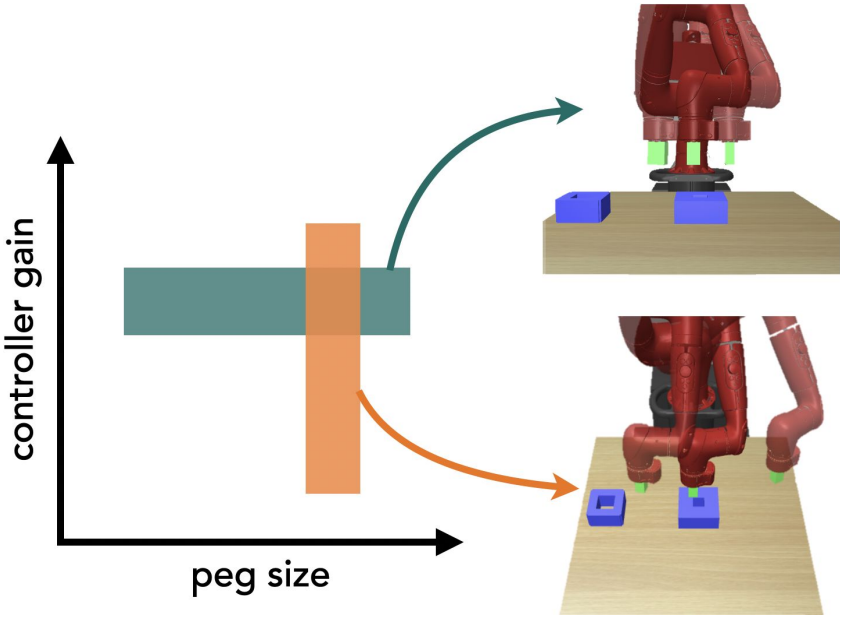
Robust Policy Learning over Multiple Uncertainty Sets

Annie Xie[1], Shagun Sodhani[2], Chelsea Finn[1], Joelle Pineau[2], Amy Zhang[2]
[1]Stanford University, [2]Facebook AI Research
ICML 2022




Motivation

Transfer to new environments presents **uncertainty**, i.e., over the context



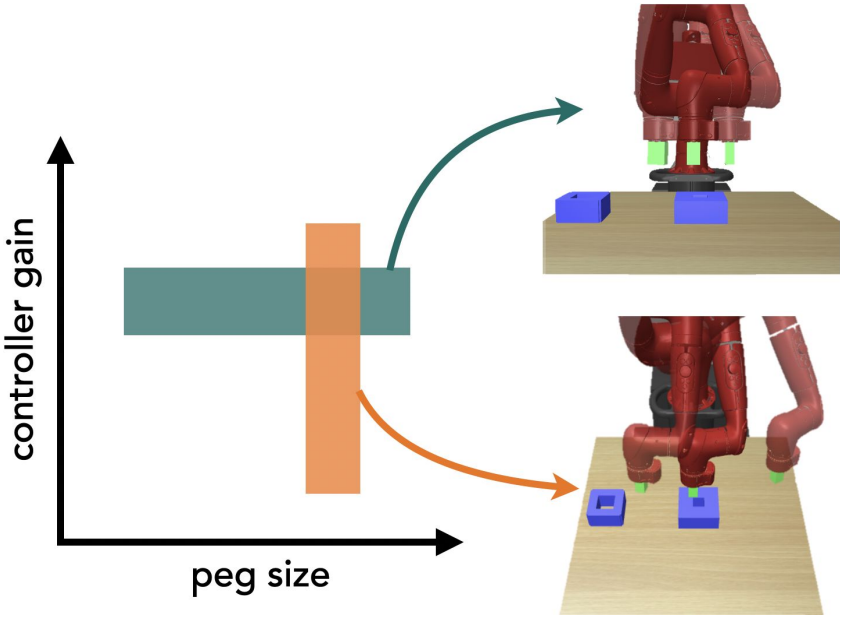
High uncertainty
over peg size

High uncertainty
over gain

context = (size, gain)

Motivation

Agent's initial estimate of the context differs across environments



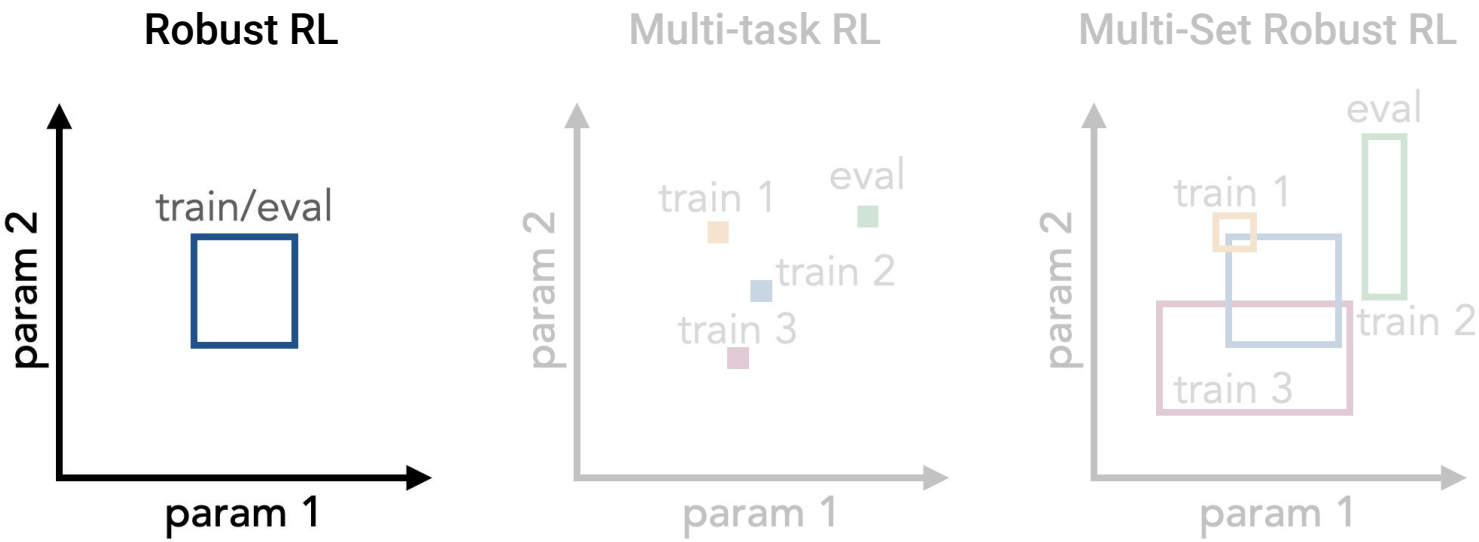
High uncertainty over peg size

High uncertainty over gain

context = (size, gain)

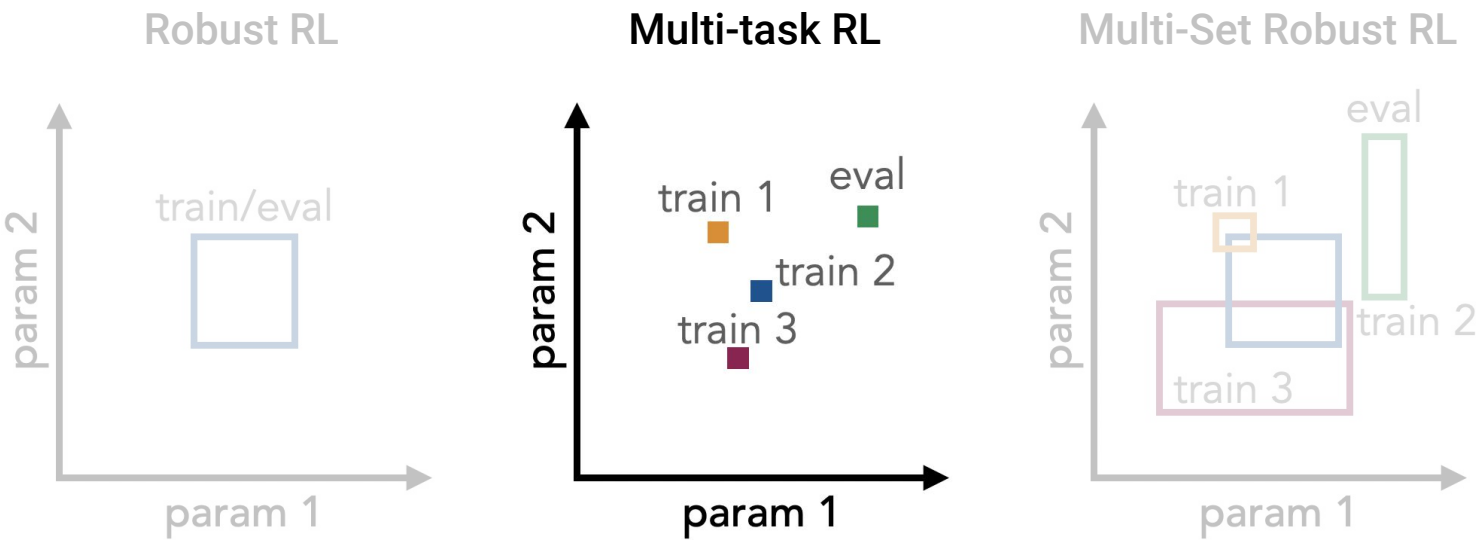
Multi-Set Robustness Problem

Robust RL: robustness with respect to a single perturbation set



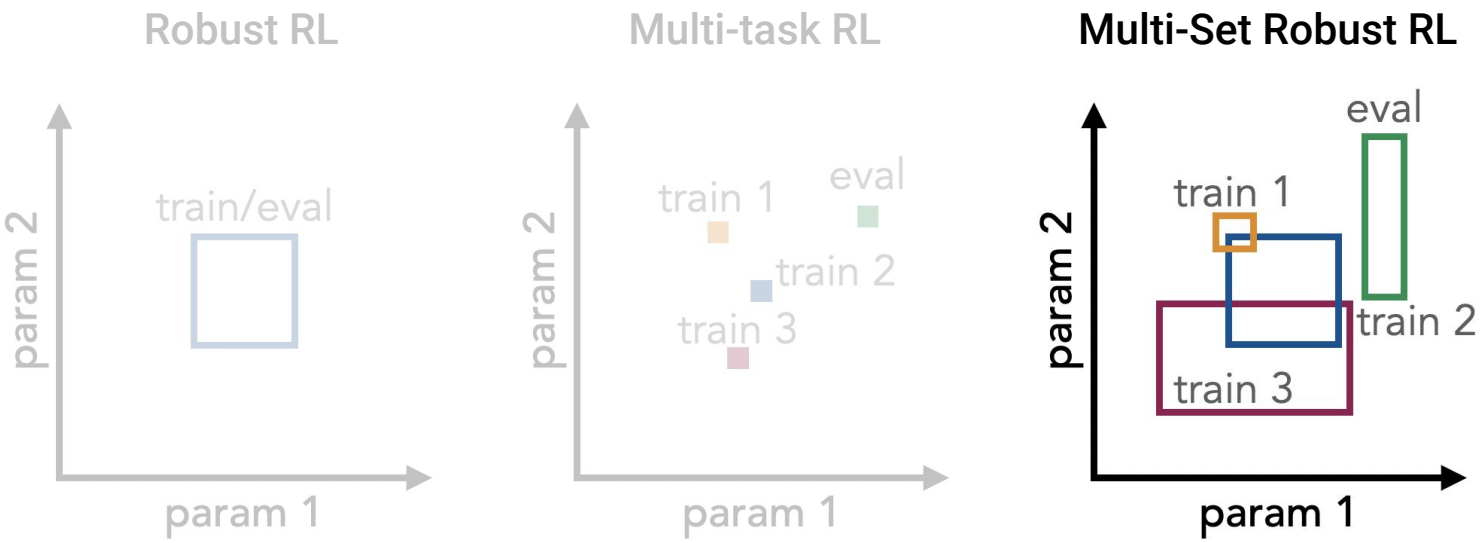
Multi-Set Robustness Problem

Multi-task RL: generalization to <u>new</u> contexts



Multi-Set Robustness Problem

Multi-set robust RL: generalization to <u>new</u> perturbation sets over contexts



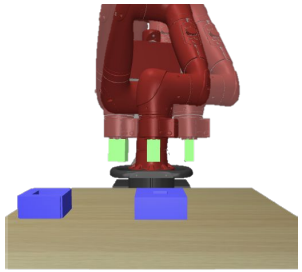
Challenges with System Identification

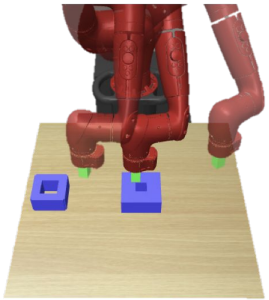
Challenges with System Identification

Context identifiability: many systems are determined by parameters that are difficult to identify from limited interaction (Dorfman & Tamar 2020)

Challenges with System Identification

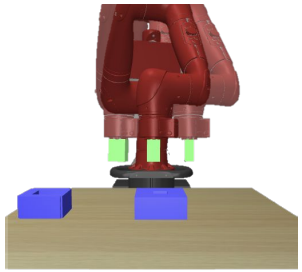
Context identifiability: many systems are determined by parameters that are difficult to identify from limited interaction (Dorfman & Tamar 2020)

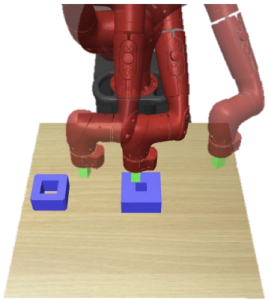




Challenges with System Identification

Context identifiability: many systems are determined by parameters that are difficult to identify from limited interaction (Dorfman & Tamar 2020)

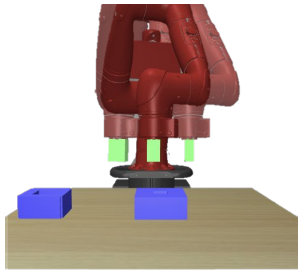




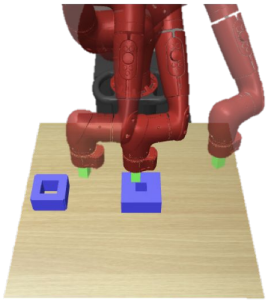
Controller gain can be inferred within a few steps

Challenges with System Identification

Context identifiability: many systems are determined by parameters that are difficult to identify from limited interaction (Dorfman & Tamar 2020)



Peg size can be inferred only when robot has tried inserting into one of the boxes



Controller gain can be inferred within a few steps

Challenges with System Identification

Context identifiability: many systems are determined by parameters that are difficult to identify from limited interaction (Dorfman & Tamar 2020)

Critical contexts: Some unidentifiable parameters are critical to the task

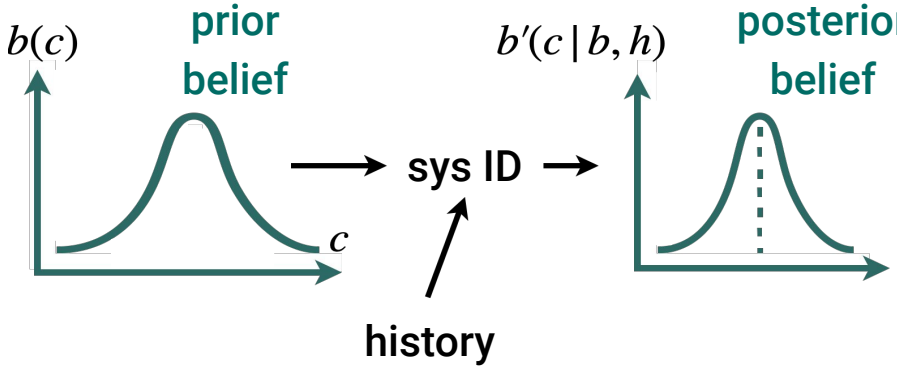
Challenges with System Identification

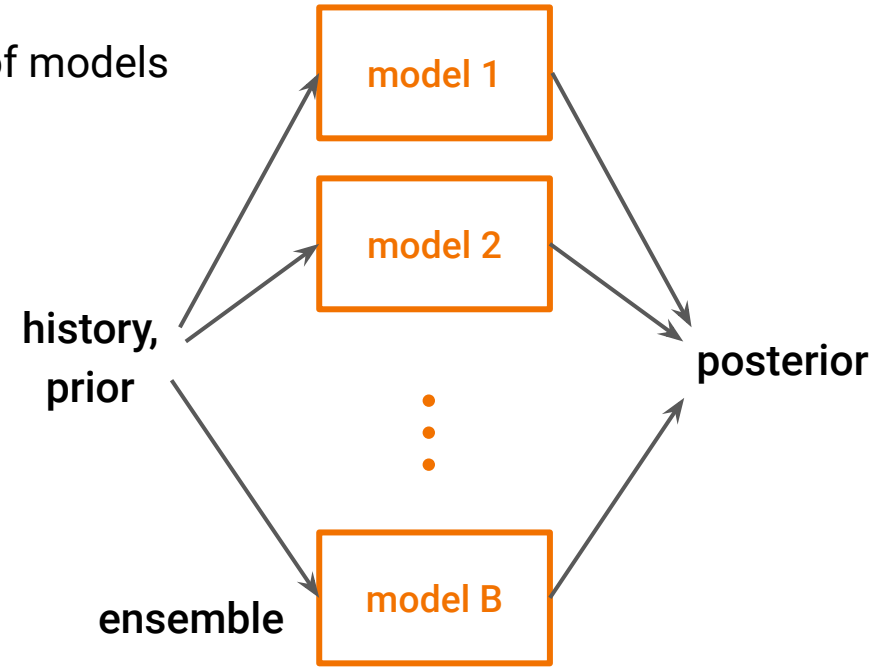
Context identifiability: many systems are determined by parameters that are difficult to identify from limited interaction (Dorfman & Tamar 2020)

Critical contexts: Some unidentifiable parameters are critical to the task ⇒ act *robustly*, i.e., with respect to the *worst* case, under uncertainty

Probabilistic System Identification

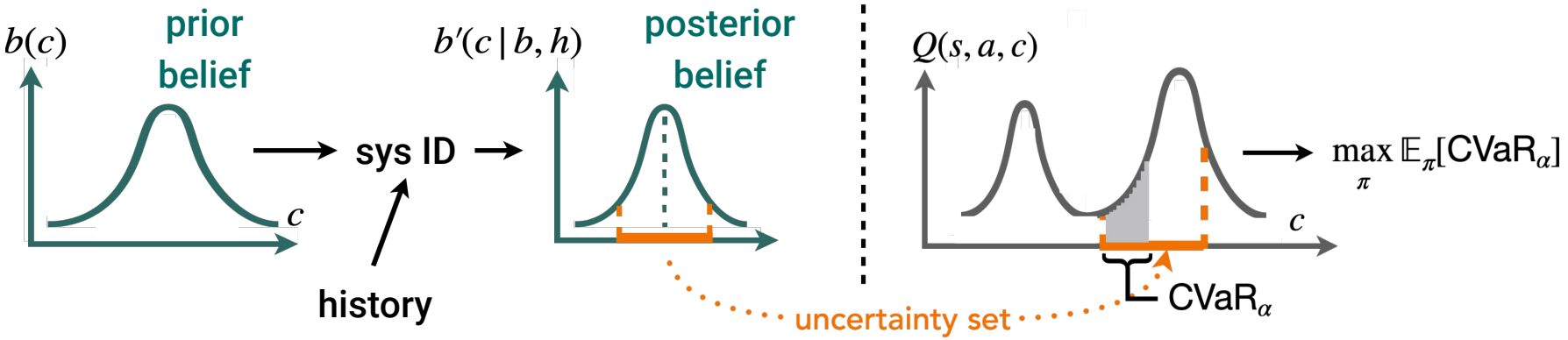
Estimate posterior distribution with ensemble of models





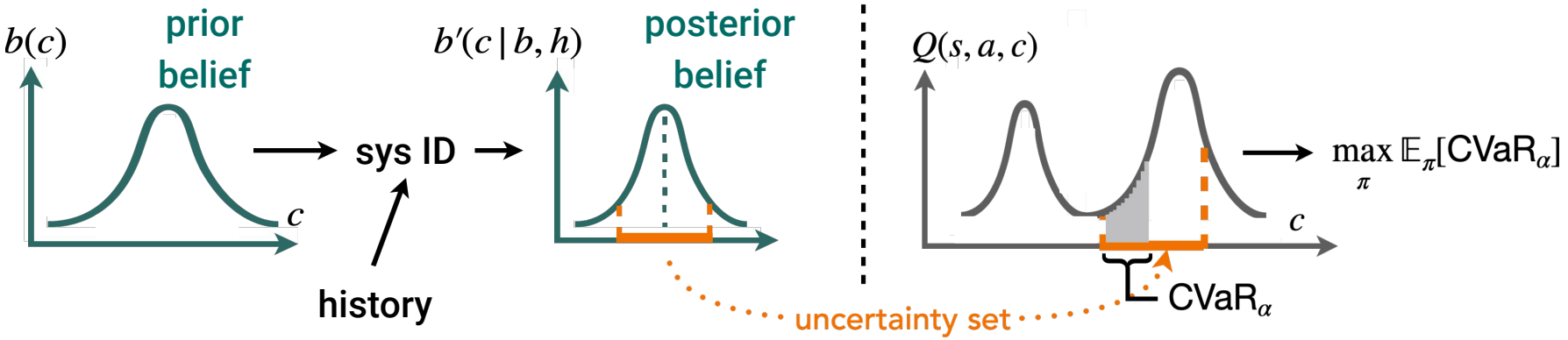
Risk-Sensitive Policy Optimization

Derive uncertainty set and **optimize** policy to maximize the CVaR return



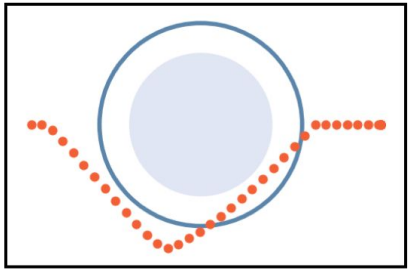
Risk-Sensitive Policy Optimization

Derive uncertainty set and **optimize** policy to maximize the CVaR return

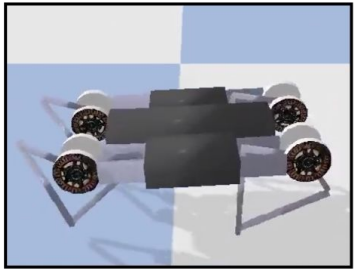


System Identification & Risk-Sensitive Adaptation

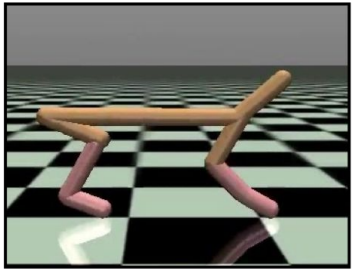
Experiments: Main Results



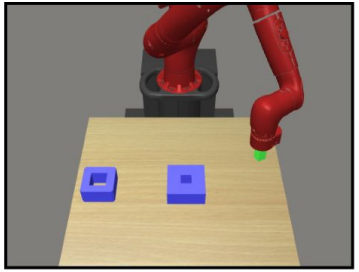
Point Mass



Minitaur



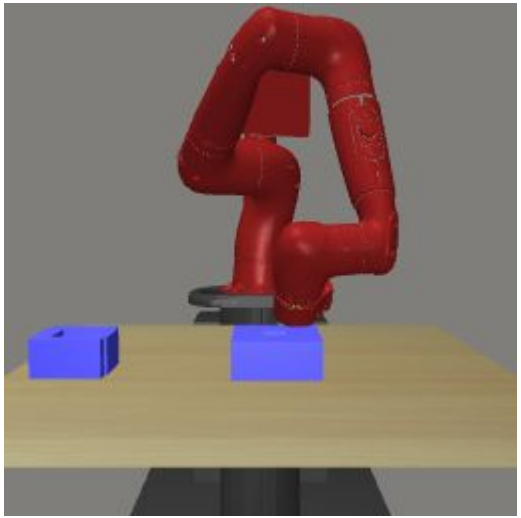
Half-Cheetah



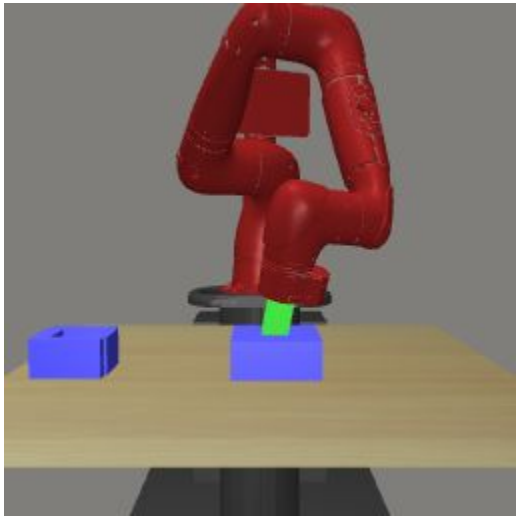
Sawyer Peg

Method	Sample Min						
Ensemble	36.8 ± 0.3	Ensemble	$\mathbf{178.0 \pm 7.1}$	Ensemble	$\mathbf{3988 \pm 75}$	Ensemble	45.2 ± 3.0
System ID	$\mathbf{37.6 \pm 0.3}$	System ID	$\mathbf{174.0 \pm 10.8}$	System ID	3774 ± 318	System ID	73.7 ± 4.3
EPOpt	36.3 ± 0.6	EPOpt	$\mathbf{172.2 \pm 7.3}$	EPOpt	2272 ± 218	EPOpt	43.2 ± 5.4
Set-EPOpt	37.1 ± 0.5	Set-EPOpt	$\mathbf{183.1 \pm 7.5}$	Set-EPOpt	3806 ± 224	Set-EPOpt	70.6 ± 3.6
WCPG	34.8 ± 0.6	WCPG	$\mathbf{165.5 \pm 17.7}$	WCPG	3747 ± 229	WCPG	33.8 ± 7.2
Set-WCPG	34.7 ± 0.7	Set-WCPG	$\mathbf{174.5 \pm 10.0}$	Set-WCPG	3871 ± 207	Set-WCPG	68.3 ± 4.6
SIRSA (Ours)	$\mathbf{37.9 \pm 0.2}$	SIRSA (Ours)	$\mathbf{187.8 \pm 7.6}$	SIRSA (Ours)	$\mathbf{4146 \pm 112}$	SIRSA (Ours)	$\mathbf{83.4 \pm 4.5}$
Oracle	38.6 ± 0.3	Oracle	172.2 ± 4.1	Oracle	4246 ± 59	Oracle	78.2 ± 3.6

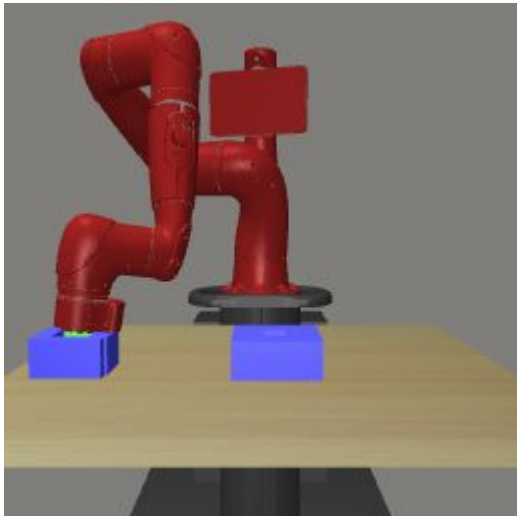
Experiments: Main Results



Set-EPOpt
(Rajeswaran et al., 2016)



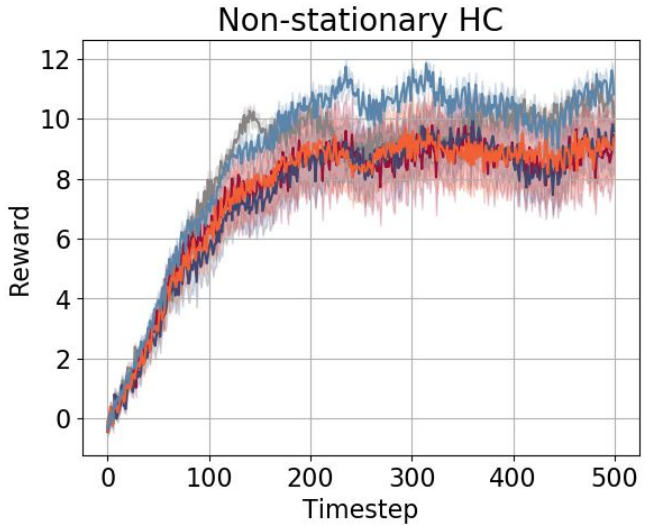
System ID
(Yu et al., 2017)



SIRSA
(Ours)

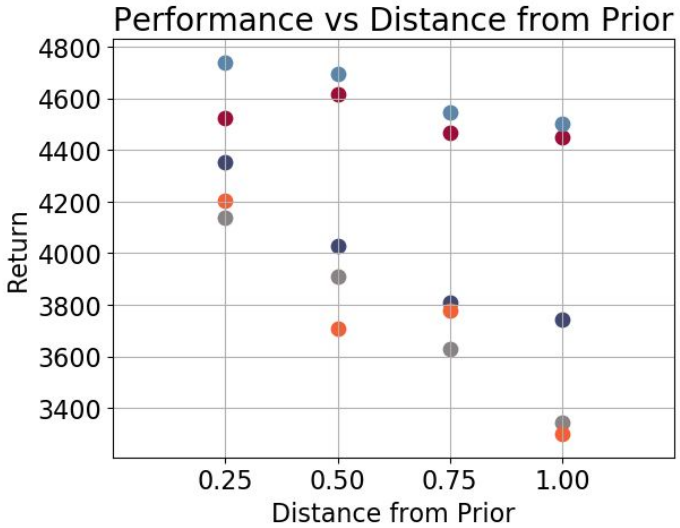
*final frame of each trial

Experiments: Generalization under Non-Stationarity



Acting robustly to the worst case is a good strategy for non-stationary environments

Experiments: Misspecified Initial Uncertainty Sets



Even when the initial uncertainty set is wrong, SIRSA and System ID can quickly find a good guess

Takeaways

- Introduced *multi-set robustness* as a more flexible & general setup for robust RL

- Designed a framework that combines probabilistic system identification with the multi-set robust RL objective

- Future work may tackle the setting where the context is not observed at training time

Takeaways

- Introduced *multi-set robustness* as a more flexible & general setup for robust RL
- Designed a framework that combines probabilistic system identification with the multi-set robust RL objective
- Future work may tackle the setting where the context is not observed at training time

Takeaways

- Introduced *multi-set robustness* as a more flexible & general setup for robust RL
- Designed a framework that combines probabilistic system identification with the multi-set robust RL objective
- Future work may tackle the setting where the context is not observed at training time