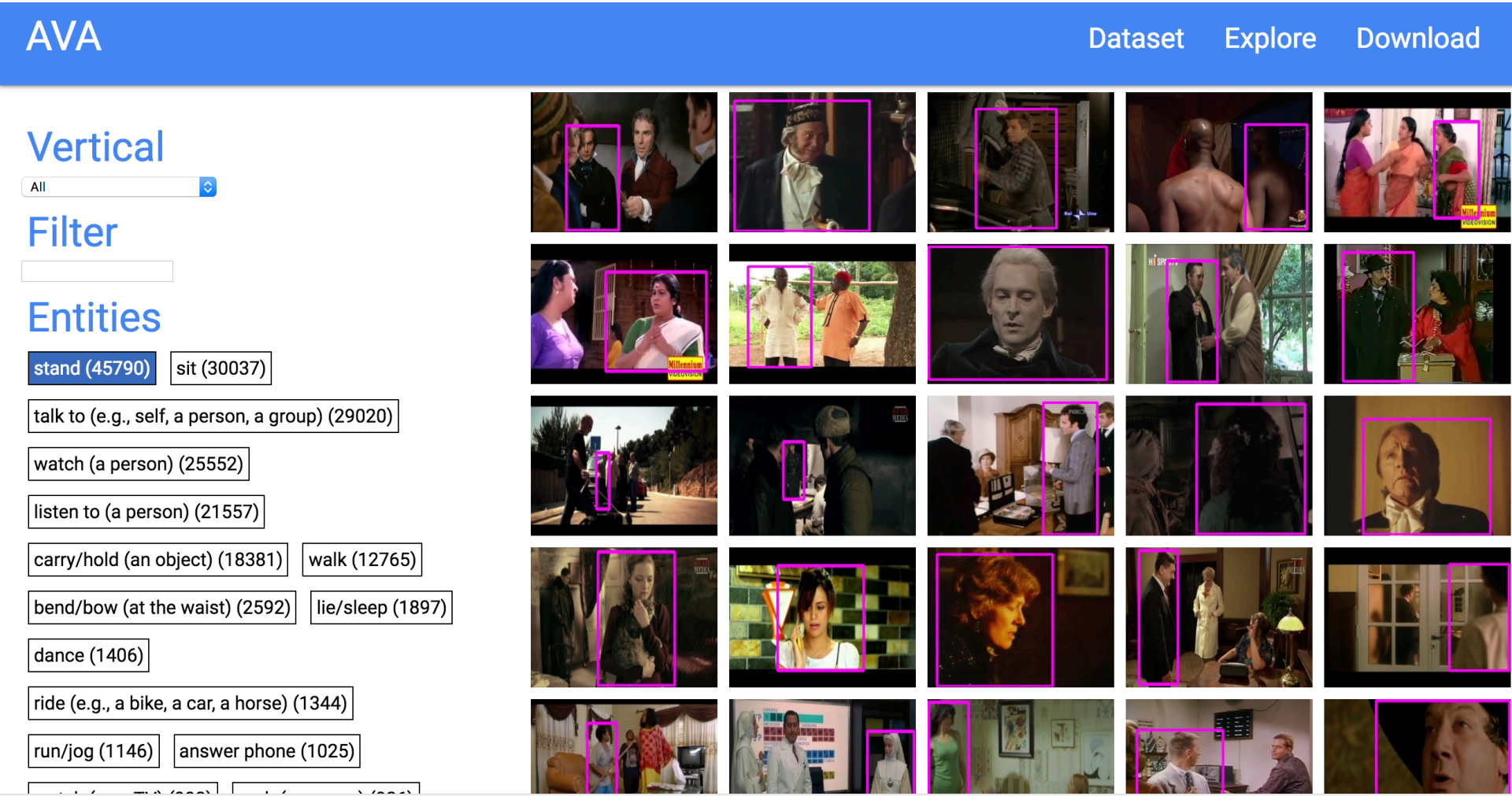
BUT


MIND MELD
First-Person Teleoperation
Zoe McCarthy
UC Berkeley, California
Pieter Abbeel, Ken Goldberg

Behavior cloning can be tedious

BUT

AVA Dataset

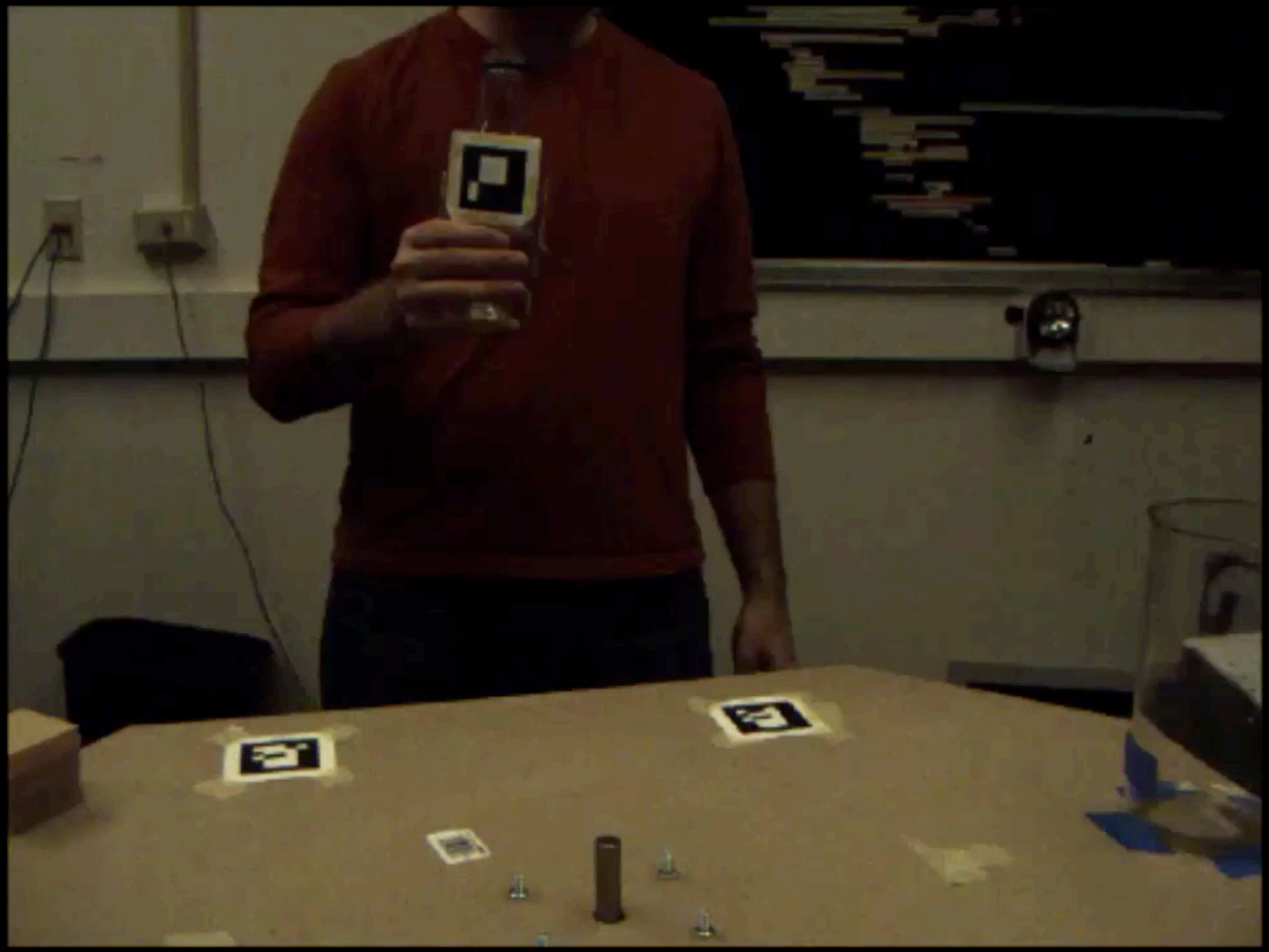


Gu et al., CVPR 2018

VLOG Dataset

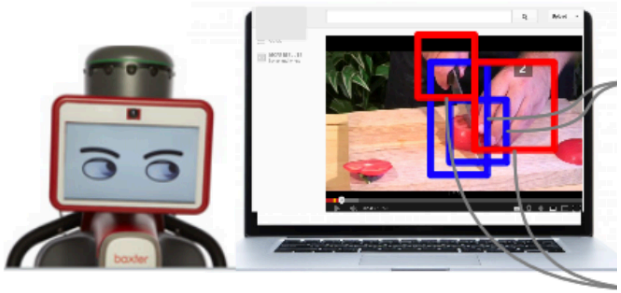
Fouhey et al., CVPR 2018

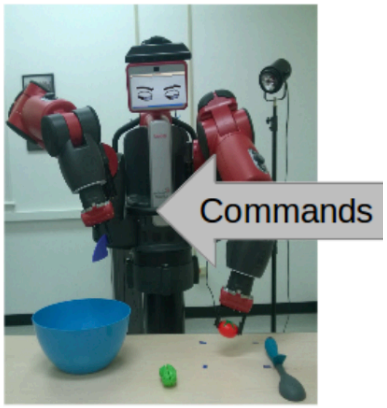
Learning by Imitating



Putting the computer vision hat

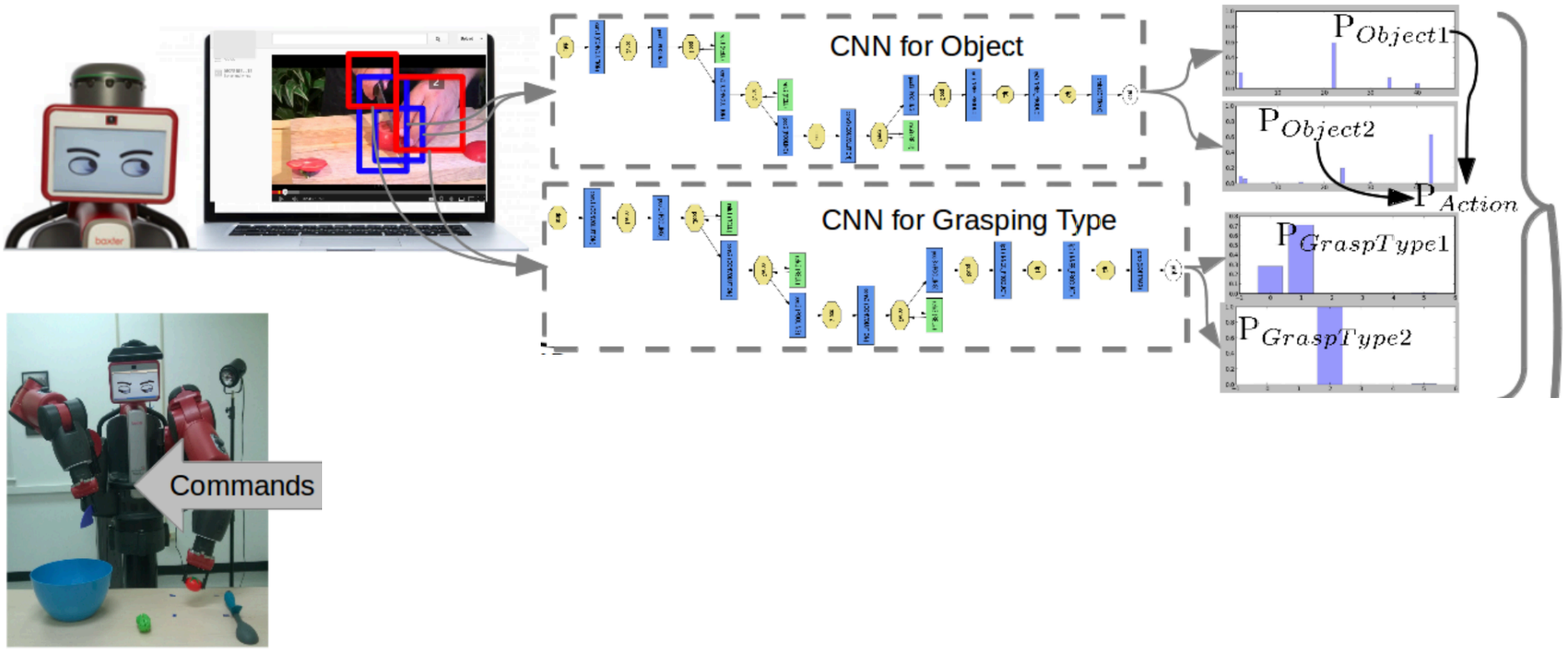
Putting the computer vision hat





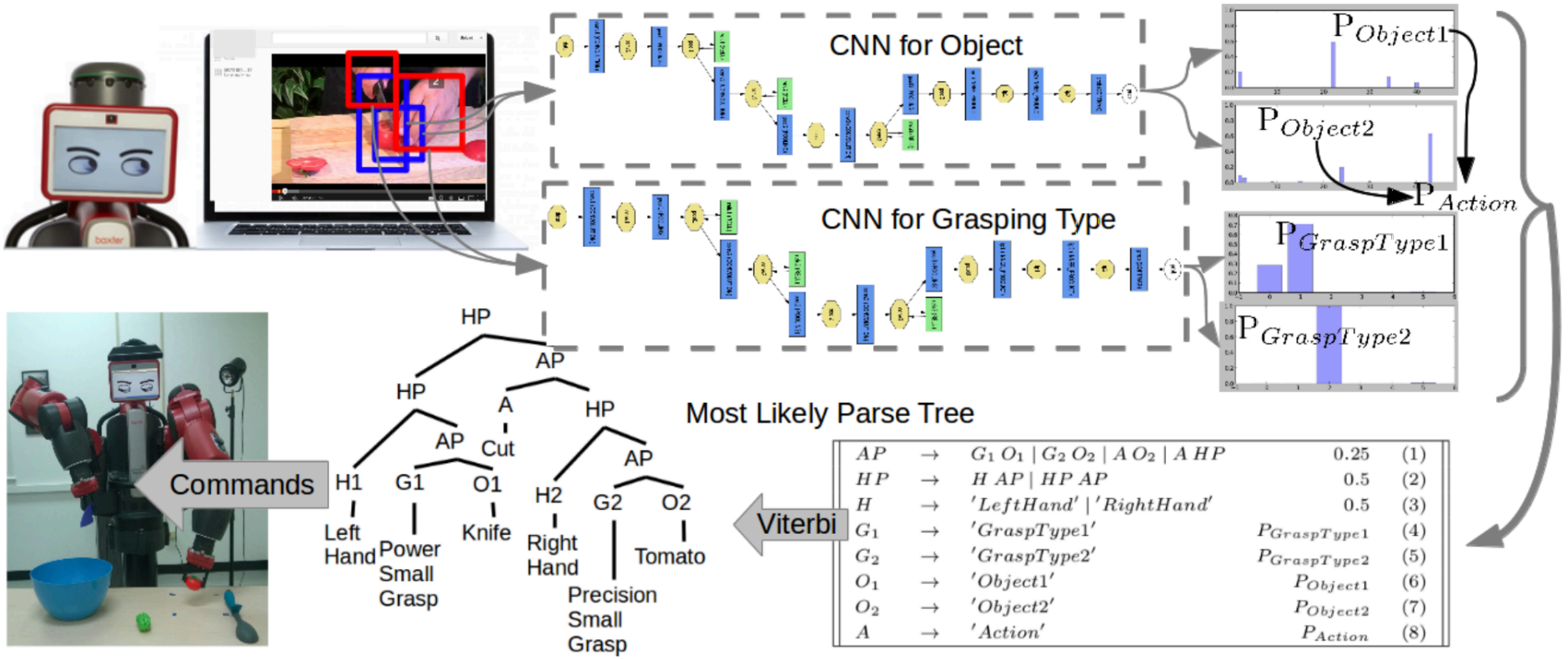
Robot Learning Manipulation Action Plans by “Watching” Unconstrained Videos from the World Wide Web
Yang et al. 2015

Putting the computer vision hat



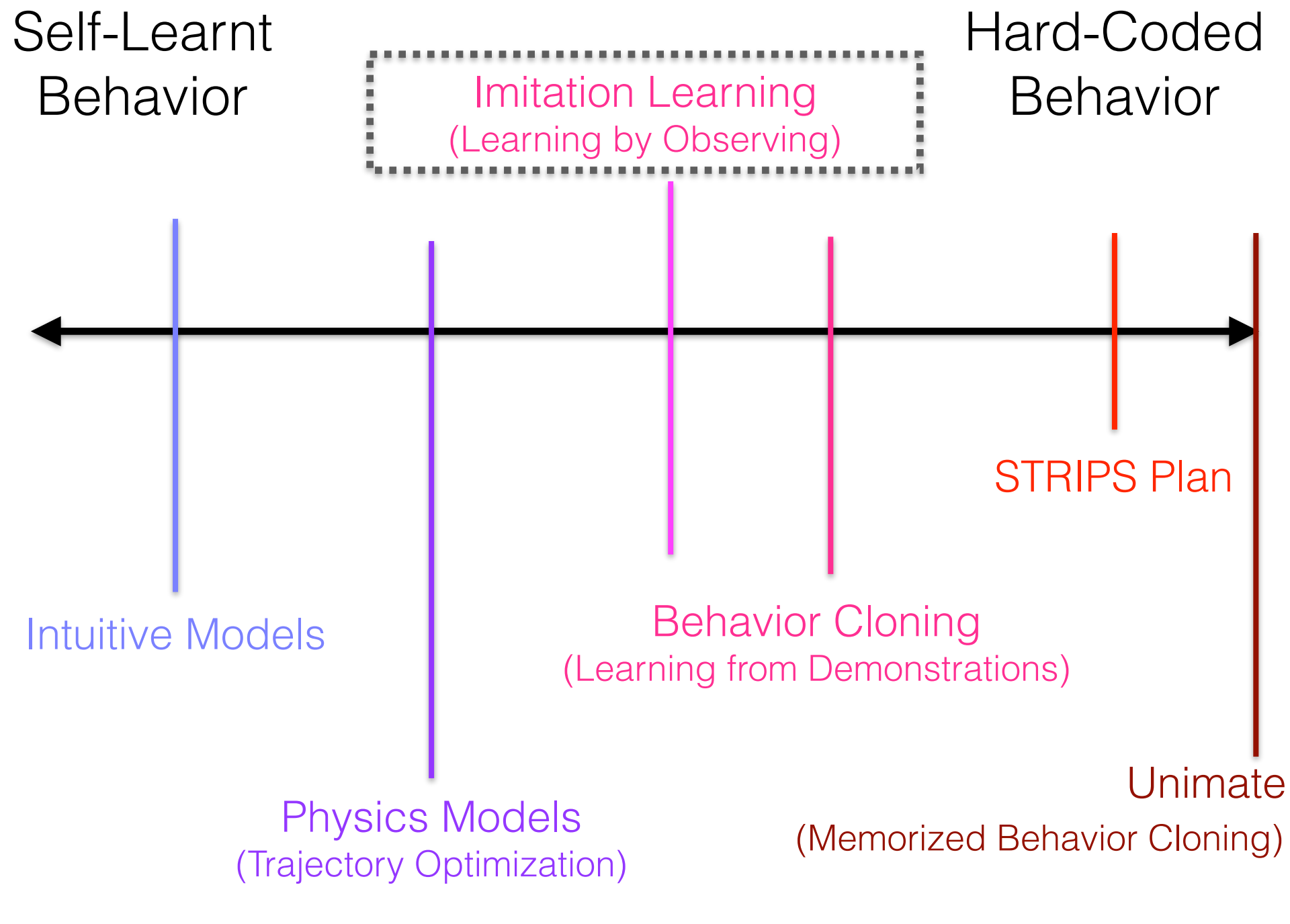
Robot Learning Manipulation Action Plans by “Watching” Unconstrained Videos from the World Wide Web
Yang et al. 2015

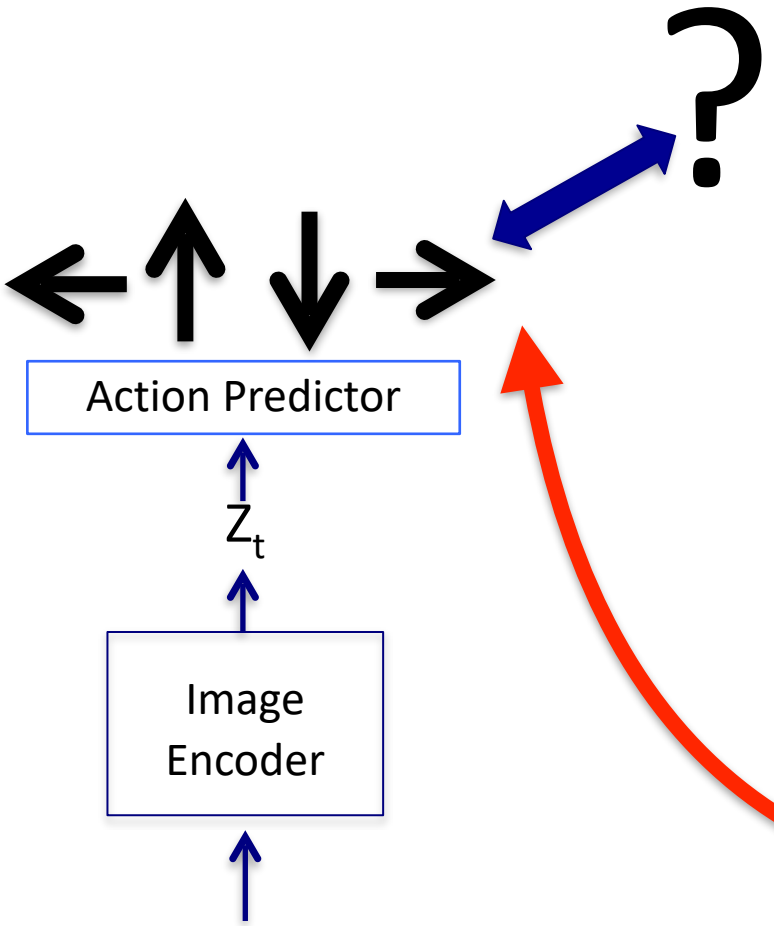
Putting the computer vision hat



Robot Learning Manipulation Action Plans by “Watching” Unconstrained Videos from the World Wide Web

Yang et al. 2015


Self-Learnt Behavior
Hard-Coded Behavior
Imitation Learning
(Learning by Observing)
Intuitive Models
Physics Models
(Trajectory Optimization)
Behavior Cloning
(Learning from Demonstrations)
STRIPS Plan
Unimate
(Memorized Behavior Cloning)



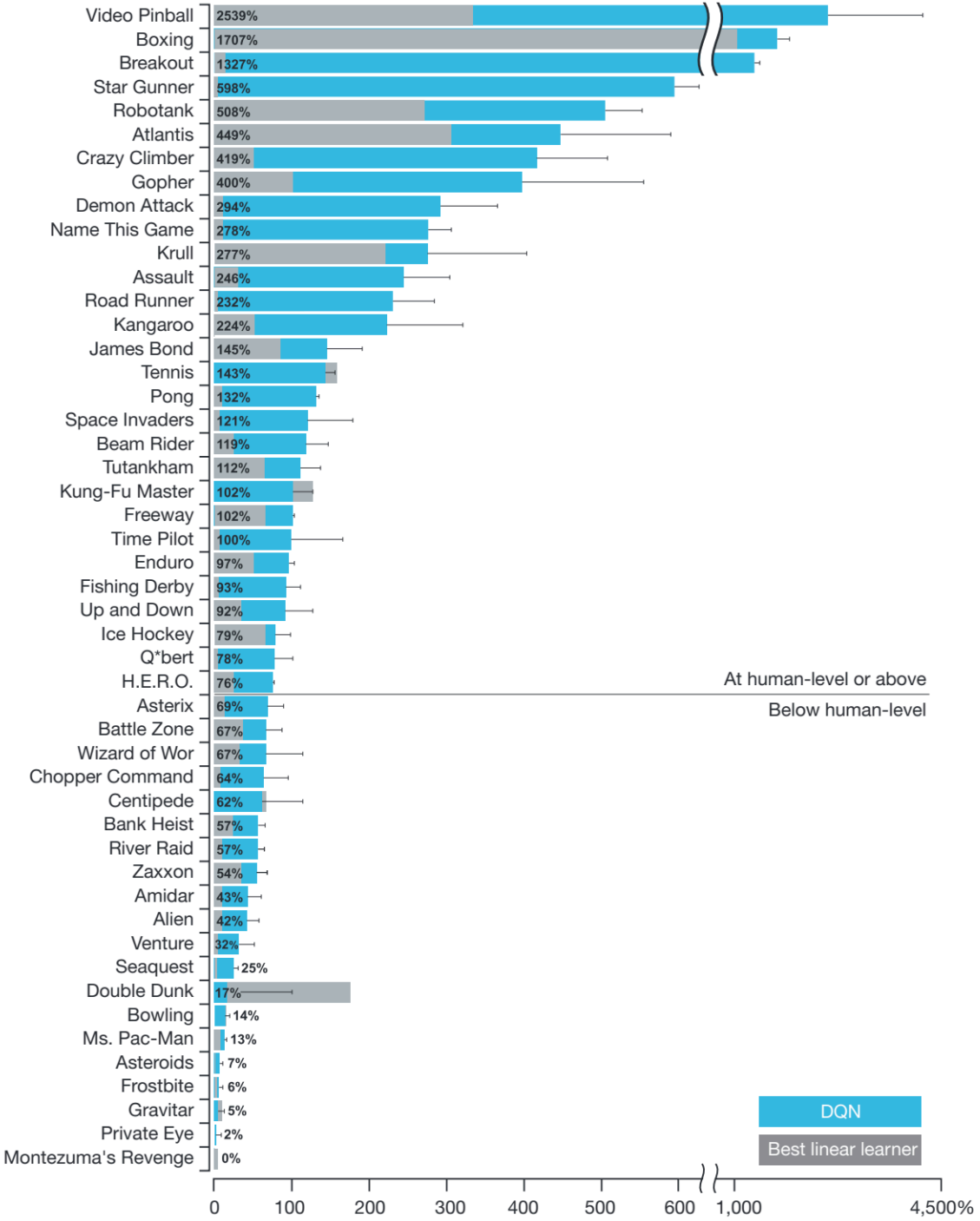
How to get the
"action" label?

Behavior Cloning

Hard to achieve

Super-Human Performance

with behavior cloning



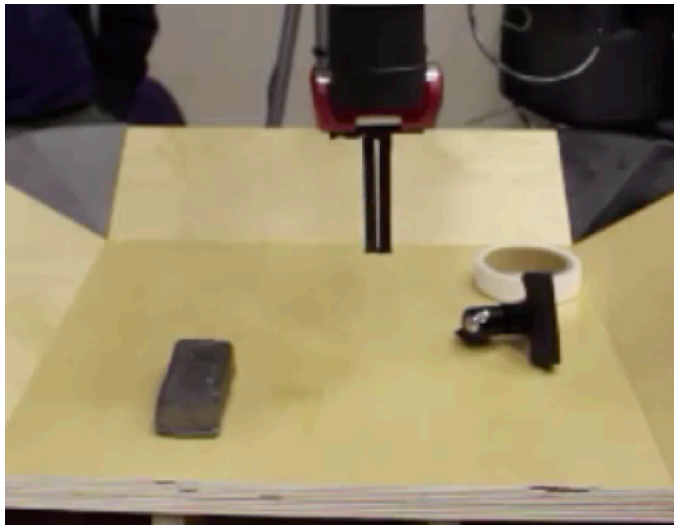
Hard to achieve

Super-Human Performance

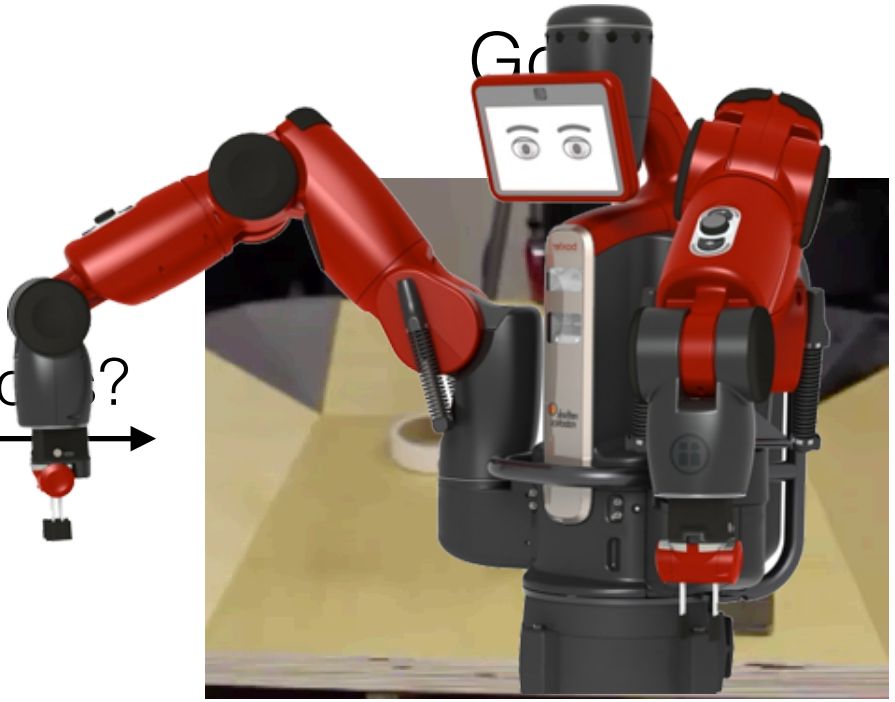
with
behavior cloning

Let the machine automatically figure out decision making rules!

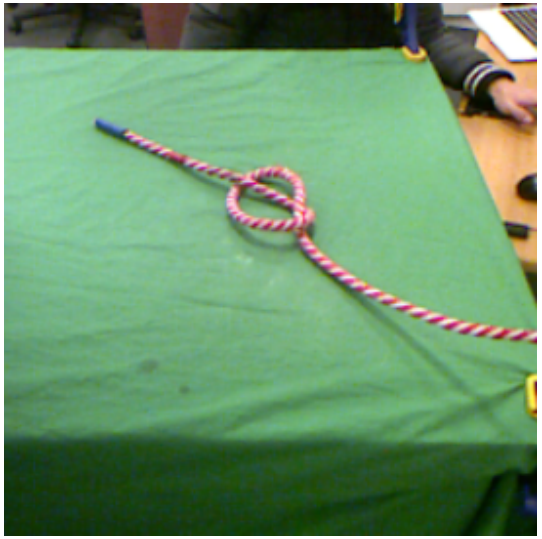
Current Observation



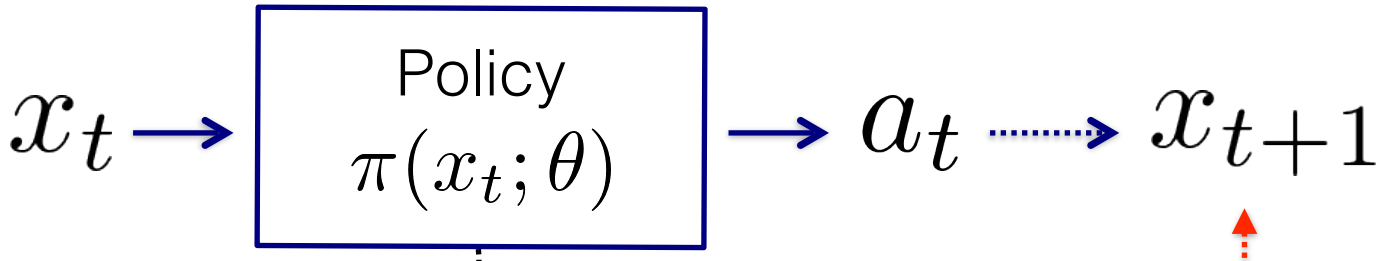
Actio?



Actions?



Brief Overview of Reinforcement Learning



initially random

(how to learn this?)

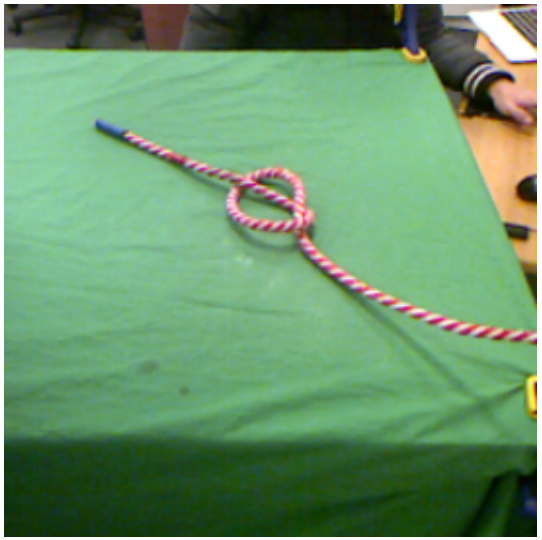
$$\max_\theta \mathbb{E}\left(\sum_{t=1}^{T} r_t\right)$$

r_{t+1}

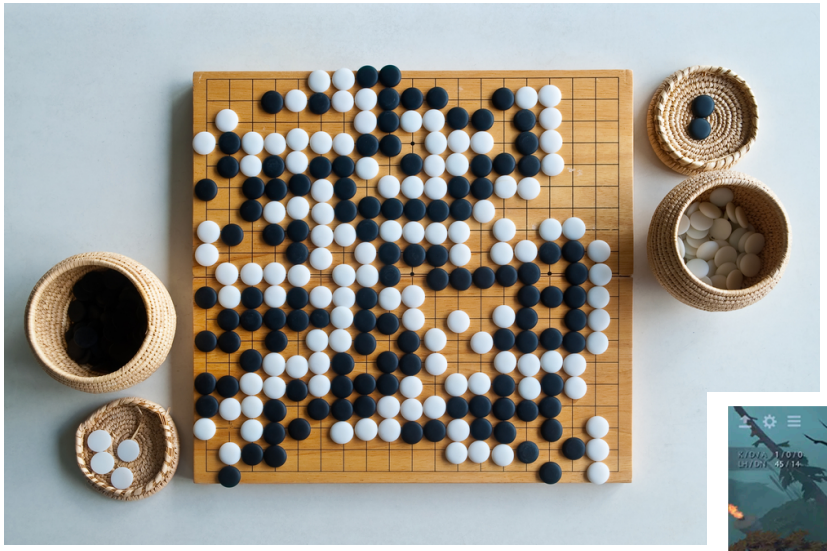
x_t

Actions?

x_G

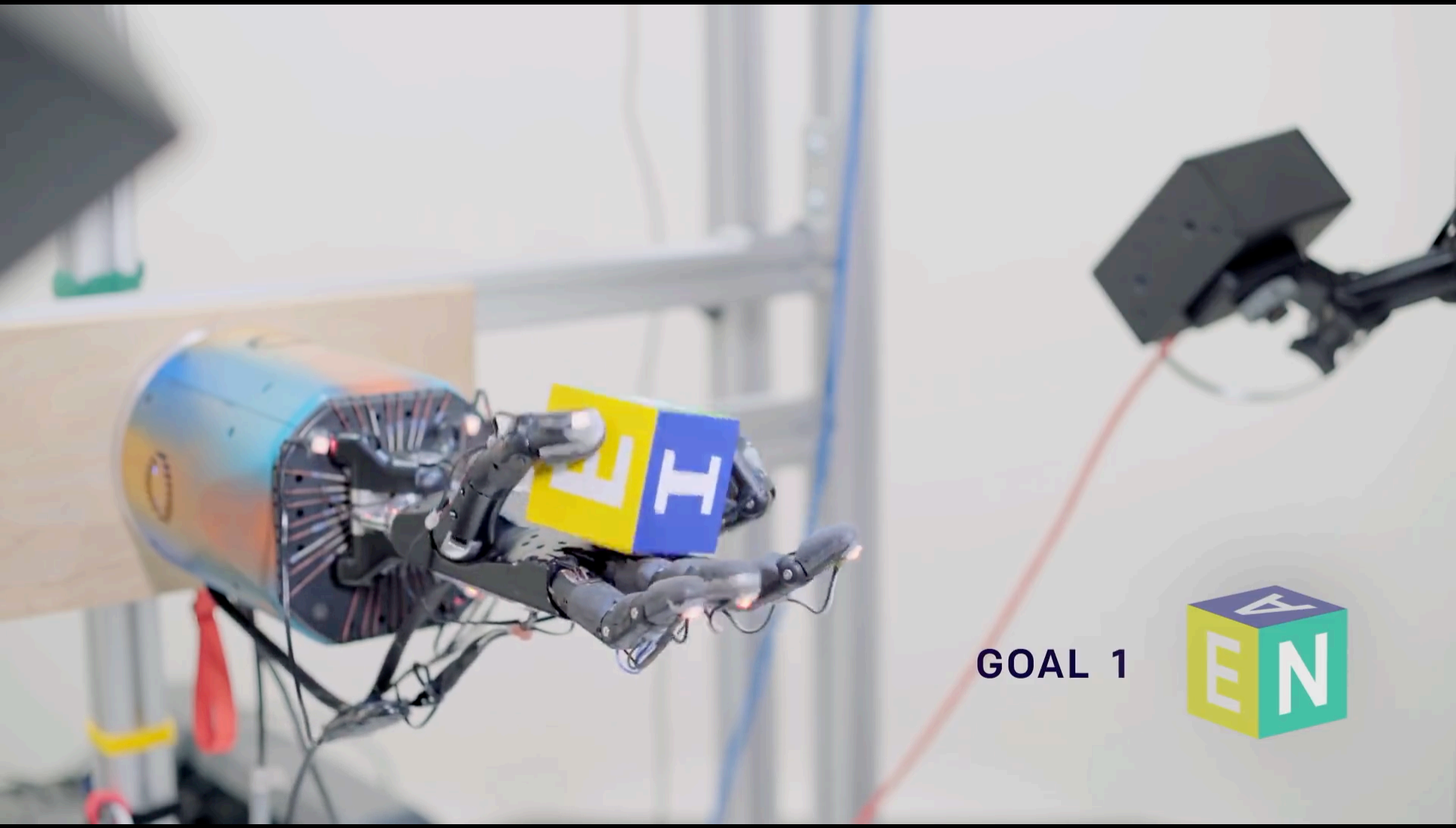


AlphaGo



Open AI Five playing DOTA

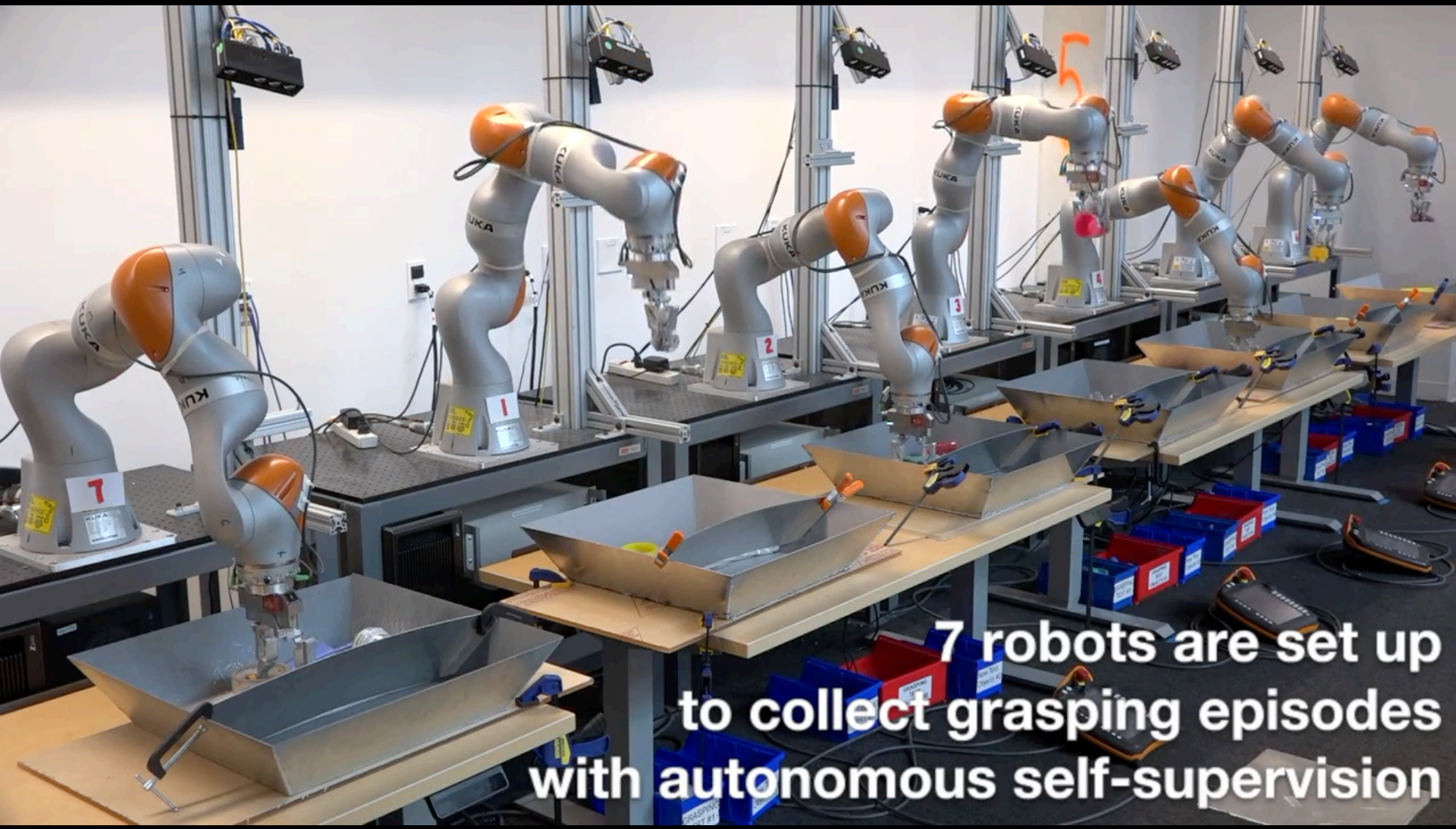
Learning Dexterity



Locomotion Strategies

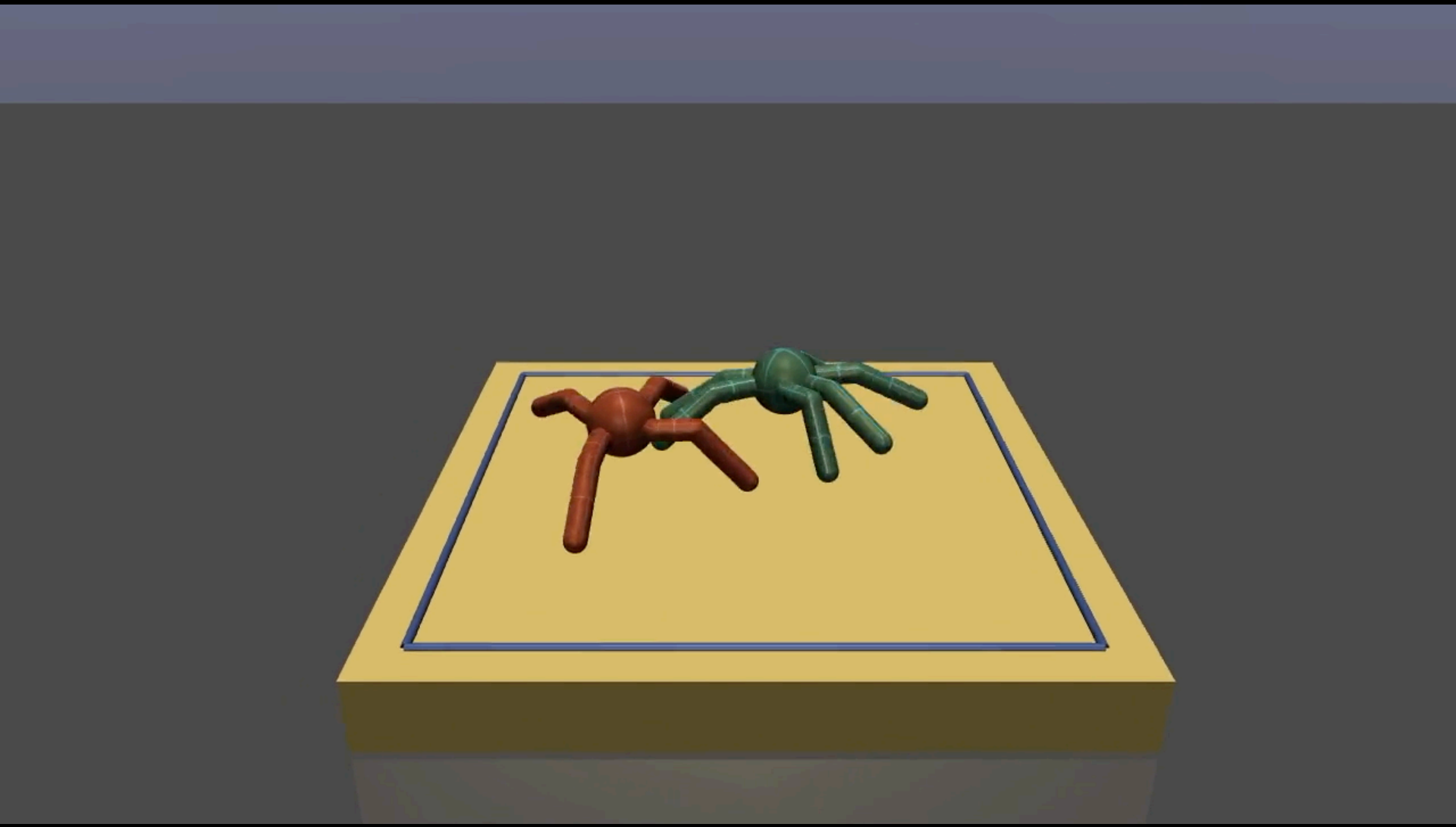
Learning Quadrupedal Locomotion over Challenging Terrain, Lee et al., 2020

Robots Learning to Grasp Objects



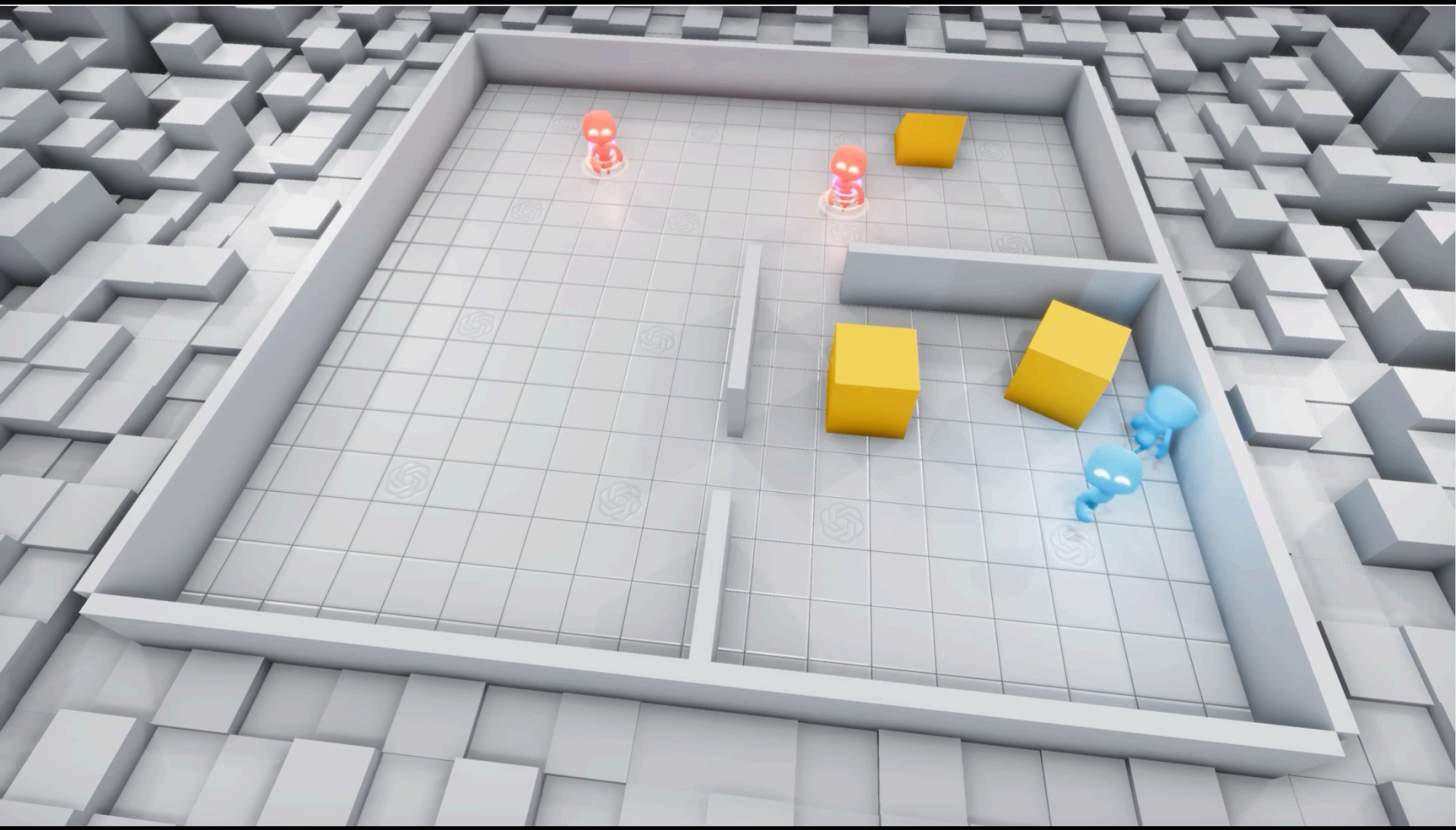
QT-OPT by Irpan et al. 2018 (Google)

Automatic Discovery of Skills



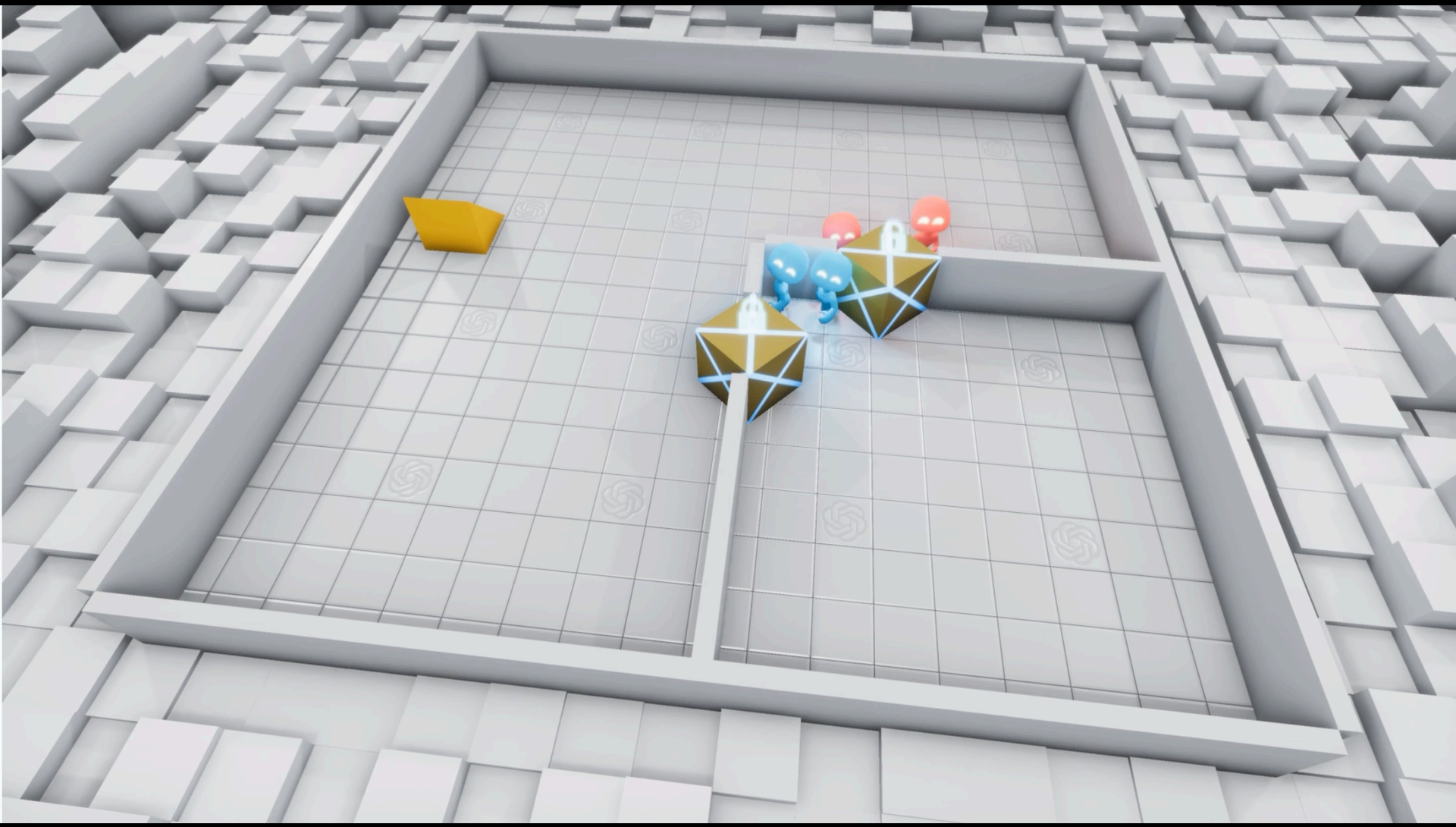
Continuous Adaptation via Meta-Learning, Al-Shedivat et al., 2018

Initial Random Exploration



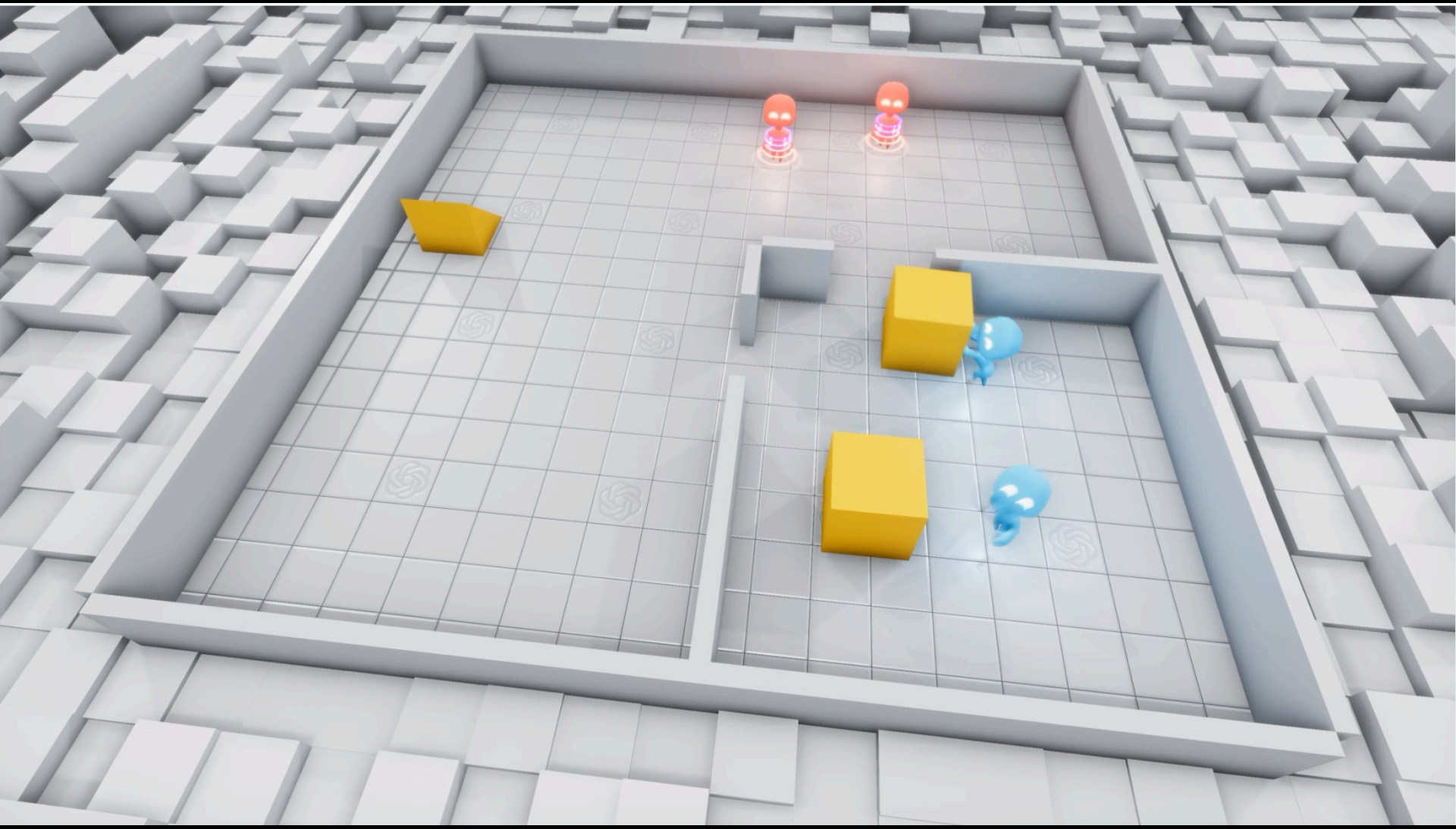
Emergent Tool Use from Multi-Agent Interaction, 2019

Use of blockers



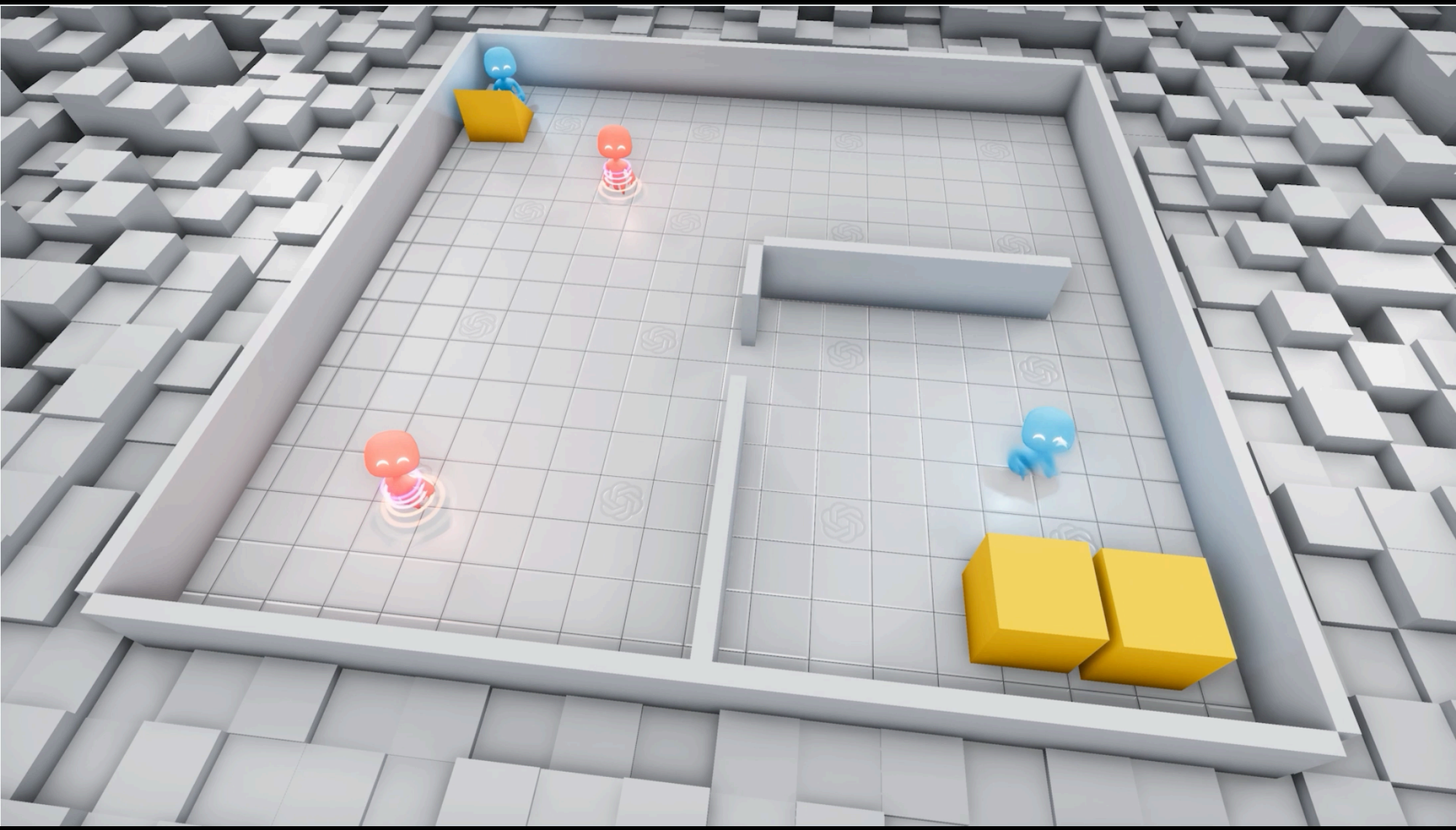
Emergent Tool Use from Multi-Agent Interaction, 2019

Using the Ramp



Emergent Tool Use from Multi-Agent Interaction, 2019

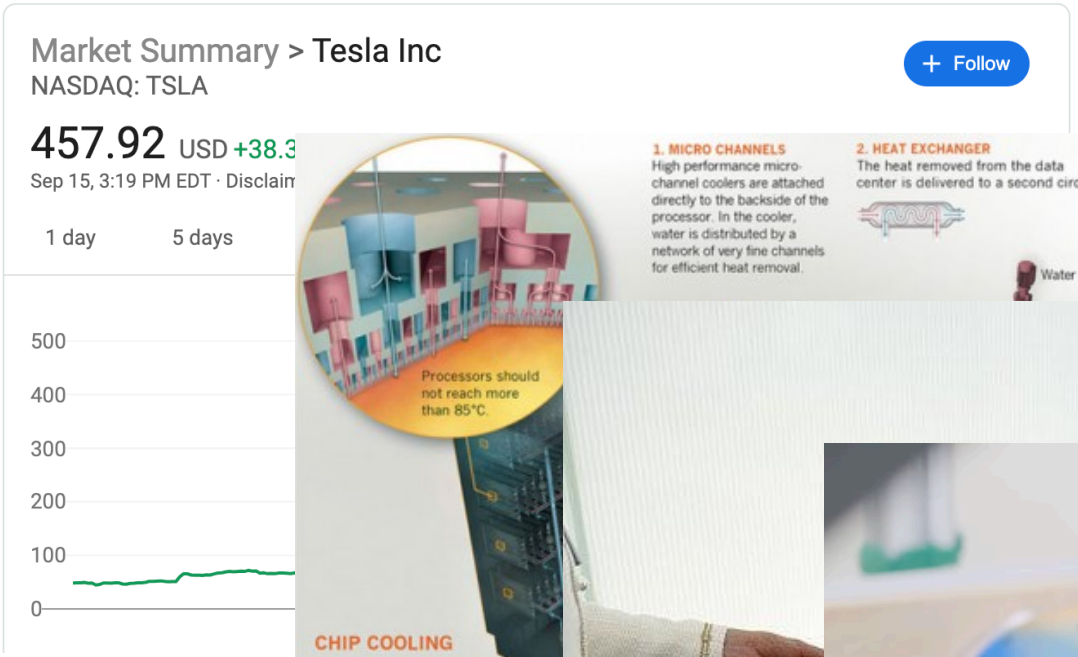
Hiding the Ramp



Emergent Tool Use from Multi-Agent Interaction, 2019

Internet to remote places: Balloon stabilization

Unknown Model



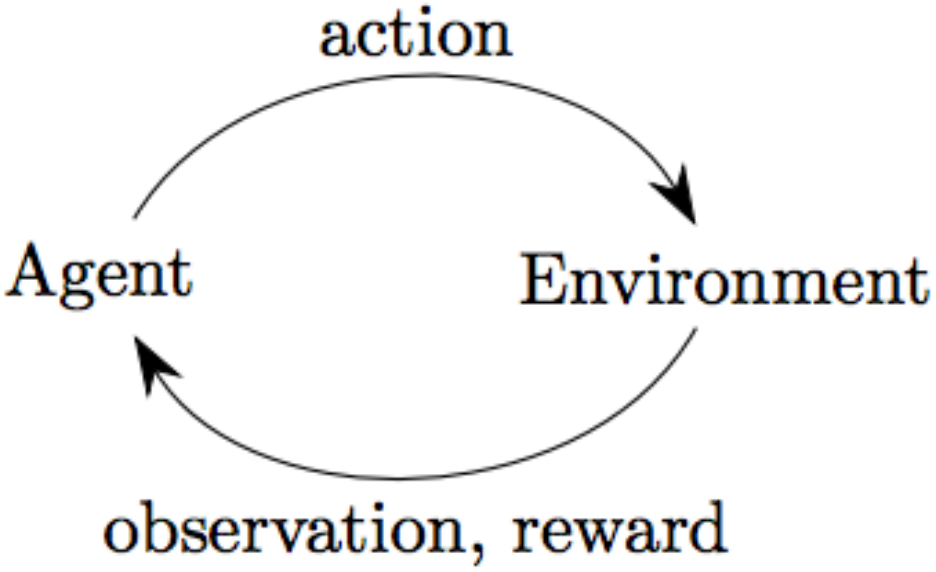
How does this all work?

Reinforcement Learning

Bandits

Contextual Bandits

Problem Formulation



Problem Formulation

Do Actions: $a_1, a_2, a_3, \ldots, a_T$

Problem Formulation

Do Actions: $a_1, a_2, a_3, \ldots, a_T$

Get Rewards: $r_1, r_2, r_3, \ldots, r_T$

Problem Formulation

Do Actions: $a_1, a_2, a_3, \ldots, a_T$

Get Rewards: $r_1, r_2, r_3, \ldots, r_T$

$$\max \sum_{t=1}^{T} r_t$$

Problem Formulation

$$\max \sum_{t=1}^{T} r_t$$

environment model

s.t. $s_{t+1} = f(s_{0:t}, a_{0:t})$

given or learnt

Model Based Methods

policy

$a_t = \pi(s_{0:t}; \theta)$

Model Free Methods

The RL Problem

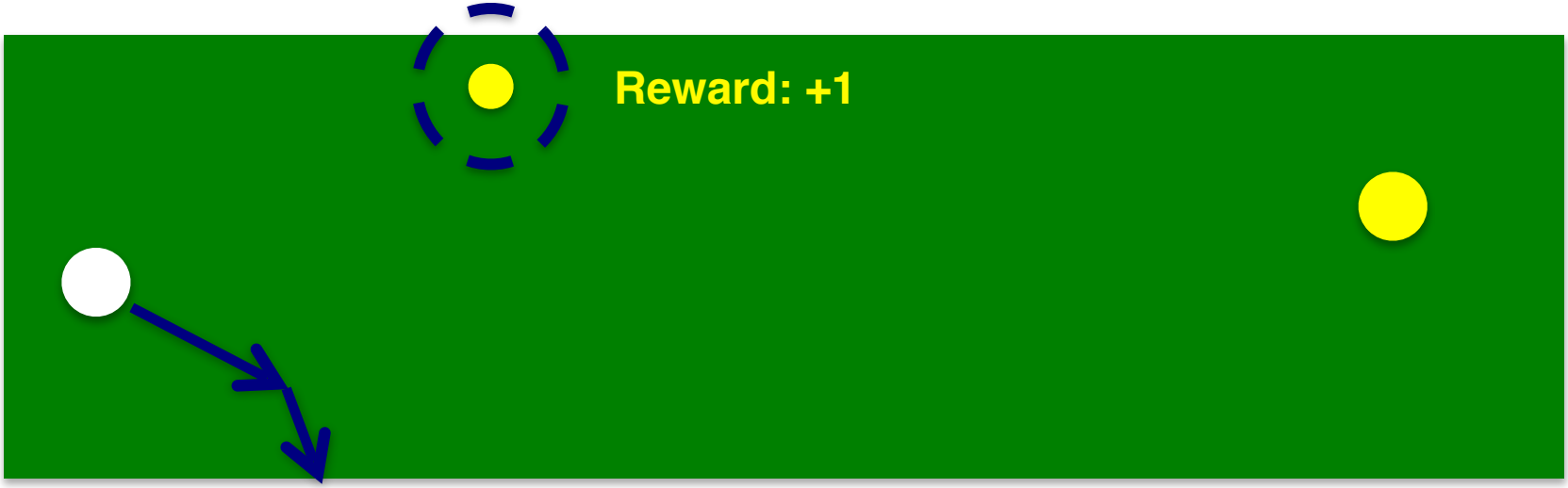
The RL Problem

The RL Problem

The RL Problem

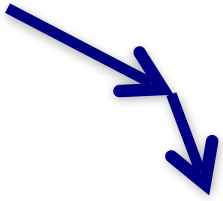
Reward: -1

Solving the MDP



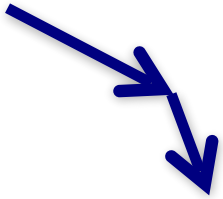
The RL Problem

Another Attempt



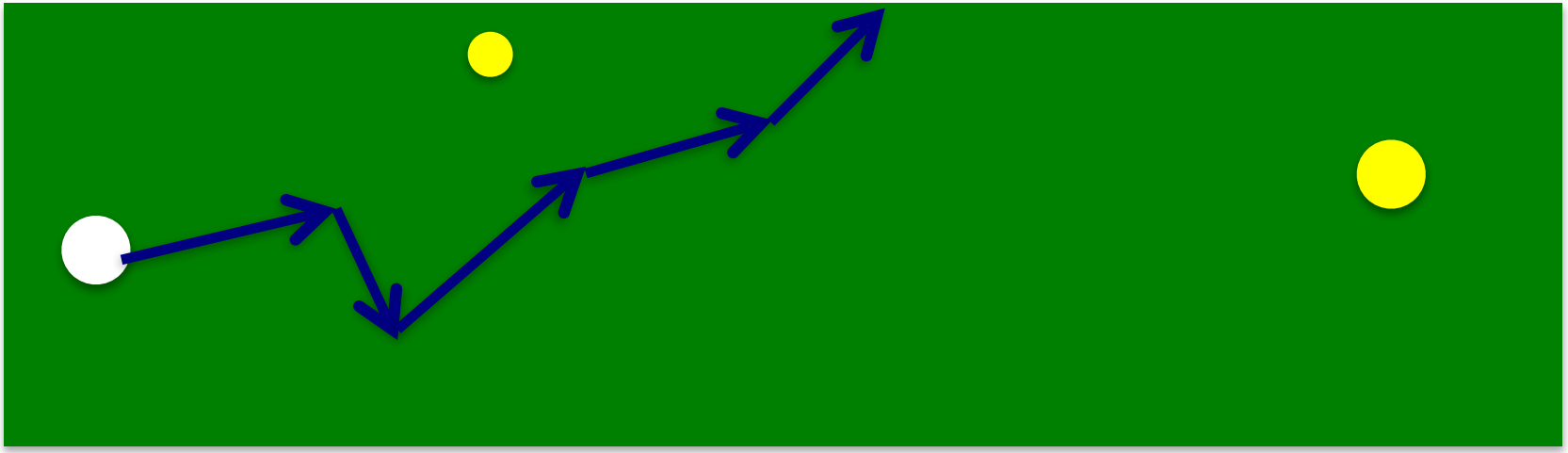
Reward: -1

Another Attempt

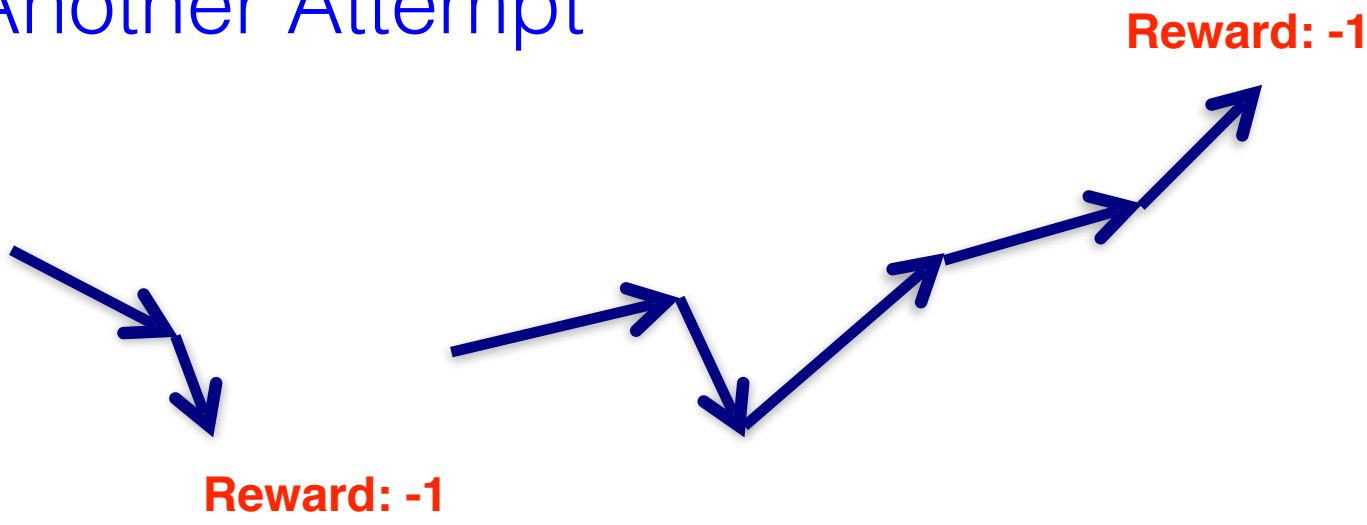


Reward: -1

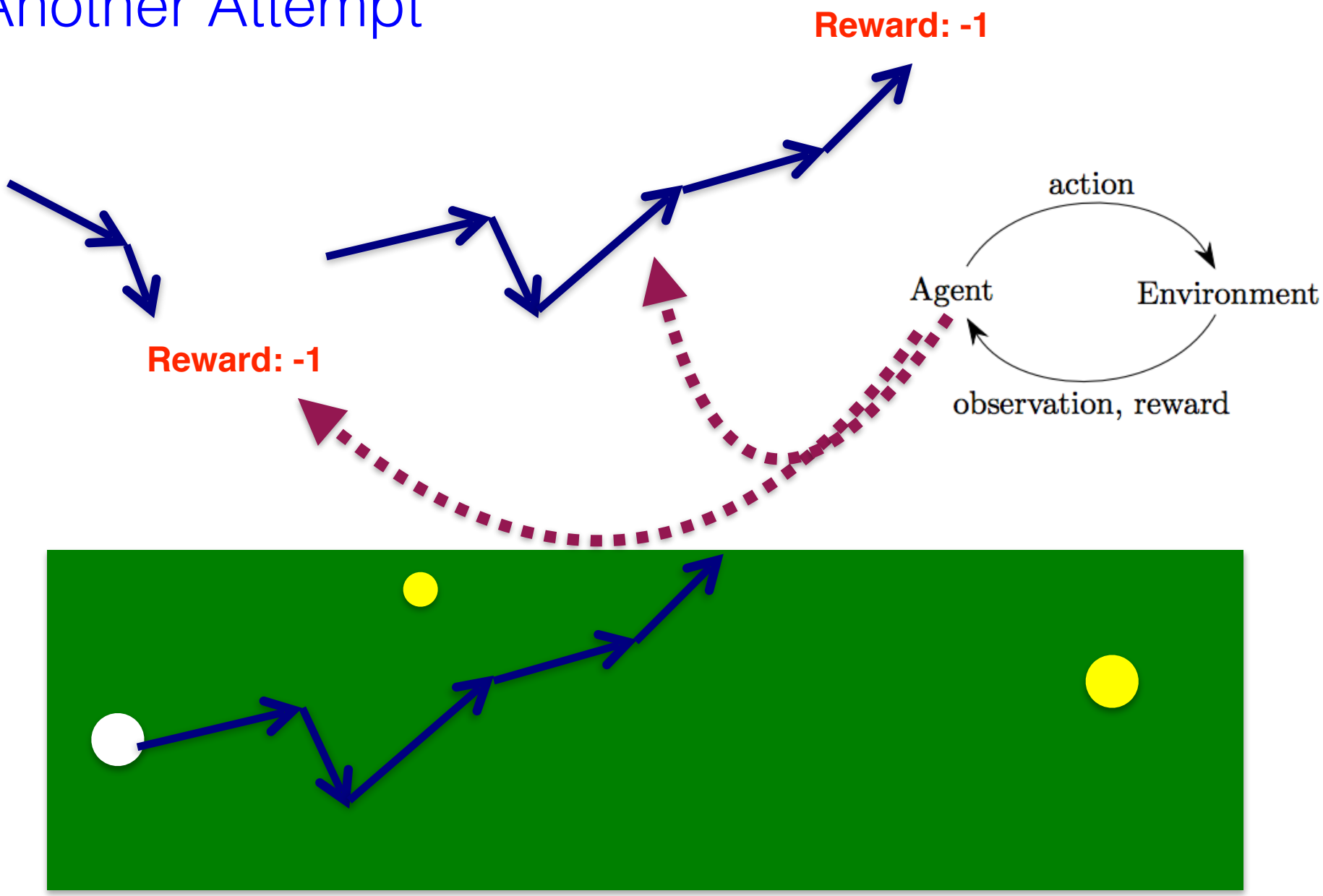
Reward: -1



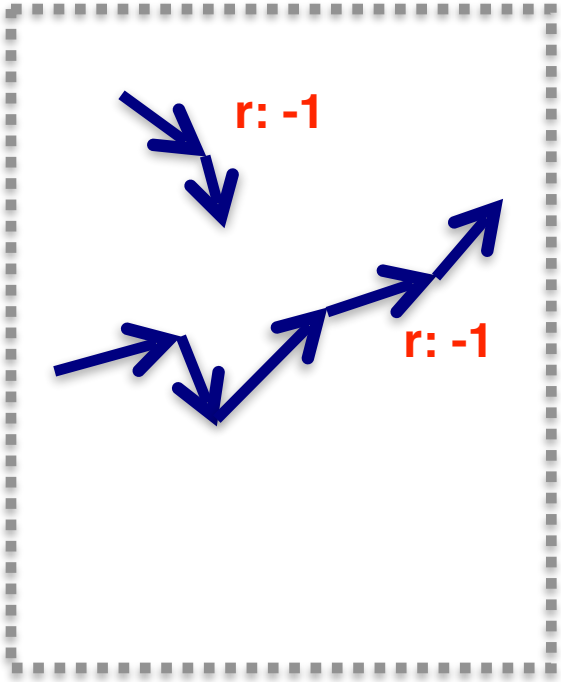
Another Attempt

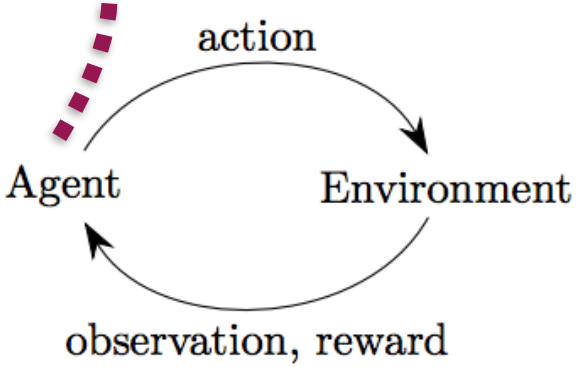


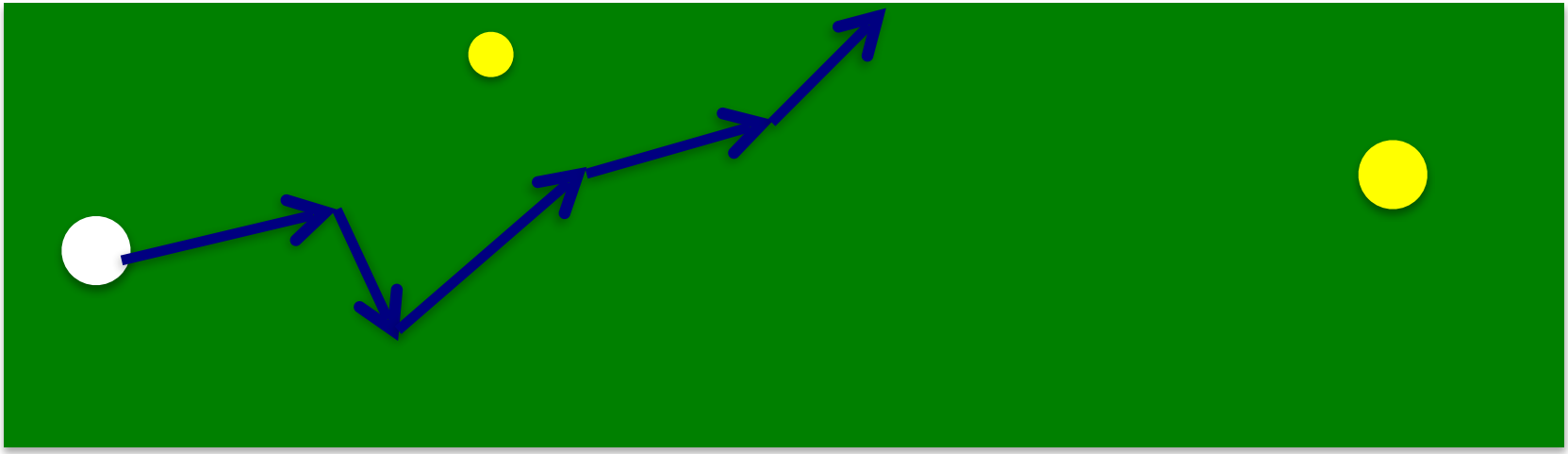
Another Attempt



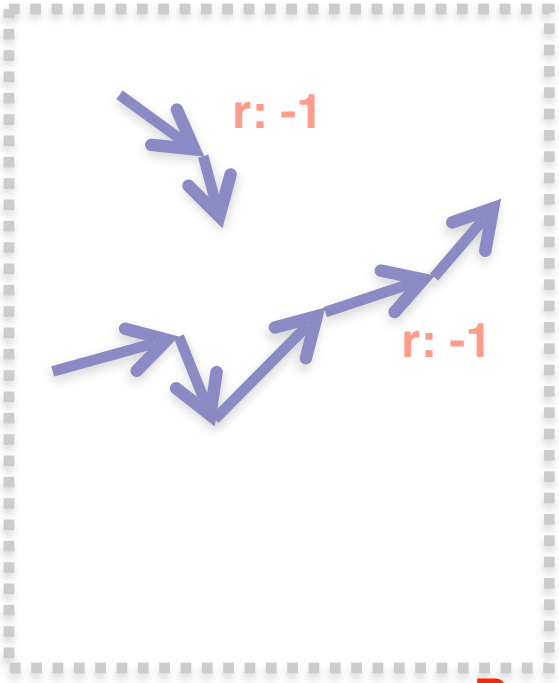
Dataset Generation

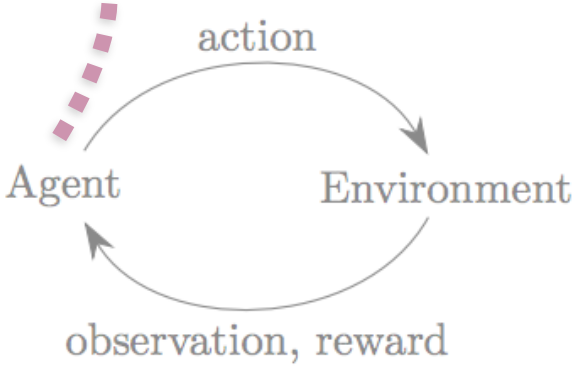






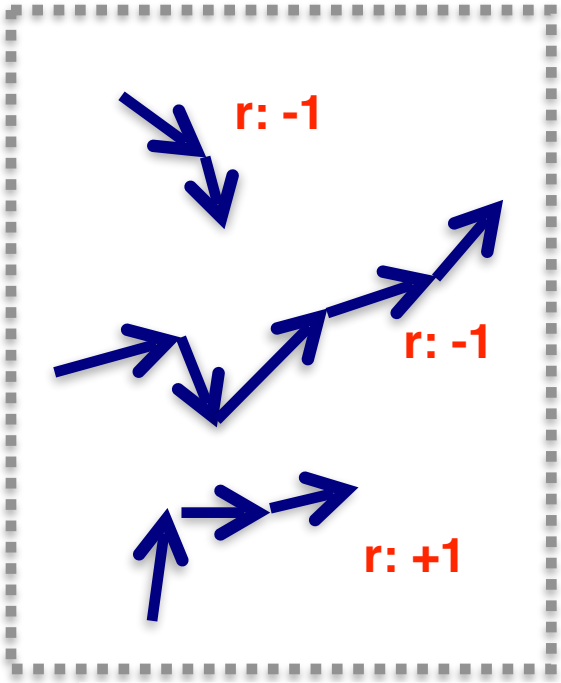
Dataset Generation

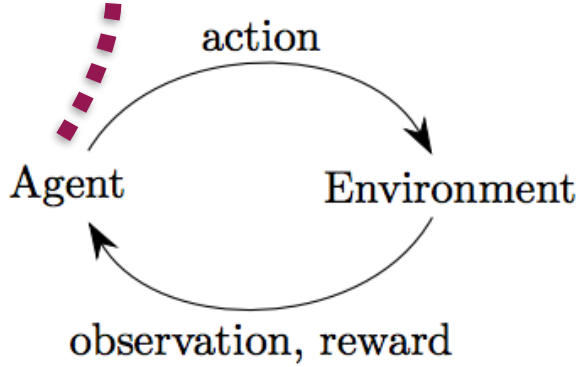




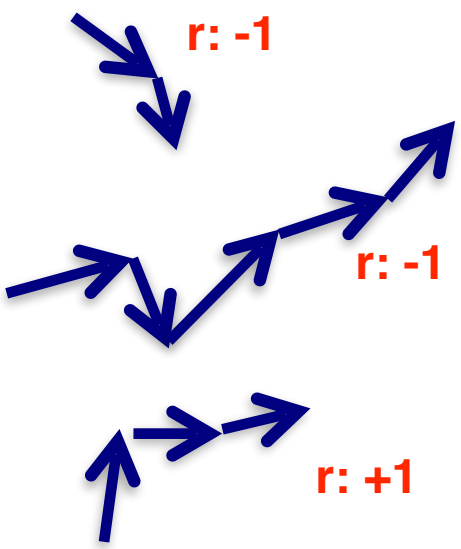
Reward: +1

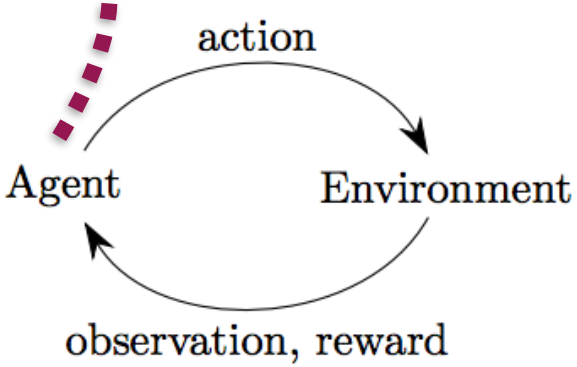
Dataset Generation





Dataset Generation





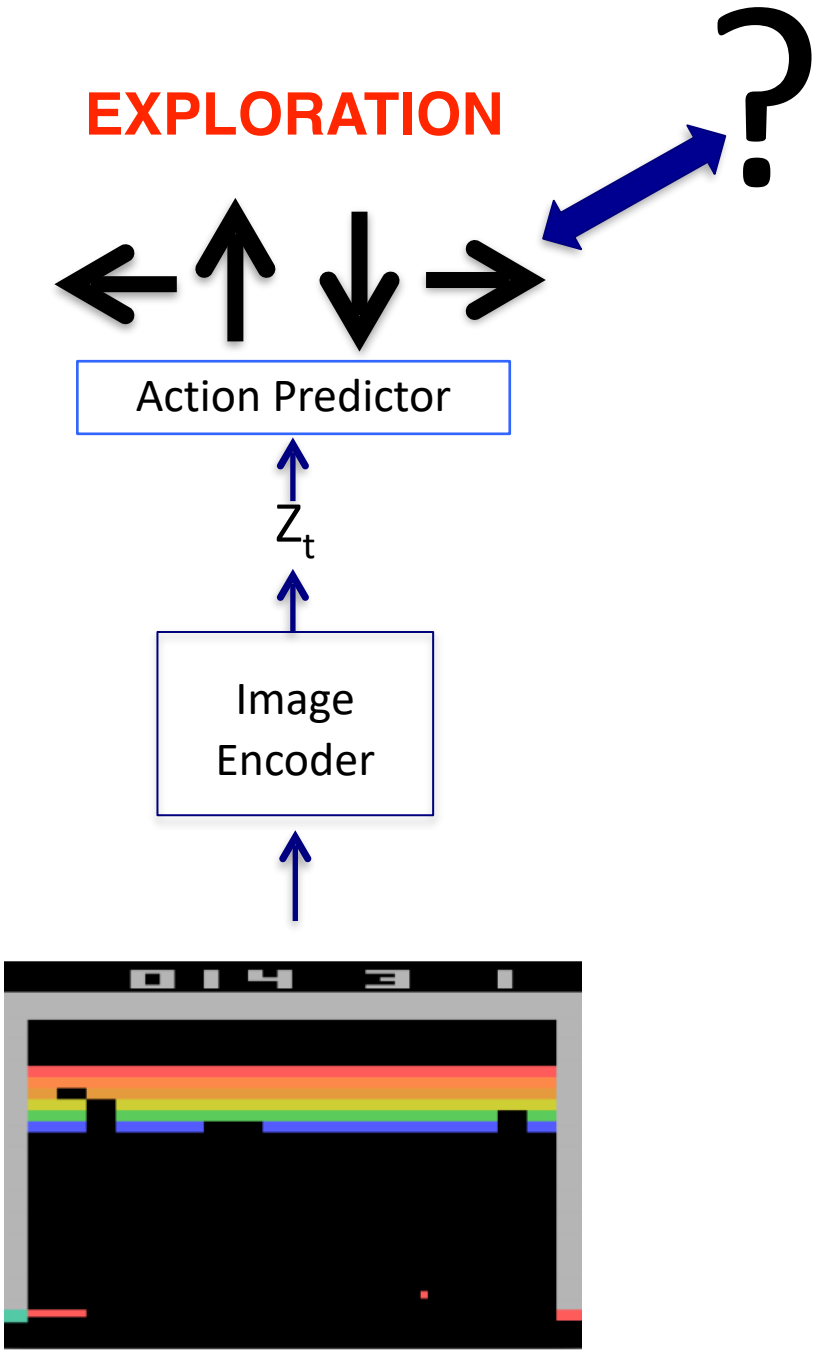
Supervised Learning

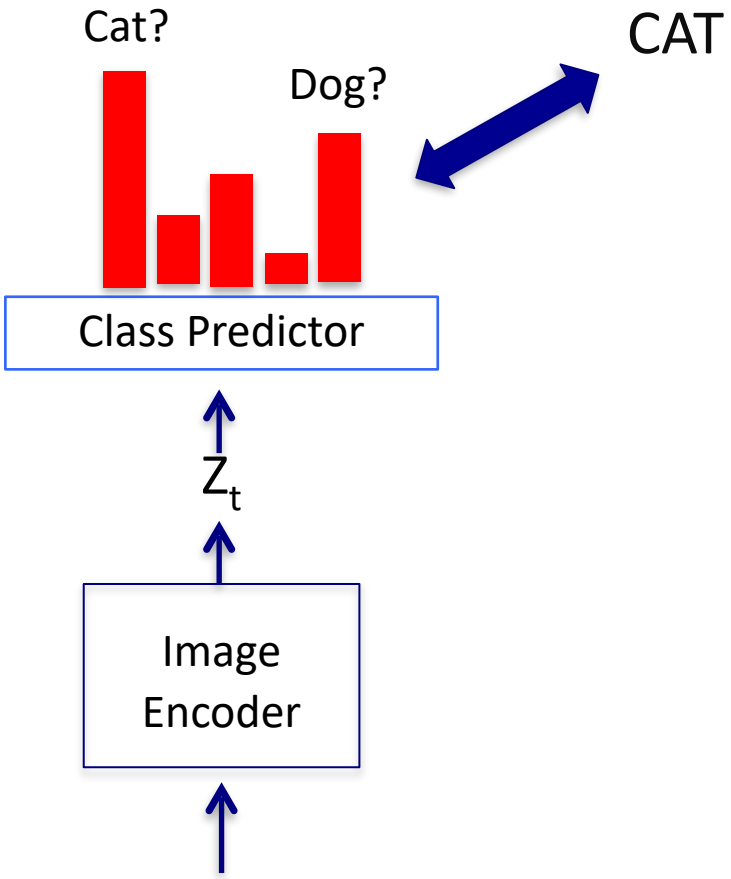
(cat)

In supervised learning, dataset is GIVEN

In reinforcement learning, the agent collects its own data

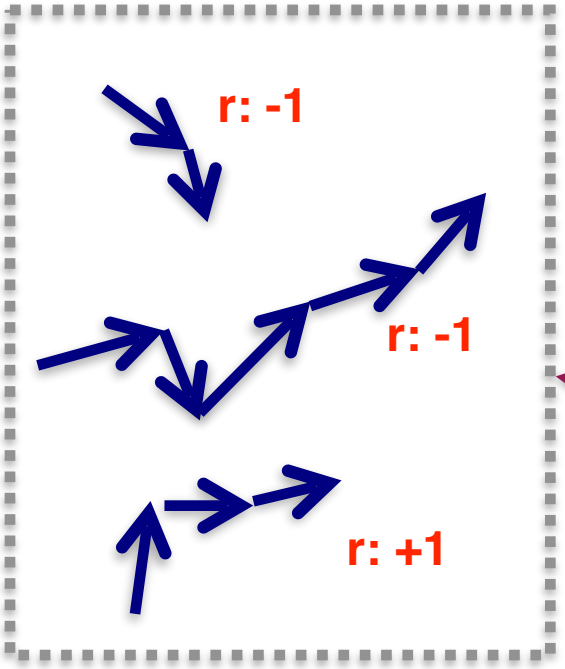
(i.e, it needs to explore)

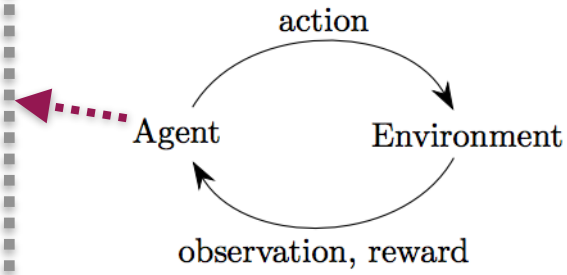

EXPLORATION
?
Action Predictor
Z_t
Image Encoder


Cat?
Dog?
CAT
Class Predictor
Z_t
Image Encoder

Dataset

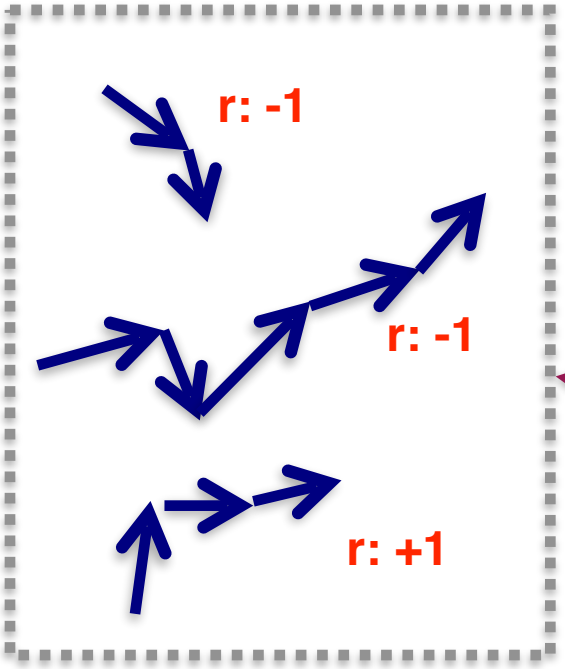
Is exploration a problem?



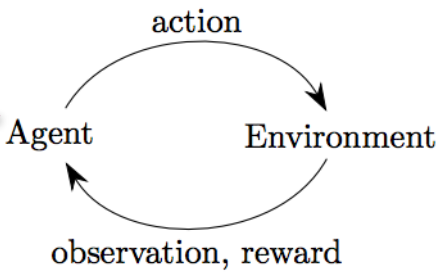


Dataset

Is exploration a problem?



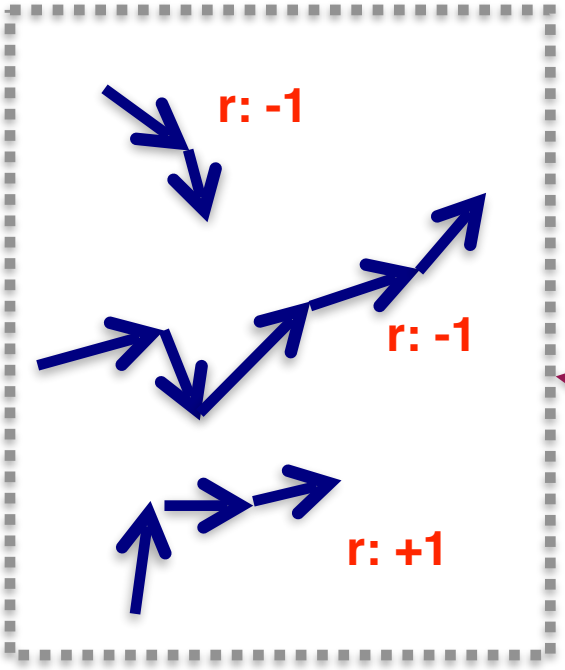
Learn



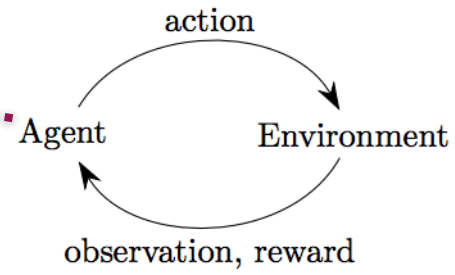
Goal

$$a_t = \pi(s_{0:t}; \theta)$$

Dataset



Learn



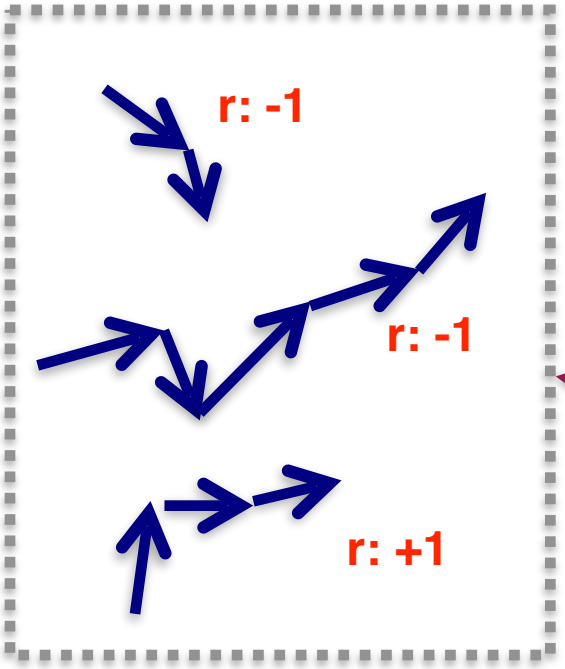
Is exploration a problem?

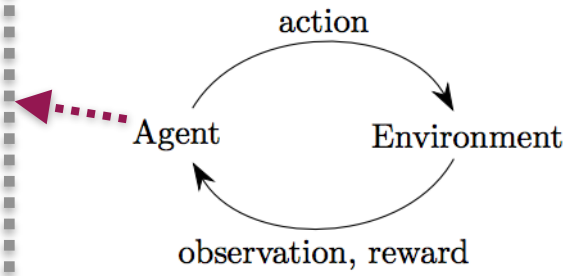
Goal

$$a_t = \pi(s_{0:t}; \theta)$$

Looks good!
(is there a problem?)

Dataset





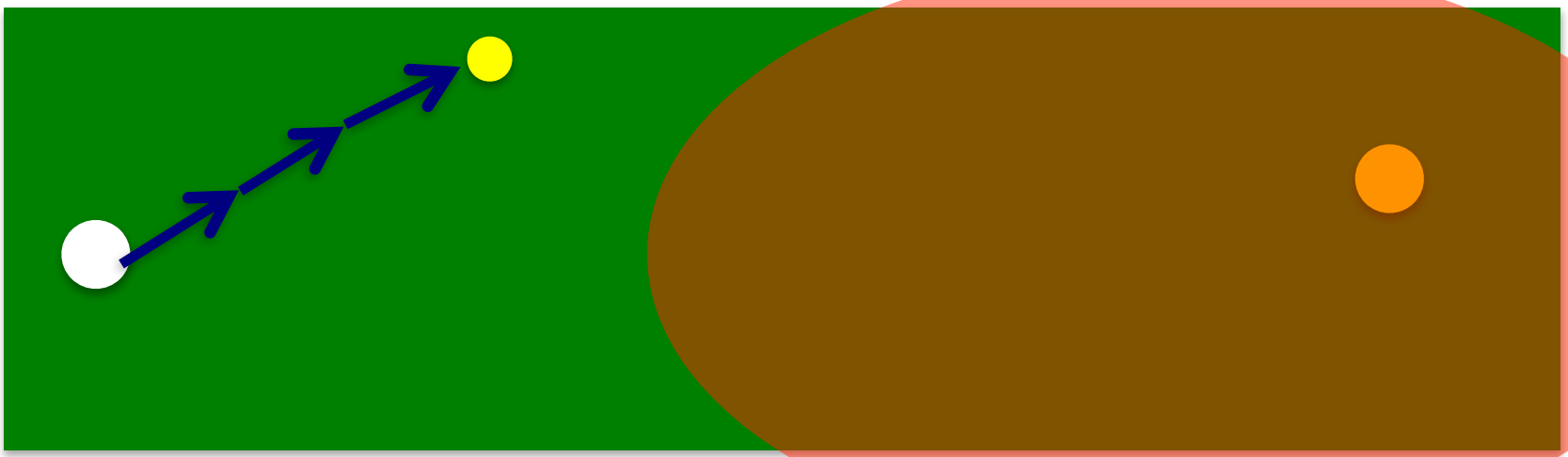
Is exploration a problem?

Goal

$$a_t = \pi(s_{0:t}; \theta)$$

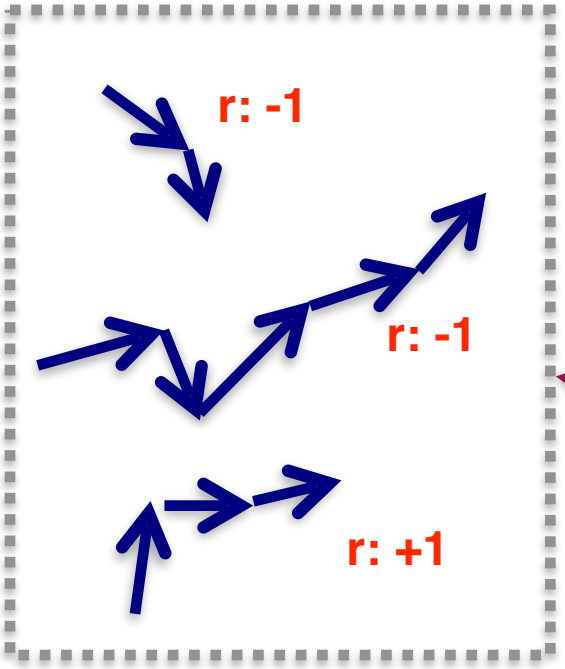
Looks good!

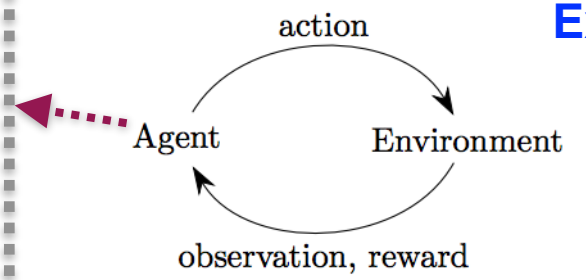
(is there a problem?)



Might not explore the state space!

Dataset

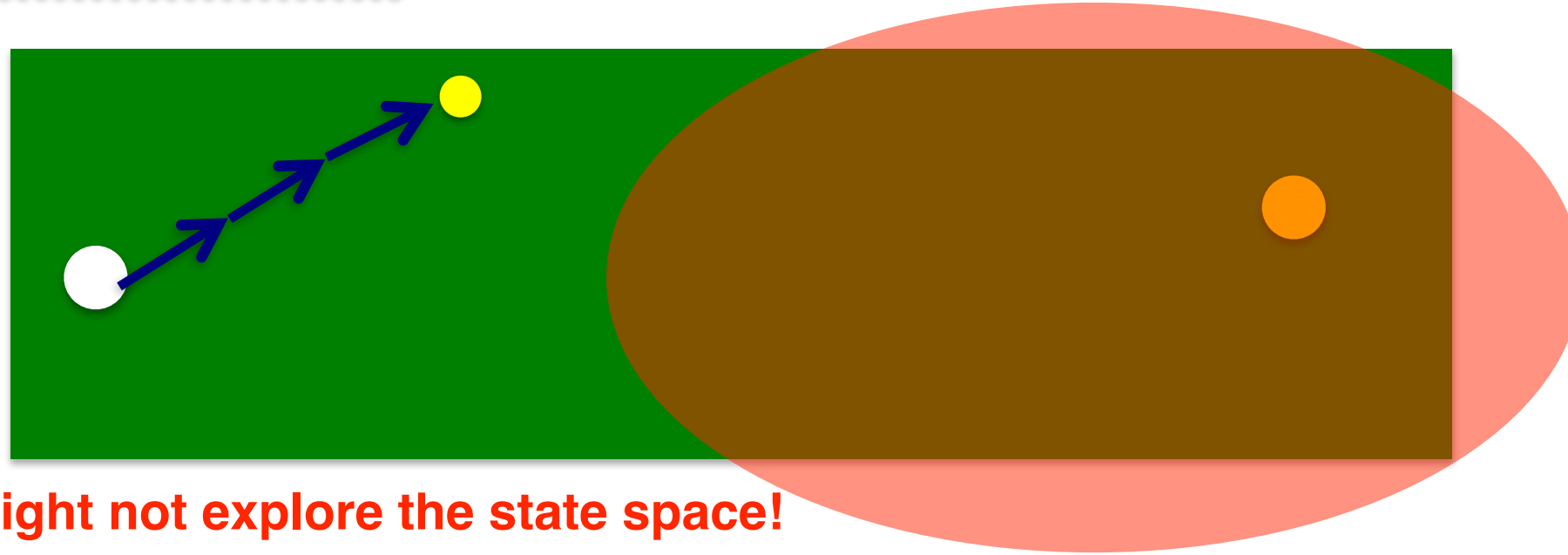




Is exploration a problem?

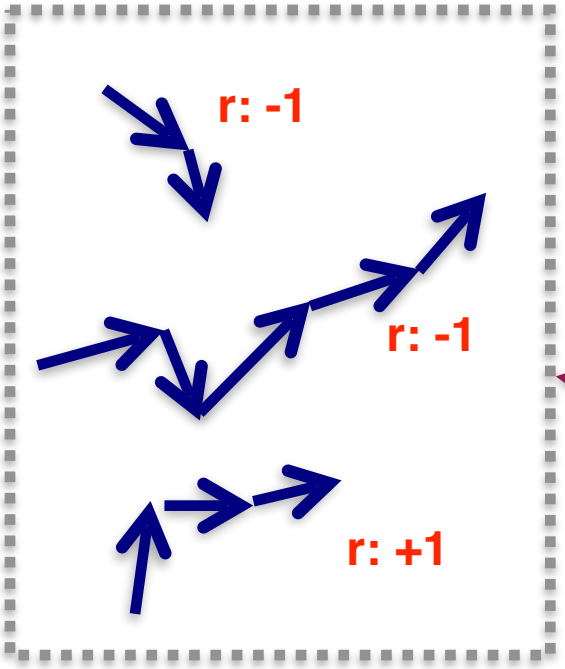
Yes!
(Can learn sub-optimal behavior)

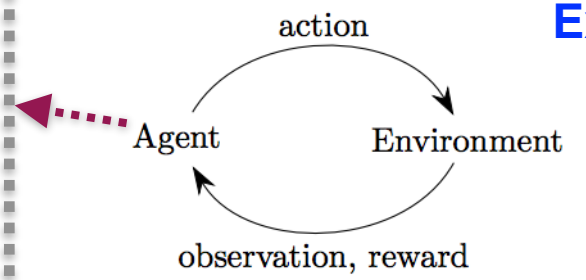
Exploration-Exploitation Dilemma



Might not explore the state space!

Dataset





Is exploration a problem?

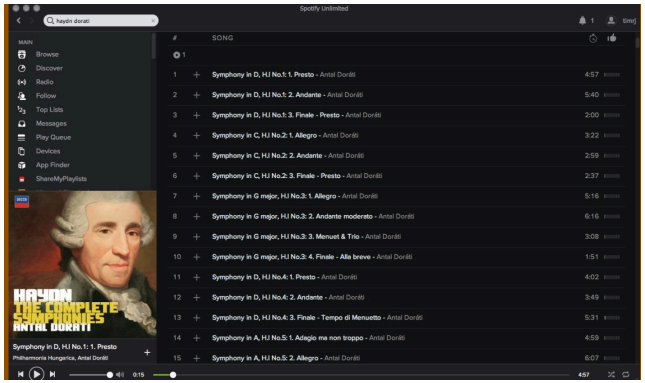
Yes!
(Can learn sub-optimal behavior)

Exploration-Exploitation Dilemma

Exploration can be quite hard!

When?

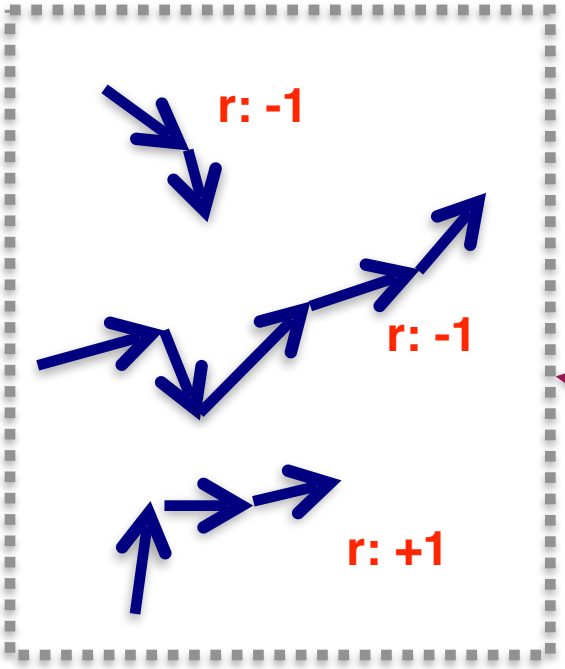
Imagine your favorite playlist

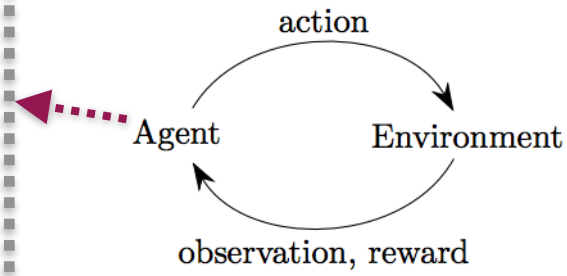


(they want you hooked)

Explore by
Suggesting other music

Dataset





What questions might be of interest?

Is there a method that will achieve the highest reward?

How fast will reach it?

(i.e., what is the overall regret)