

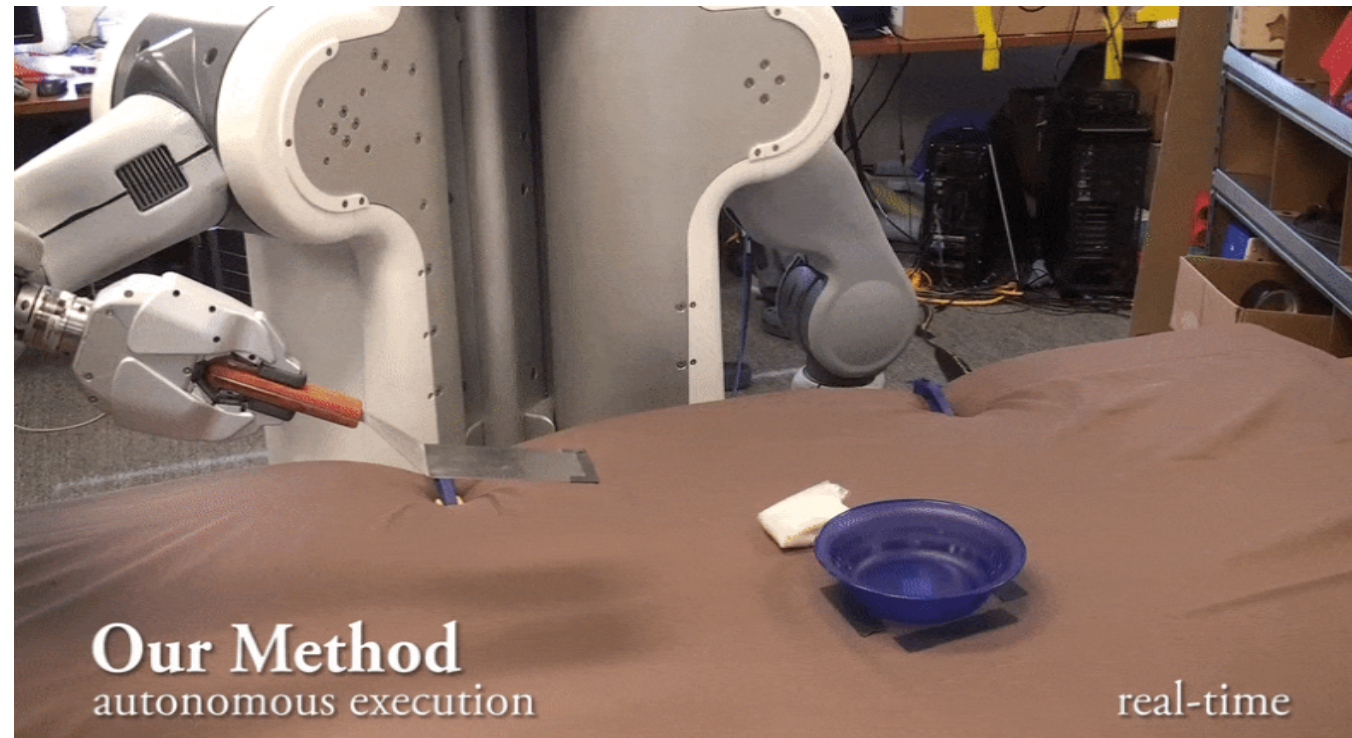
Tackling Distribution Shift through Pessimism, Adaptation, and Anticipation

Chelsea Finn



Stanford

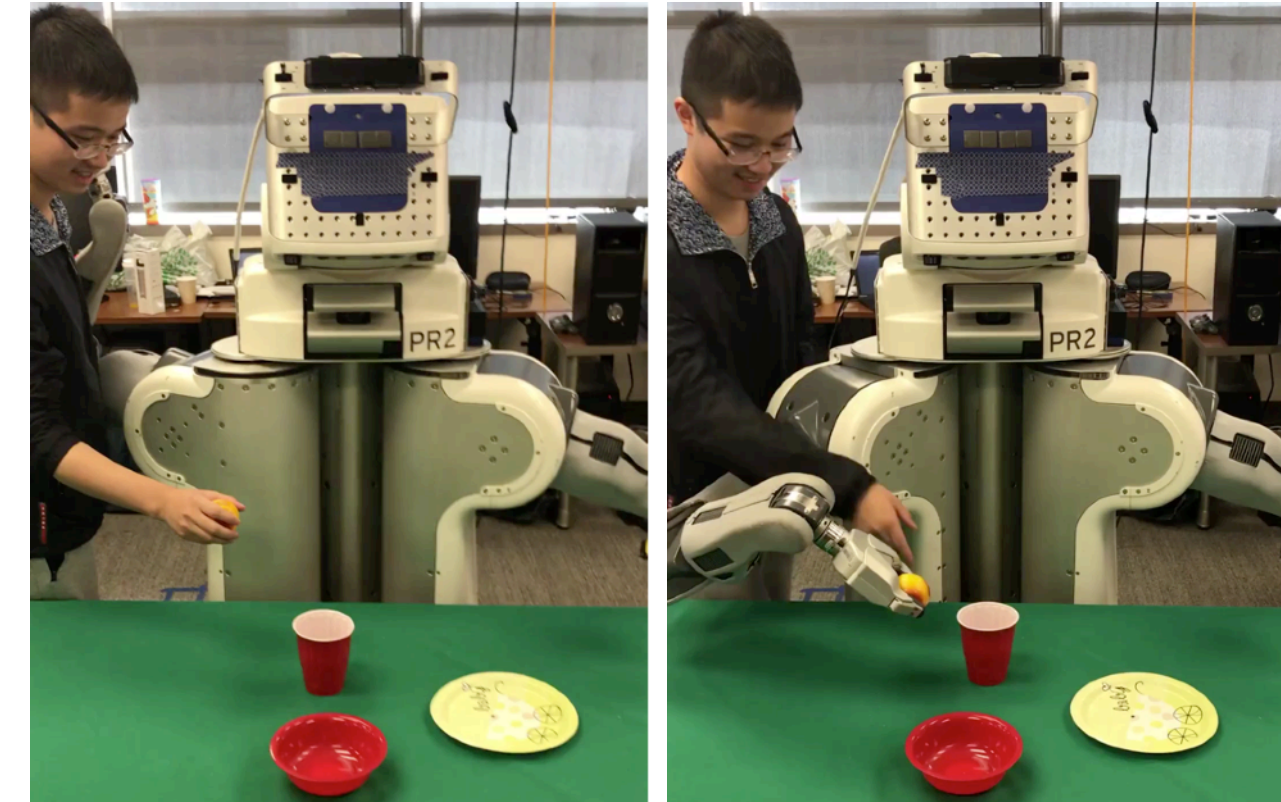
Can robots develop broadly intelligent behavior through learning & interaction?



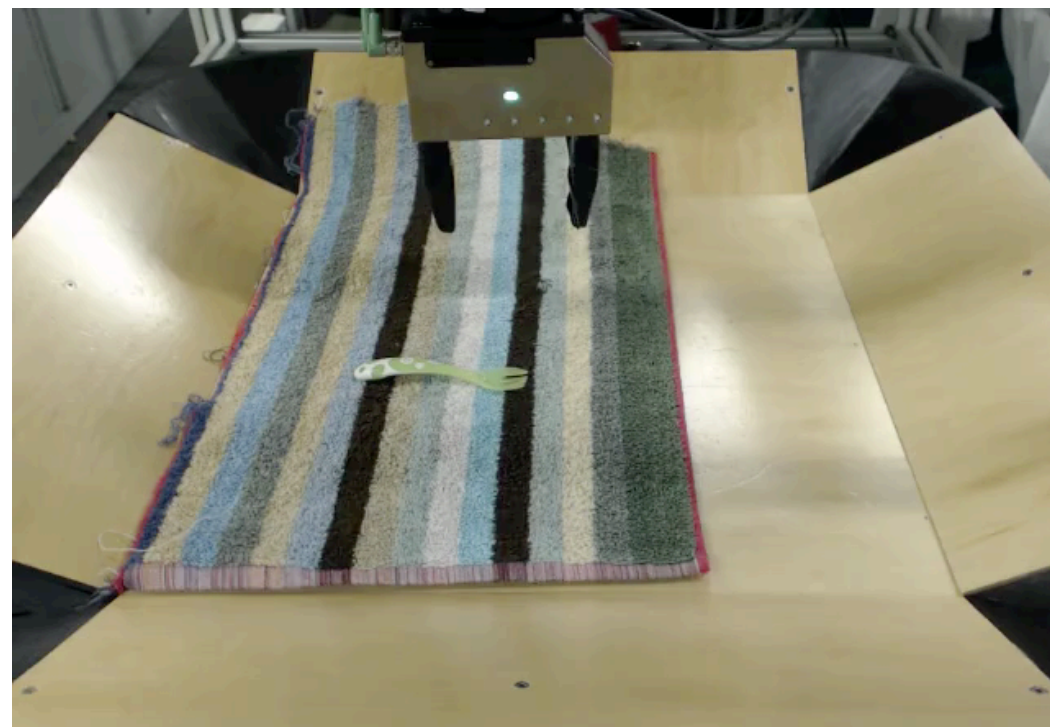
Finn, Tan, Duan, Darrell, Levine, Abbeel.
ICRA '16



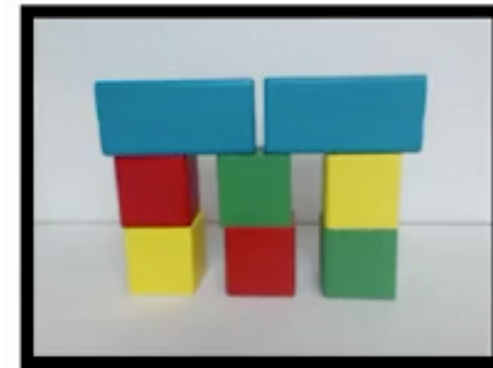
Levine*, Finn*, Darrell, Abbeel.
JMLR '16



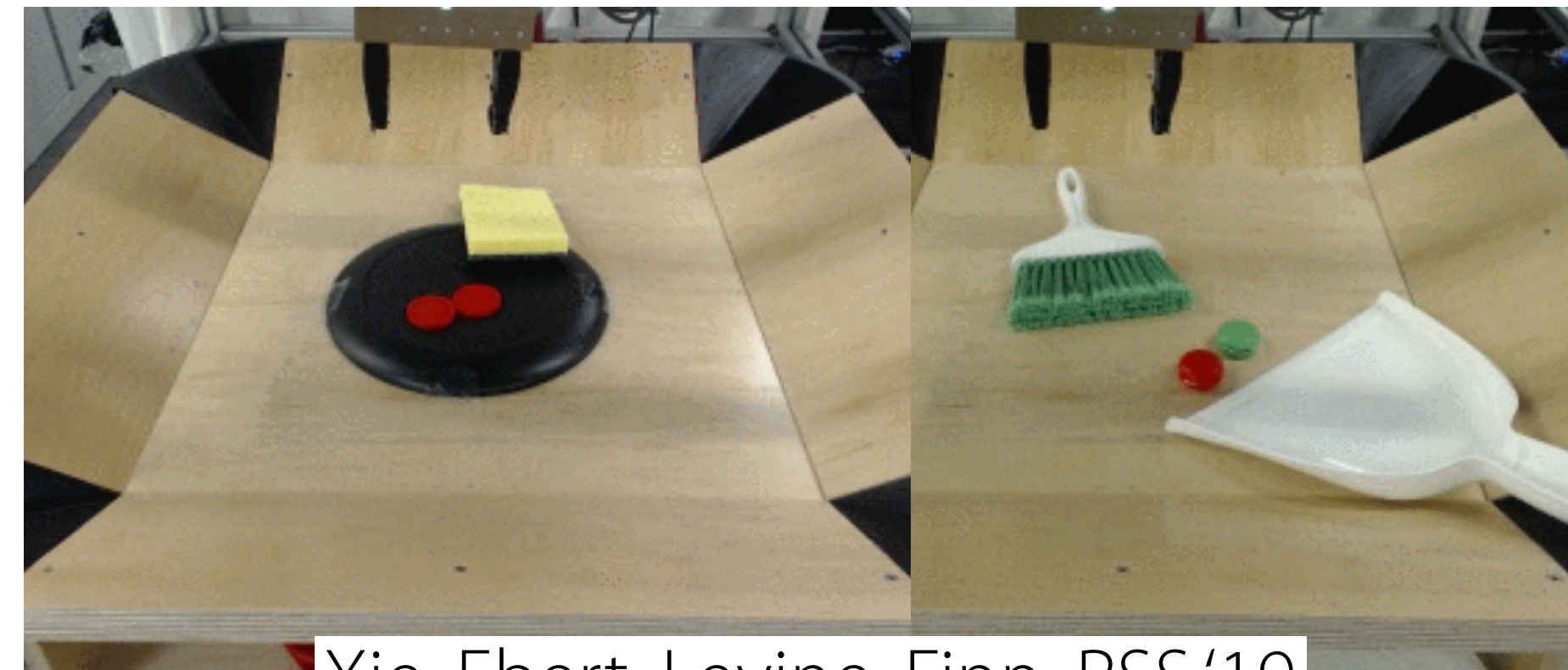
Yu*, Finn*, Xie, Dasari, Zhang,
Abbeel, Levine, RSS '18



Ebert*, Finn*, Dasari, Xie,
Lee, Levine. '18



Janner, Levine, Freeman,
Tenenbaum, Finn, Wu. ICLR '19



Xie, Ebert, Levine, Finn. RSS '19

Machine learning works



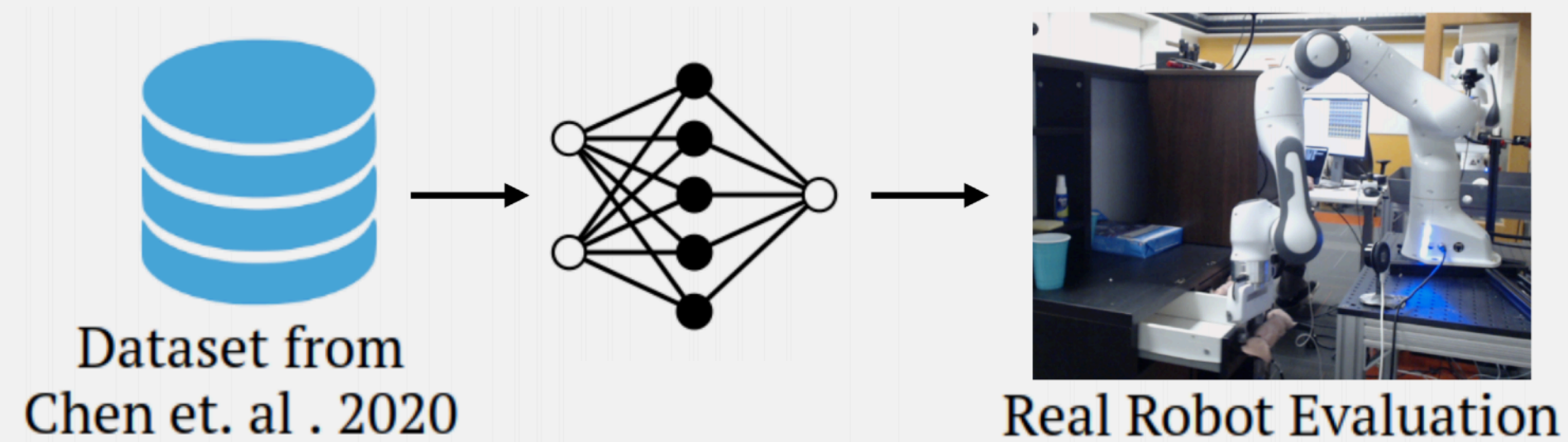
on the training data distribution

Core assumption

$$P_{\text{train}} = P_{\text{test}}$$

Examples of distribution shift: offline RL and temporal shifts

RL from offline datasets



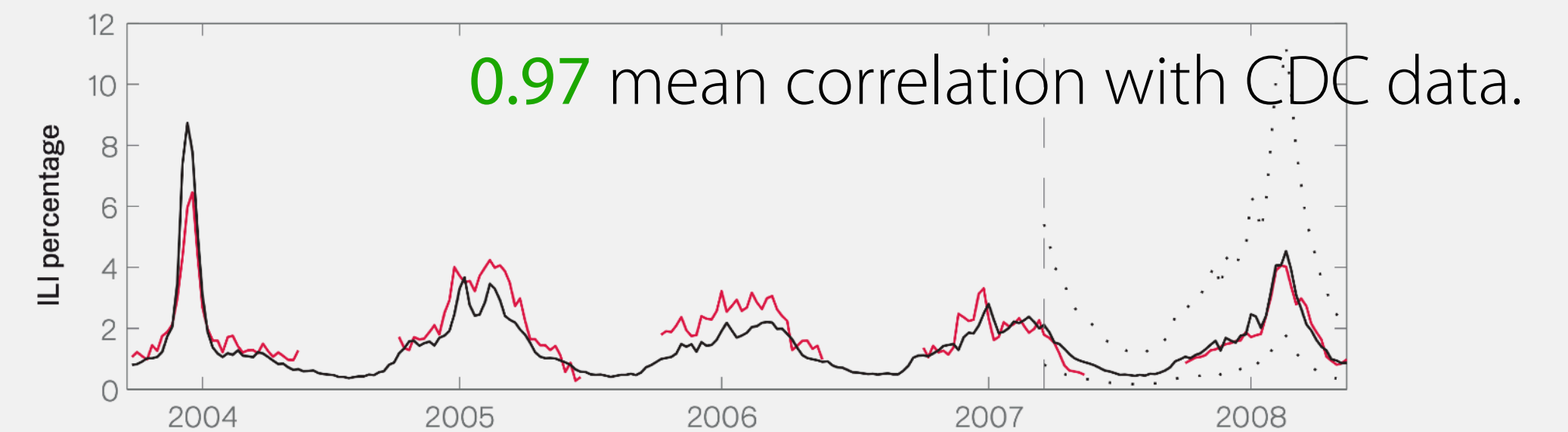
Distribution shift between **policy in the dataset** and the **policy being optimized**.

If you don't account for this shift:



Shift over time

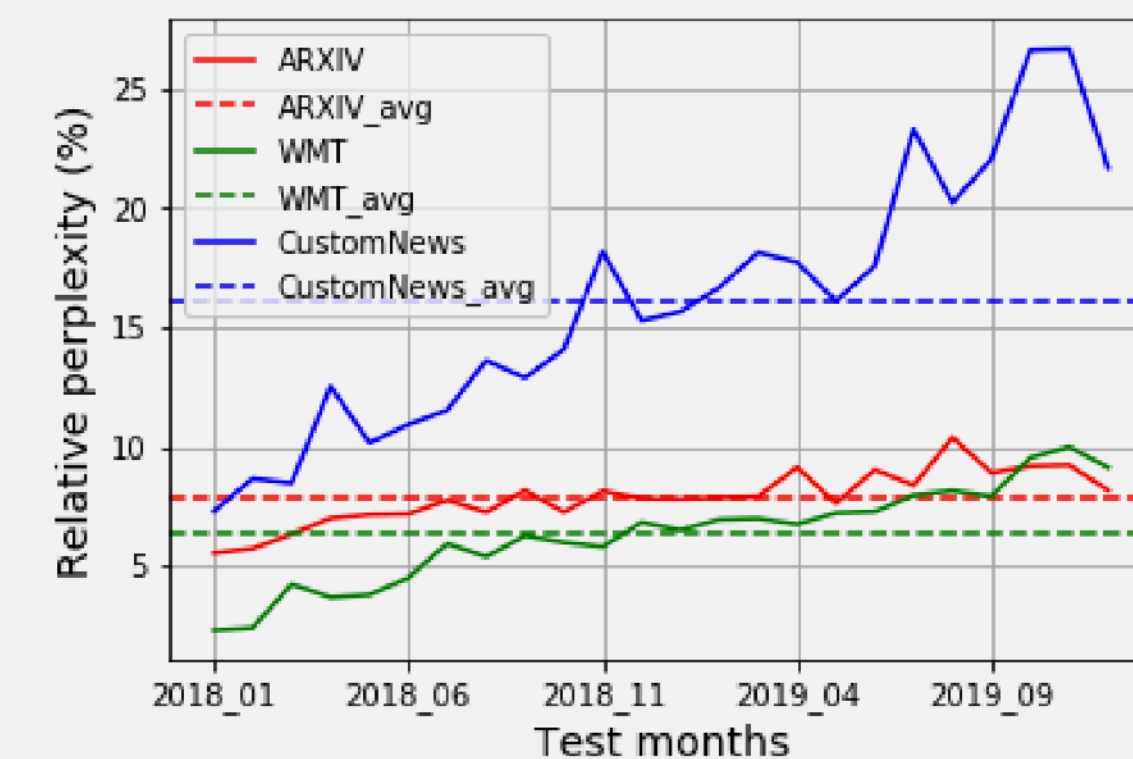
Predicting flu incidence from search queries



Ginsberg et al. *Detecting influenza epidemics using search engine query data*. Nature '09

Feb 2013: predicting **double** the incidence

Language model perplexity over time.



Lazaridou et al. *Pitfalls of Static Language Modeling*. '21

Examples of distribution shift: domains & subpopulations

Online content moderation (Borkan et al. 2019)

Comment: "I doubt that anyone cares whether you believe it or not" → toxic / not toxic

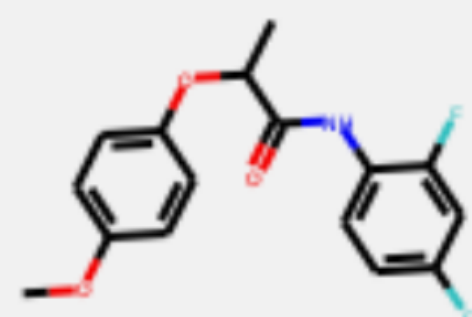
92.2% average test accuracy

Demographic	Test accuracy on non-toxic comments
Male	87.3 (0.7)
Female	89.0 (0.6)
LGBTQ	74.6 (0.5)
Christian	92.1 (0.2)
Muslim	80.9 (1.0)
Other religions	86.1 (0.1)
Black	69.2 (1.3)
White	71.2 (1.4)

69.2% on non-toxic comments mentioning Black demographic

Molecular Property Prediction (Hu et al. 2020)

Molecule:



→ (0,1,1,0,0,...)
biological activity prediction

34.4% average precision on test molecules from training scaffolds

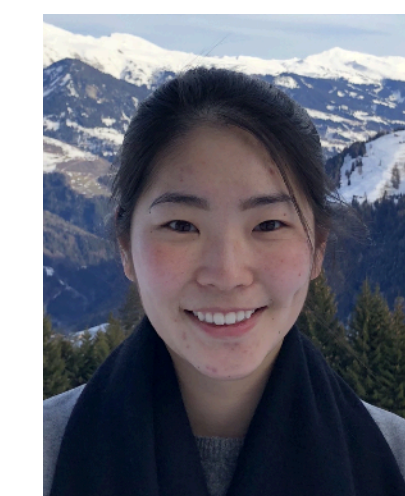
26.8% average precision on test molecules from held-out scaffolds

WILDS

WILDS has 5+ more datasets with distribution shift, ranging from ecological conservation to medical imaging.



Pang Wei Koh



Shiori Sagawa

Koh*, Sagawa*, Marklund, Xie, Zhang, Balsubramani, Hu, Yasunaga, Phillips, Gao, Lee, David, Stavness, Guo, Earnshaw, Haque, Beery, Leskovec, Kundaje, Pierson, Levine, Finn, Liang. **WILDS: A Benchmark of in-the-Wild Distribution Shifts**. arXiv 2020.

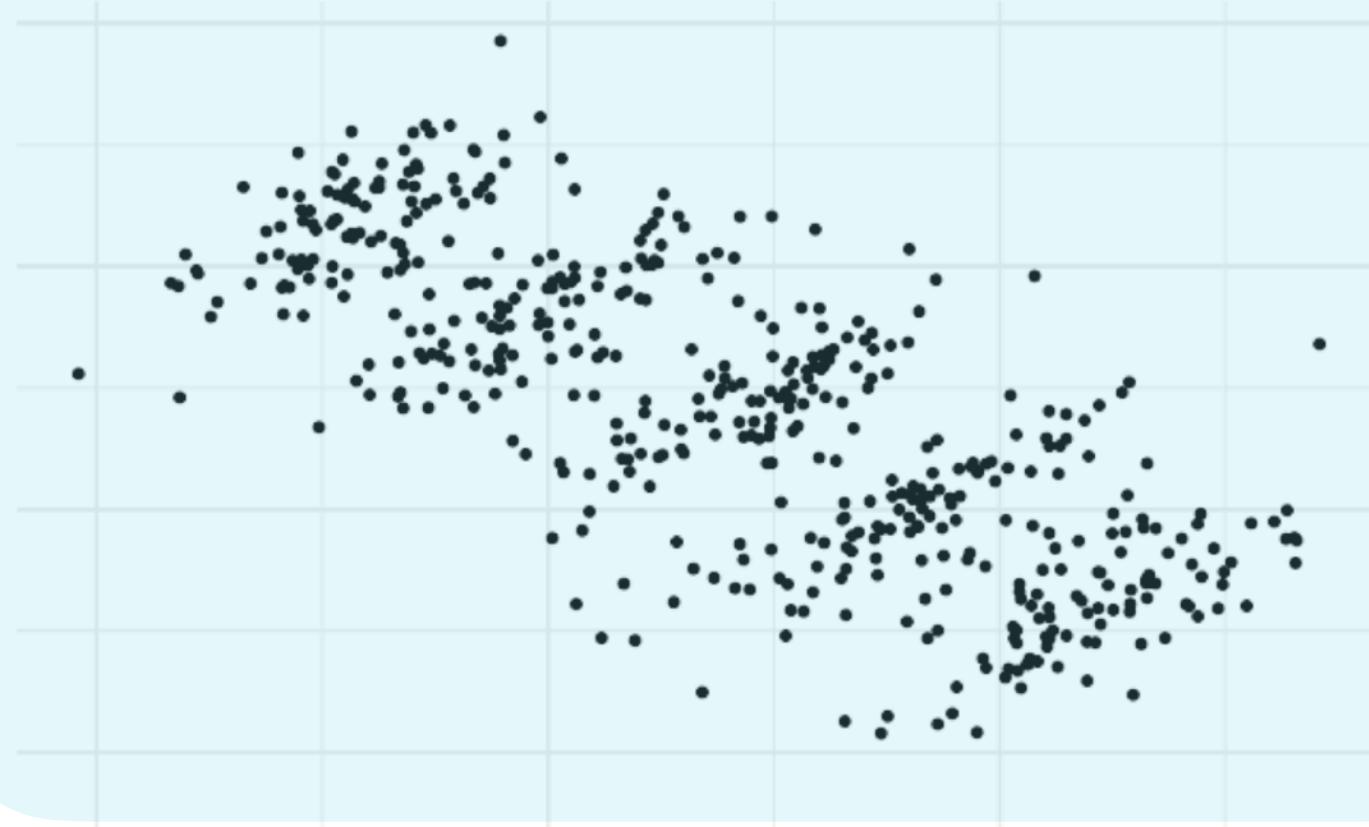
wilds.stanford.edu

Tools for tackling distribution shift

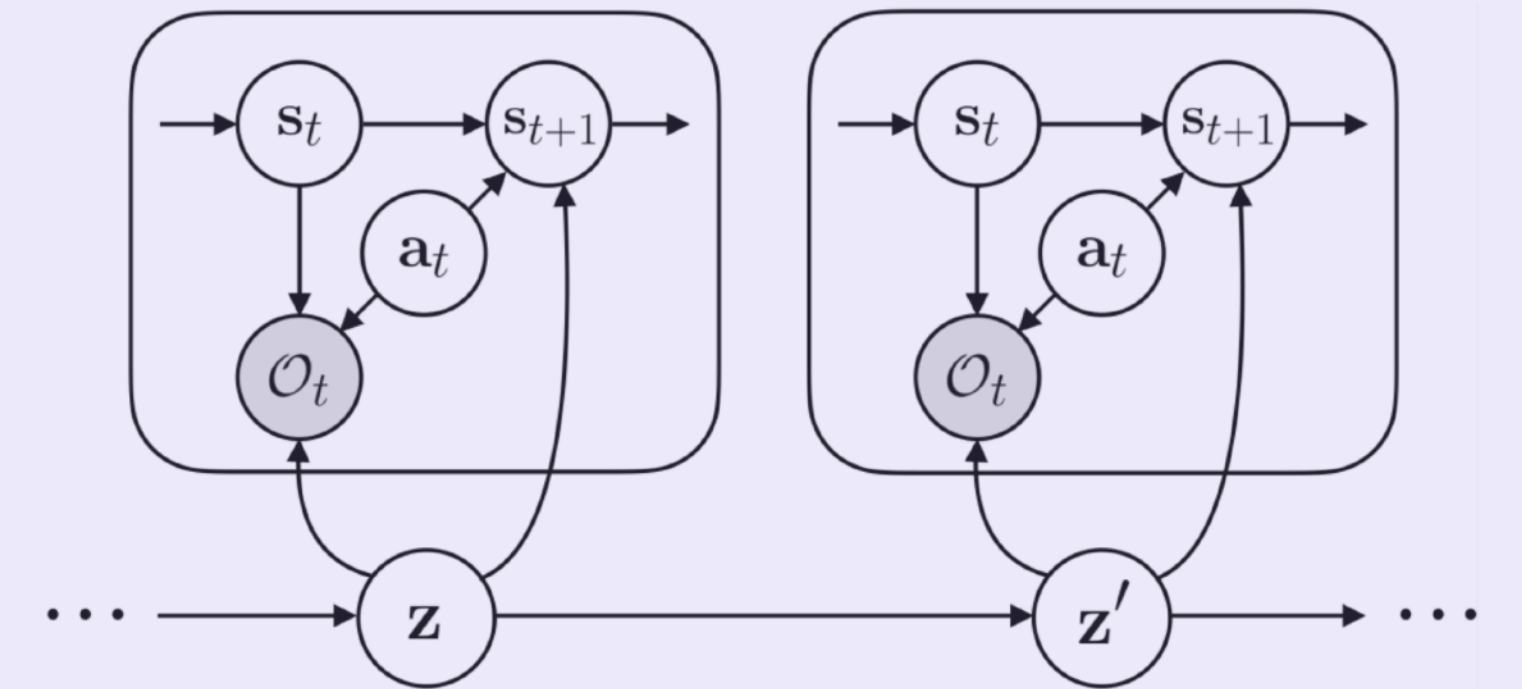
Pessimism

$$\min_{\theta} \sup_{Q \in U(P)} \mathbb{E}_Q[\ell(x, y; \theta)]$$

Adaptation



Anticipation



Introducing
more assumptions

The Principle of Pessimism

“prepare for the worst”

Distributionally robust optimization

(Ben-Tal et al. '13, Duchi et al '16)

$$\min_{\theta} \sup_{Q \in U(P)} \mathbb{E}_Q[\ell(x, y; \theta)]$$

- Adversarial training is a special case (but doesn't prepare the model for natural shifts)
- Common choices for $U(P)$:

+ beautiful, principled framework

Wasserstein DRO: $Q : W_p(P, Q) \leq \epsilon$

CVaR DRO: all distributions over α portion of the data

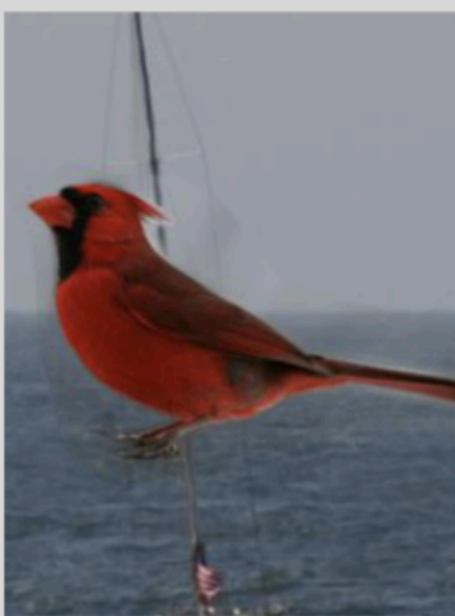
Group DRO: $|\mathcal{D}|$ distributions with mass only over each domain in \mathcal{D}

} very large uncertainty set
(too pessimistic)
<or>


} requires detailed knowledge during training

Do CVaR DRO and Group DRO produce models robust to **spurious correlations**?

Waterbirds

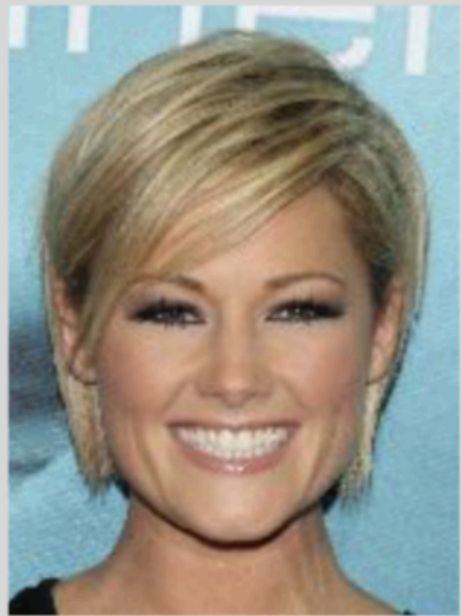


y: land bird
a: in water

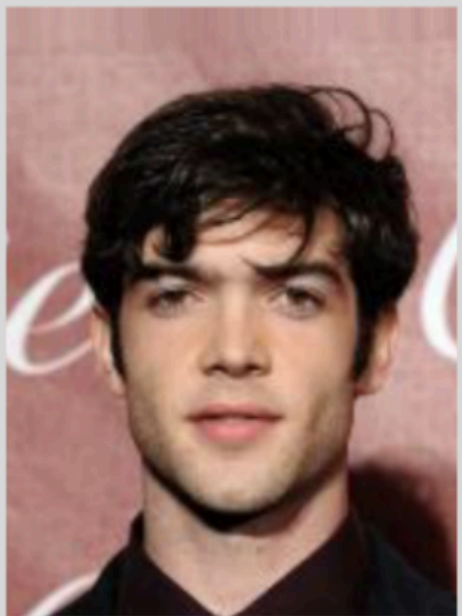


y: land bird
a: on land

CelebA



y: blond
a: female



y: not blond
a: male

MultiNLI

S1: How do you know? All this is their information again.
S2: This information belongs to them.
y: entailment a: no negation

S1: Vrenna and I both fought him and he nearly took us.
S2: Neither Vrenna nor myself have ever fought him.
y: contradiction a: has negation

CivilComments-WILDS

She hates men because that's what her mother taught her.
y: toxic
a: male, female

I doubt that anyone cares whether you believe it or not.
y: non-toxic
a: none

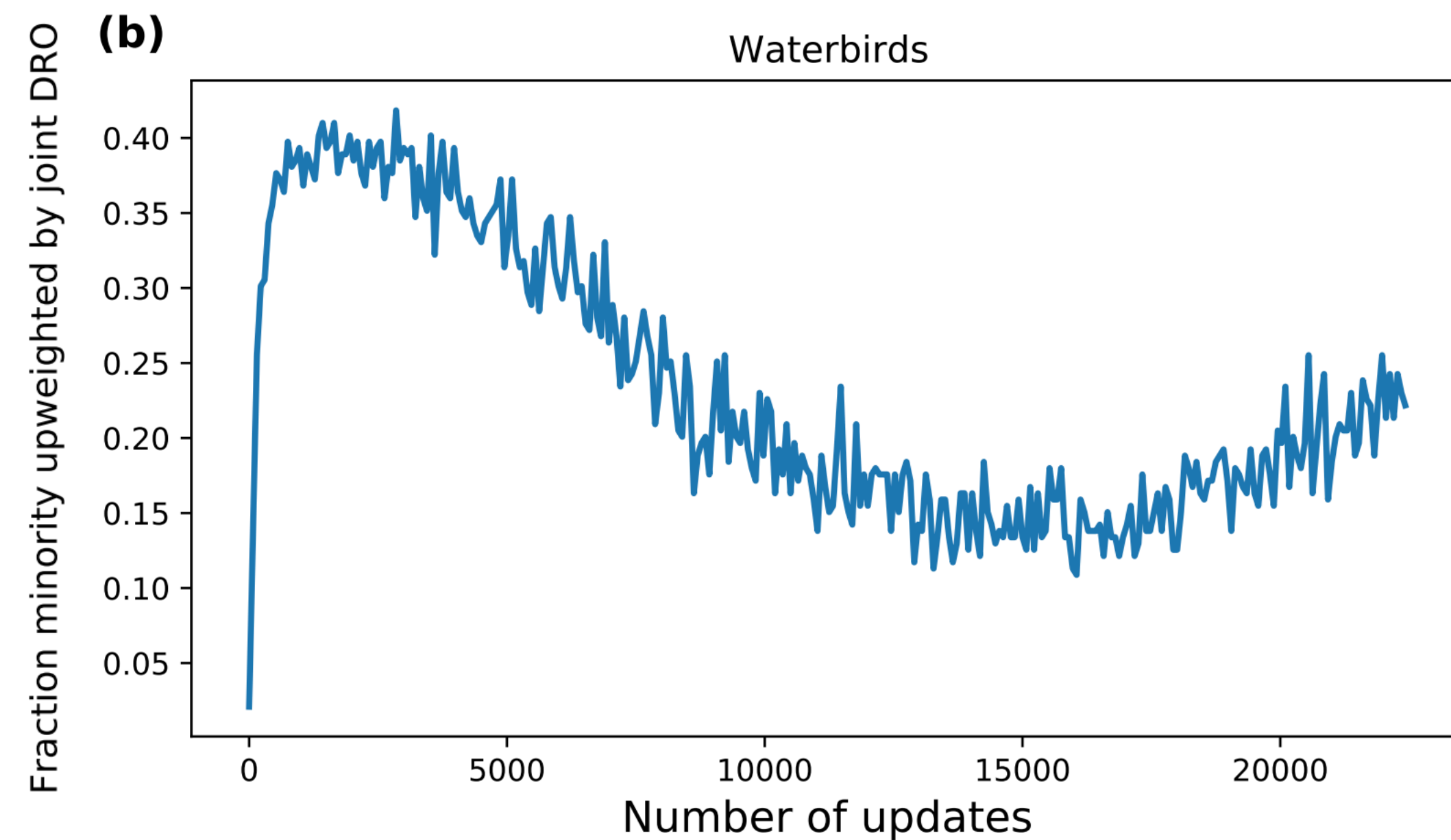
Group labels during training?		Waterbirds		CelebA		MultiNLI		CivilComments-WILDS	
		Average Acc.	Worst-group Acc.	Average Acc.	Worst-group Acc.	Average Acc.	Worst-group Acc.	Average Acc.	Worst-group Acc.
ERM	No	85.86%	72.59%	95.55%	47.22%	82.41%	67.94%	92.55%	59.38%
Joint DRO	No	88.51%	<u>69.47%</u>	82.35%	<u>64.44%</u>	81.95%	<u>68.03%</u>	92.46%	<u>56.63%</u>
Group DRO	Yes	89.47%	<u>85.72%</u>	92.97%	<u>87.22%</u>	80.31%	<u>75.26%</u>	88.88%	<u>69.92%</u>

Group DRO does well 😊, but requires group labels during training 😞

Joint DRO shows little gains over ERM 😞

Why doesn't joint DRO work for spurious correlations?

Joint DRO objective is *less directed at minority groups* than group DRO



—> doesn't sufficiently prioritize examples where spurious correlations don't hold



Evan Liu



Behzad Haghighi



Annie Chen

High-Level Approach

Stage 1: Automatically identify examples where spurious correlations don't hold

Recall: ERM performs poorly on these examples.

Stage 2: Prioritize these examples.

Complete algorithm

1. Train *identification model* $g_\phi(x)$ via ERM.
2. Compute *error set* as misclassified examples.

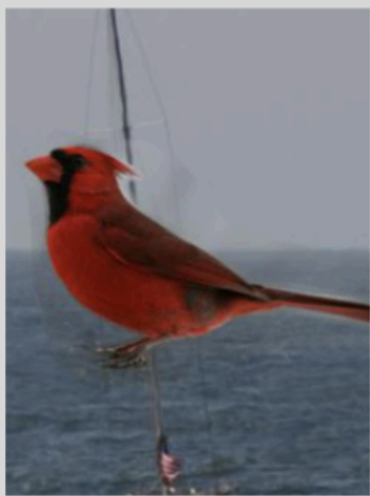
$$E = \{(x, y) : g_\phi(x) \neq y\}$$

3. Upsample examples from E in train set.
4. Train *final model* f_θ on upsampled dataset


“just train twice” (JTT)

Experimental Results

Waterbirds

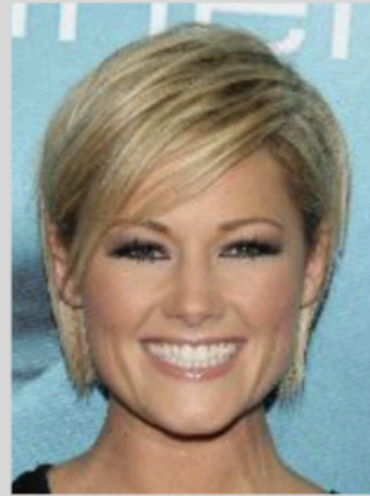


y: land bird
a: in water




y: land bird
a: on land

CelebA



y: blond
a: female



y: not blond
a: male

MultiNLI

S1: How do you know? All this is their information again.
S2: This information belongs to them.

y: entailment a: no negation

S1: Vrenna and I both fought him and he nearly took us.
S2: Neither Vrenna nor myself have ever fought him.

y: contradiction a: has negation

CivilComments-WILDS

She hates men because that's what her mother taught her.

y: toxic
a: male, female

I doubt that anyone cares whether you believe it or not.

y: non-toxic
a: none

Comparing to:

- ERM
- Joint DRO
- LfF: representative recent approach with strong results

All methods tuned w.r.t. worst-group val loss.

Method	Group labels during training?	Waterbirds		CelebA		MultiNLI		CivilComments-WILDS	
		Average Acc.	Worst-group Acc.	Average Acc.	Worst-group Acc.	Average Acc.	Worst-group Acc.	Average Acc.	Worst-group Acc.
ERM	No	85.86%	72.59%	95.55%	47.22%	82.41%	67.94%	92.55%	59.38%
Joint DRO	No	88.51%	69.47%	82.35%	64.44%	81.95%	68.03%	92.46%	56.63%
LfF (Nam et al., 2020)	No	91.56%	75.23%	85.96%	70.56%	80.77%	70.20%	92.52%	58.77%
JTT (Ours)	No	90.33%	86.03%	87.96%	81.11%	80.38%	72.29%	91.07%	69.31%
Group DRO	Yes	89.47%	85.72%	92.97%	87.22%	80.31%	75.26%	88.88%	69.92%

>10% improvement in worst-group accuracy on 3 of 4 datasets

on 2 datasets, JTT is comparable to group DRO

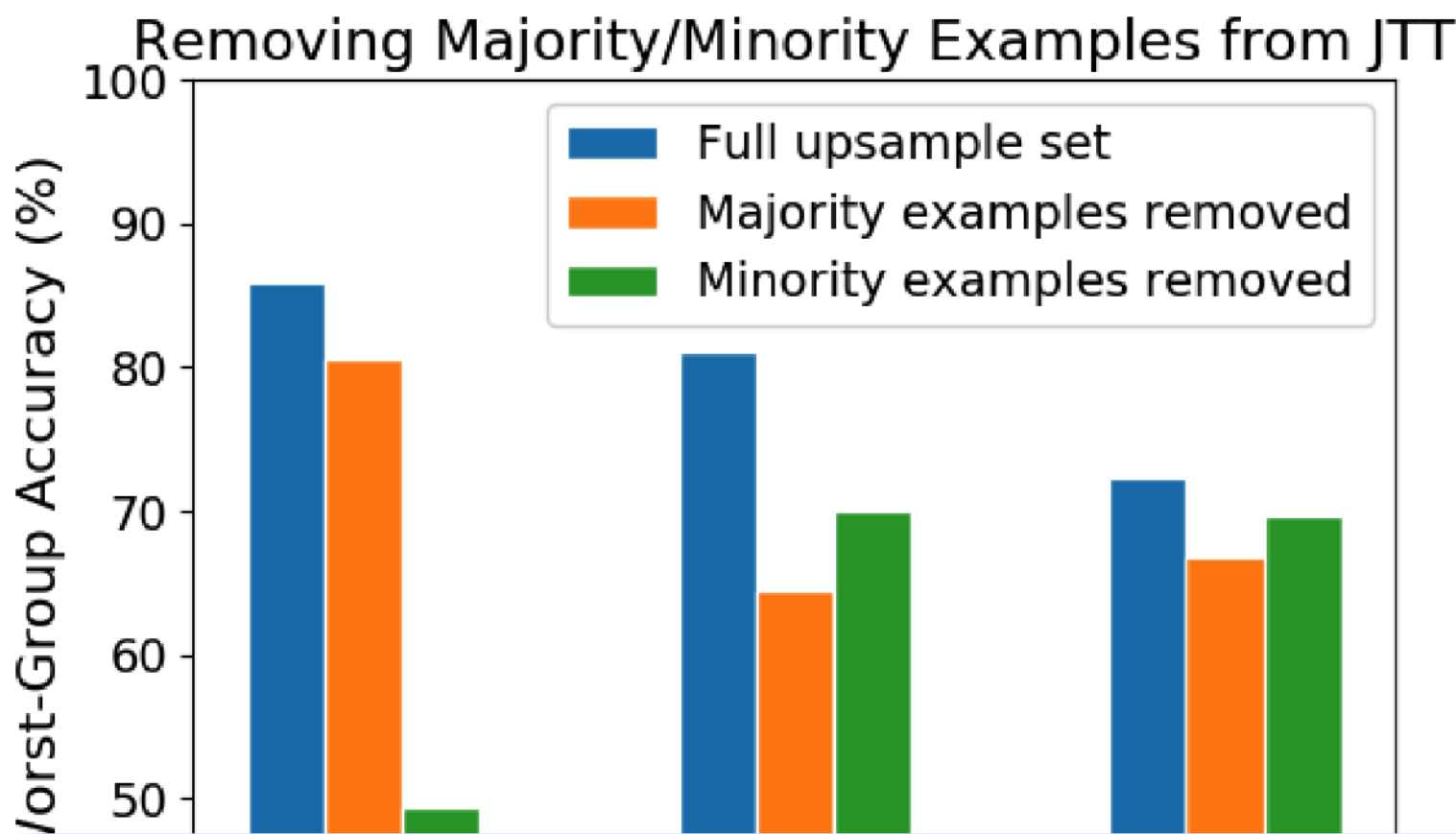
What datapoints does JTT identify?

What portion of the error set is minority examples?

Dataset	Minority-group Precision	Minority-group Empirical Rate
Waterbirds	45.5%	5.0%
CelebA	23.2%	44.9%
MultiNLI	38.4%	28.6%
CivilComments	72.2%	10.7%

—> much higher rate of minority points in the error set vs. empirical distribution

What if we remove the majority or minority points from the error set?



—> both majority and minority points are helpful!

water bird + water background

In error set



Not in error set



Takeaway: Can achieve robustness to spurious correlations by prioritizing hard datapoints.

ous attribute is more ambiguous in identified points.

Aside: What about pessimism for reinforcement learning?

Key idea Be pessimistic about value of OOD states, actions.
—> avoid visiting out-of-distribution states

Minimize $Q(s, a)$ for OOD actions a .

Kumar, Zhou, Tucker, Levine. Conservative Q-Learning for Offline RL. NeurIPS '20

Reward penalty when model is uncertain

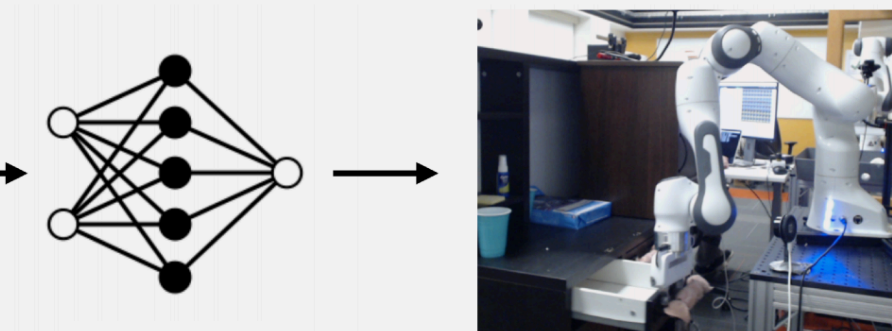
Yu*, Thomas*, Yu, Ermon, Zou, Levine, Finn, Ma. MOPO: Model-based Offline Policy Optimization. NeurIPS '20

Minimize $Q(s, a)$ on OOD states & actions (s, a)

Yu*, Kumar*, Rafailov, Rajeswaran, Levine, Finn. COMBO: Conservative Offline Model-Based Policy Optimization. '21

All come with theoretical performance guarantees!

Offline RL



Distribution shift between **policy in the dataset** and the **policy being optimized**.

By incorporating pessimism:



76% success rate

Rafailov*, Yu*, Rajeswaran, Finn. *Offline Reinforcement Learning from Images with Latent Space Models*, arXiv '20

Tools for tackling distribution shift

Pessimism

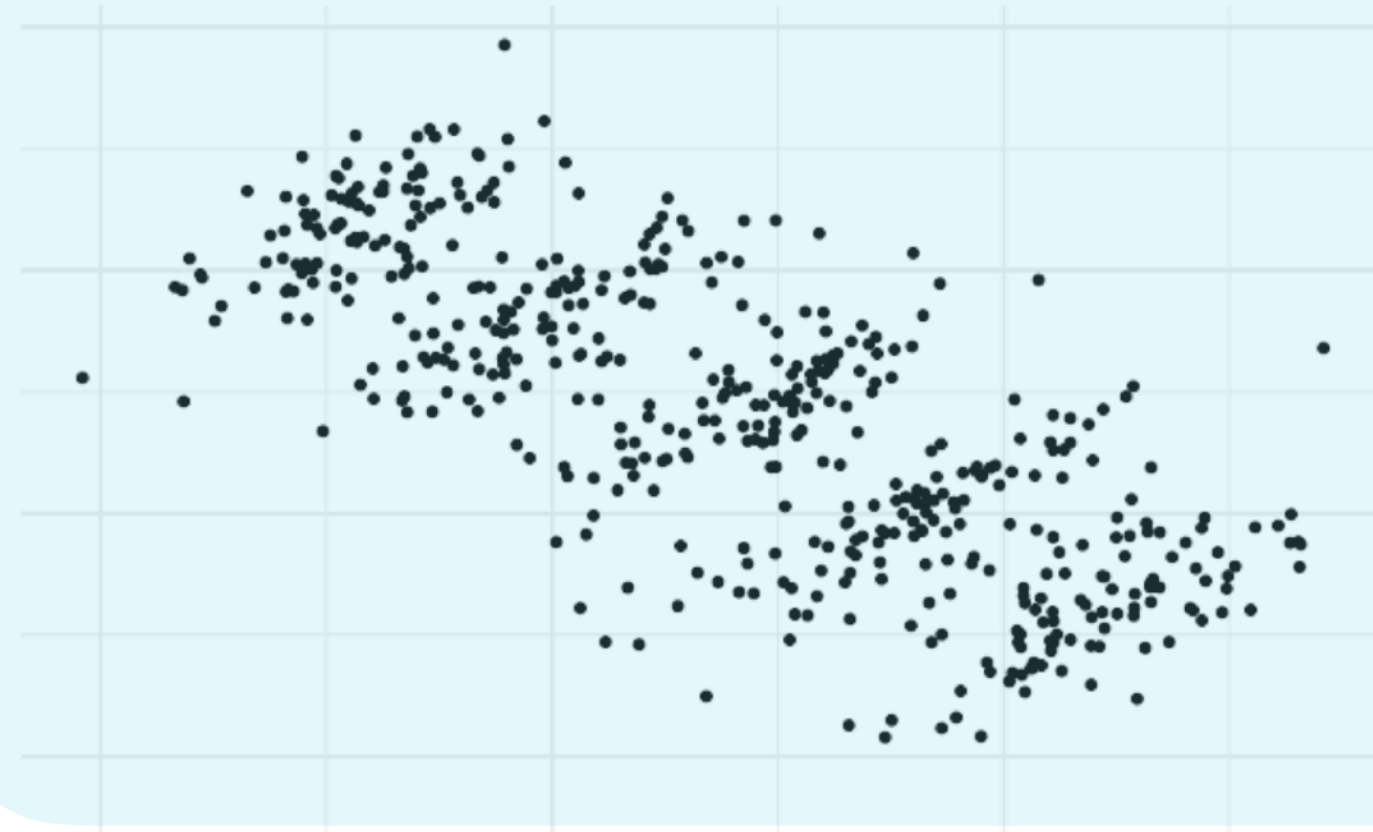
$$\min_{\theta} \sup_{Q \in U(P)} \mathbb{E}_Q[\ell(x, y; \theta)]$$

+ powerful tool for addressing **spurious correlations** and **policy distribution shift**

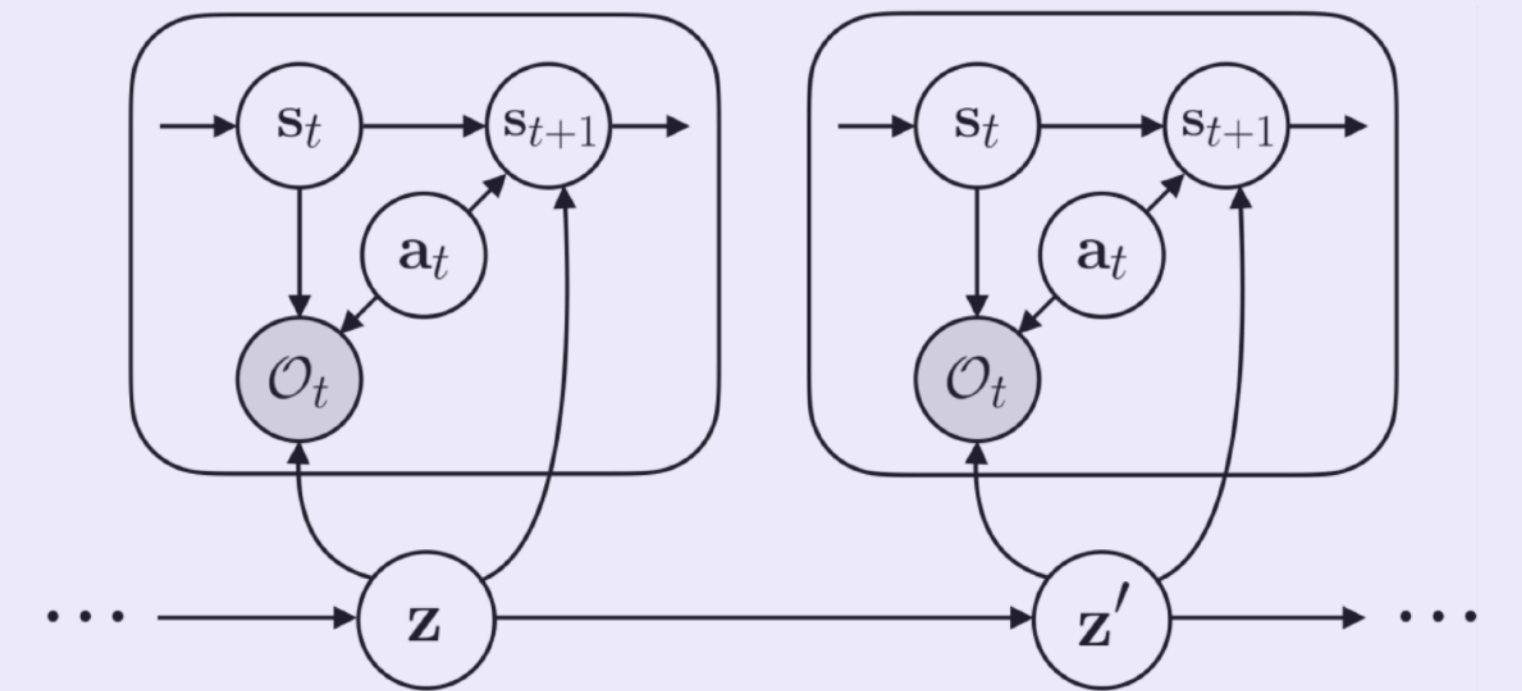
+ makes few assumptions

+ often possible to analyze theoretically

Adaptation

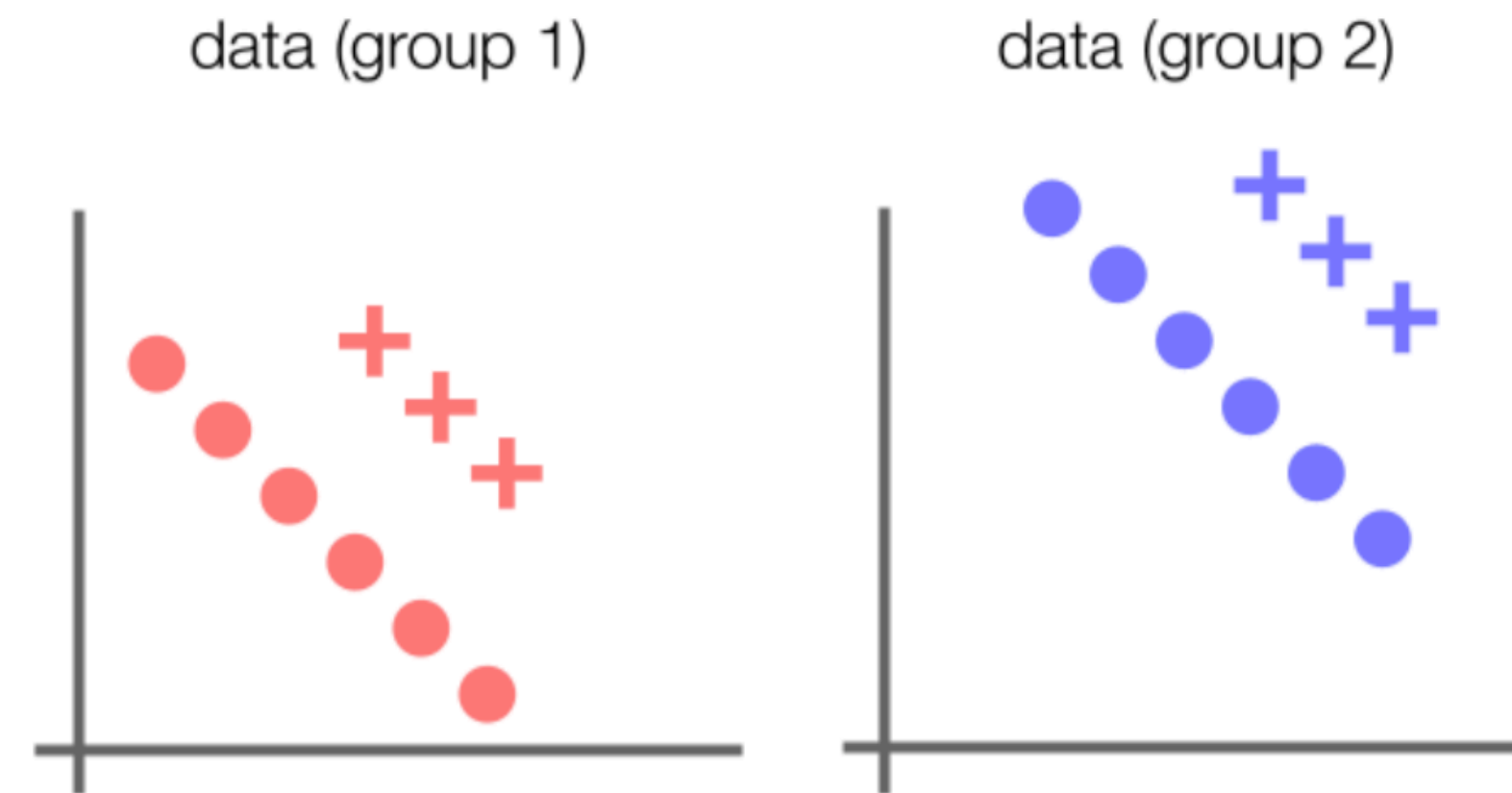


Anticipation



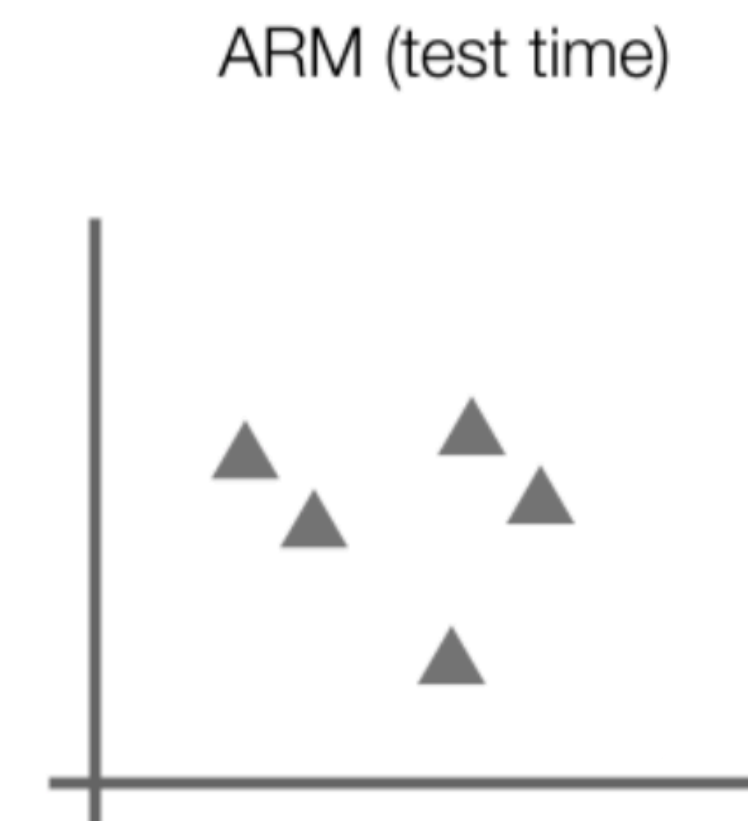
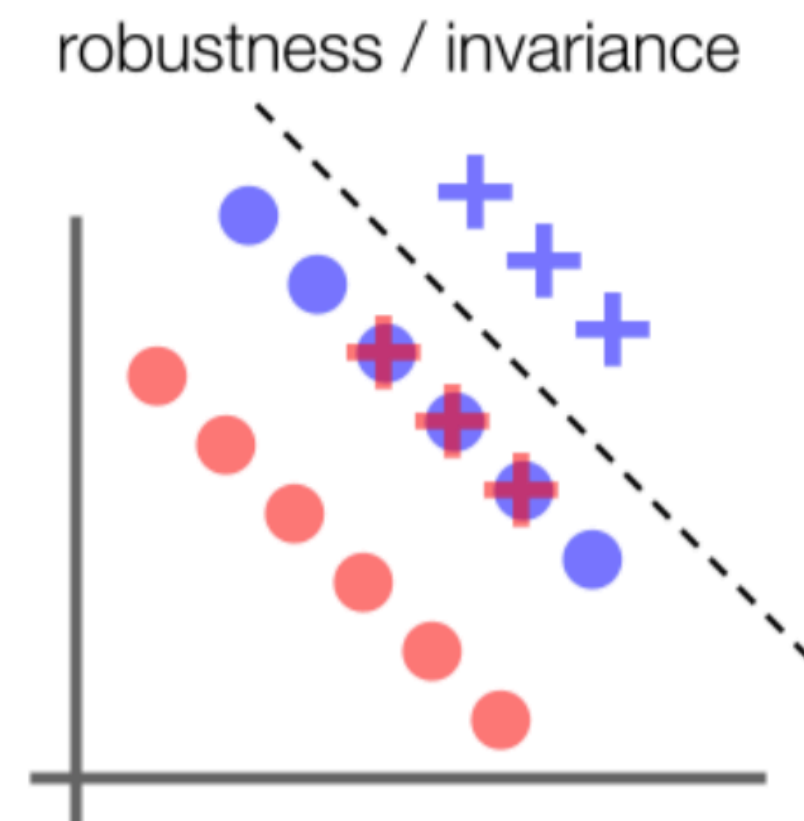
Introducing
more assumptions

Why adapt? A simple example



Robustness approaches cannot solve this form of distribution shift.

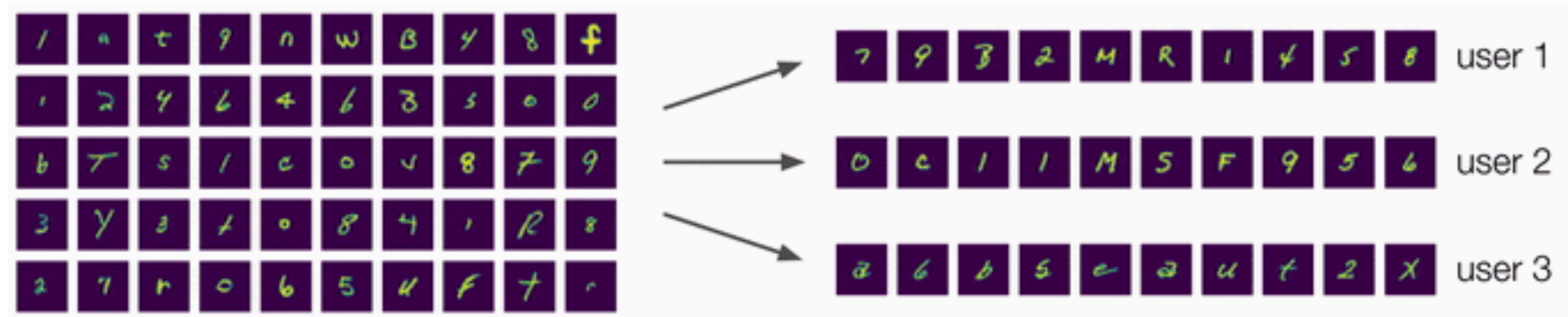
If we see enough groups during training & see unlabeled points at test time:



Potential solutions: domain adaptation, transfer learning, meta-learning

Motivating problem setting: federated learning

e.g. federated
handwriting
recognition



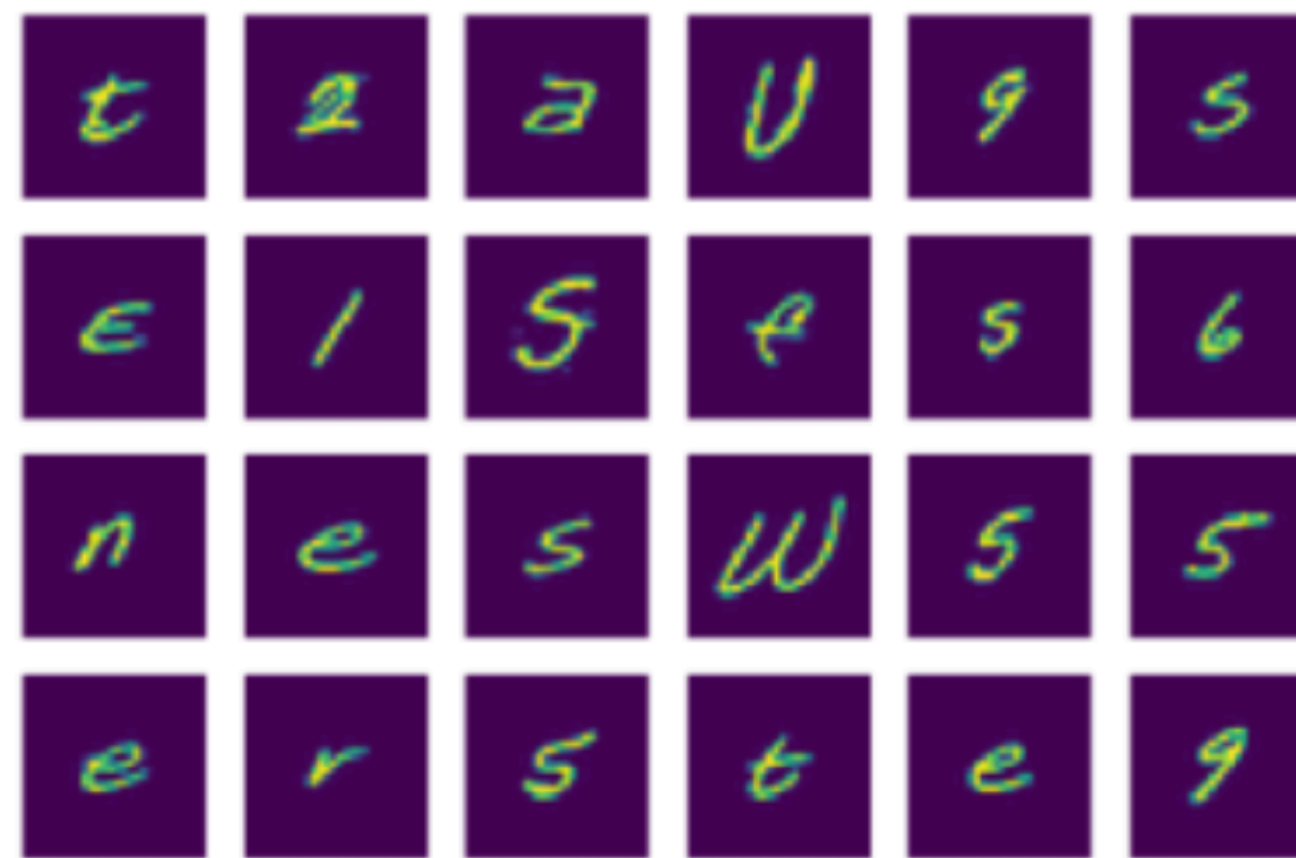
+ possibly many different target domains

+ want to adapt on the fly with minimal data, labels, and compute

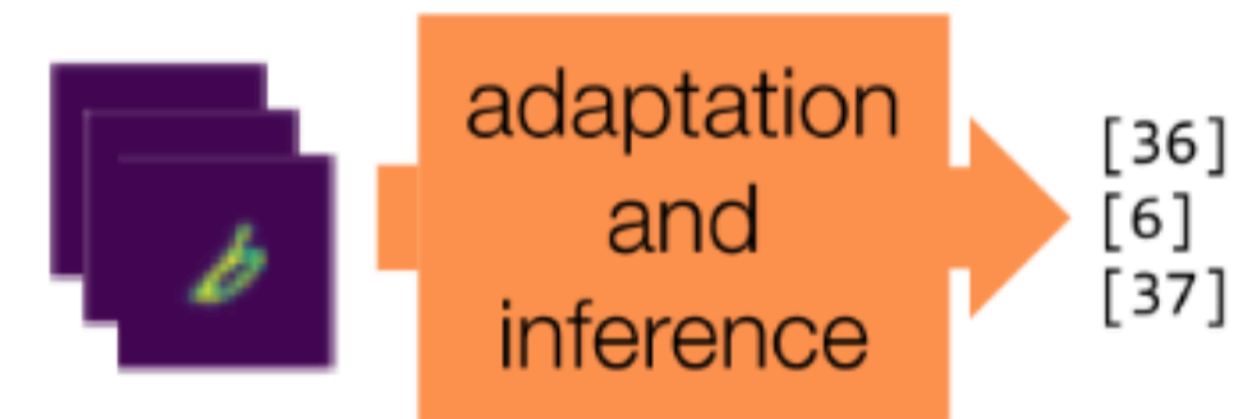
Adaptive Risk Minimization

Test time

unlabeled data from new test domain
(e.g. new user, different time-of-day, new place)



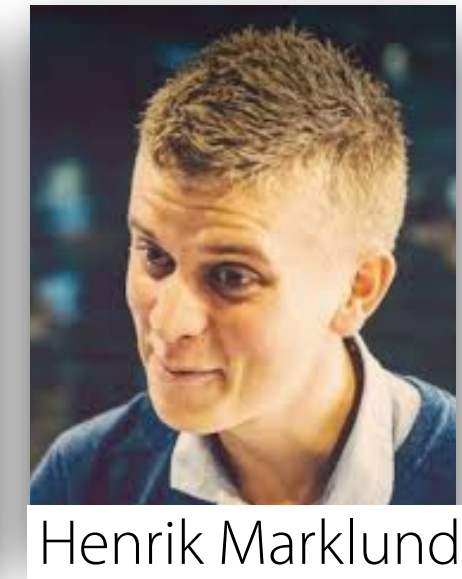
adapt model & infer labels



Assumption: test inputs from one group available in a batch or streaming.



Marvin Zhang



Henrik Marklund



Nikita Dhawan

Adaptive Risk Minimization

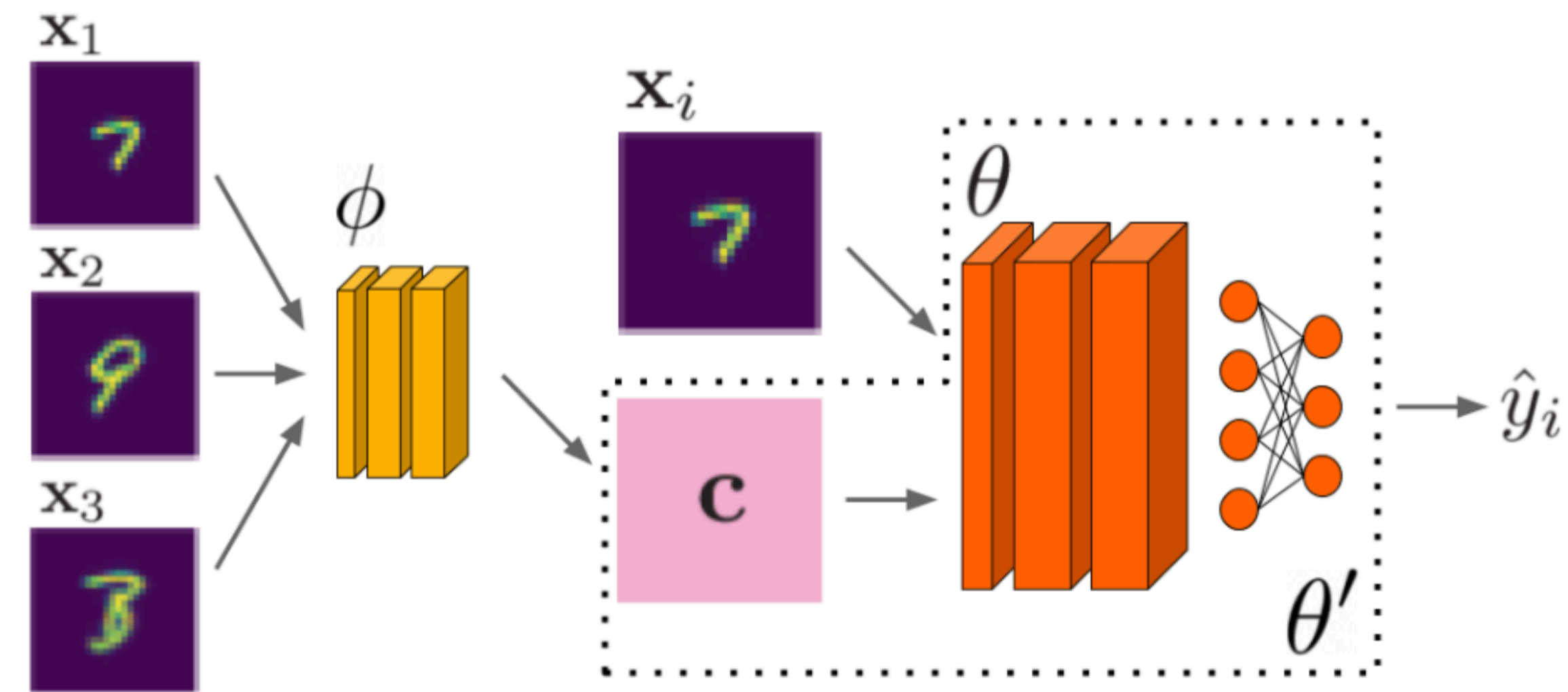
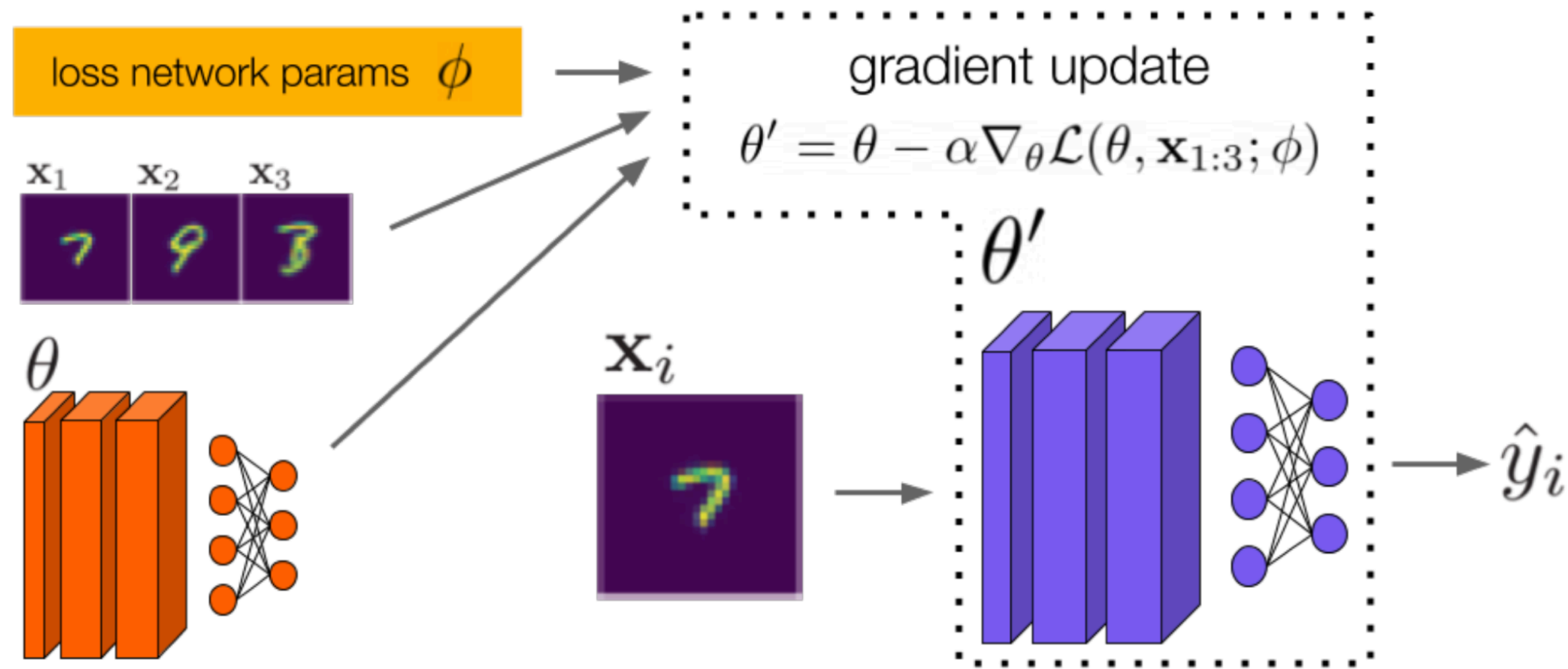
Train time

1. Separate training data into domains
2. Train for model that can adapt with only unlabeled examples.

How to adapt with unlabeled data?

MAML with learned loss

or meta-learning with context variable



Simplest setting: context = BN statistics

Experimental Comparisons

ERM - standard deep network training

DRNN - distributional robustness
(Sagawa, Koh et al. ICLR '20)

UW - ERM but upweight groups to the uniform distribution

ARM - adaptive risk minimization

ARM-CML - adapt with context variable

ARM-BN - adapt using batch norm stats

ARM-LL - adapt with learned loss

Experiment 1. Federated Extended MNIST (Cohen et al. 2017, Caldas et al. 2019)

Distribution shift: adapt to *new* users with only unlabeled data

Method	FEMNIST	
	WC	Avg
ERM	62.9 \pm 1.9	80.1 \pm 0.9
UW*	61.8 \pm 0.9	80.1 \pm 0.3
DRNN	58.1 \pm 0.7	74.4 \pm 0.8
<i>q</i> -FedAvg [37]	58.2 \pm 1.0	80.8 \pm 0.3
ARM-CML	67.8 \pm 1.3	85.7 \pm 0.3
ARM-BN	72.6 \pm 0.3	85.7 \pm 0.1
ARM-LL	69.6 \pm 2.1	85.6 \pm 0.5

+ 5% improvement in average accuracy
+ 10% improvement in worst-case accuracy

ARM - adaptive risk minimization

DRNN - distributional robustness
(Sagawa, Koh et al. ICLR '20)

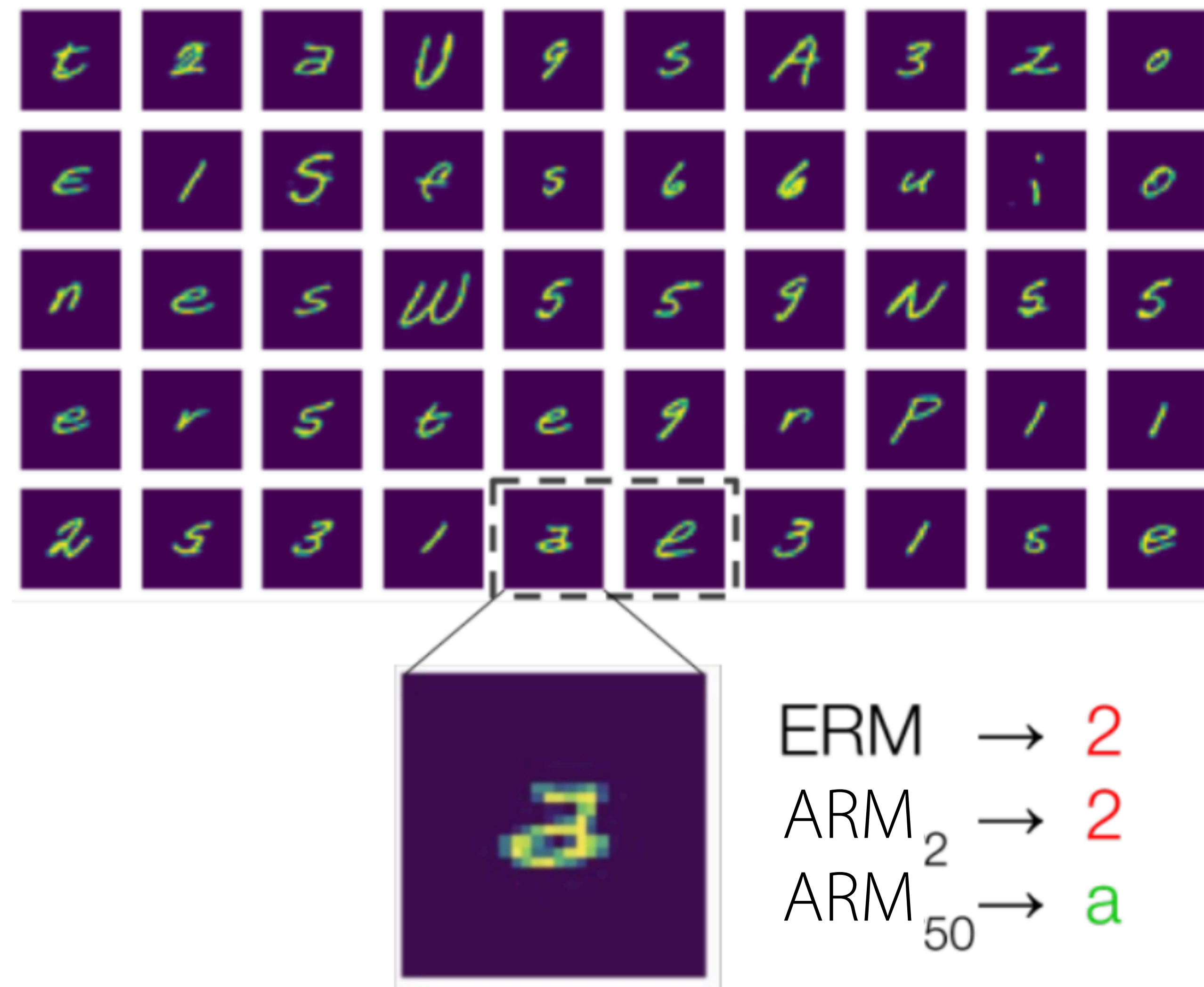
ERM - standard deep network training

UW - ERM but upweight groups to
the uniform distribution

***q*-FedAvg** (Li et al. 2020) - federated learning method

Experiment 1. Federated Extended MNIST (Cohen et al. 2017, Caldas et al. 2019)

Distribution shift: adapt to *new* users with only unlabeled data



Experiment 2. CIFAR-C, TinyImageNet-C (Hendrycks & Dietterich, 2019)

Distribution shift: adapt to *new* image corruptions
(train using 56 corruptions, test using 22 disjoint corruptions)

Method	CIFAR-10-C		Tiny ImageNet-C	
	WC	Avg	WC	Avg
ERM	49.6 ± 0.1	69.8 ± 0.4	19.3 ± 0.5	41.4 ± 0.2
UW*	—	—	—	—
DRNN	44.5 ± 0.5	70.7 ± 0.6	19.9 ± 0.3	41.6 ± 0.2
ARM-CML	67.7 ± 0.5	79.2 ± 0.3	21.4 ± 0.2	43.3 ± 0.4
ARM-BN	71.1 ± 0.1	80.9 ± 0.2	27.7 ± 0.2	44.9 ± 0.2
ARM-LL	66.9 ± 0.2	75.7 ± 0.3	27.1 ± 0.3	44.2 ± 0.4

+ 3-10% improvement in average accuracy

+ 8-21% improvement in worst-case accuracy

ARM - adaptive risk minimization

DRNN - distributional robustness

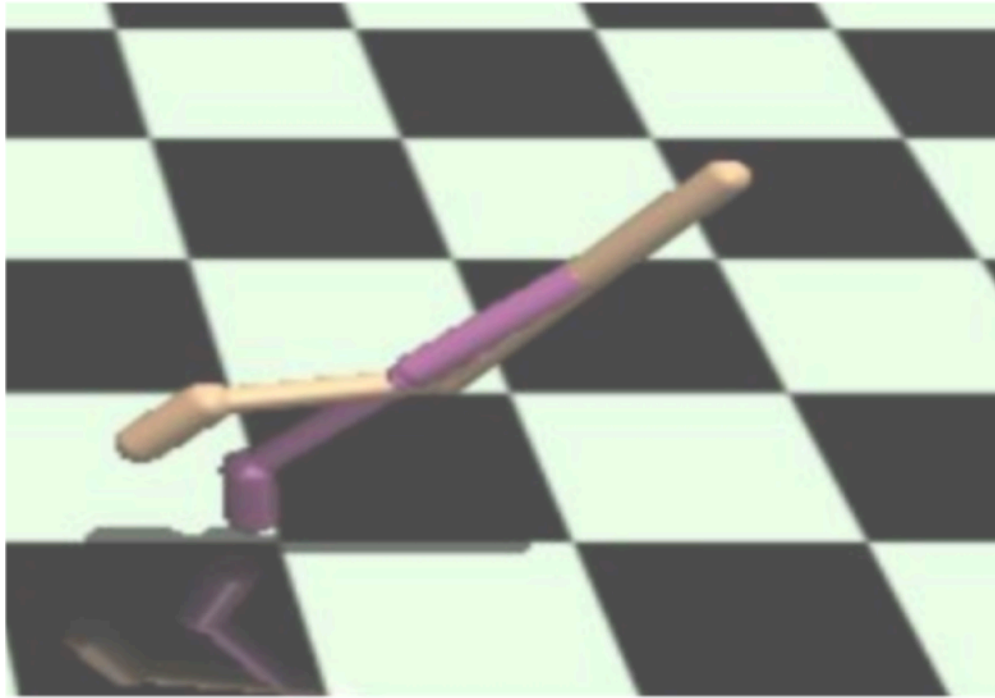
ERM - standard deep network training

UW - ERM but upweight groups to the uniform distribution

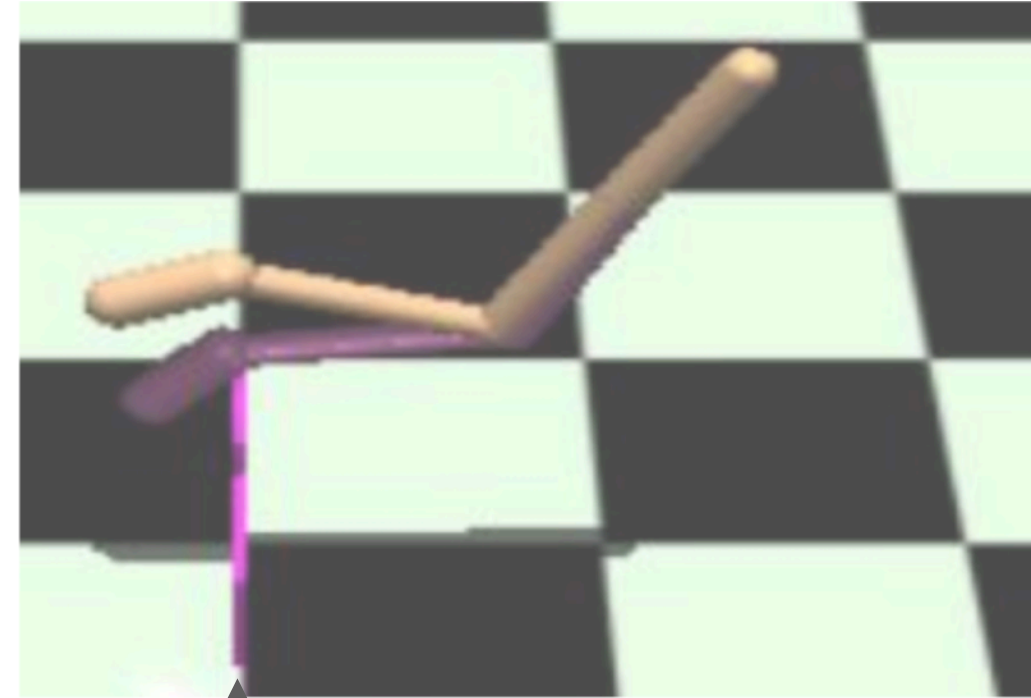
Takeaway: Small amount of unlabeled data provides leverage for distribution shift.

Can you learn to adapt without known training groups? (time-permitting)

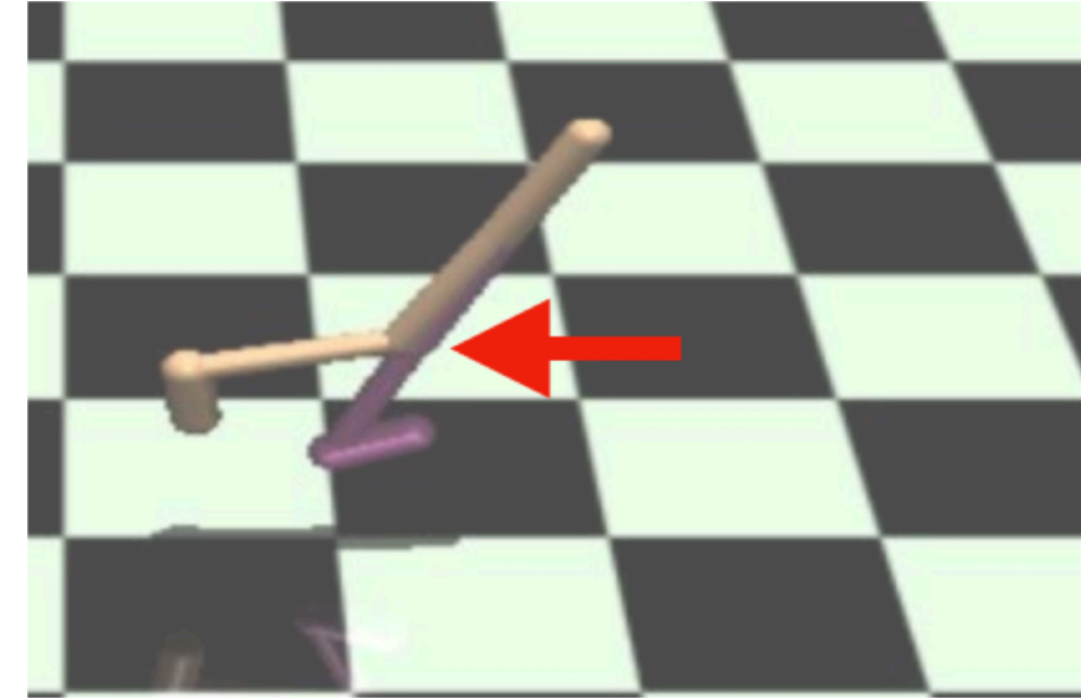
one training
environment M_{train}



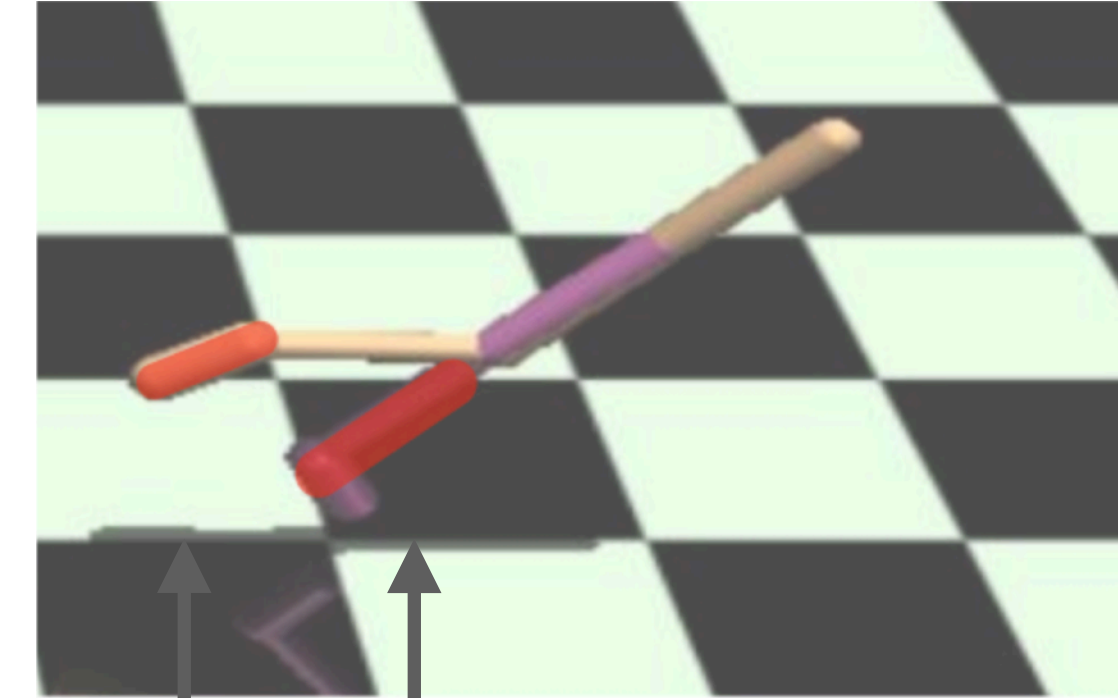
new test environments M_{test}



obstacle



force perturbation

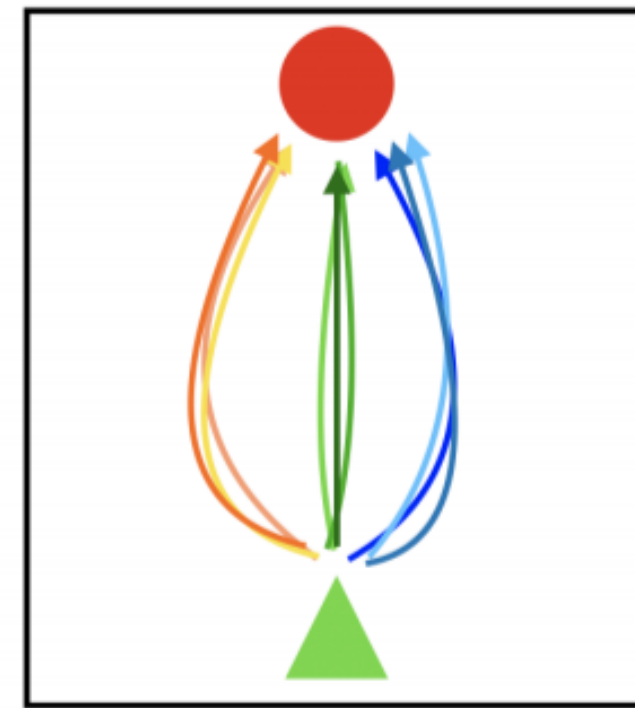


disabled joints

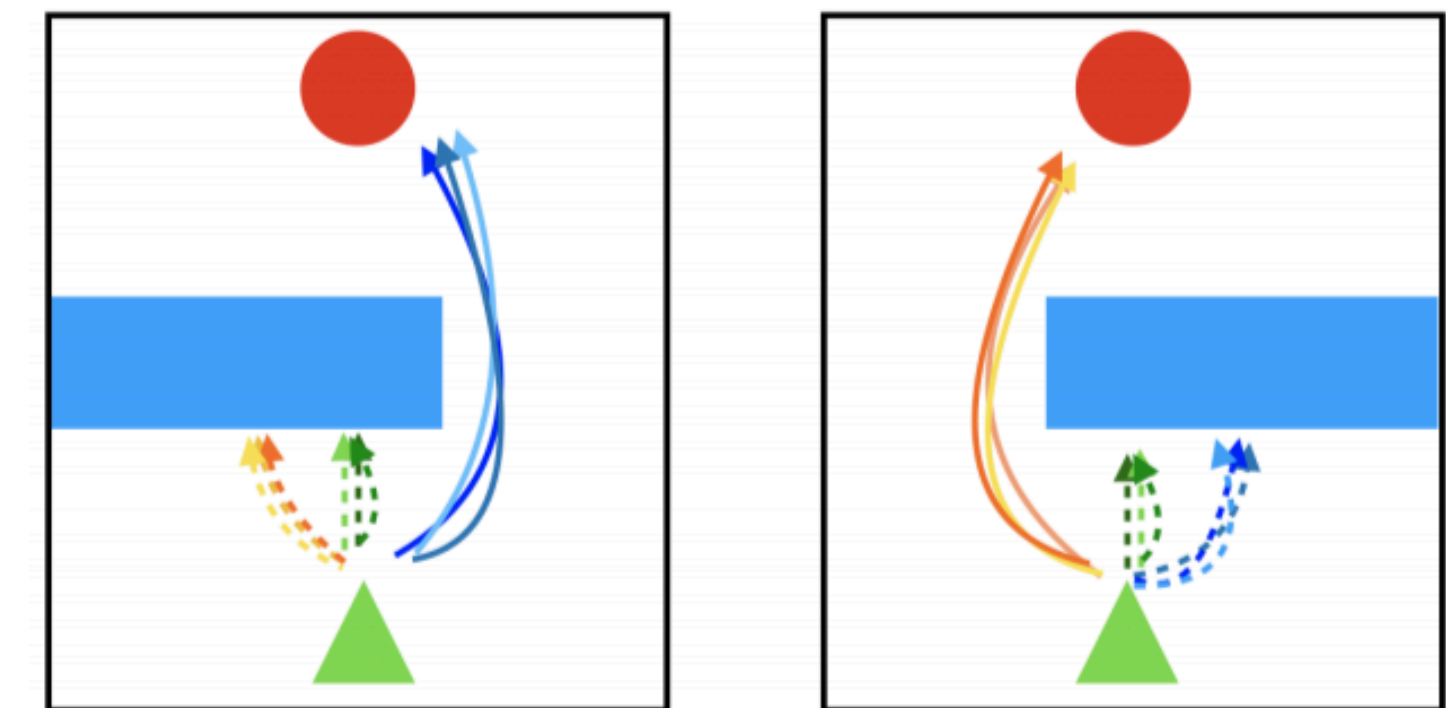
Can you learn to adapt without known training groups?

Simple idea:

Learn & *remember* multiple solutions to M_{train}



Adapt solution set to M_{test}



Assumption #1: ability to adapt with modest amount of data

Assumption #2: changes to the environment are local such that the optimal policy in M_{test} also does well in M_{train}

e.g., few-shot robustness to local changes in obstacles, terrains, friction, etc



Saurabh Kumar

How to learn multiple solutions?

Learn controllable space of diverse policies that achieve return with ϵ of optimal

using latent variables

$$\pi_{\theta}(a | s, z)$$

constrained optimization

Train time:

$$\arg \max_{\theta} \sum_{t=1}^T \underbrace{I(s_t; z)}_{\mathcal{H}(s) - \mathcal{H}(s | z)} \quad \text{s.t.} \quad \forall z, \underline{R_{\mathcal{M}}(\pi_{\theta}) \geq R_{\mathcal{M}}(\pi_{\mathcal{M}}^*) - \epsilon}$$

Test time: Roll-out K policies with different z . Return $\pi_{\theta}(a | s, z_i)$ for best performing z_i .

“structured maximum entropy RL” (SMERL)

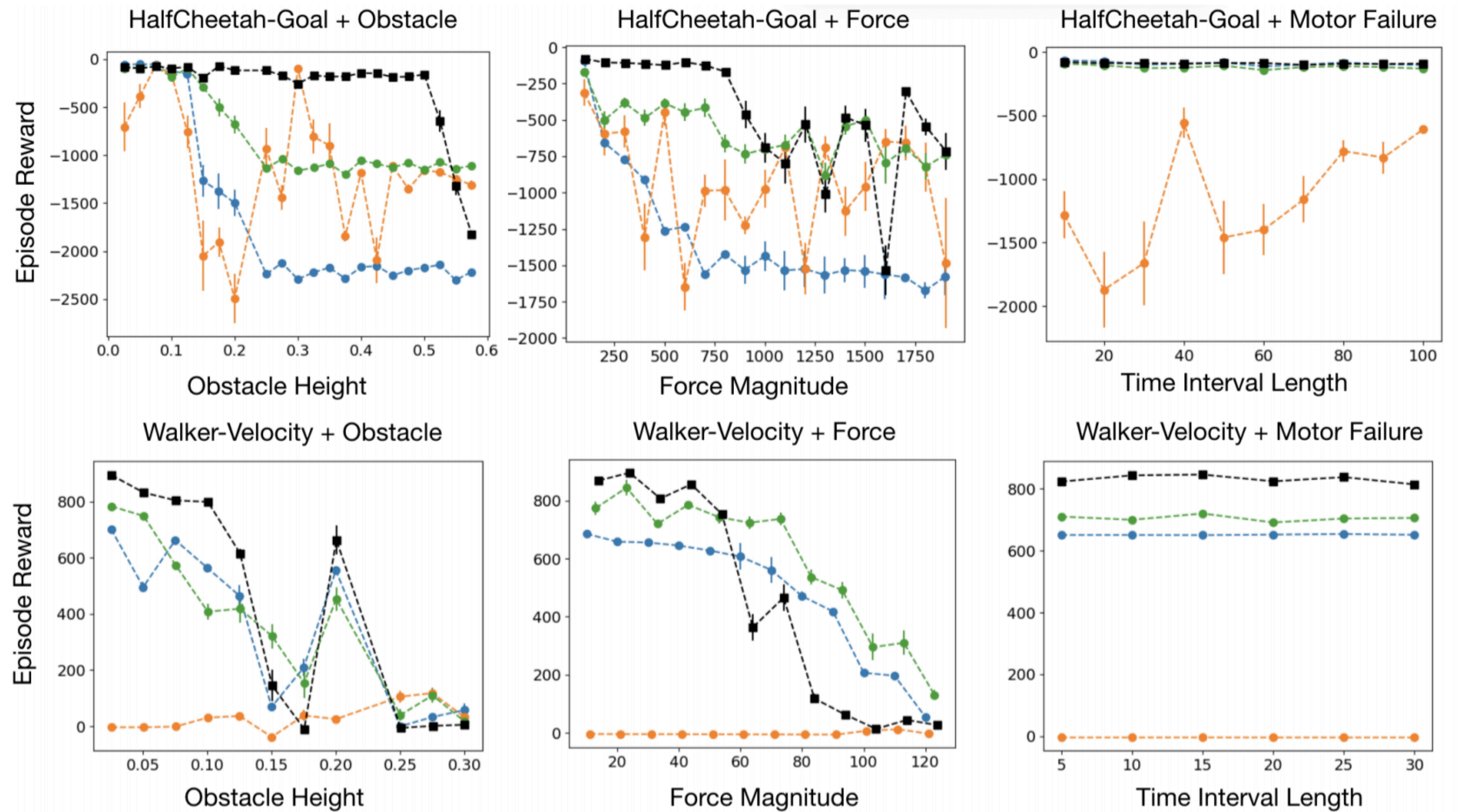
Testing Robustness to Obstacles, Perturbations, and Motor Failures

Compare:



Measuring 5-shot generalization.

performance

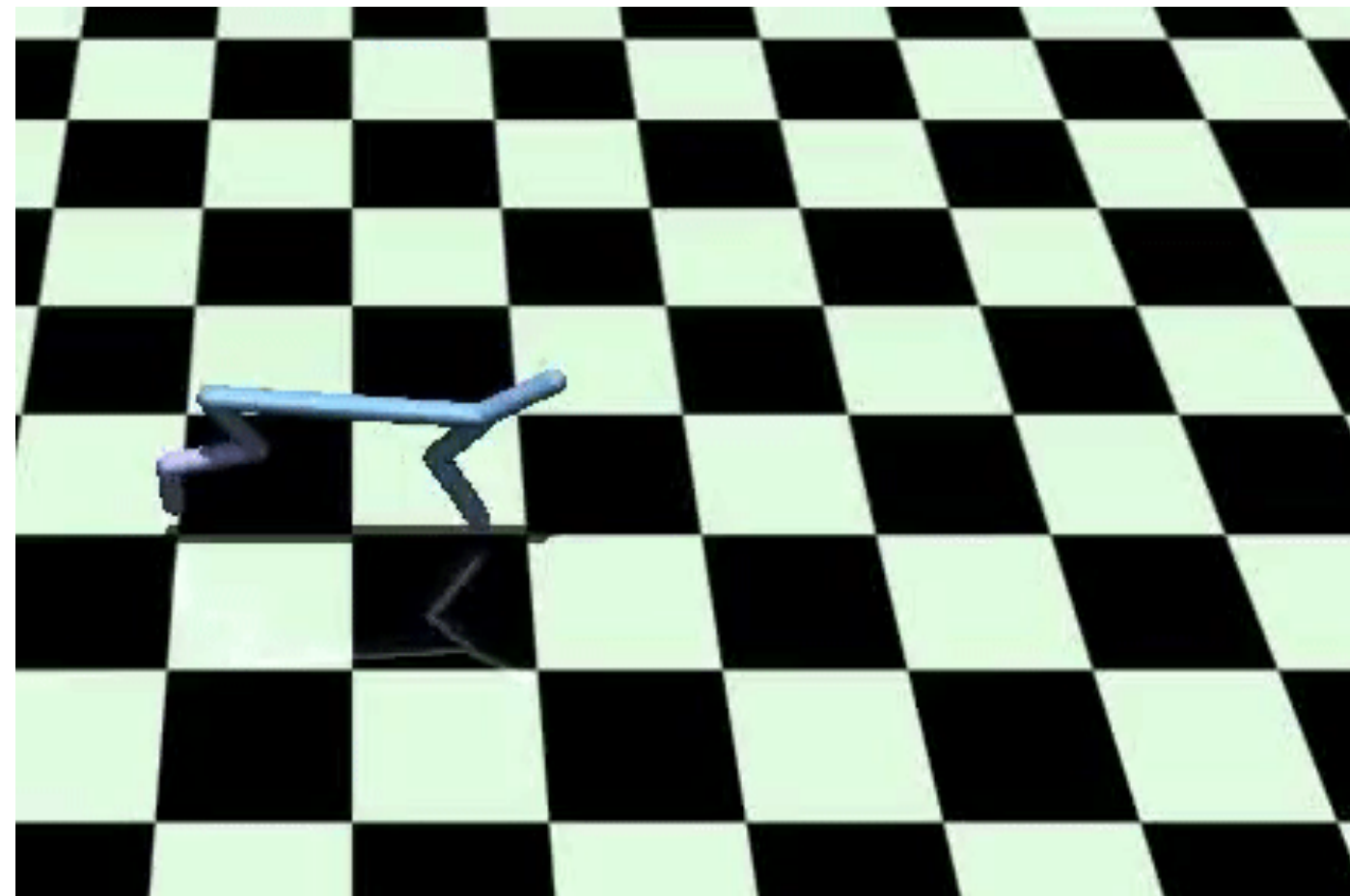


degree of environment change

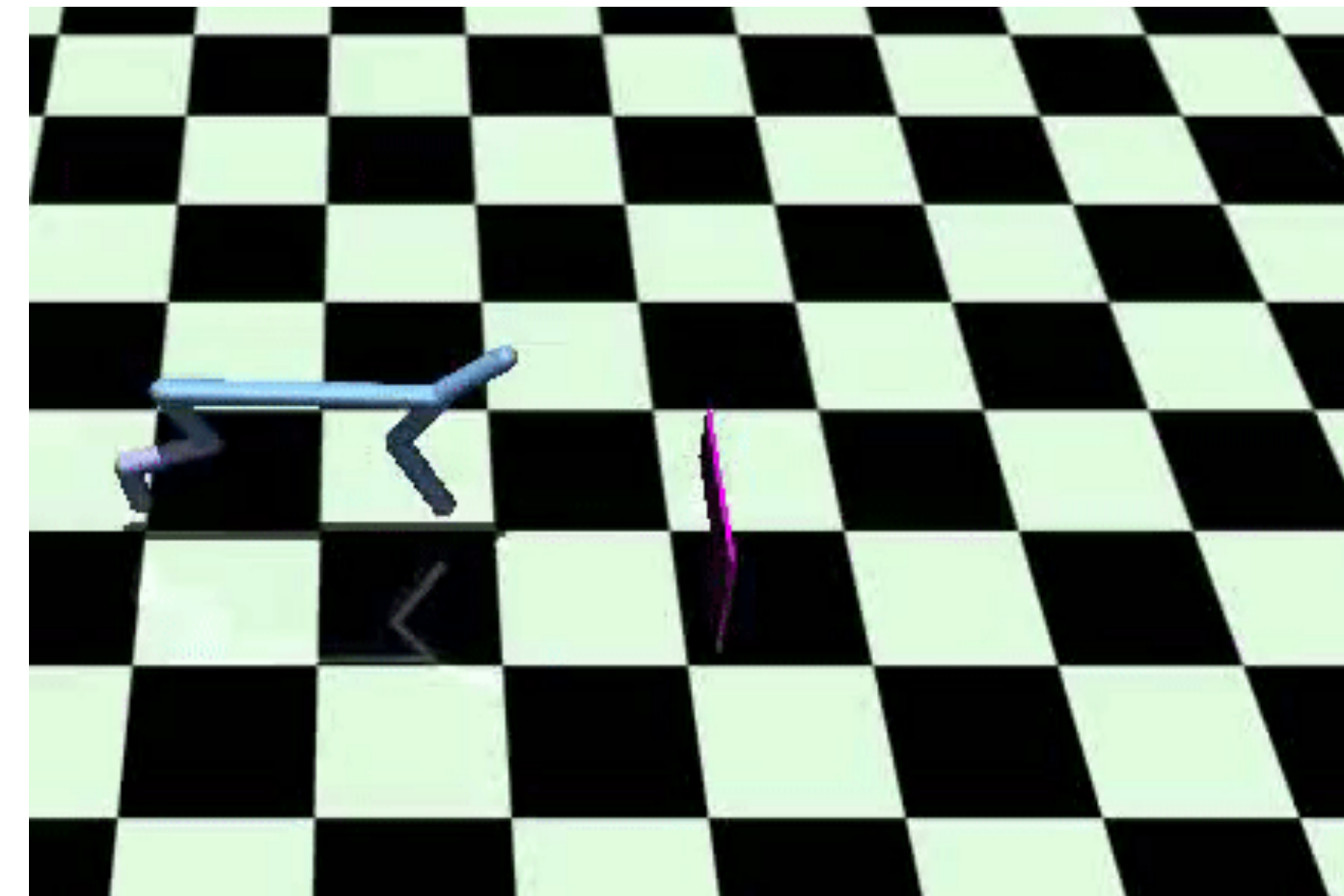
Pinto, Davidson, Sukthankar, Gupta. *Robust Adversarial Reinforcement Learning*, ICML '17

S. Kumar, A. Kumar, Levine, Finn. *One Solution is Not All You Need: Few-Shot Extrapolation via Structured MaxEnt RL*, NeurIPS '20

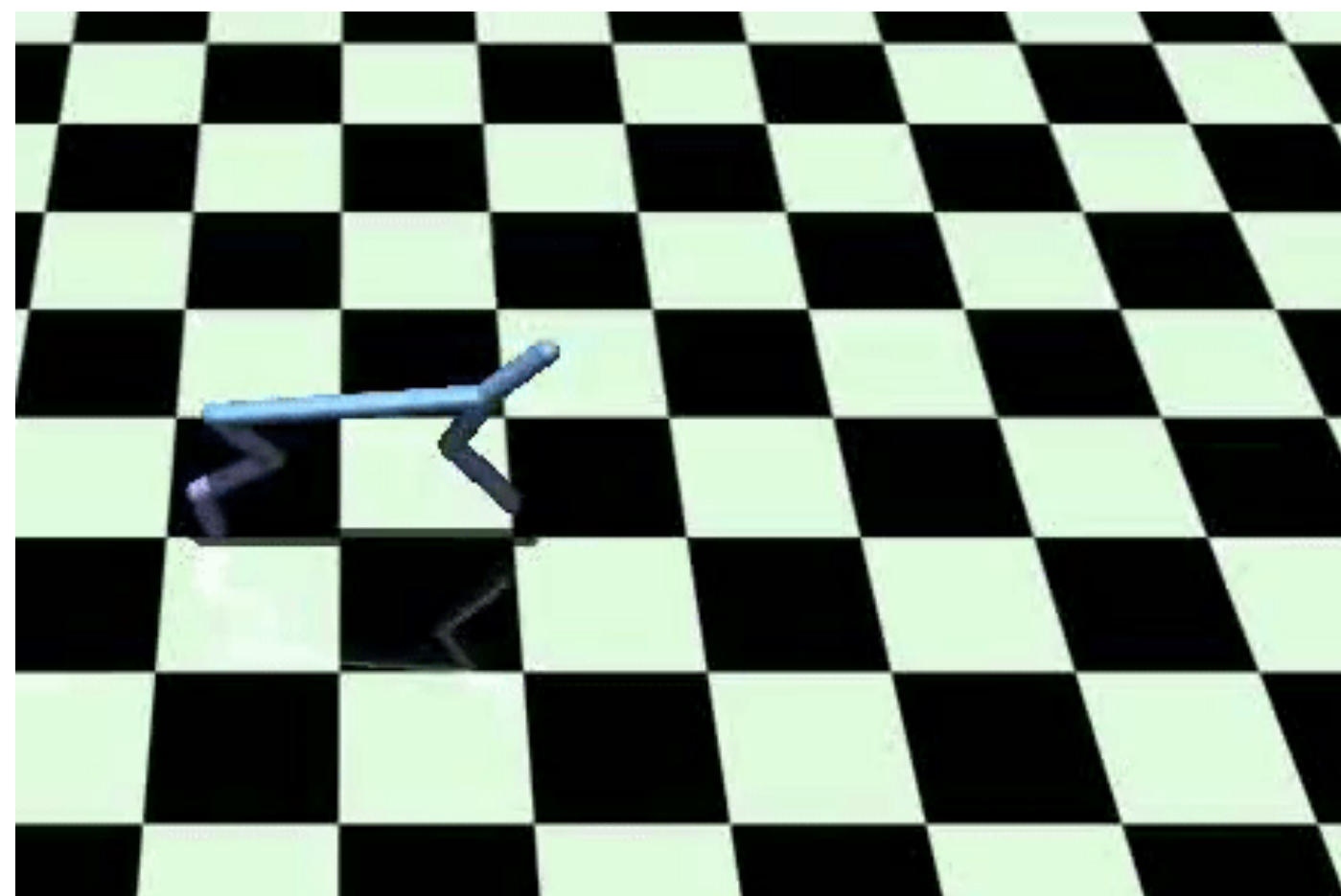
SAC policies at train time.



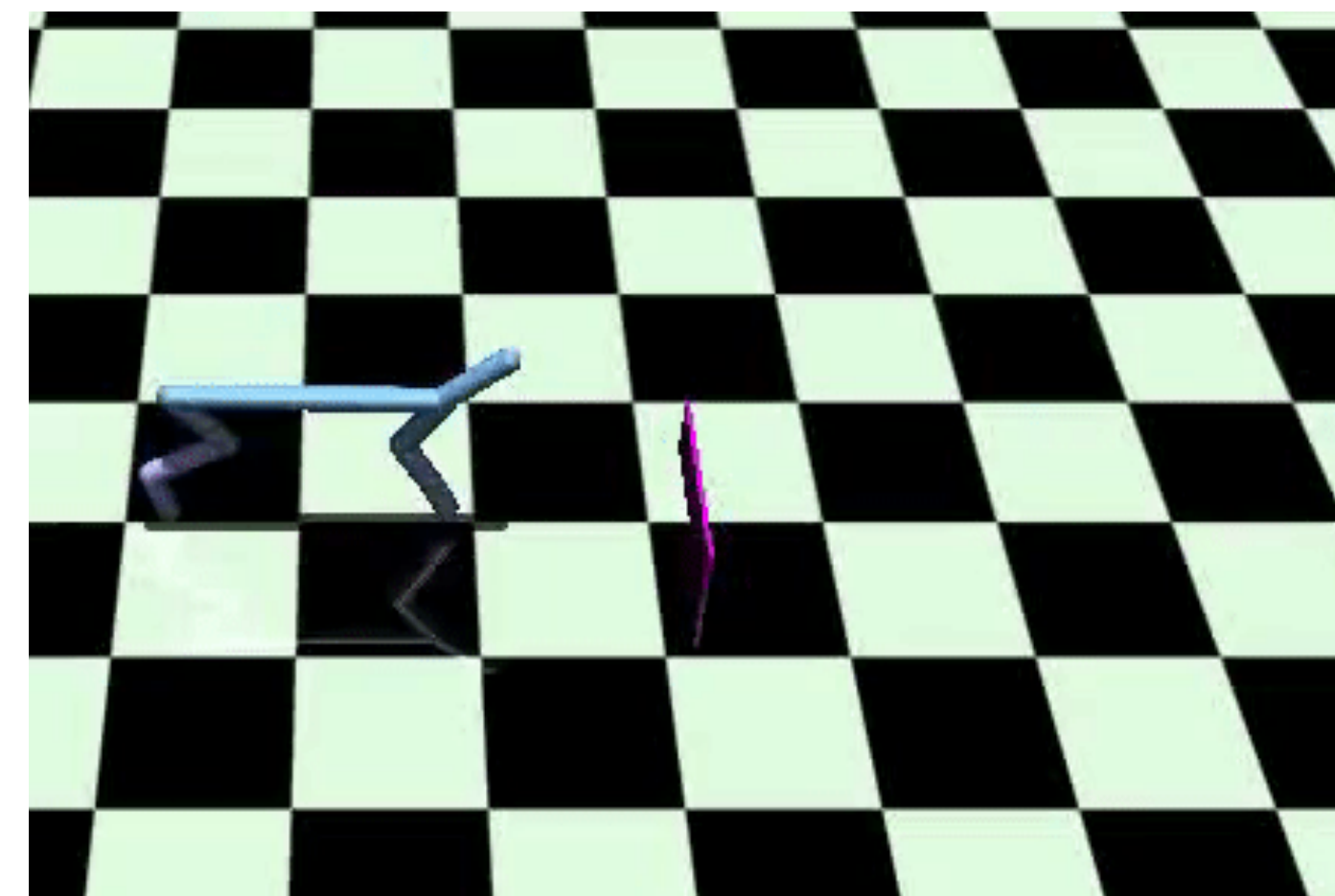
Best **SAC** policy at test time.



SMERL policies at train time.



Best **SMERL** policy at test time.



Tools for tackling distribution shift

Pessimism

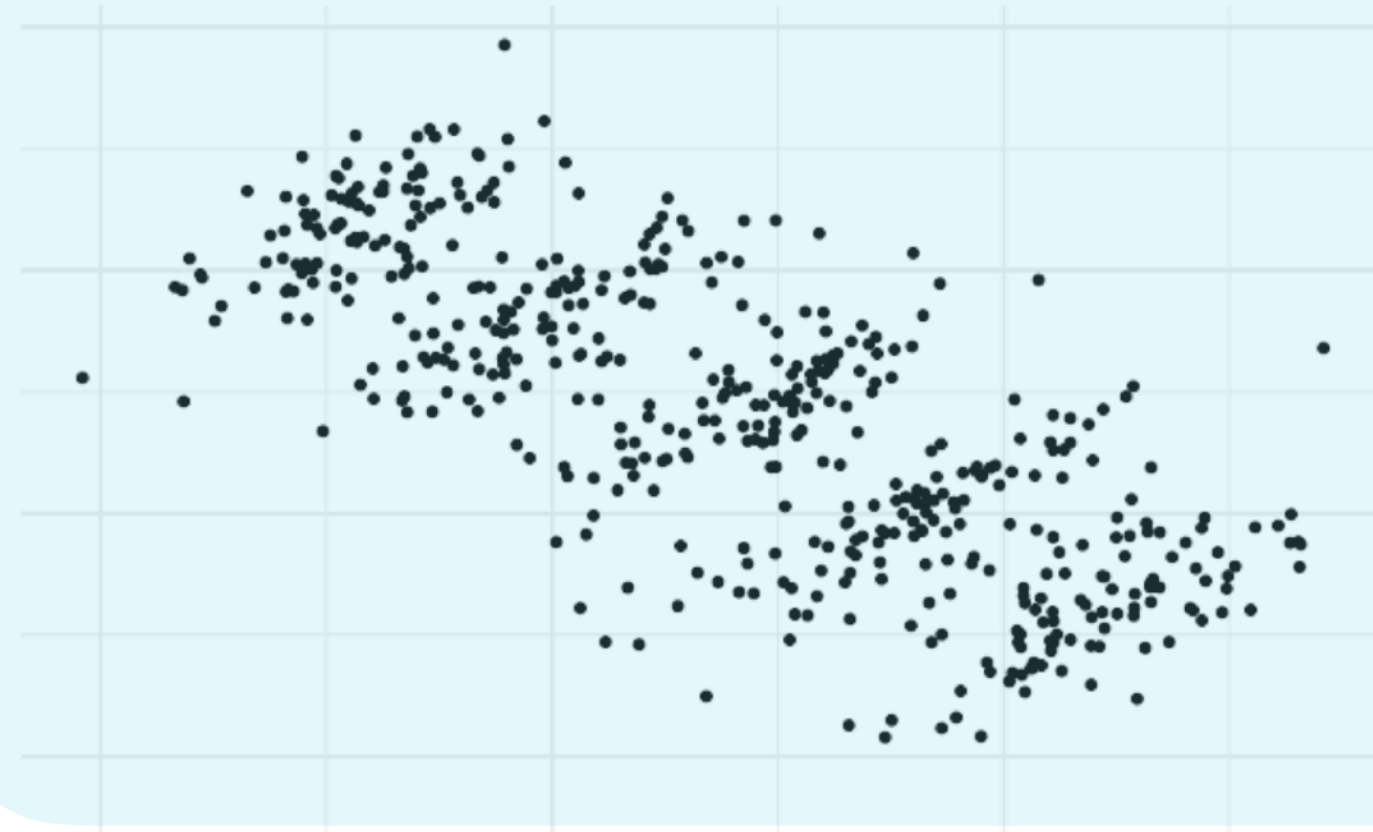
$$\min_{\theta} \sup_{Q \in U(P)} \mathbb{E}_Q[\ell(x, y; \theta)]$$

+ powerful tool for addressing **spurious correlations** and **policy distribution shift**

+ makes few assumptions

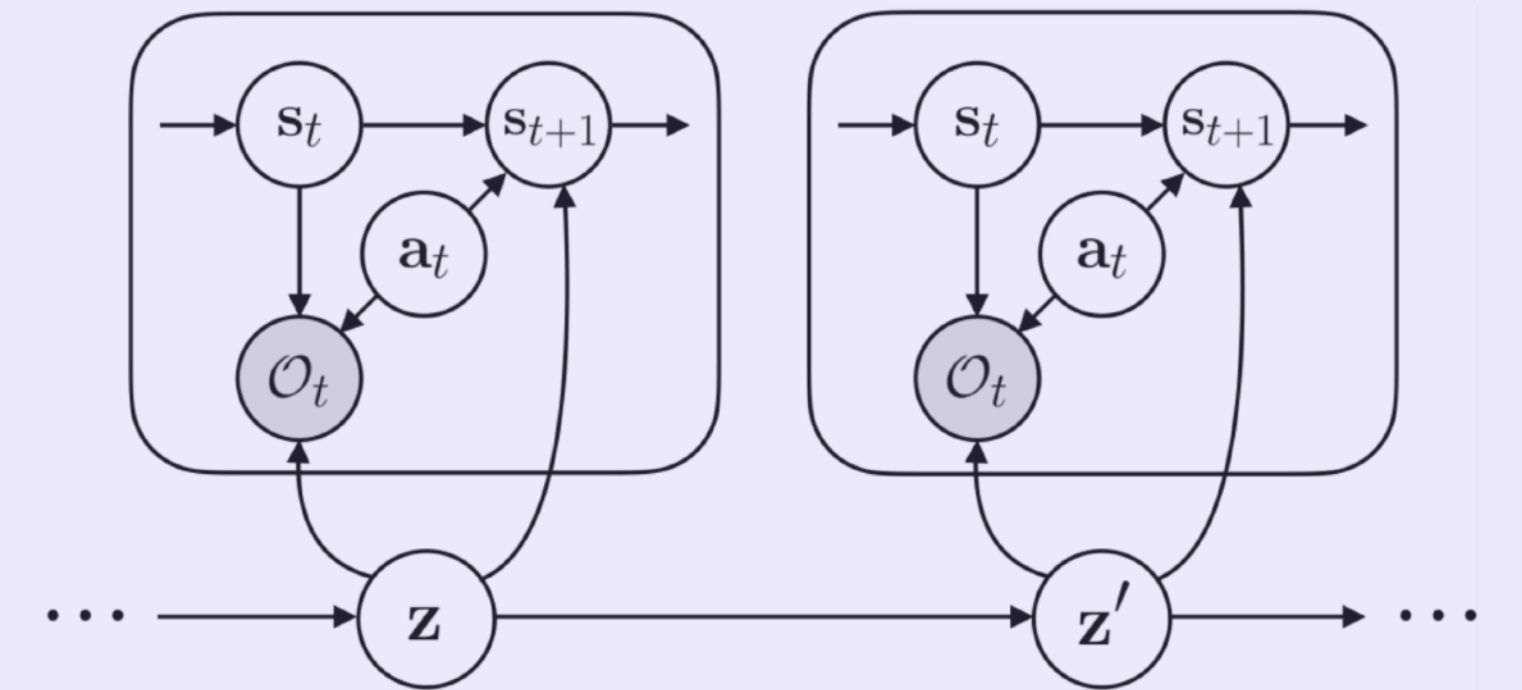
+ often possible to analyze theoretically

Adaptation



+ small amount of data can provide large amount of leverage

Anticipation

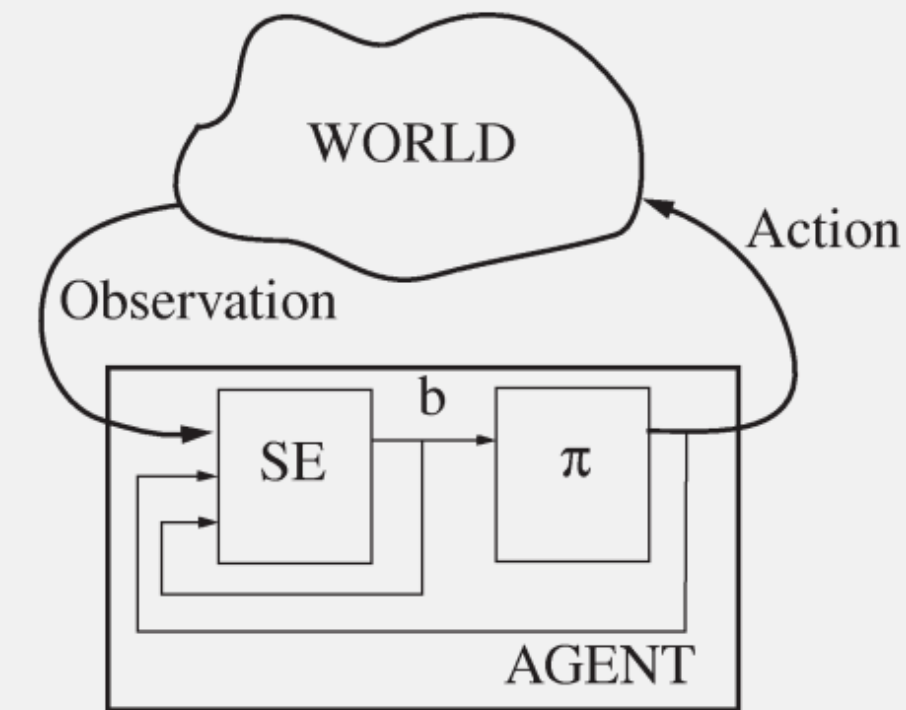


Introducing
more assumptions

Handling continuous distribution shifts in RL

POMDPs (Kaelbling et al. '98)

- + covers this setting
- perhaps **too** general



Prior work?

BAMDP (Duff & Barto et al. '02)

HiP-MDP (Doshi-Velez & Konidaris et al. '16)

- + **hidden parameters** underlying transitions, rewards
- assume hidden parameters are **stationary**

Handling continuous distribution shifts in RL

POMDPs (Kaelbling et al. '98)

- + covers this setting
- perhaps **too** general

BAMDP (Duff & Barto et al. '02)

HiP-MDP (Doshi-Velez & Konidaris et al. '16)

- + **hidden parameters** underlying transitions, rewards
- assume hidden parameters are **stationary**

Dynamic parameter MDP (DP-MDP)

- > **hidden parameters** underlying transition, rewards
(fixed within an episode)
- > parameters systematically **shift across episodes**

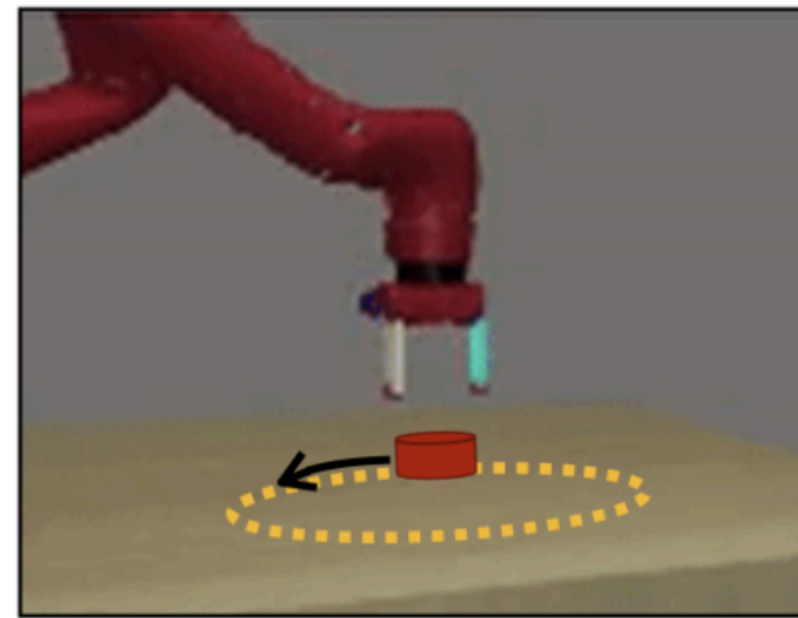


Annie Xie

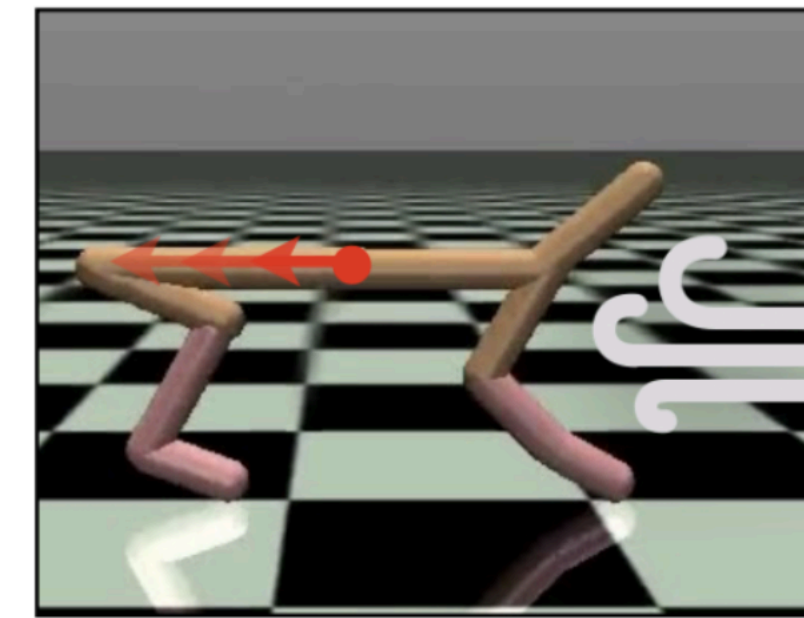
Handling continuous distribution shifts in RL

How well do existing algorithms perform with such shifts?

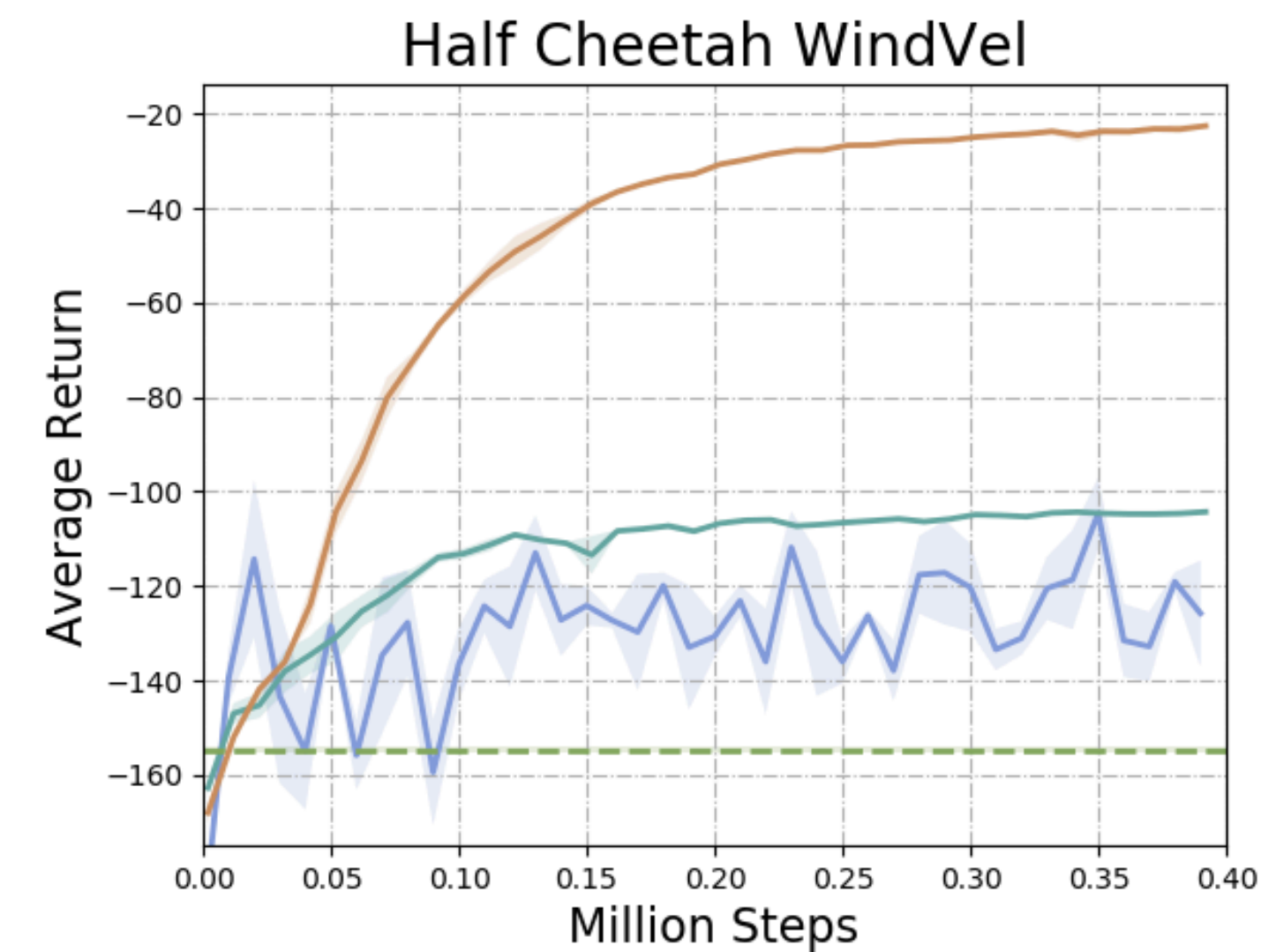
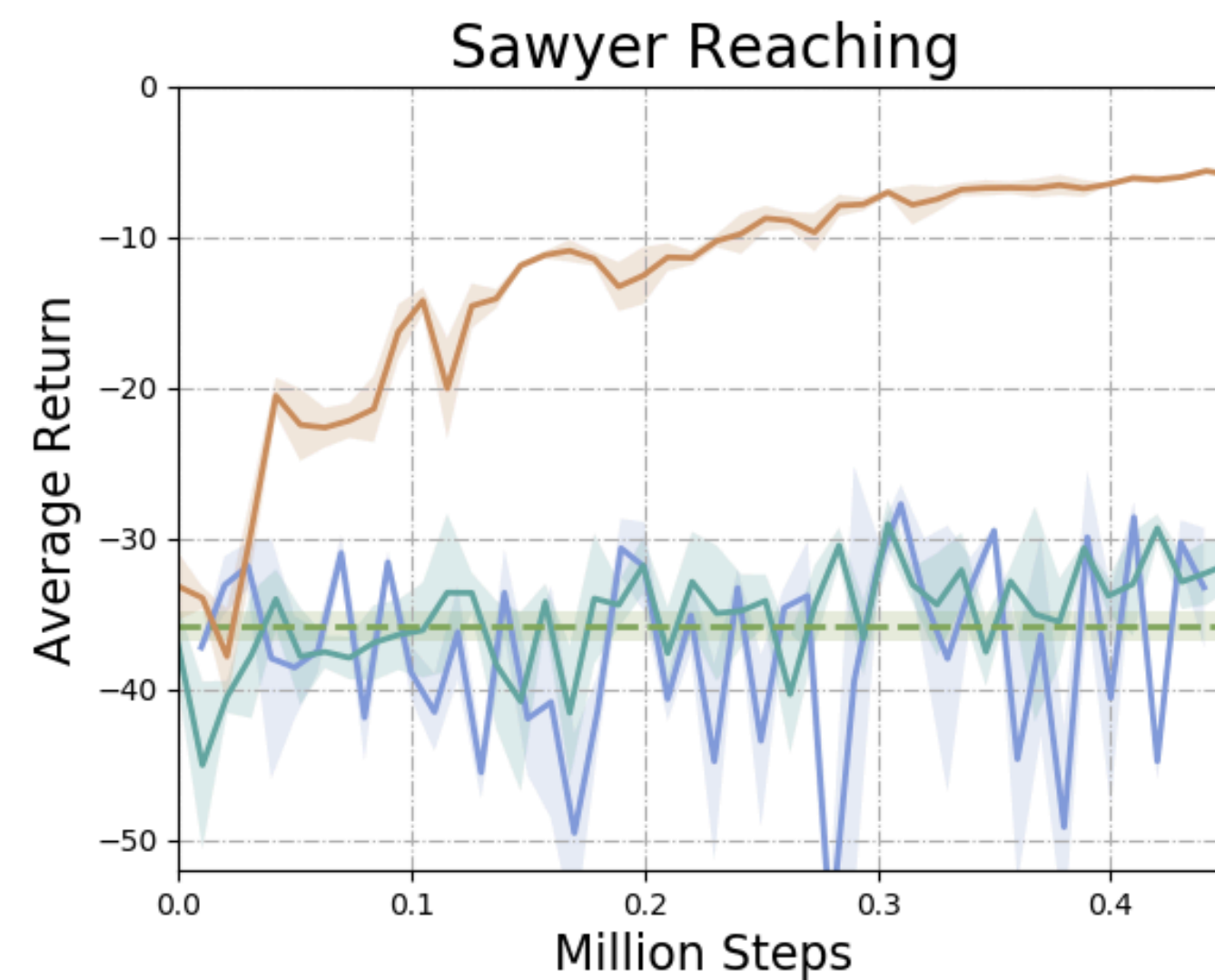
two simple
settings:



shifting goal



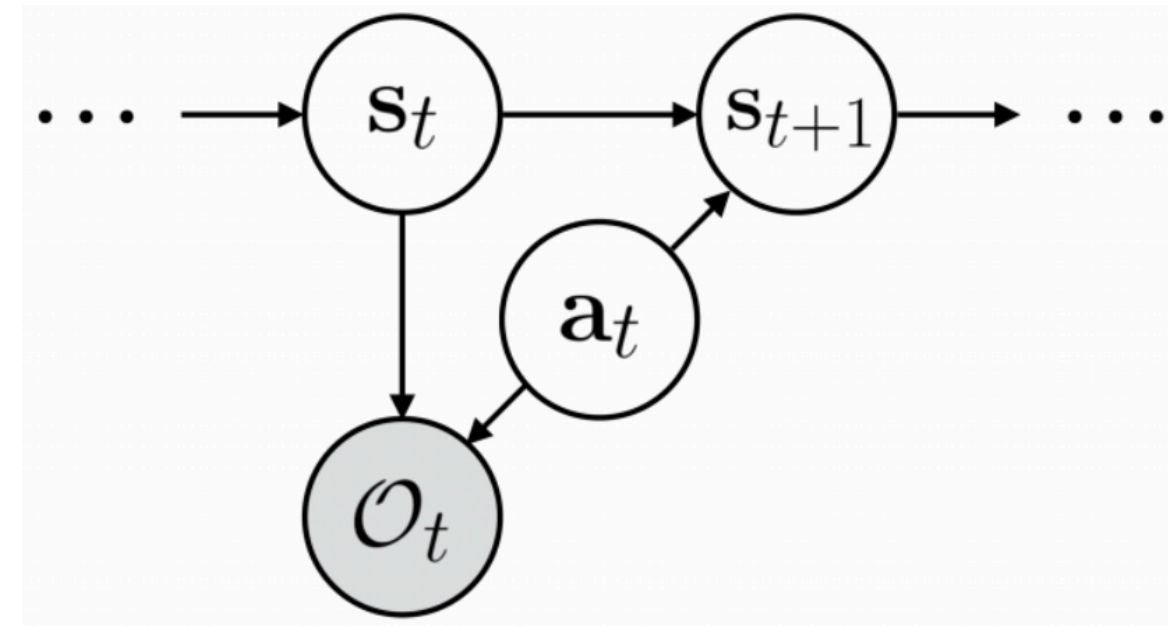
shifting wind + goal velocity



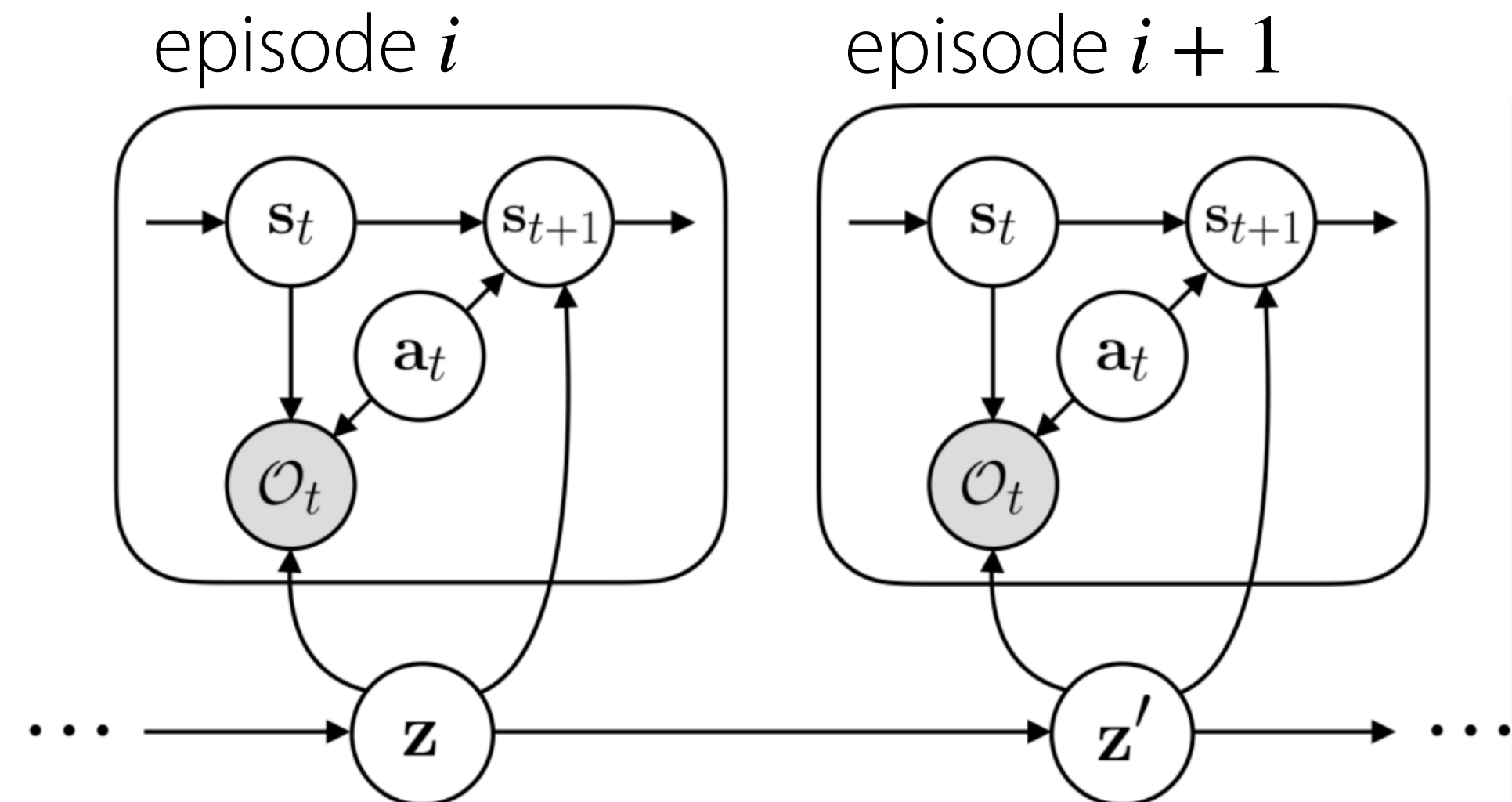
— SLAC — SAC — PPO — Oracle

standard RL “as inference”

(Todorov '08, Toussaint '09, Levine '18)



with **dynamic latent parameter \mathbf{z}**



key assumption: predictability

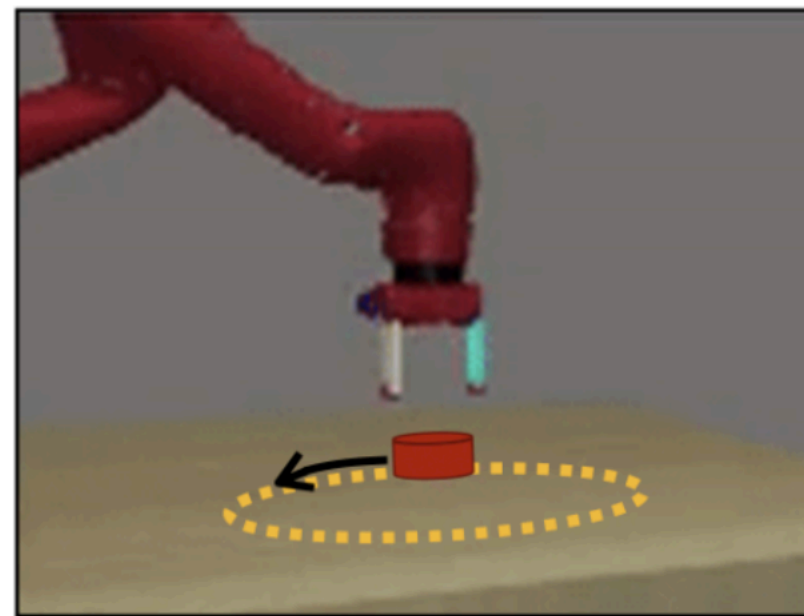
Resulting objective

$$\log p(\tau^{1:i-1}, \mathcal{O}_{1:T}^i = 1) \geq \mathbb{E}_q \left[\underbrace{\sum_{i'=1}^i \sum_{t=1}^T \log p(\mathbf{s}_{t+1}, r_t | \mathbf{s}_t, \mathbf{a}_t; \mathbf{z}^{i'})}_{\text{model dynamics \& reward}} - \underbrace{D_{\text{KL}}(q(\mathbf{z}^{i'} | \tau^{i'}) || p(\mathbf{z}^{i'} | \mathbf{z}^{i'-1}))}_{\text{model latent variable shifts}} \right] + \underbrace{\mathbb{E}_{\substack{p(\mathbf{z}^i | \tau^{1:i-1}) \\ \pi(\mathbf{a}_t | \mathbf{s}_t, \mathbf{z}^i)}} \left[\sum_{i=1} \underbrace{r(\mathbf{s}_t, \mathbf{a}_t; \mathbf{z}^i) - \log \pi(\mathbf{a}_t | \mathbf{s}_t, \mathbf{z}^i)}_{\text{entropy regularized RL}} \right]}_{\text{entropy regularized RL}}$$

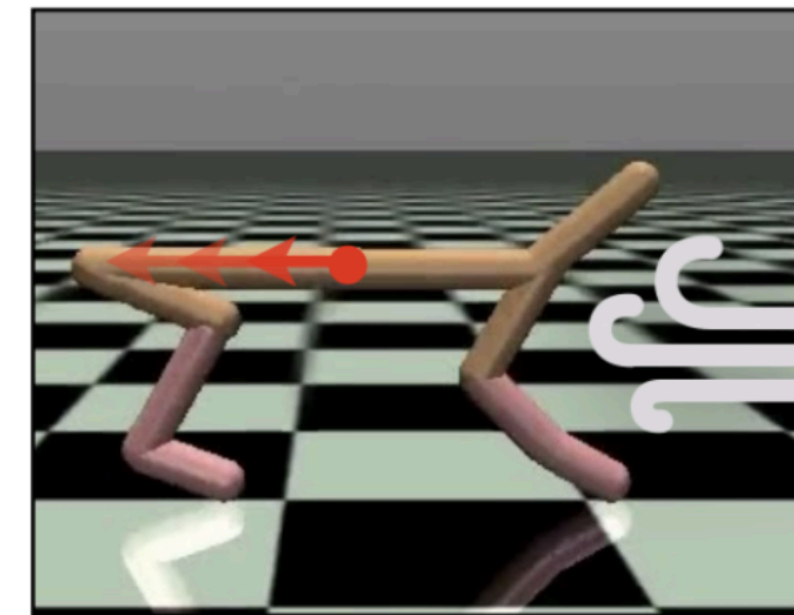
lifelong latent actor-critic (LILAC)

Experiments

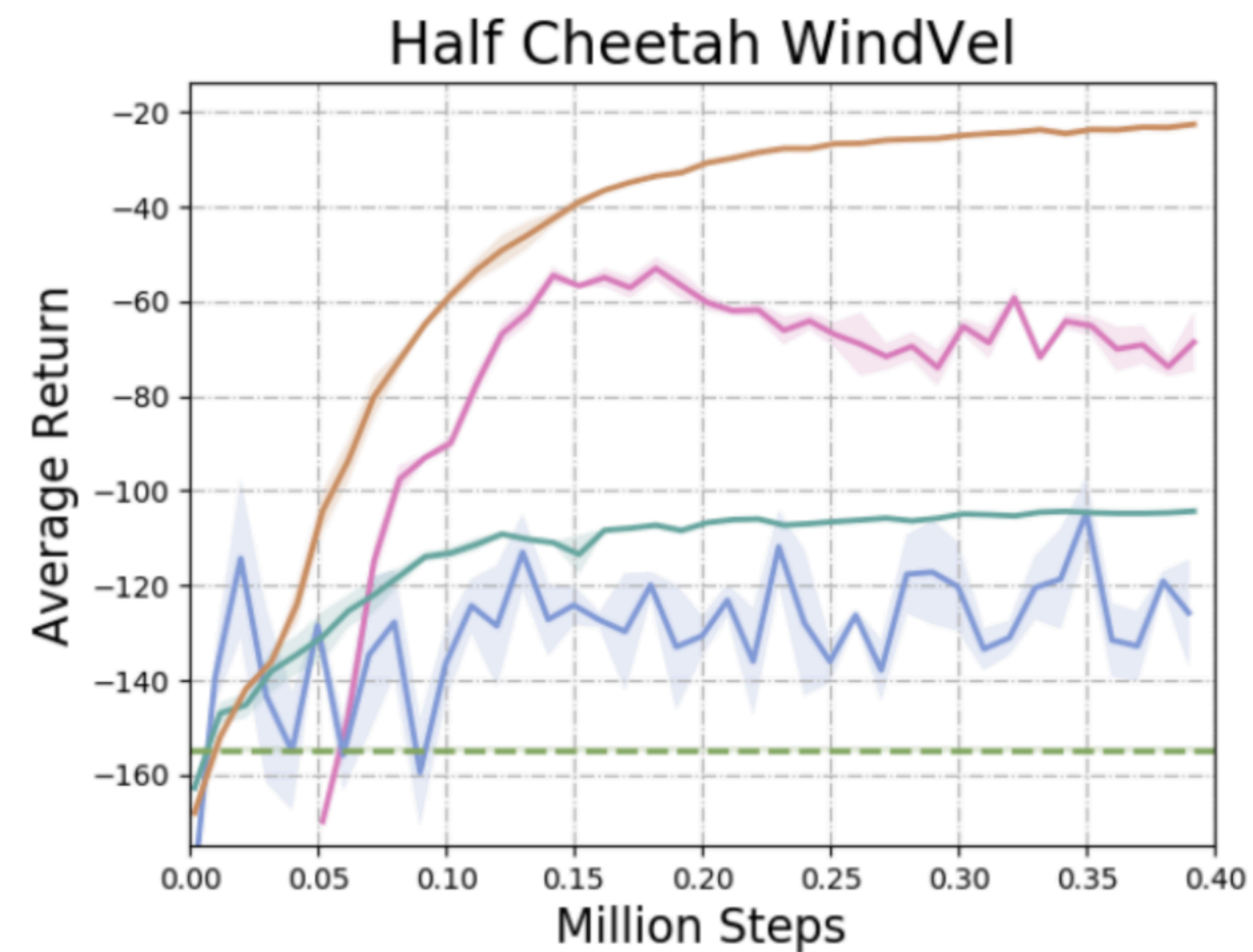
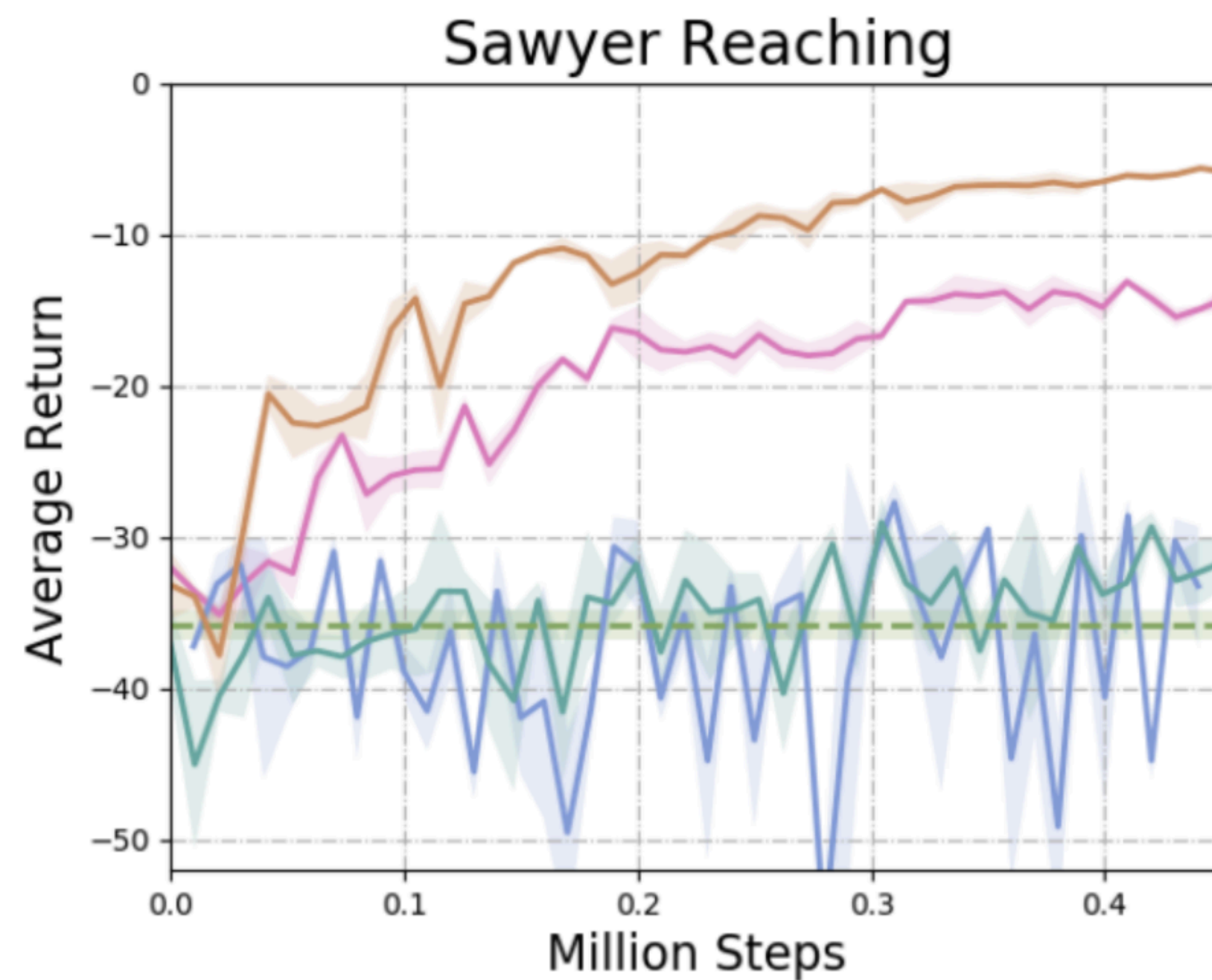
two simple settings:



shifting goal



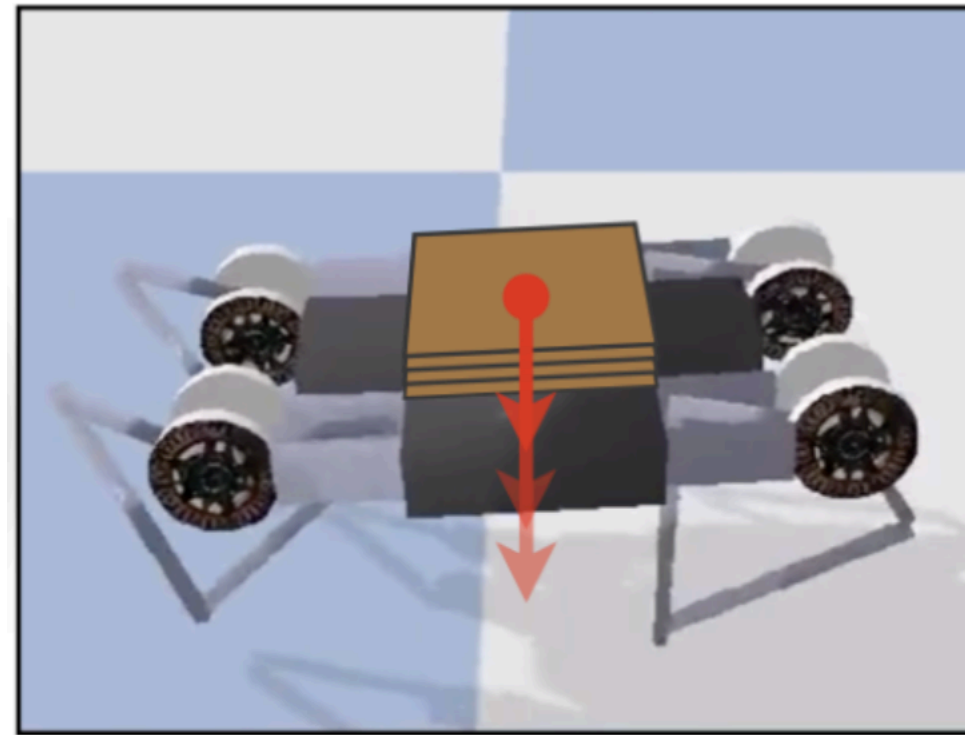
shifting wind + goal velocity



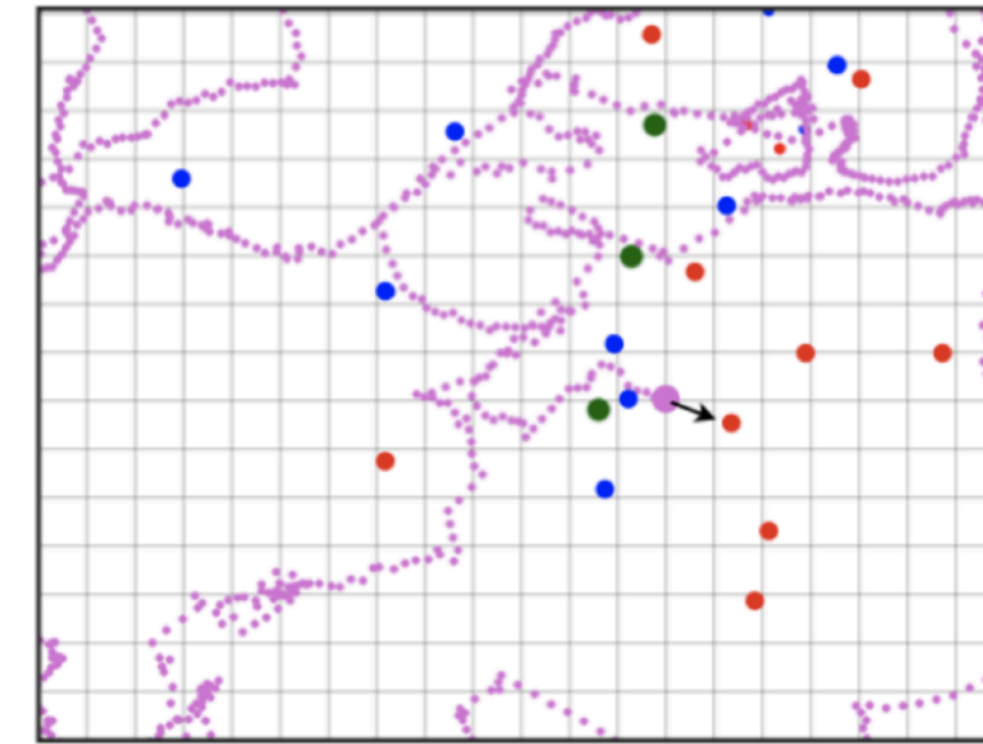
— LILAC (Ours) — SLAC — SAC — PPO — Oracle

Experiments

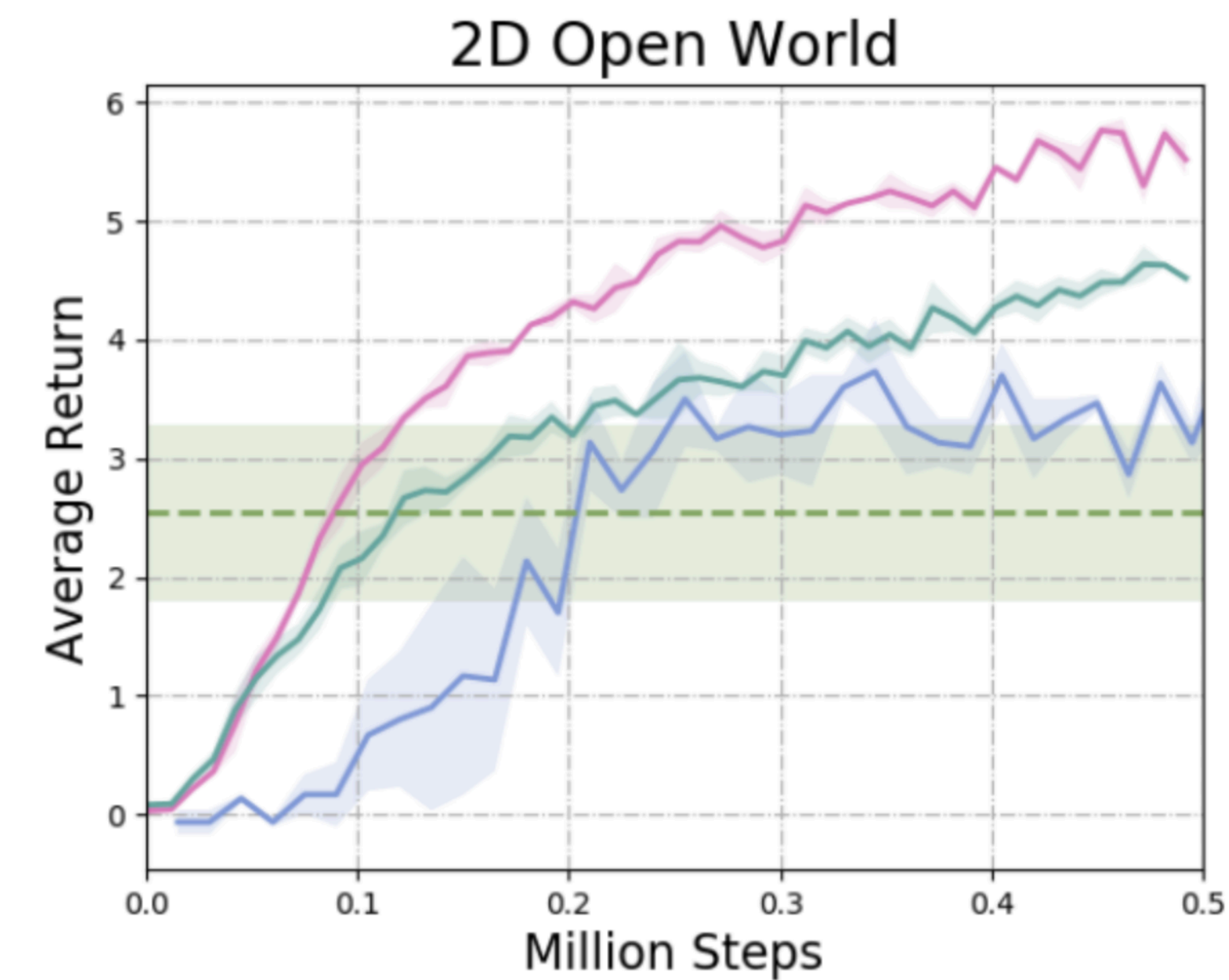
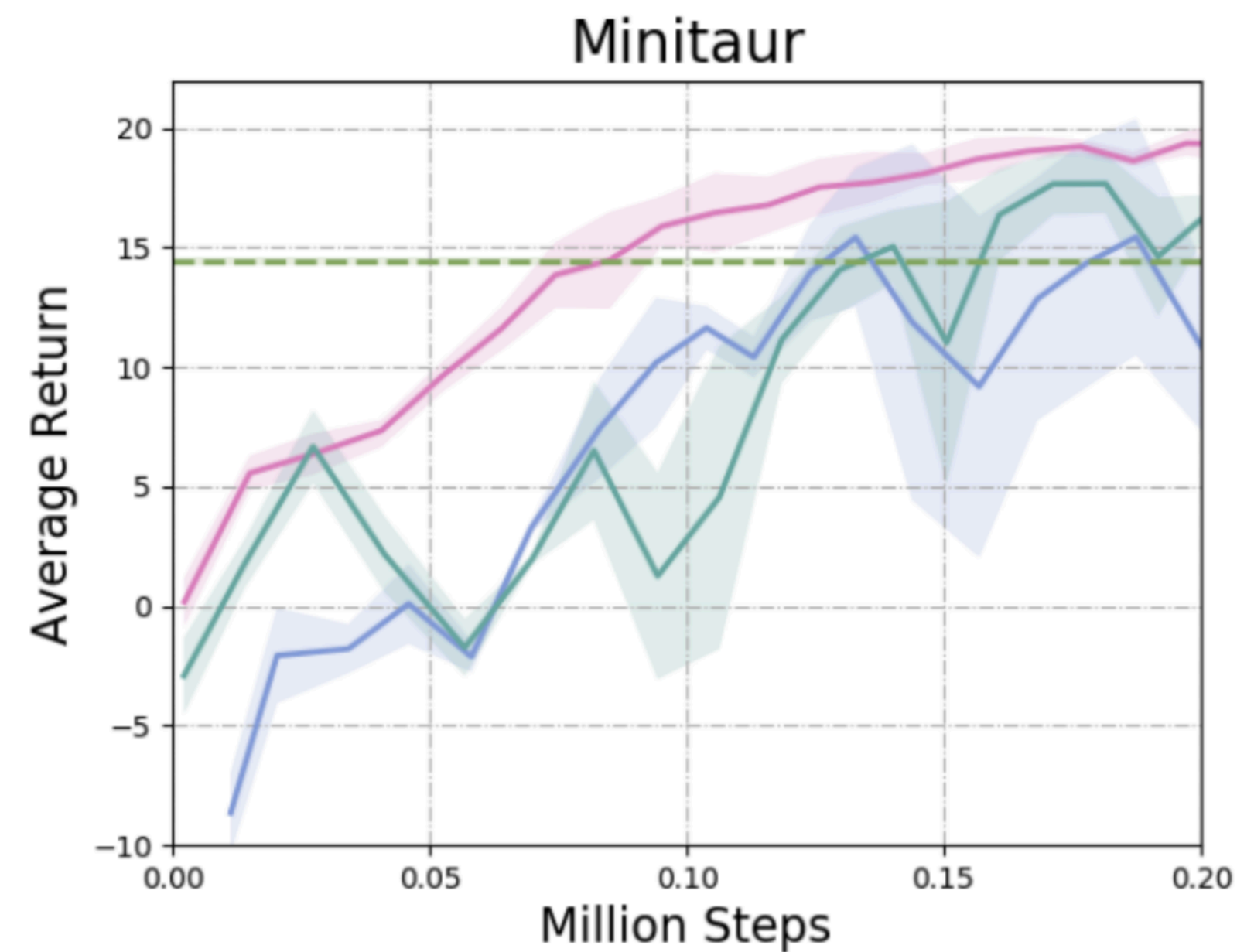
more challenging
settings:



continuously varying mass



continuously varying wind + **no resets**



— LILAC (Ours) — SLAC —

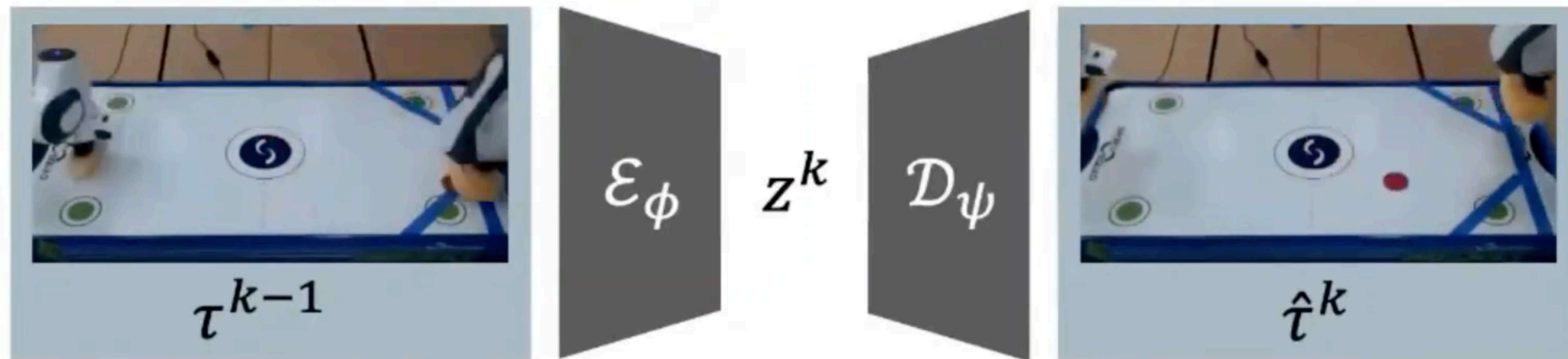
Takeaway: By modeling and anticipating change, we can learn amidst non-stationarity.

Extension to distribution shift caused by other agents.

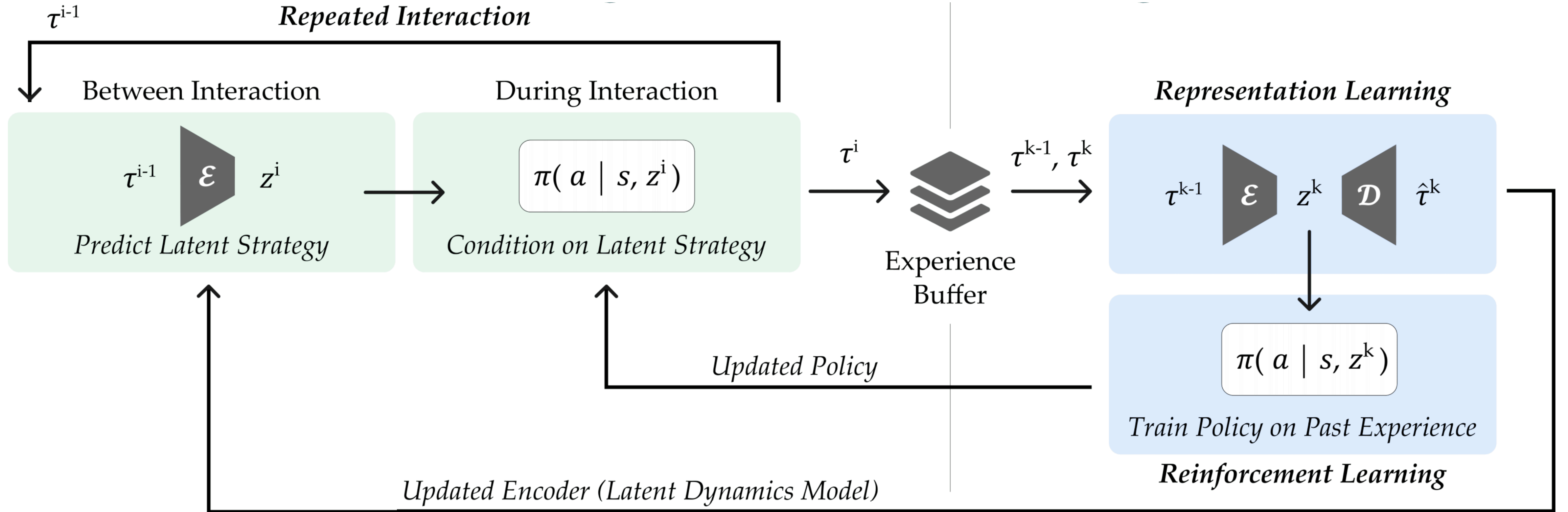
Key conceptual change: now agent's action can influence latent variable z

z captures other agent's strategy

Predict future z not only from past z , but from entire past trajectory.



Extension to distribution shift caused by other agents.



Maximize rewards *within* interactions to *anticipate* change.

Maximize rewards *across* interactions to *influence* change.

$$\max_{\theta} \sum_{i=1}^{\infty} \gamma^i \mathbb{E}_{\pi_{\theta}(a|s, z^i)} \left[\sum_{t=1}^H R(s, z^i) \right]$$

2x speed

SAC: 2 hours of training

Xie, Losey, Tolsma, Finn, Sadigh. *Learning Latent Representations to Influence Multi-Agent Interaction*, CoRL '20

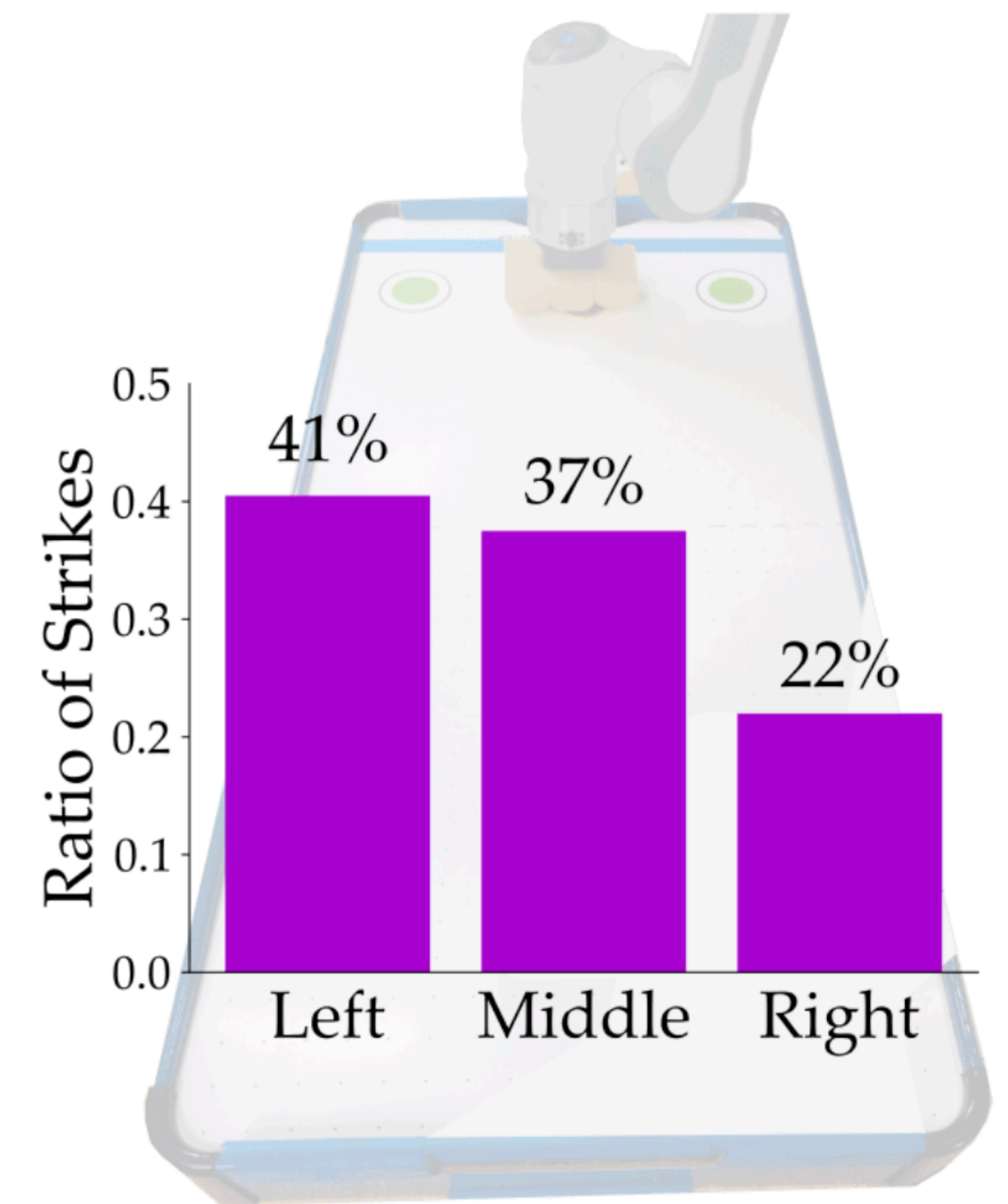
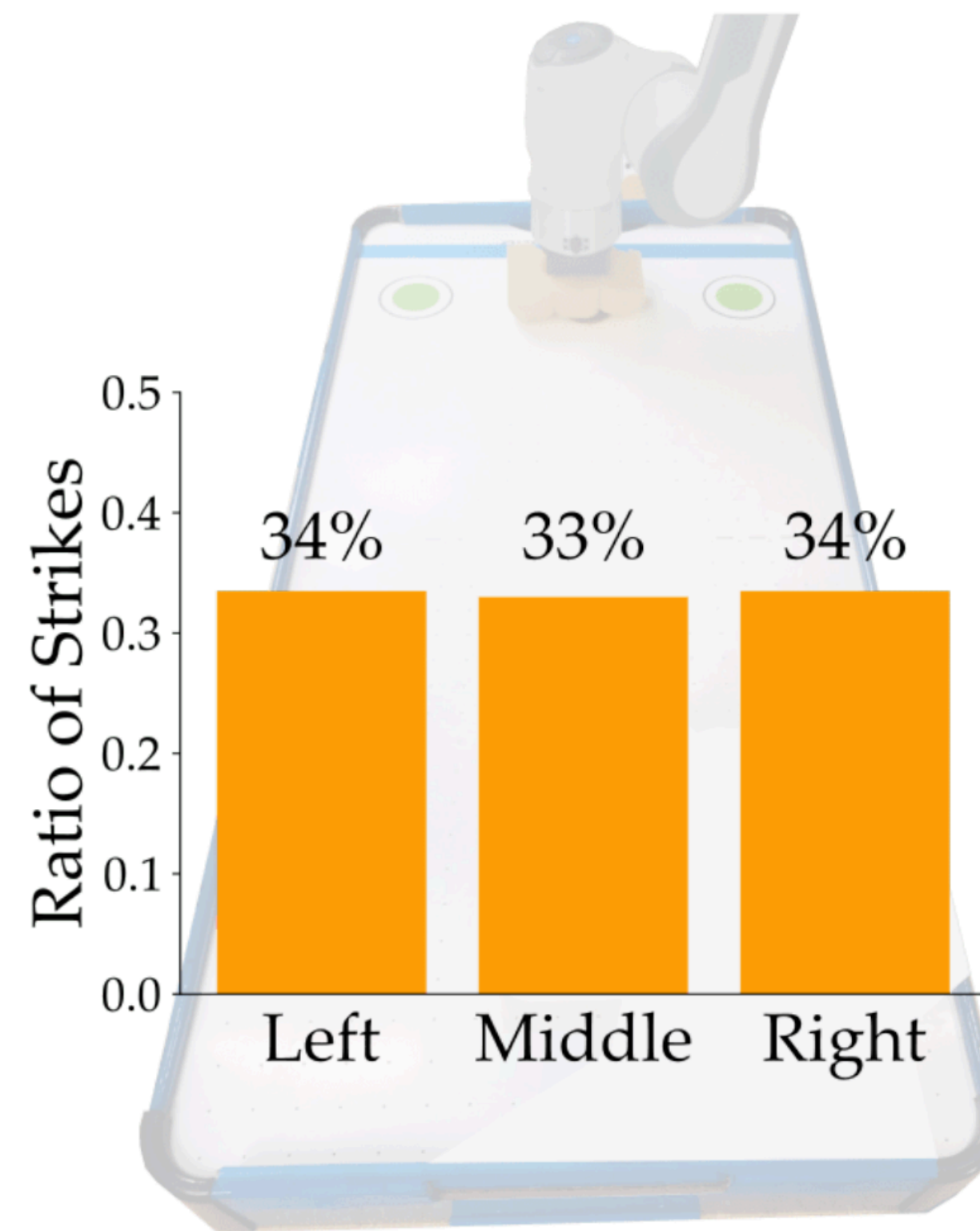
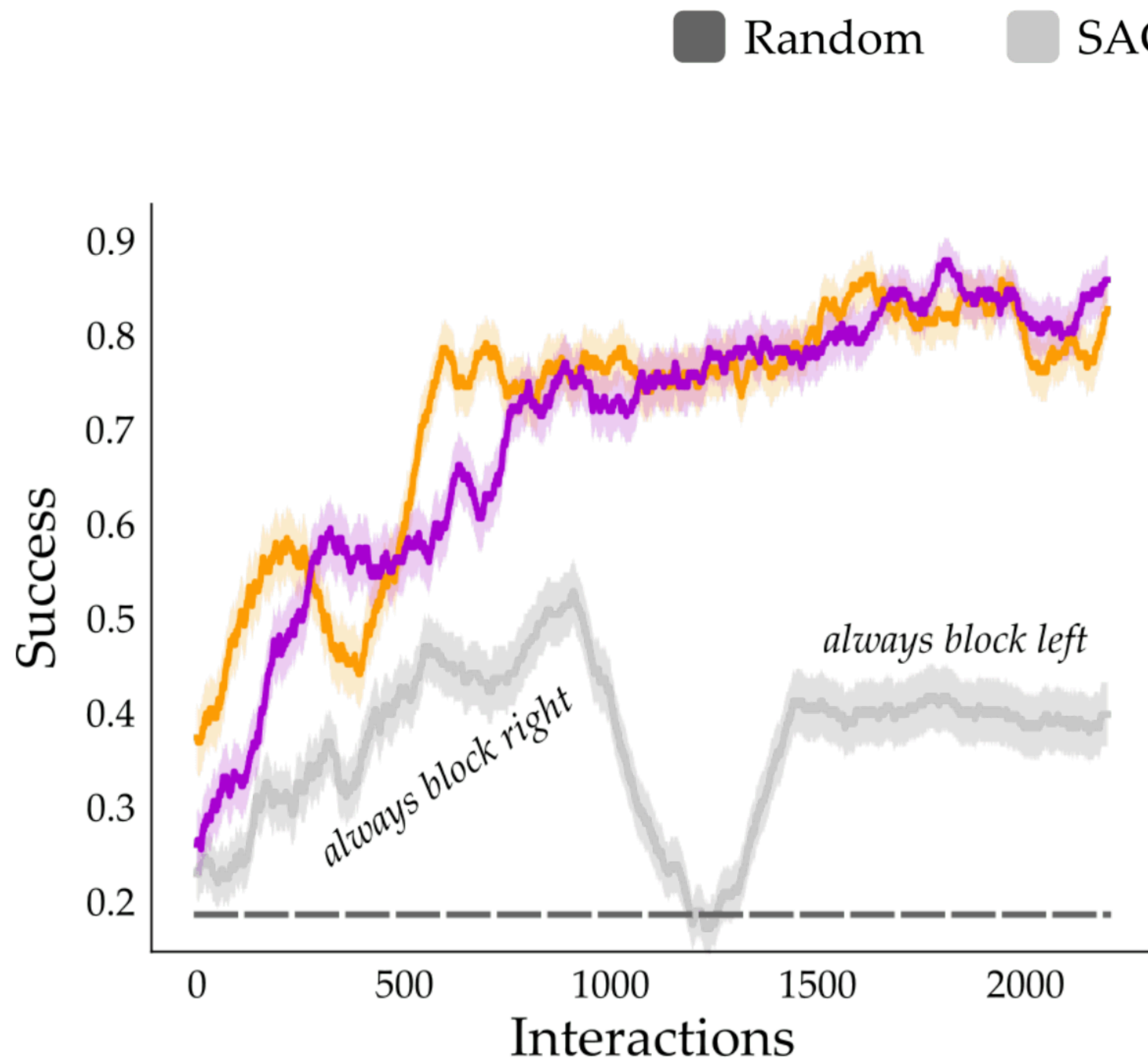
2x speed



LILI (ours): 2 hours of training

Xie, Losey, Tolsma, Finn, Sadigh. *Learning Latent Representations to Influence Multi-Agent Interaction*, CoRL '20

Air Hockey Quantitative Results



Tools for tackling distribution shift

Pessimism

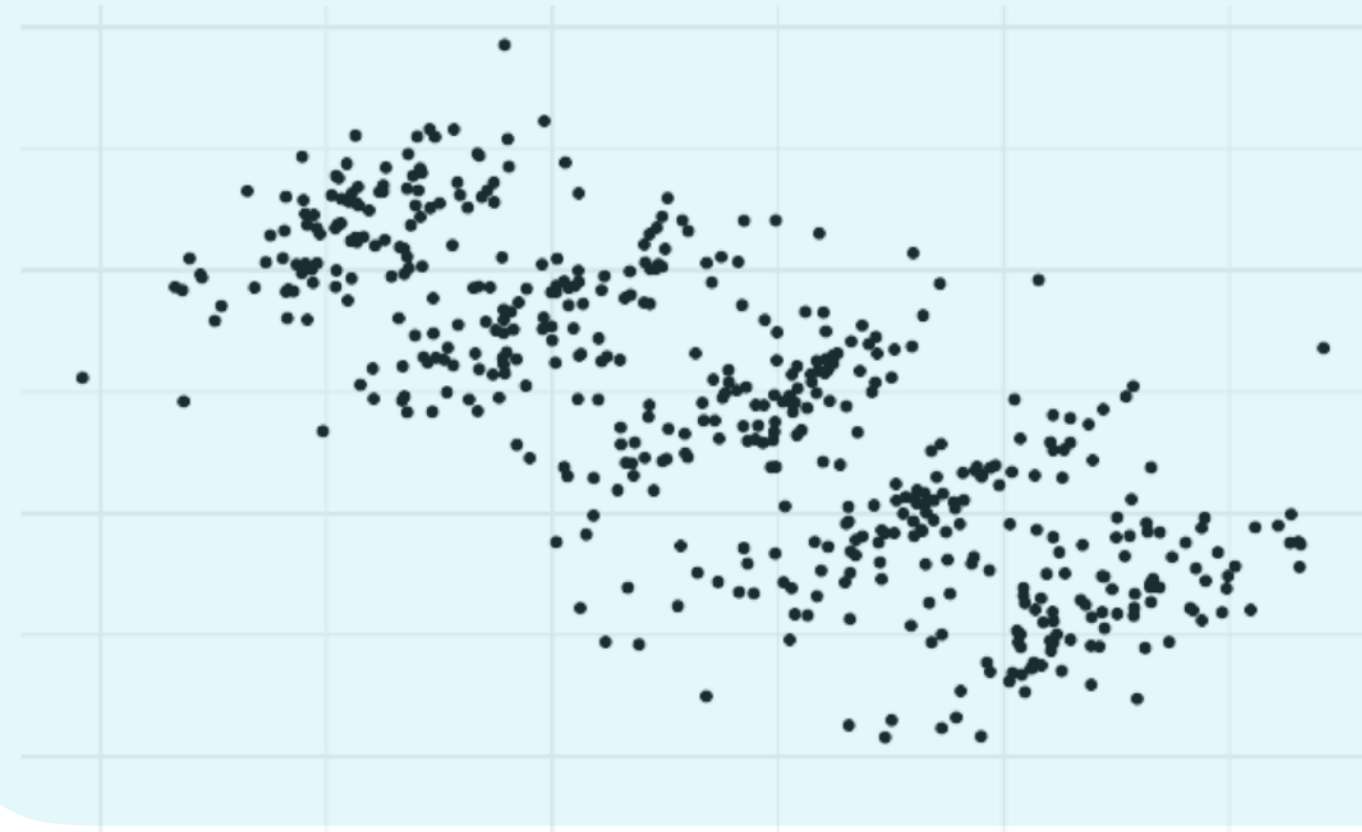
$$\min_{\theta} \sup_{Q \in U(P)} \mathbb{E}_Q[\ell(x, y; \theta)]$$

+ powerful tool for addressing **spurious correlations** and **policy distribution shift**

+ makes few assumptions

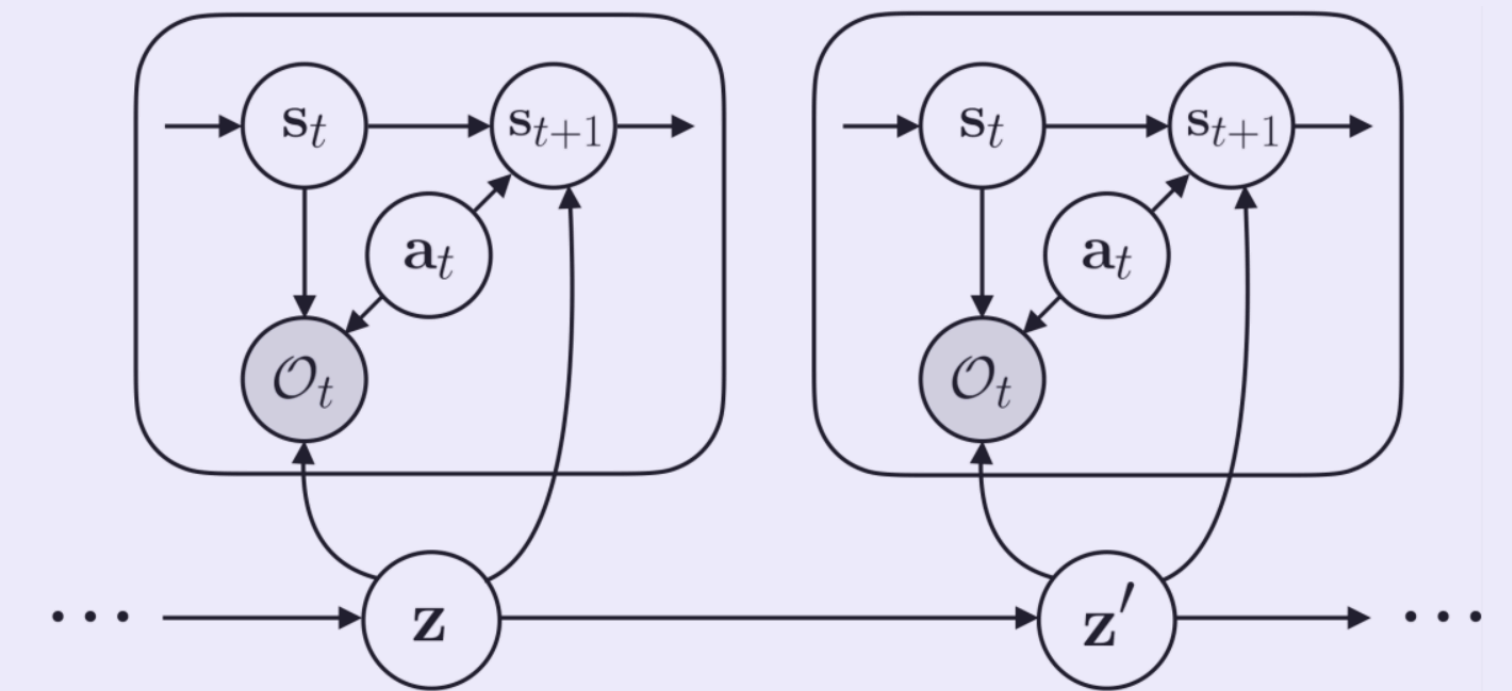
+ often possible to analyze theoretically

Adaptation



+ small amount of data can provide large amount of leverage

Anticipation



+ can **get ahead** of the shift in **predictably** changing environments

+ can even **influence** the shift in some environments

Introducing
more assumptions



Working on distribution shift?

WILDS

Benchmark with distribution shifts
arising in real-world applications.

wilds.stanford.edu

Questions?