

Robots Learning (Through) Interactions

April 5th 2022

International AI Doctoral Academy

Jens Kober



Who should help whom?



verdict.com

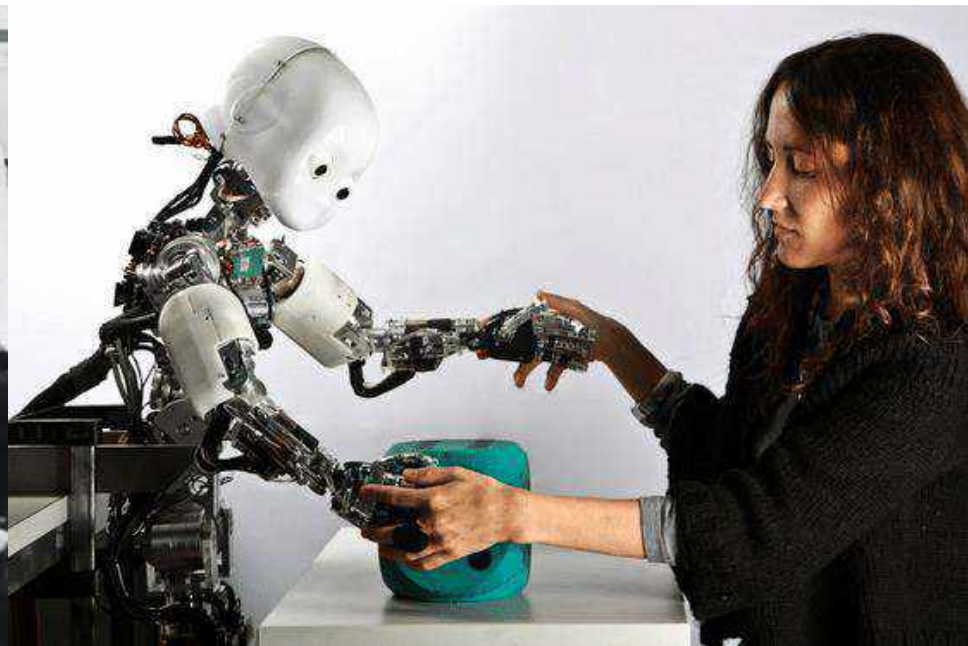


starwarsnewsnet.com

Who should teach whom?



timeshighereducation.com



ROBOSKIN project

Supervised Learning

- Teacher demonstrates skill, student tries to mimic



testmyprep.com

Reinforcement Learning

- Practice, practice, practice...



cnn.com



incaavalanche.com

What Makes Tasks Hard?

- Complex dynamics
- Uncertainties and variations
 - Objects
 - Environment
 - Tasks
 - Human behaviour
- Occurring in all robot domains!

Learning Interactions



CC BY-SA Mixalbest



CC BY-NC-SA
Brandon Heyer

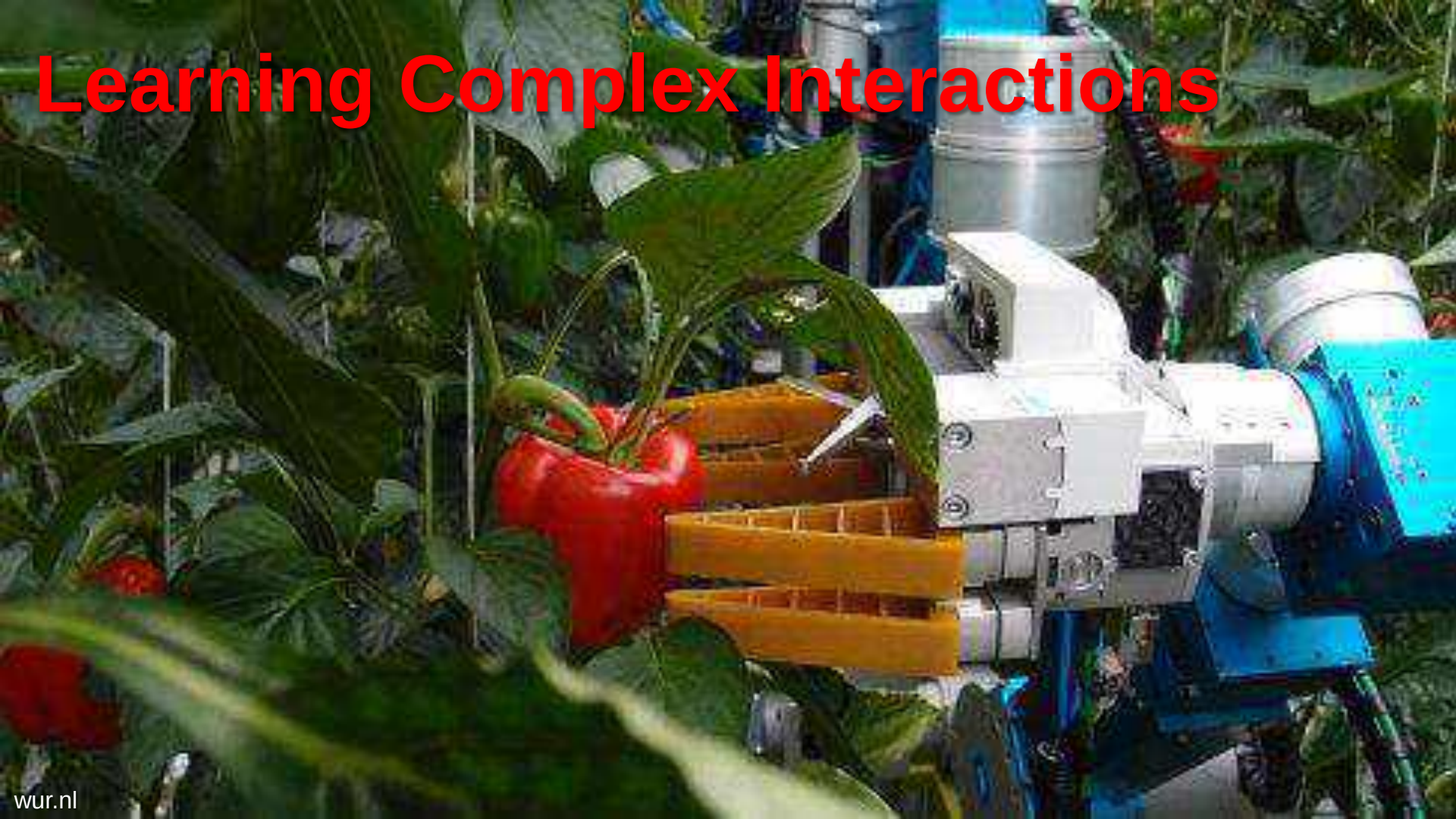
How to Learn?

- Continued student-teacher interaction
 - Additional demonstrations
 - Intermittent feedback
- Largely missing in robot learning!
- Benefits
 - Speed-up
 - Complex tasks
 - Intuitive

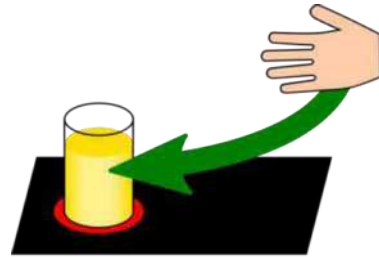
Learning Through Interactions



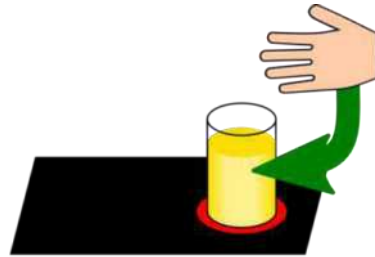
Learning Complex Interactions



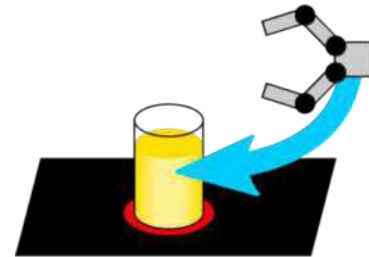
Imitation Learning



a) Demonstration 1

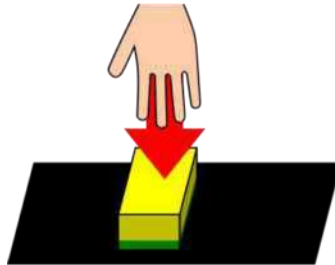


b) Demonstration 2

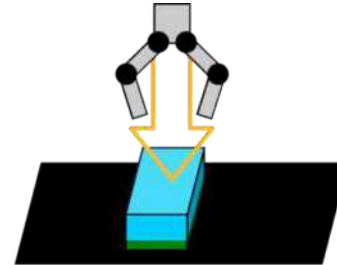


c) Generalization 1

- Combinatorial explosion
- Force $\rightarrow$ Position? Position $\rightarrow$ Force?



a) Demonstration



b) Generalization

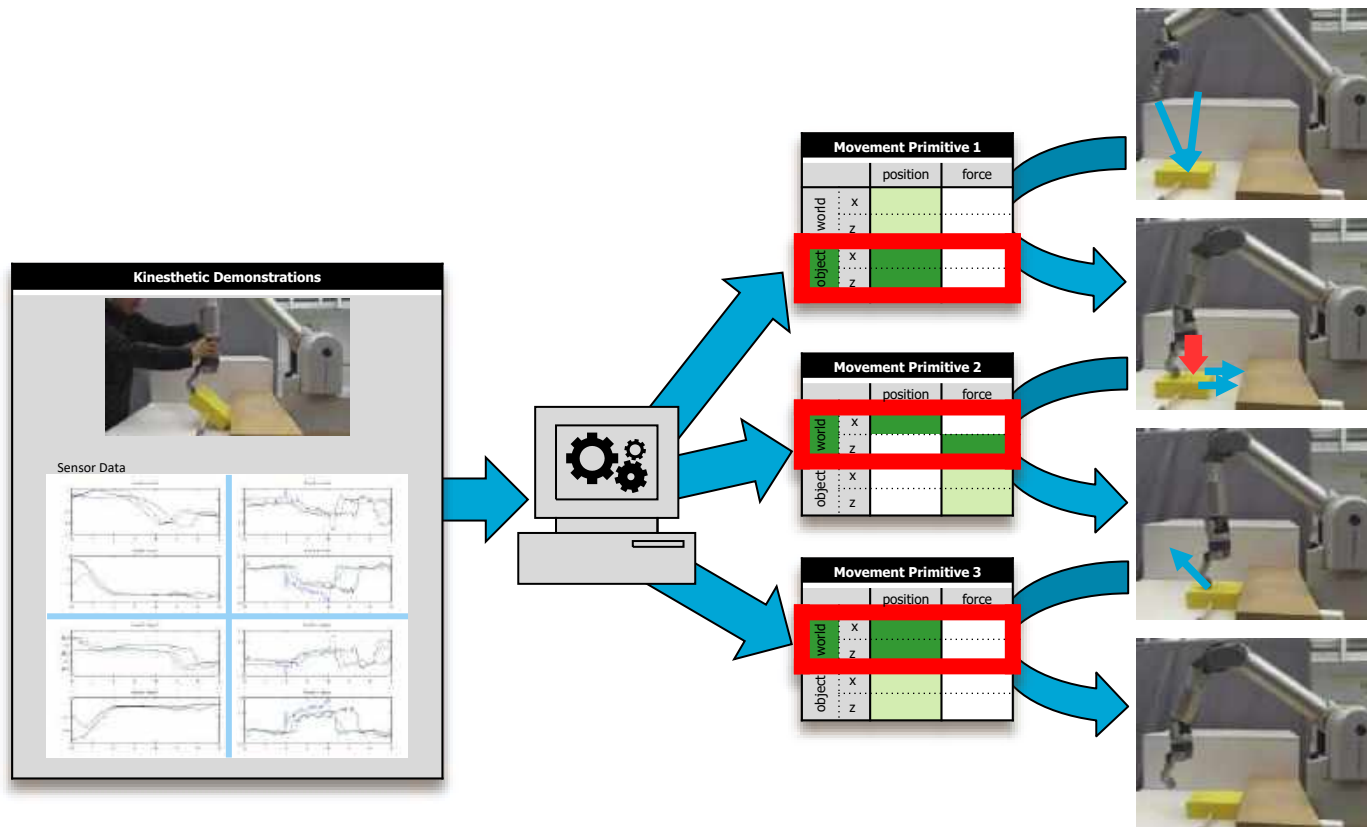
Kinesthetic teaching



Skill reproduction



Primitives within a Sequence



Primitives within a Sequence

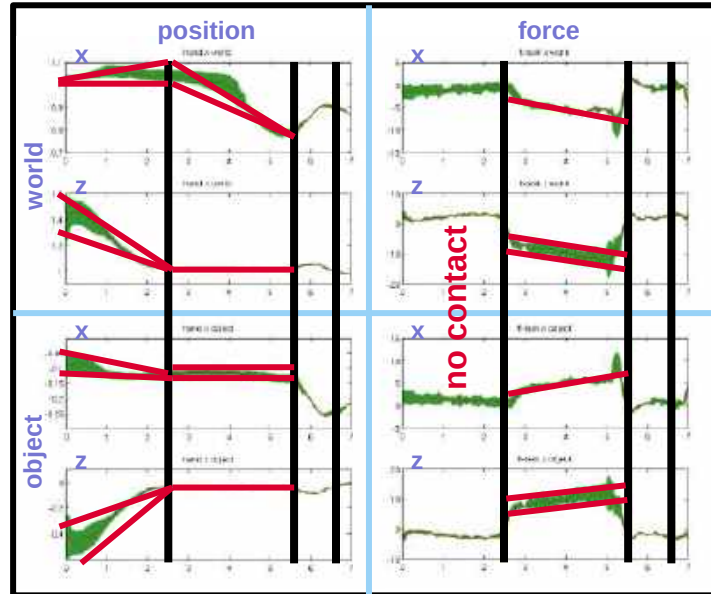
Score per MP/frame/component/modality based on statistics & contact



converging/attractor

constant

diverging



Scores: Movement Primitive 2			
		position	force
world	x	90.0	20.0
	z	10.0	16.0
object	x	8.0	20.0
	z	10.0	16.0



Scores: Movement Primitive 2			
		position	force
world	x		
	z		
object	x		
	z		





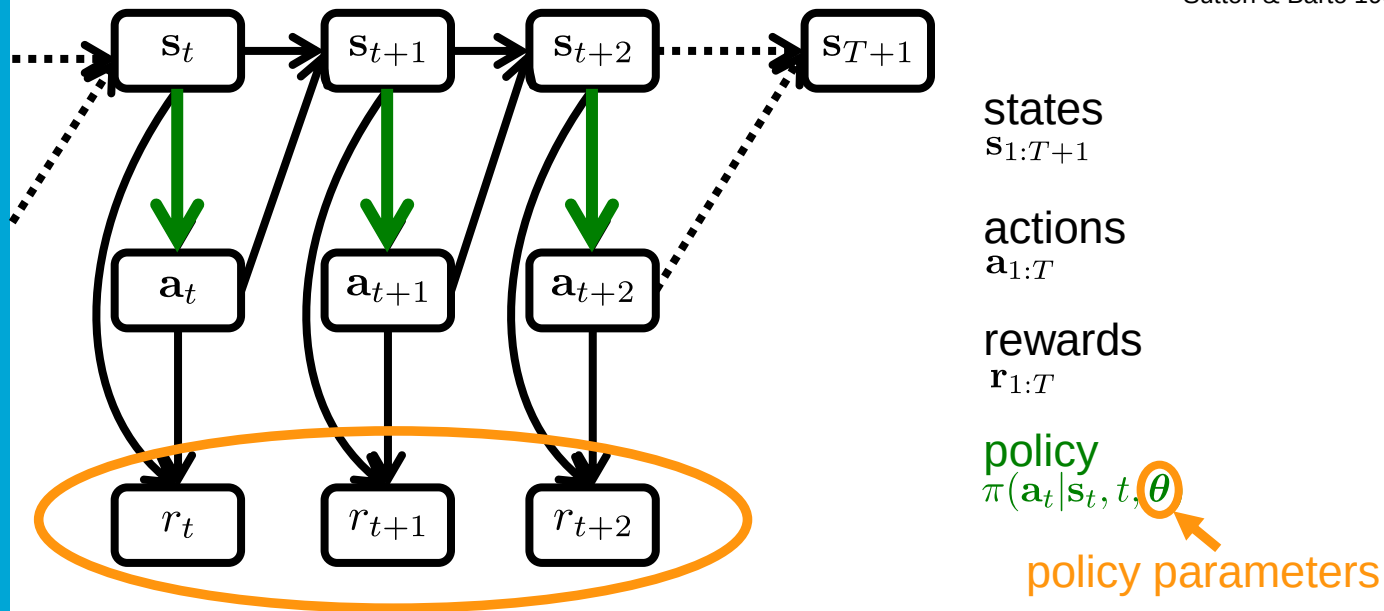
Simon Manschitz



Kinesthetic Demonstration

Reinforcement Learning

Sutton & Barto 1998



Objective: maximize expected return $J(\theta)$



The first arm posture



Jihong Zhu

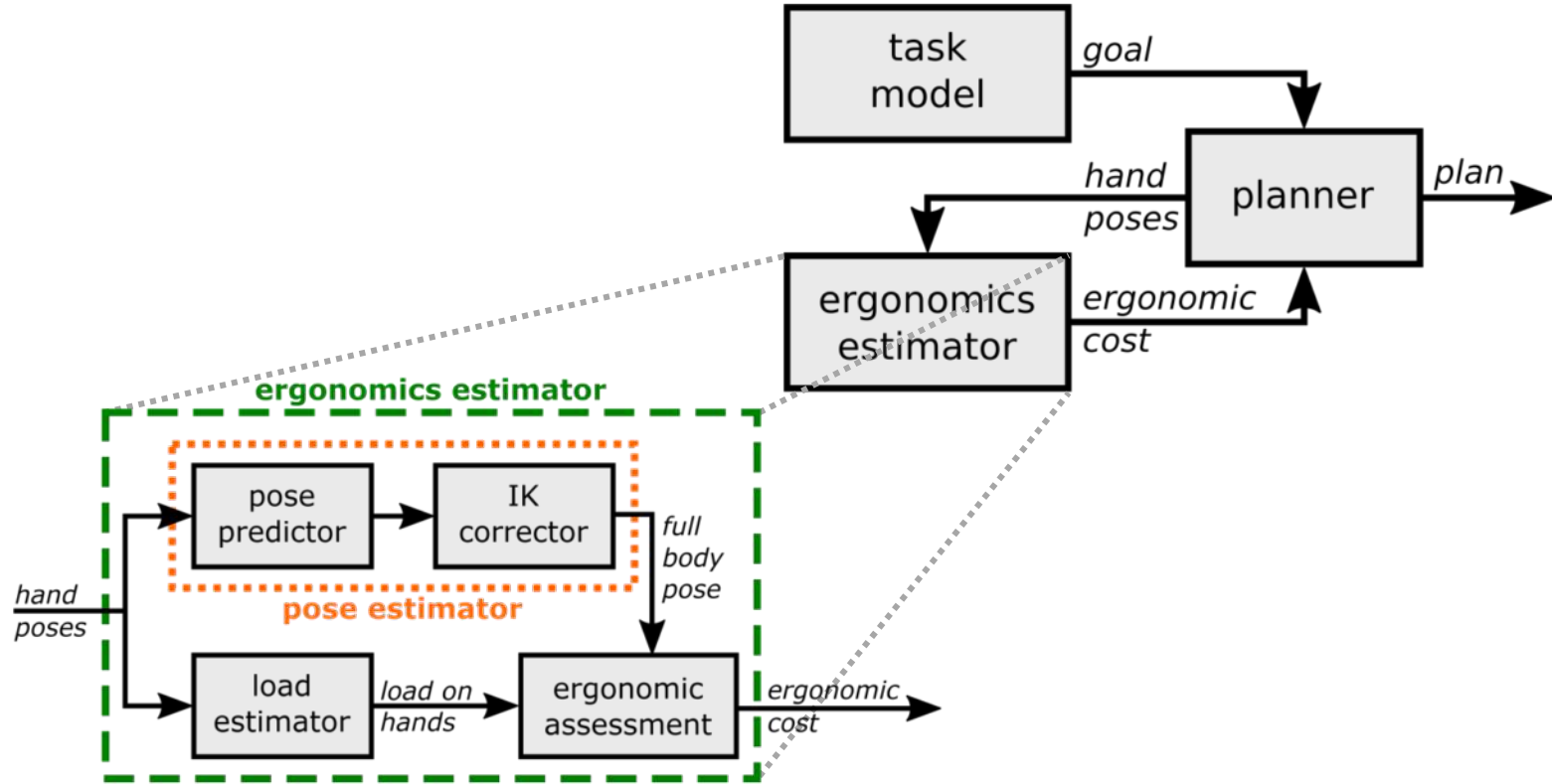


*rotate box 180 degrees
counterclockwise*



Linda
van der Spaa

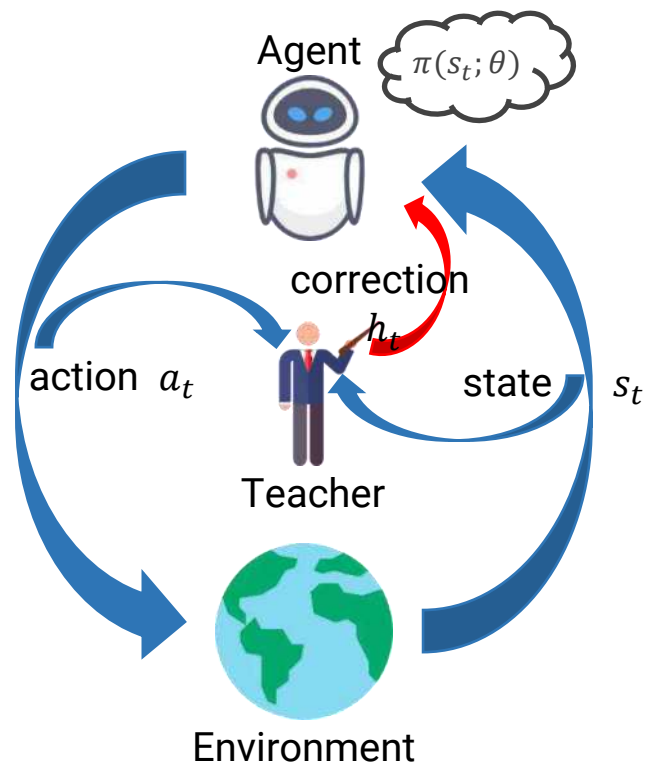
Planning for Ergonomic Poses



Learning Through Interactions



Interactive Learning



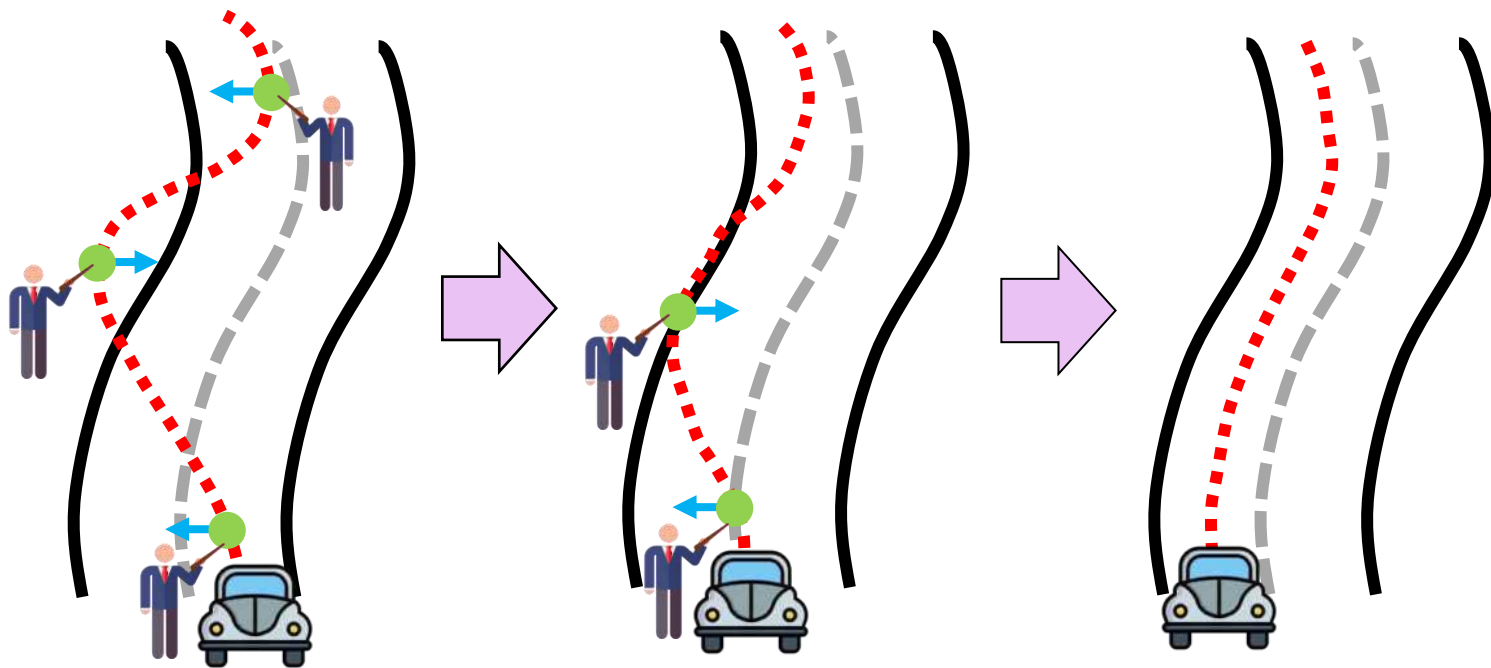
Interactive Imitation Learning





Carlos Celemin

COACH



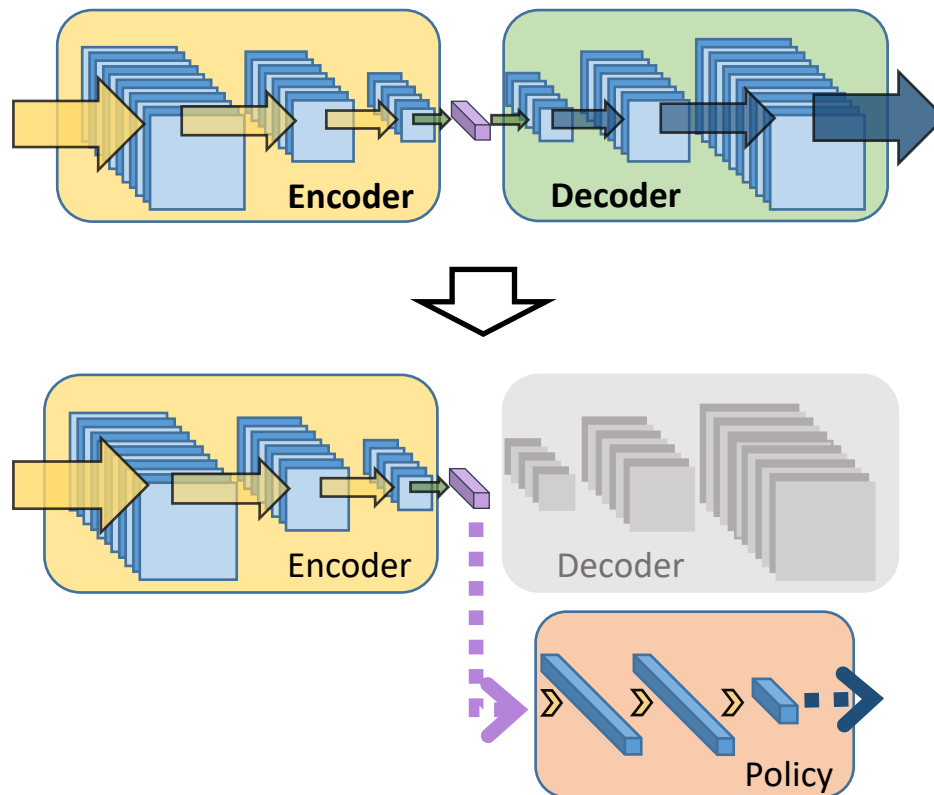


Rodrigo Pérez

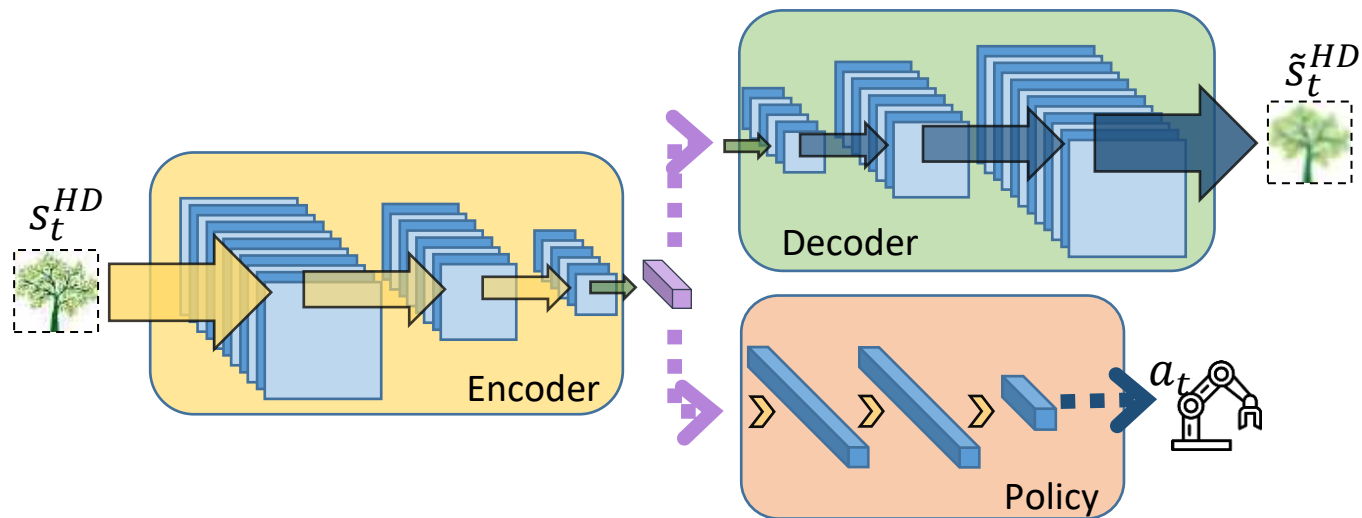


Carlos Celemin

Deep COACH - Offline



Deep COACH - Online







Rodrigo Pérez

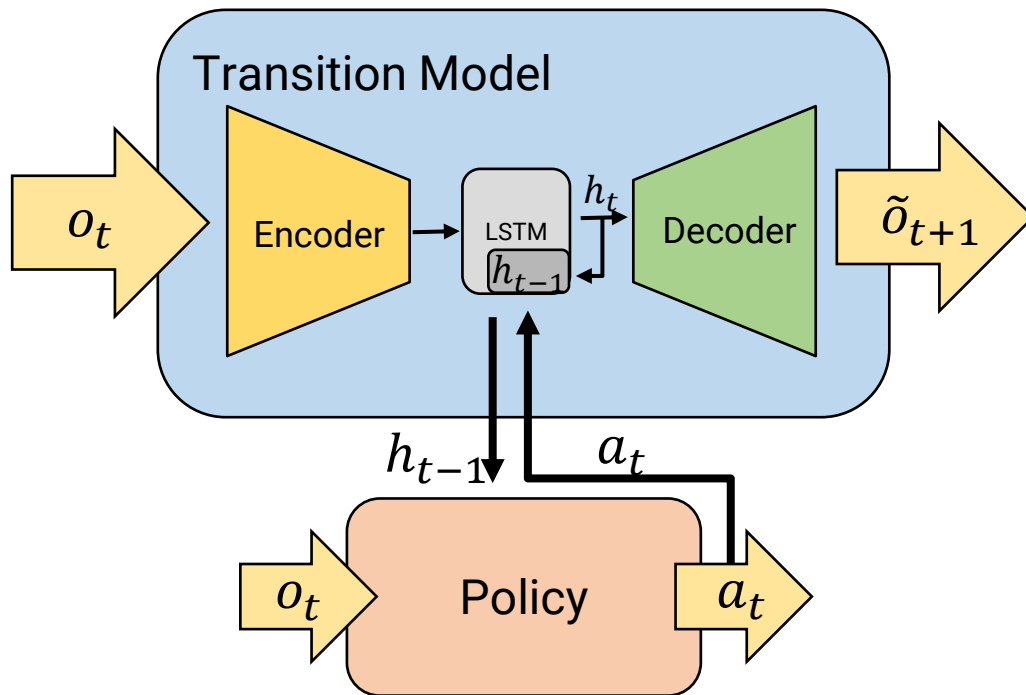


Giovanni Franzese



Carlos Celemin

Deep COACH - Memory



Simultaneous Policy and Model Learning using D-COACH and SRL

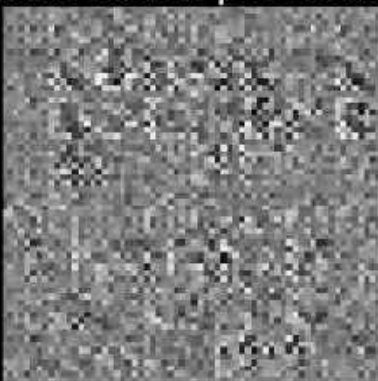
x4

Mountain Car

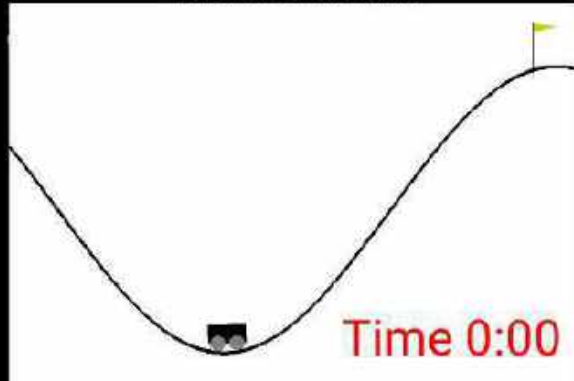
Network Input



Observation prediction



Environment

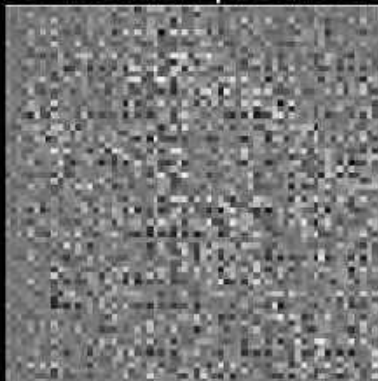


Swing-up Pendulum

Network Input

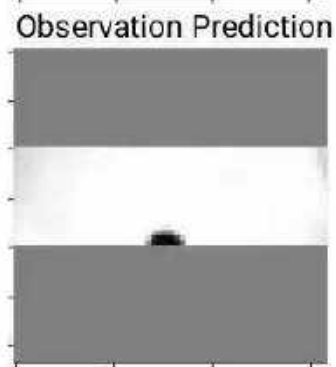
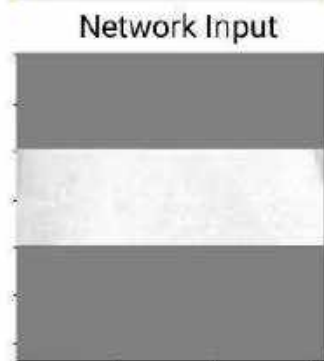
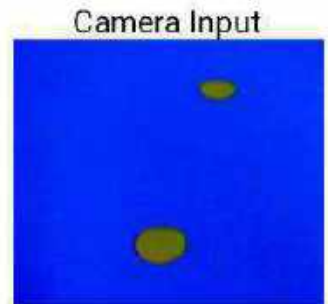


Observation prediction



Environment





x4

Start Training

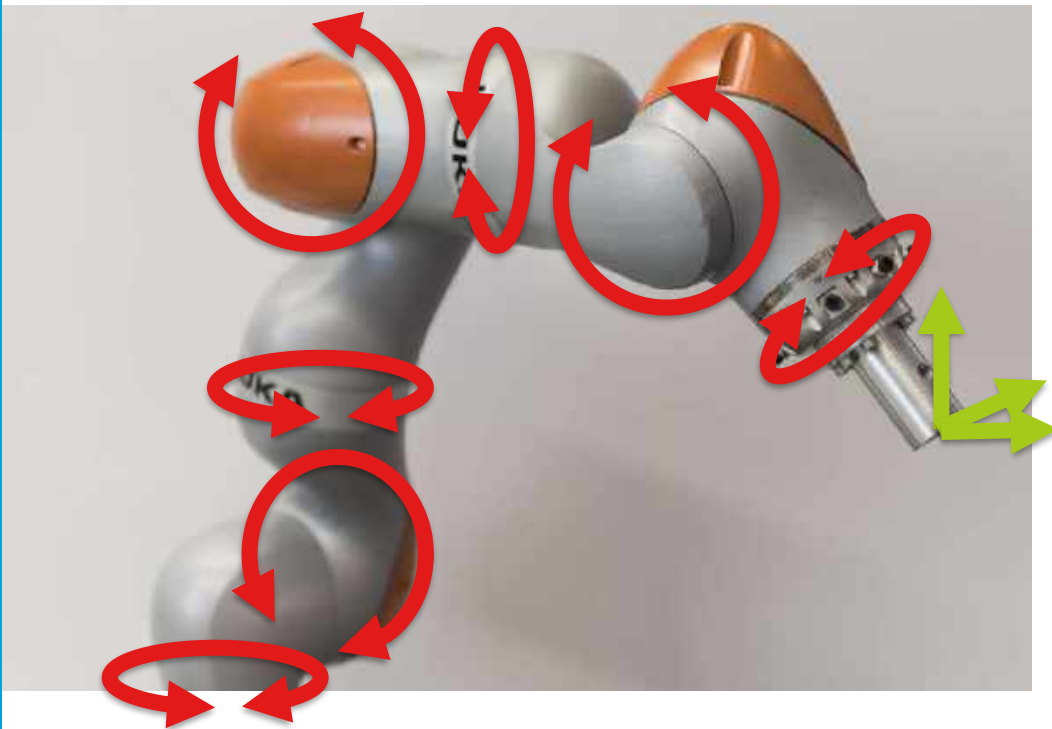


Snehal Jauhri



Carlos Celemin

Feedback in State-Space

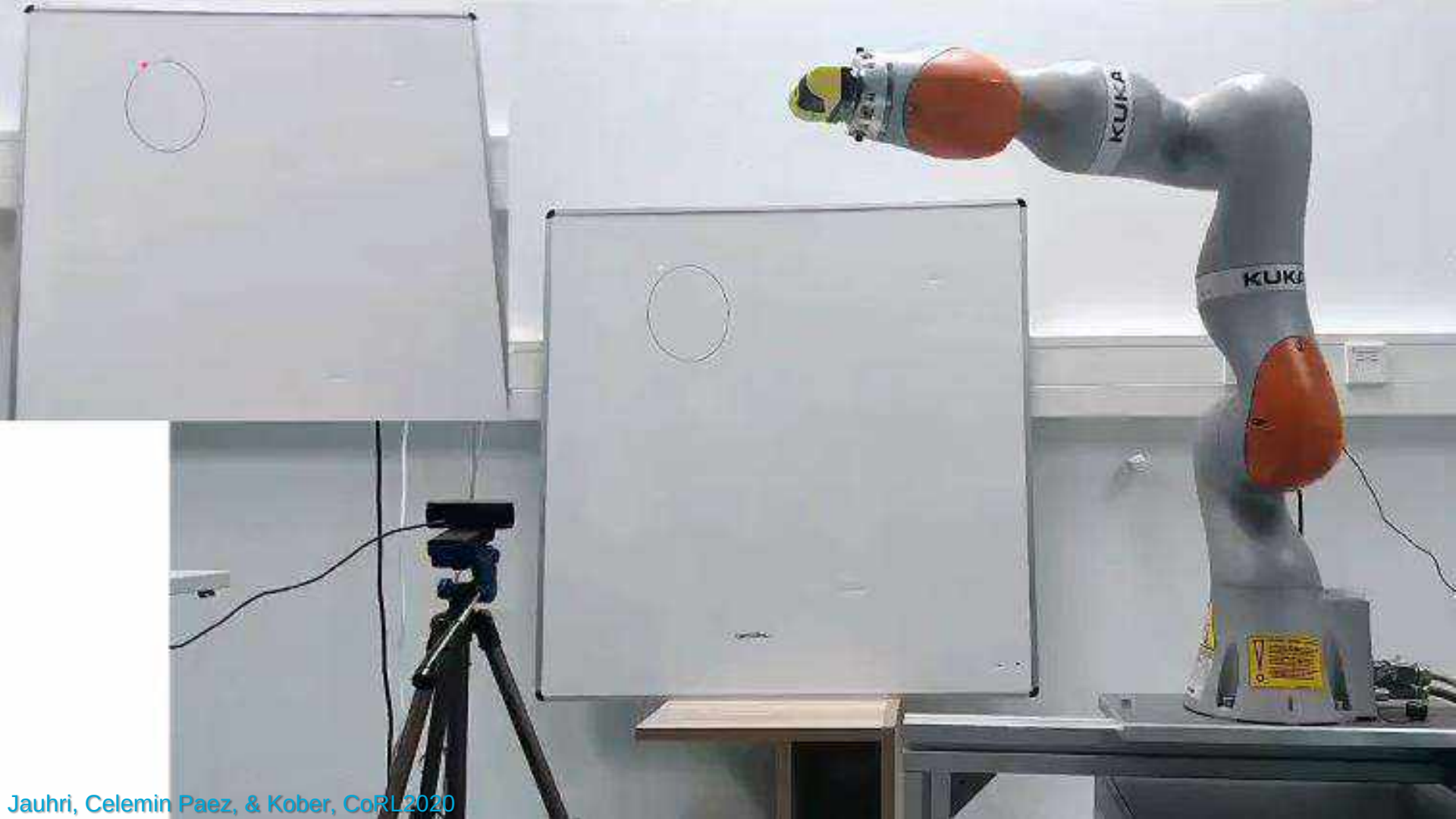


action:

- ✓ direct
- ⚡ high-dim
- ⚡ non-intuitive

state:

- ⚡ needs inverse model



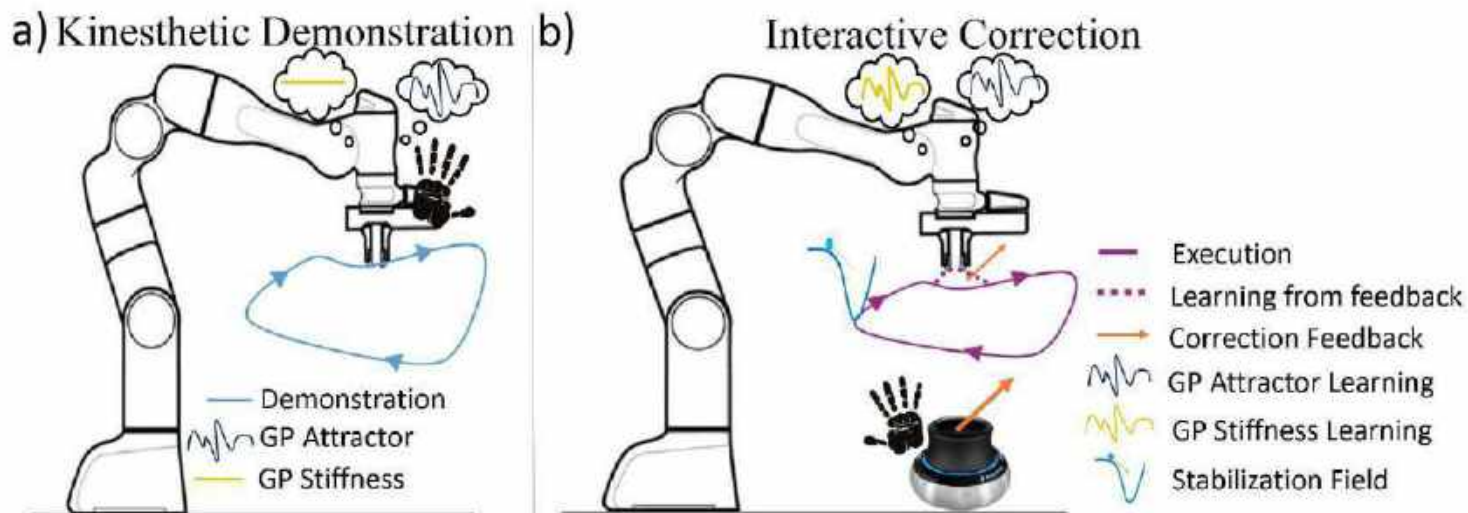


Giovanni Franzese



Anna Mészáros

Interactive Learning of Stiffness and Attractors (ILoSA)

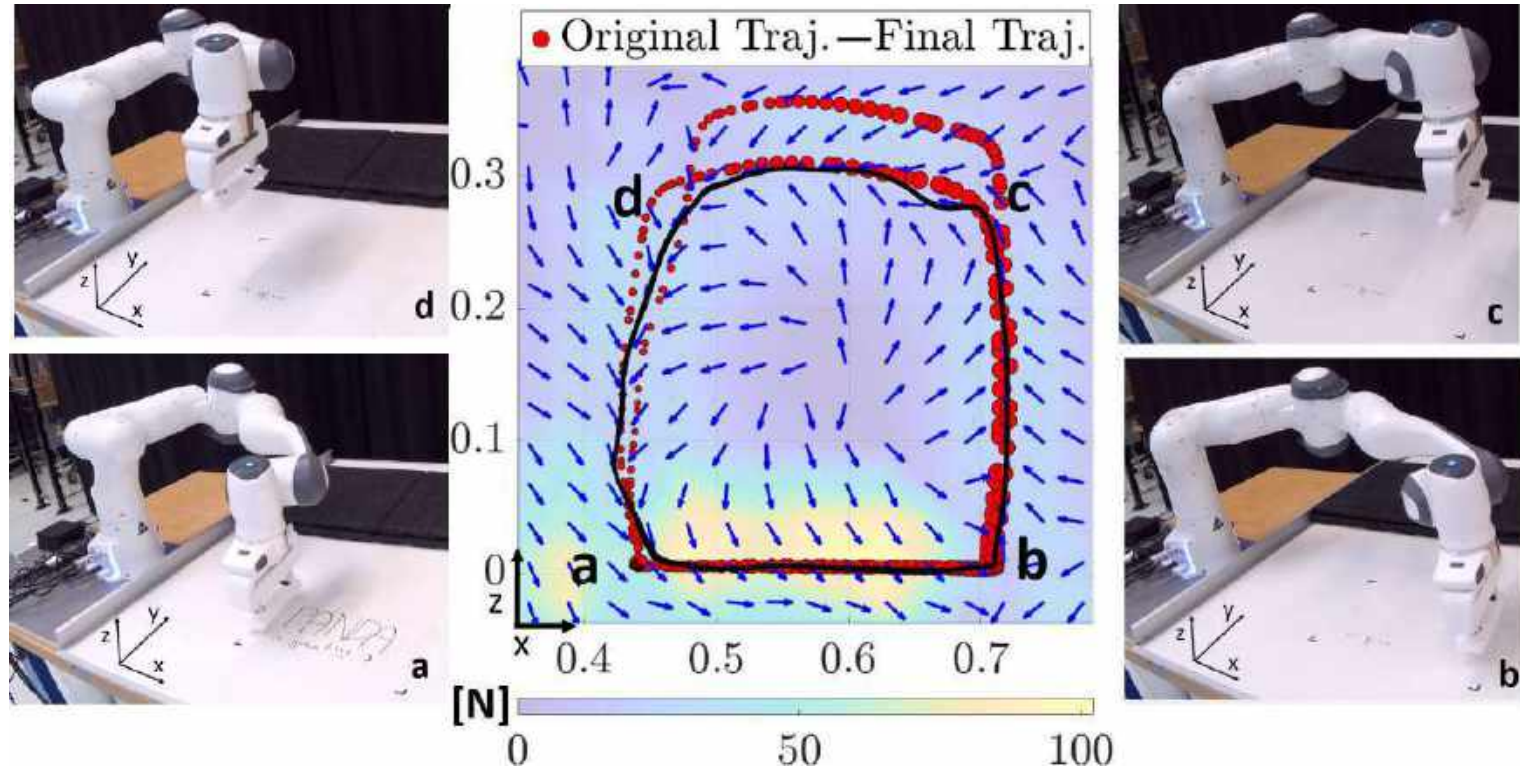


2x



Kinesthetic Demonstration

Learned Trajectory





Kinesthetic Demonstration



Giovanni Franzese



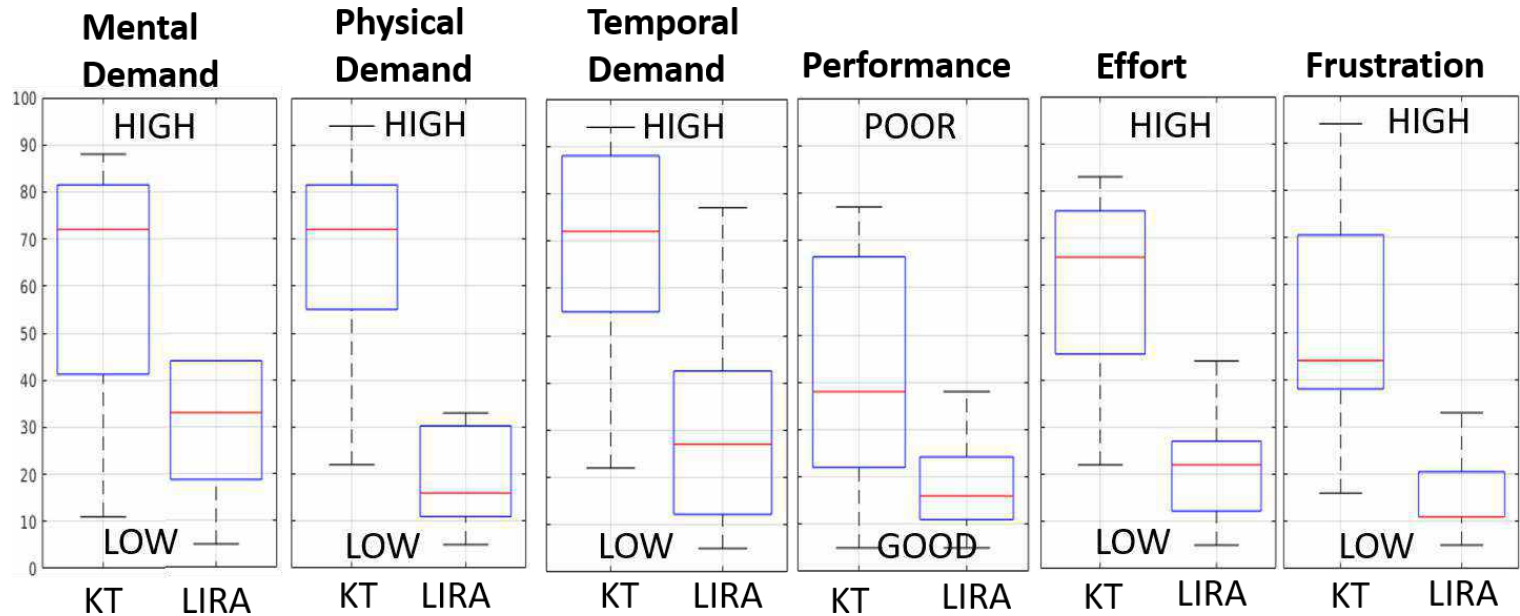
Carlos Celemin

Resolving Ambiguities



Task 1: Fruit sorting in crates

NASA-TLX



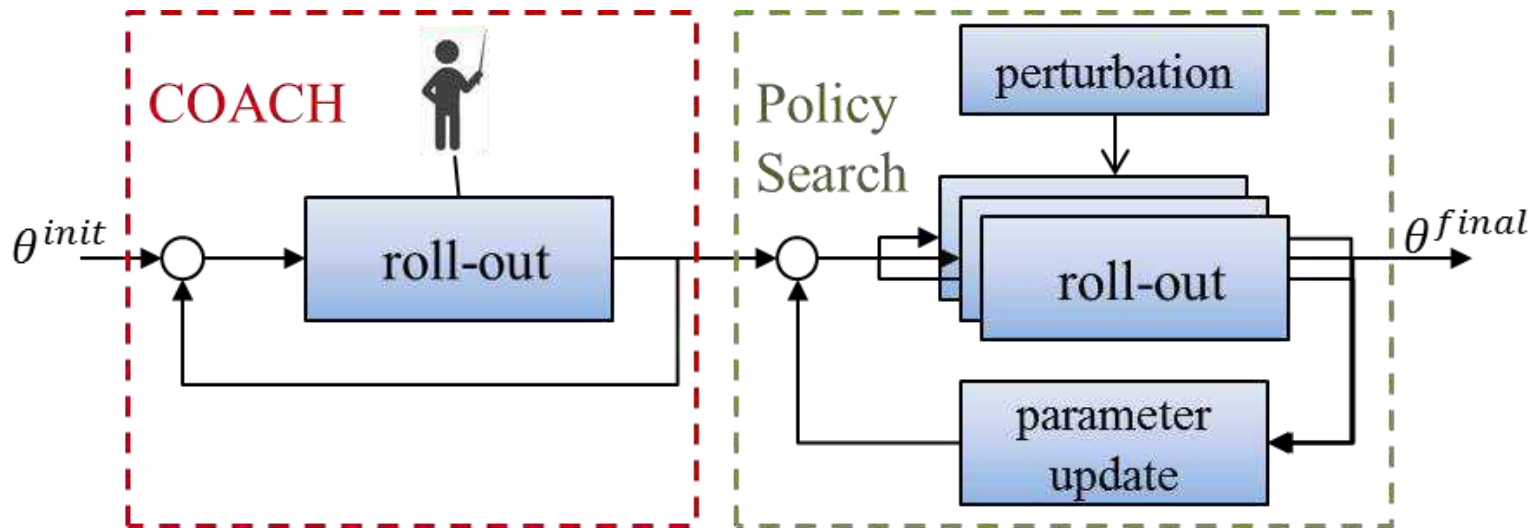
Interactive Reinforcement Learning



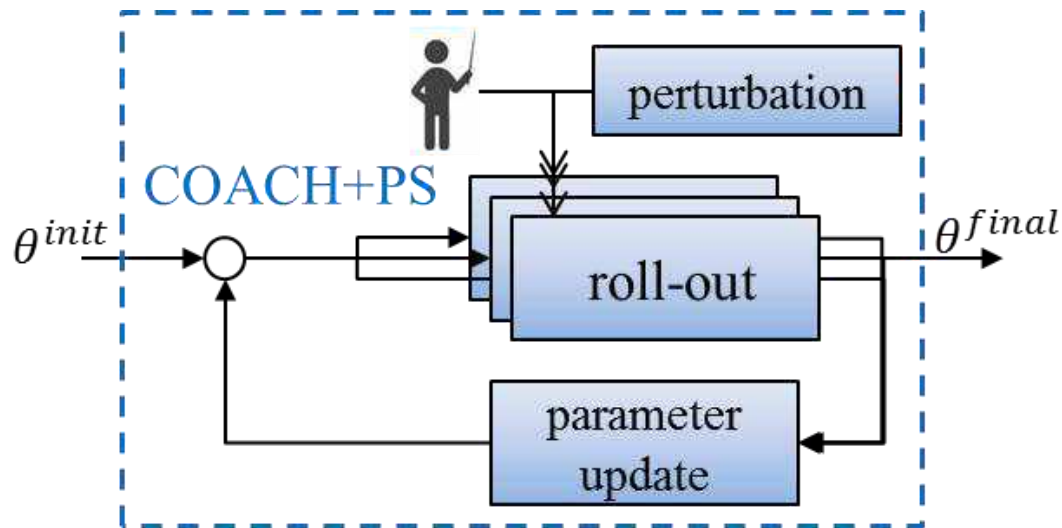


Carlos Celemin

IL + RL: Sequential

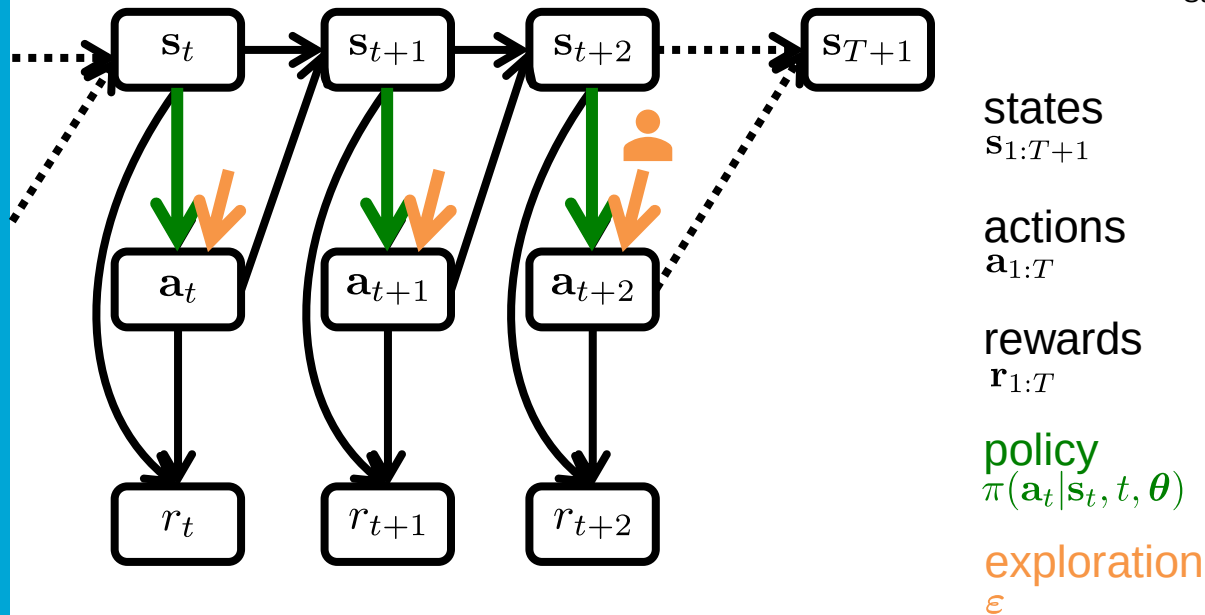


IL + RL: Simultaneous



Reinforcement Learning: Exploration & Human Advice

Sutton & Barto 1998



Model human advice as exploration

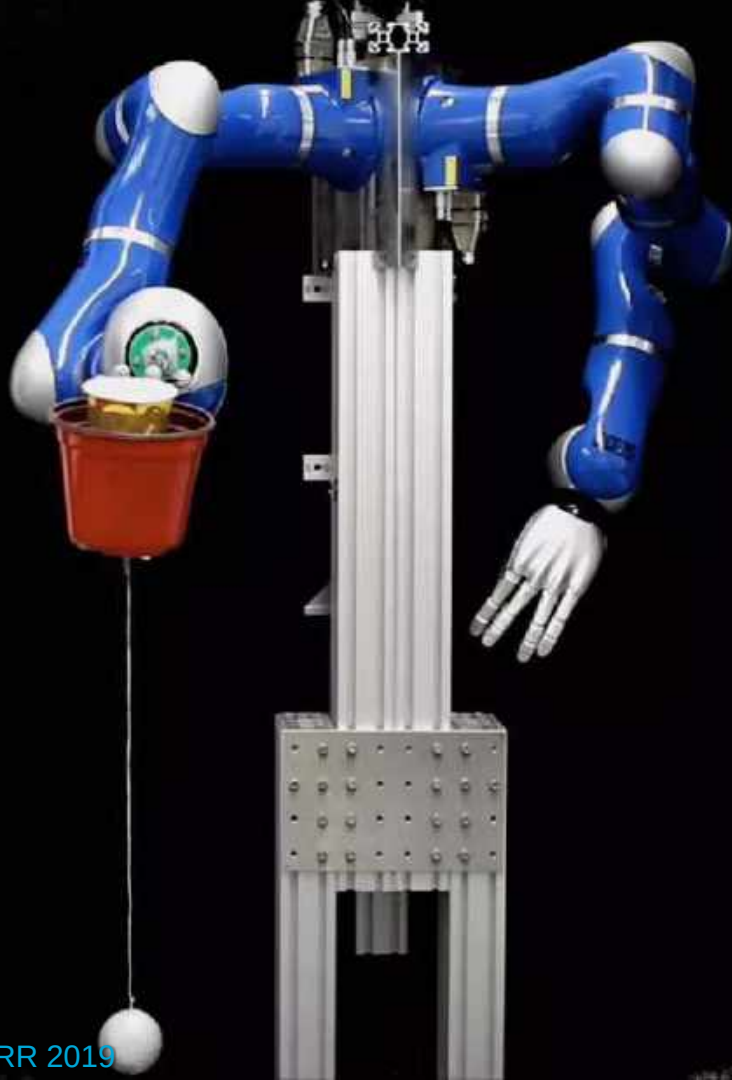
Interactive Policy Search

Algorithm 4 Policy Search with Simultaneous Human Guidance.

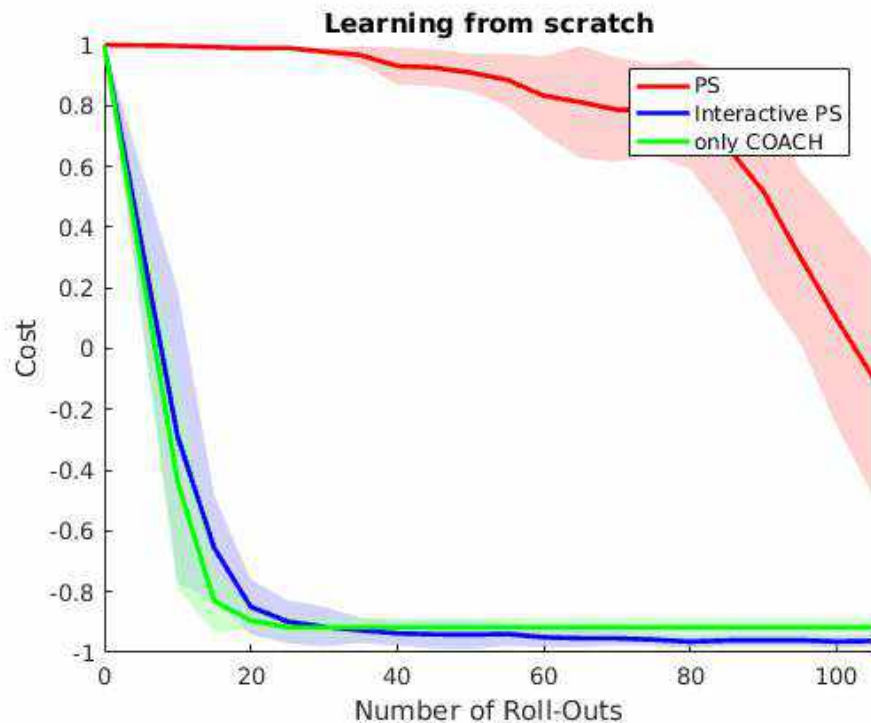
```
1: repeat
2:   Explore: first roll-out with the current policy  $\pi_k$ 
       $w_1 = w^{[k]}$ 
3:   for  $m = 1 \dots M$  do
4:     if HumanGuidance==True do
5:        $[w_m]_t \leftarrow \text{RunRollOut}_{COACH}([w_m]_1)$ 
6:        $[w_{m+1}]_1 \leftarrow [w_m]_T$ 
7:     else
8:        $[w_m]_t \leftarrow PS\_exploration(w^{[k]})$ 
9:        $\text{RunRollOut}([w_m]_t)$ 
10:   Evaluate: cost of each roll-out
       $[R_m]_t = \phi_T + \sum_{t=0}^T r_t^{[m]}$ 
11:   Update: Compute new policy parameters using
       $w^{[k+1]} \leftarrow \text{Update}([w_m]_t, [R_m]_t)$ 
12: until Policy converges  $\pi_{k+1} \approx \pi_k$ 
```



Carlos Celemin



Results



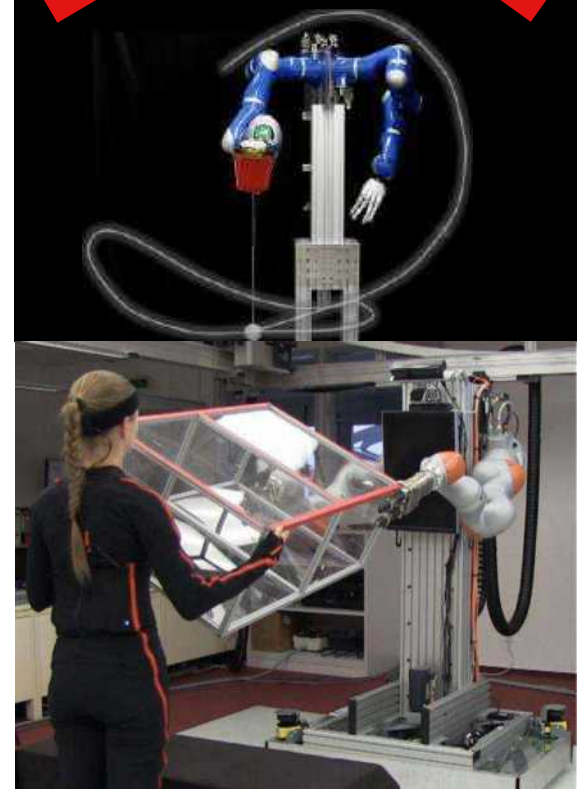
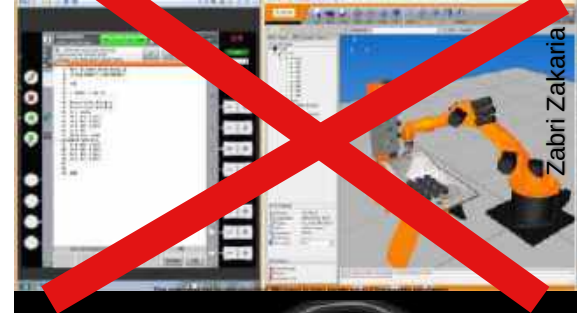


Walt Disney Studios – Real Steel

Conclusion

Conclusion

- Programming → Teaching
- Learning from Teachers
+ Self-learning
- Need to understand
Environment & Humans



Conclusion

- Big data
 - Unlabeled data = cheap
 - Annotated data = expensive
 - Real world interaction = extremely expensive
- Tractability through
 - Combining learning/teaching paradigms
 - Prior knowledge



Questions?

j.kober@tudelft.nl

www.jenskober.de

