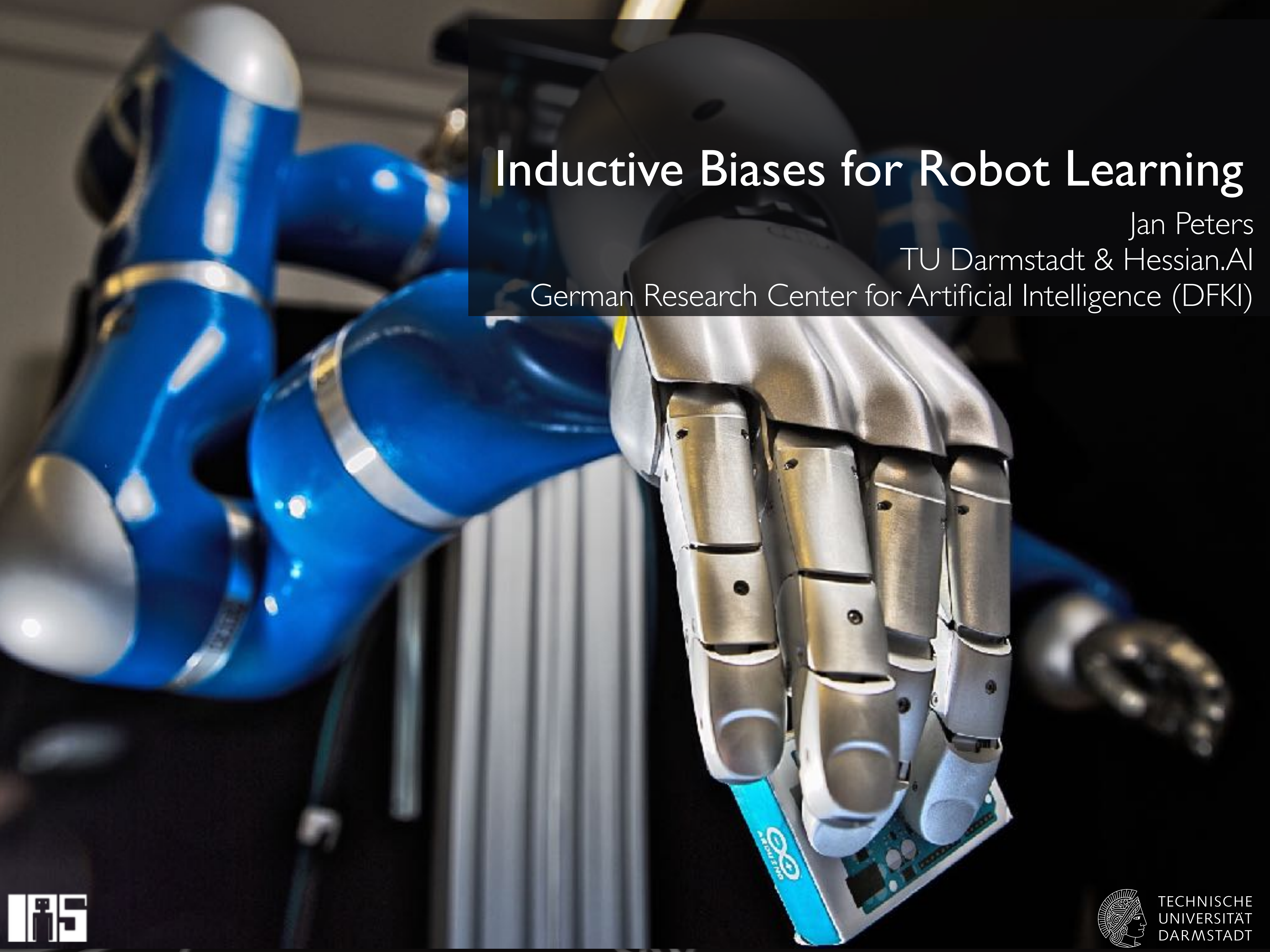


Inductive Biases for Robot Learning

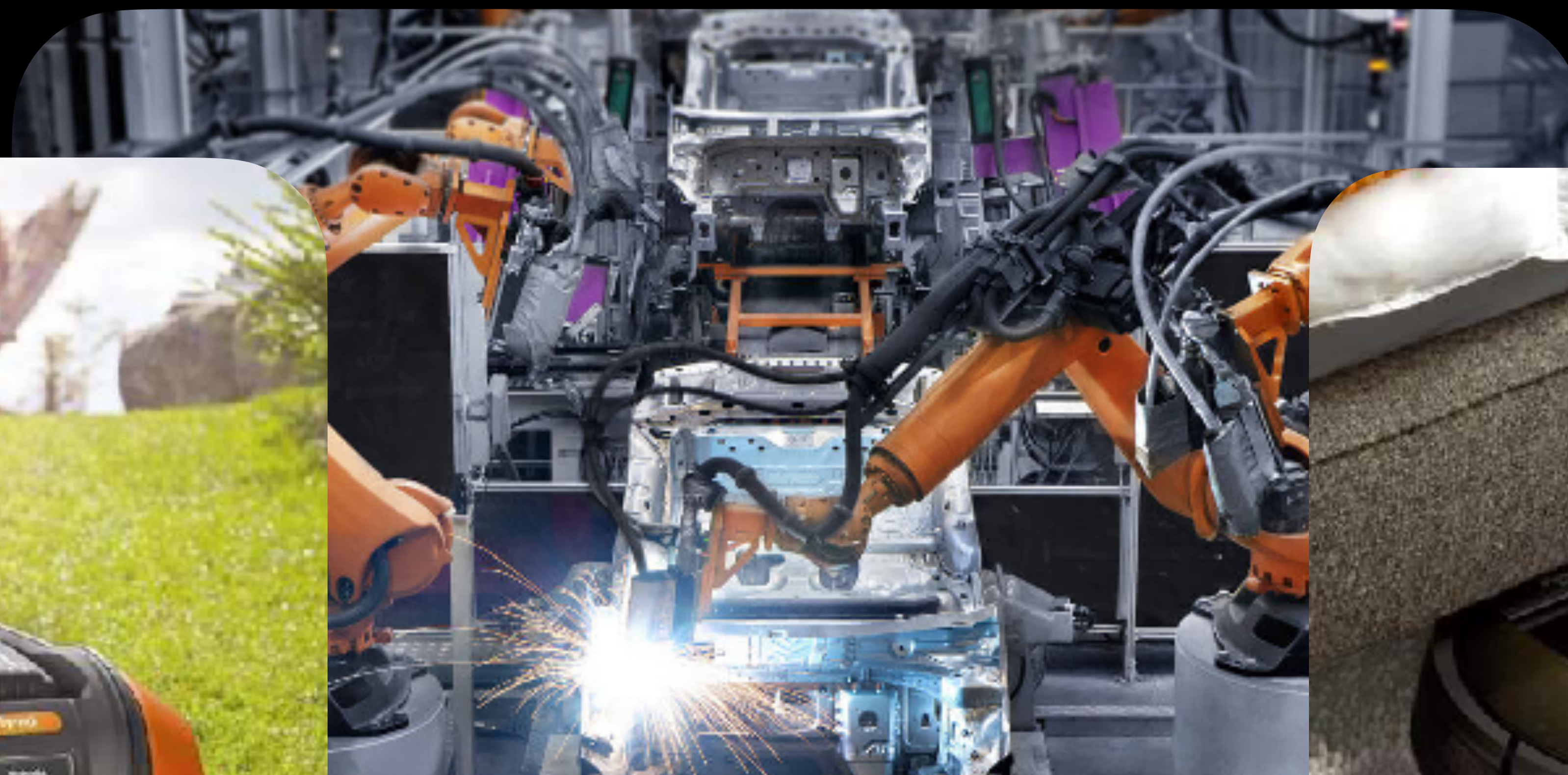
Jan Peters
TU Darmstadt & Hessian.AI
German Research Center for Artificial Intelligence (DFKI)

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Why don’t we have such robots?

Source: British TV (1969)

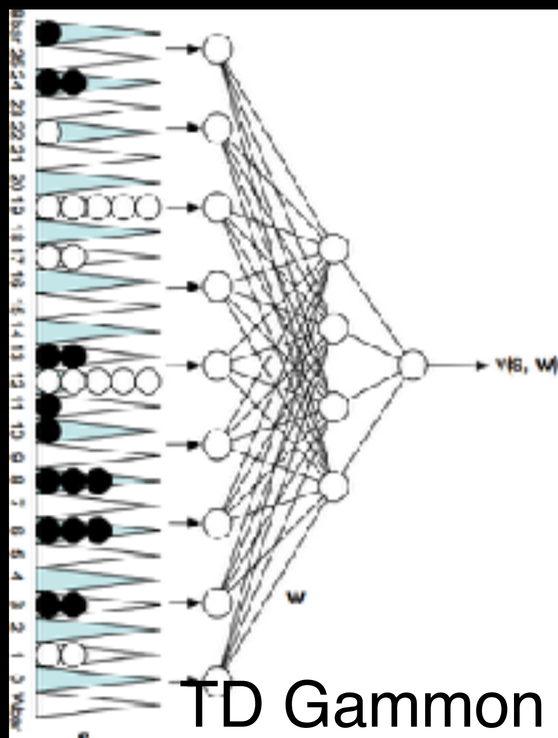
So where have robots been successful?



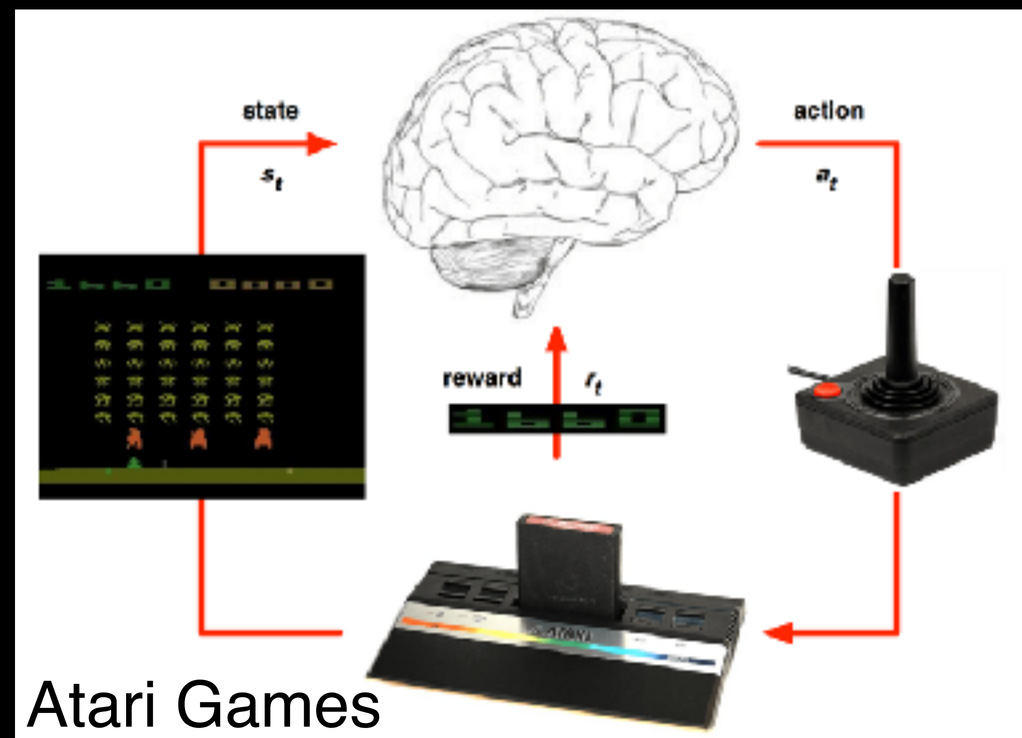
Whenever we adapt tasks to robots!

We need to adapt robots to tasks!

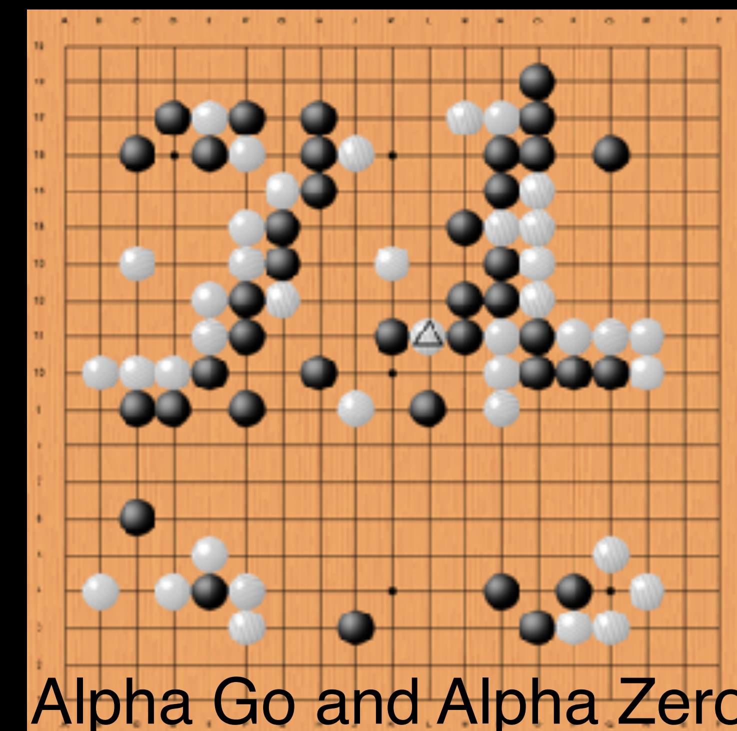
Adapting Robots to Tasks?

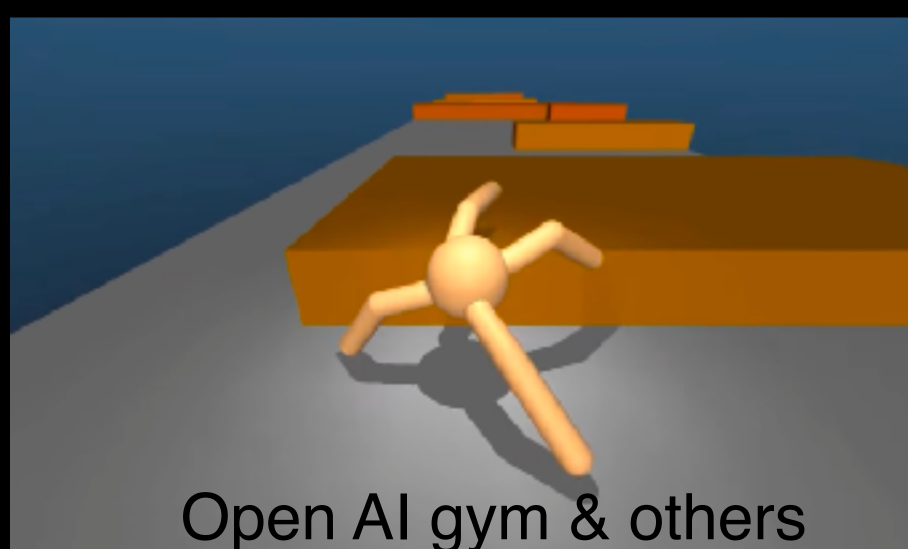


TD Gammon



Atari Games

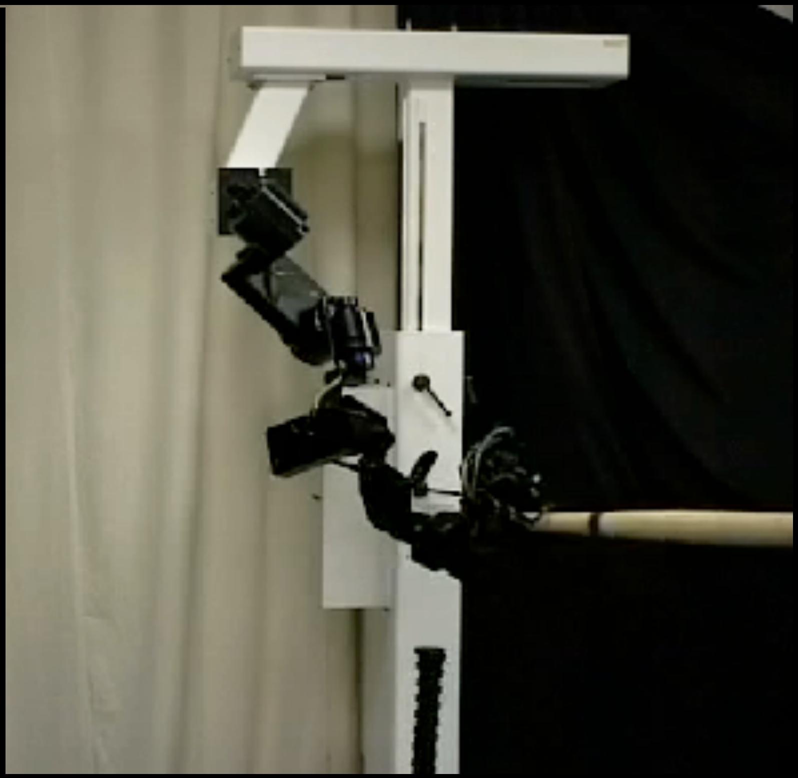

Alpha Go and Alpha Zero


Open AI gym & others

Open AI

Do Robot Learning?

Robot Learning…



- Classical robot engineering is really good at adapting tasks to the robot!
- Data is very expensive!
- If you can learn in a robot simulator, you don't need RL!
- Things break!
- Generalization is often impossible: "Golf does not help Hockey!" (John Milton)
- Learning in Real-Time
- Computation, communication and energy limitations…

…is not a straightforward answer!

Robot Reinforcement Learning requires…

Robotics Inductive Biases

An inductive bias allows a learning algorithm to prioritize one solution (or interpretation) over another, independent of the observed data. [...] Inductive biases can express assumptions about either the data-generating process or the space of solutions.

(Mitchell, 1980; Battaglia et al., 2018)

Robot Reinforcement Learning requires…

Robotics Inductive Biases

What inductive biases does robotics offer? How can we use them for improving robot reinforcement learning?

Outline

1. Inductive Bias
2. Inductive Bias
3. Inductive Bias
4. Inductive Bias
5. Inductive Bias
6. Inductive Bias

Imitation Learning is always easier than Reinforcement Learning

Imitation Learning

Model-Based Behavioral Cloning (Englert et al.)

Objective: Policy Similarity

$$\max_{\pi,\mu^\pi} J(\pi) = \sum_{s,a} \mu^\pi(s)\pi(a|s) \log \frac{\mu^\pi(s)\pi(a|s)}{q(s,a)}$$

Model-Free Behavioral Cloning (Michie & Chambers, Sammut et al.)

Constraints: Assumptions on the Policy

$$\mu^\pi(s') = \sum_{s,a} \mathcal{P}^a_{ss'} \mu^\pi(s)\pi(a|s)$$

$$1 = \sum_{s,a} \mu^\pi(s)\pi(a|s)$$

Dual Problem

Putermann (1998) implies:
IRL is harder than MBC!

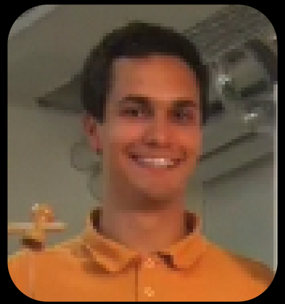
Dual Function for Minimal Physics

Inverse Reinforcement Learning

(Ziebart et al.; Boularias et al.)

Solve for the optimal parametric policy class: Motor primitives

(Schaal et al; Kober et al; Paraschos et al; Gomez-Gonzalez)


Jens Kober

Learning Perception-adapted Probabilistic Motor Primitives

Learning from human demonstrations

Sebastian Gomez-Gonzalez

Gomez-Gonzalez, S.; Neumann, G.; Schölkopf, B.; Peters, J. (2020). Adaptation and Robust Learning of Probabilistic Movement Primitives, IEEE Transactions on Robotics (T-Ro), 36, 2, pp.366-379.

Reinforcement Learning Problem

Dual: RL by Linear Programming

Objective: Expected Returns

$$\max_{\pi,\mu^\pi} J(\pi) = \sum_{s,a} \mu^\pi(s)\pi(a|s)\mathcal{R}_{sa}$$

Constraints: Assumptions on the Policy

$$\mu^\pi(s') = \sum_{s,a} \mathcal{P}^a_{ss'}\mu^\pi(s)\pi(a|s)$$

$$1 = \sum_{s,a} \mu^\pi(s)\pi(a|s)$$

Dual Problem

Bellman's Principle of Optimality

Primal: RL by Linear Programming

"Bellman Equation":

$$V^*(s) = \max_a E_{s'}\left\{r(s,a,s') + \gamma V(s')\right\}$$

No natural notion of data!

1. Inductive Bias

Robotics
Inductive
Bias

Stay close to your training data

Relative Entropy Policy Search

Dual: RL by Linear Programming

Objective: Expected Returns

$$\max_{\pi,\mu^\pi} J(\pi) = \sum_{s,a} \mu^\pi(s)\pi(a|s)\mathcal{R}_{sa}$$

Constraints: Assumptions on the Policy

$$\mu^\pi(s') = \sum_{s,a} \mathcal{P}^a_{ss'}\mu^\pi(s)\pi(a|s)$$

$$1 = \sum_{s,a} \mu^\pi(s)\pi(a|s)$$

Further Constraint: Policy Similarity

$$\epsilon \geq \sum_{s,a} \mu^\pi(s)\pi(a|s)\log\frac{\mu^\pi(s)\pi(a|s)}{q(s,a)}$$

Objective from Behavioral Cloning

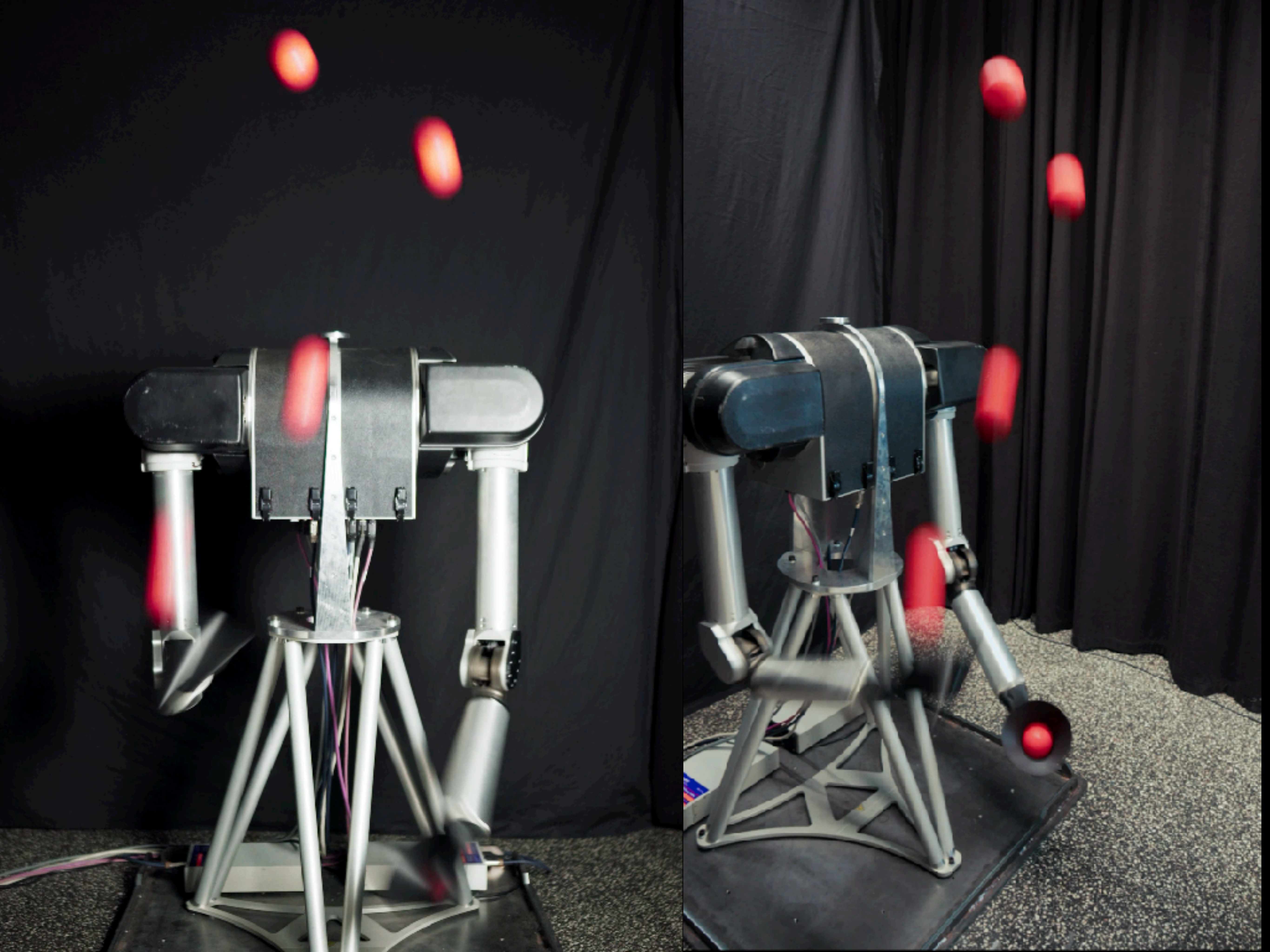
Peters (2007). Relative Entropy Policy Search, Tech. Rep.
Peters, Muelling, Altun (2010). Relative Entropy Policy Search, AAAI

Different *q* yield analytical solution, mellow/softmax, entropy regularization, SAC, …

Natural policy gradients/NAC/TRPO are approximations!

Ploeger, K.; Lutter, M.; Peters, J. (2020). High Acceleration Reinforcement Learning for Real-World Juggling with Binary Rewards, Proceedings of the 4th Conference on Robot Learning (CoRL).

Ploeger, K.; Lutter, M.; Peters, J. (2020). High Acceleration Reinforcement Learning for Real-World Juggling with Binary Rewards, Proceedings of the 4th Conference on Robot Learning (CoRL).



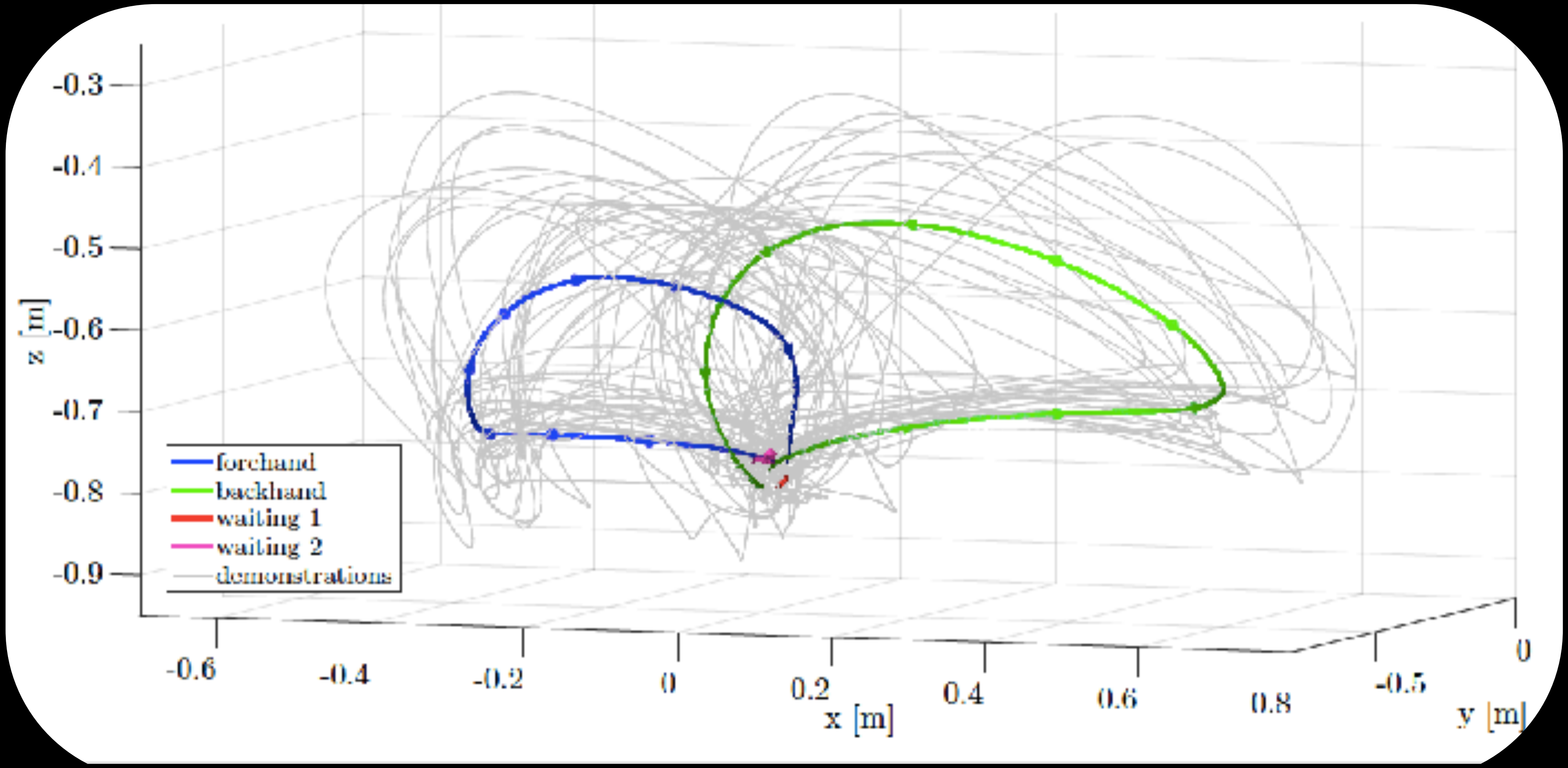
Outline

1. Inductive Bias: Stay close to your training data!
2. Inductive Bias
3. Inductive Bias
4. Inductive Bias
5. Inductive Bias
6. Inductive Bias

(Robot) Movement is composed only of strokes or rhythmic behavior

Learning from a single long demonstration…

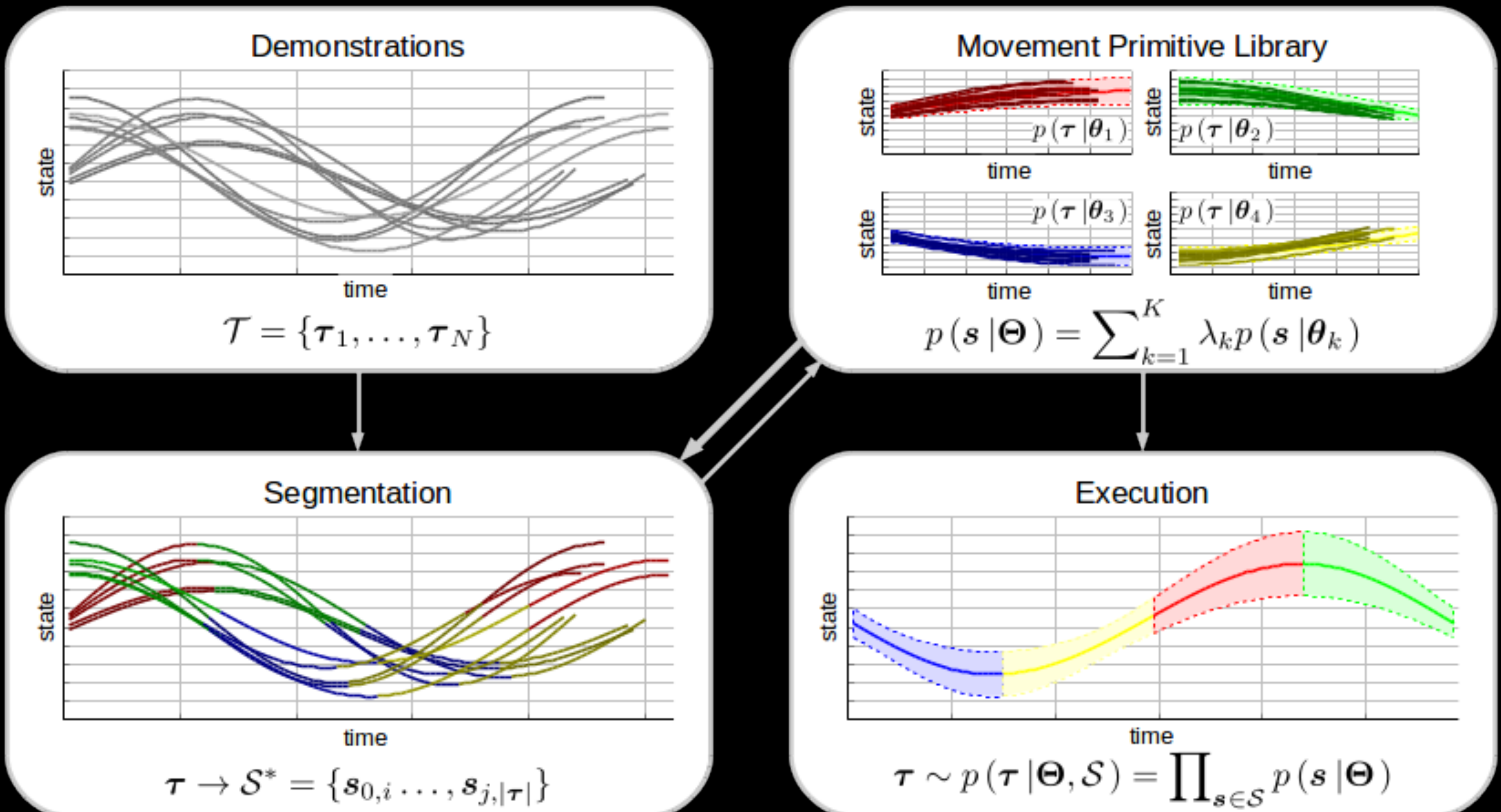
Rudolf Lioutikov



Lioutikov, R.; Neumann, G.; Maeda, G.; Peters, J. Learning Movement Primitive Libraries through Probabilistic Segmentation, International Journal of Robotics Research

…generates modular movement libraries!

Rudolf Lioutikov



Lioutikov, R.; Neumann, G.; Maeda, G.; Peters, J. Learning Movement Primitive Libraries through Probabilistic Segmentation, International Journal of Robotics Research (IJRR).

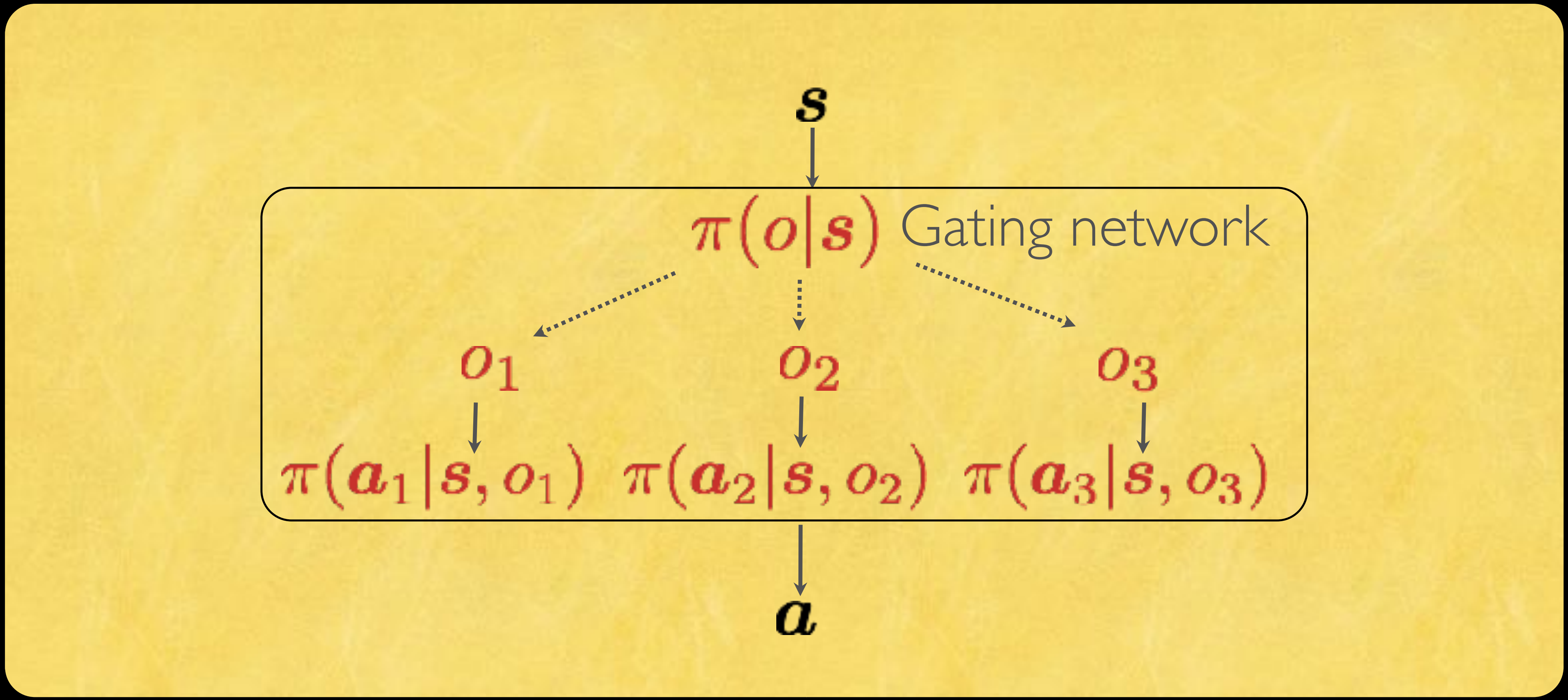
TECHNISCHE UNIVERSITÄT DARMSTADT

2. Inductive Bias

Robotics
Inductive
Bias

Use modular policy structure for composition!

Modular Control Policies



Mülling, K.; Kober, J.; Kroemer, O.; Peters, J. Learning to Select and Generalize Striking Movements in Robot Table Tennis, International Journal on Robotics Research.

Daniel, Neumann & Peters. Hierarchical Relative Entropy Policy Search, JMLR

“Naïve” Extension of REPS

Relative Entropy Policy Search (REPS)

$$\max_{\pi,\mu^\pi} J(\pi) = \sum_{s,a} \mu^\pi(s)\pi(a|s)\mathcal{R}_{sa}$$ Maximize reward

$$1 = \sum_{s,a} \mu^\pi(s)\pi(a|s)$$ Probability distribution

$$\mu^\pi(s') = \sum_{s,a} \mathcal{P}^a_{ss'}\mu^\pi(s)\pi(a|s)$$ Follow system dynamics

$$\epsilon \geq \sum_{s,a} \mu^\pi(s)\pi(a|s) \log \frac{\mu^\pi(s)\pi(a|s)}{q(s,a)}$$ Close to training data (no wild exploration)

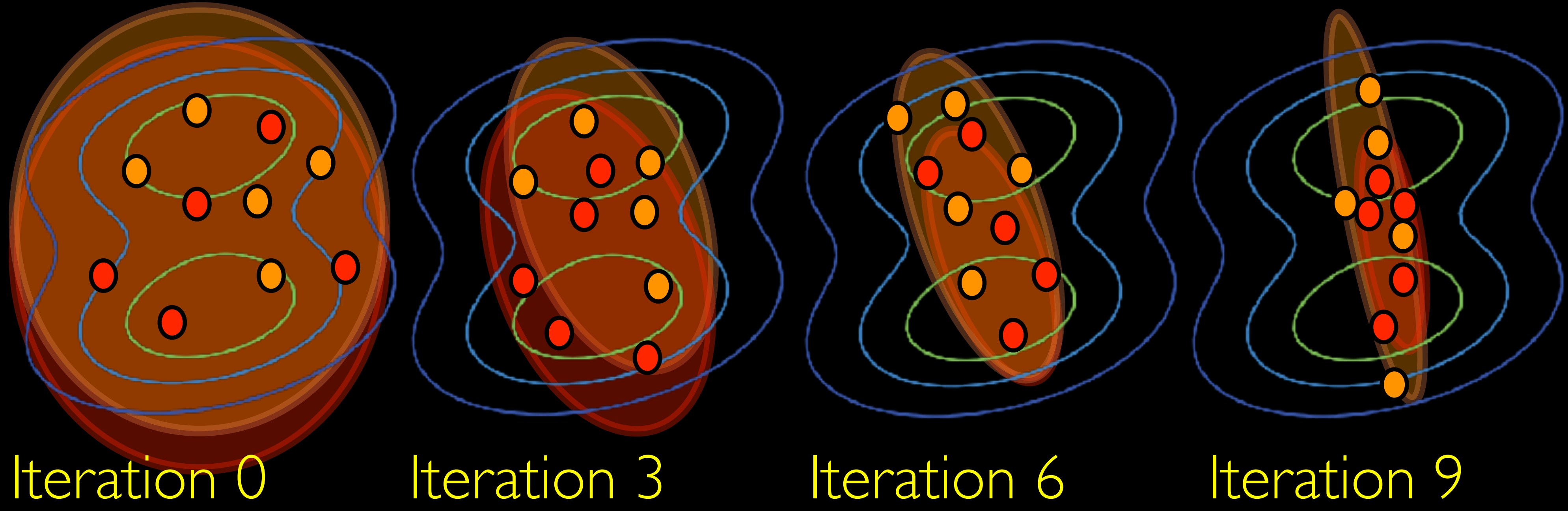
Mülling, K.; Kober, J.; Kroemer, O.; Peters, J. Learning to Select and Generalize Striking Movements in Robot Table Tennis, International Journal on Robotics Research.

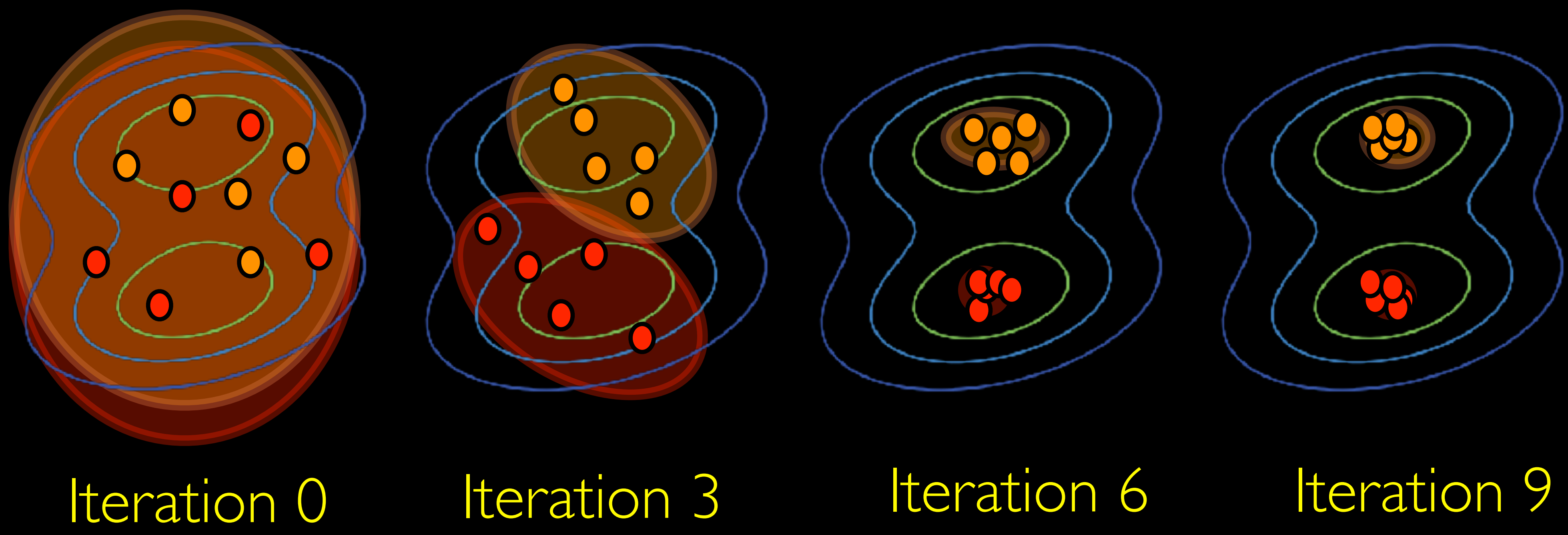
Daniel, Neumann & Peters. Hierarchical Relative Entropy Policy Search, JMLR


K. Muelling, Z. Wang, J. Peters of TU Darmstadt and MPI Intelligen

Problems with Naïvety

Christian Daniel





$$\kappa \geq \mathbb{E}_{s,a}\left[\sum_{o} -p(o|s,a)\log p(o|s,a)\right]$$

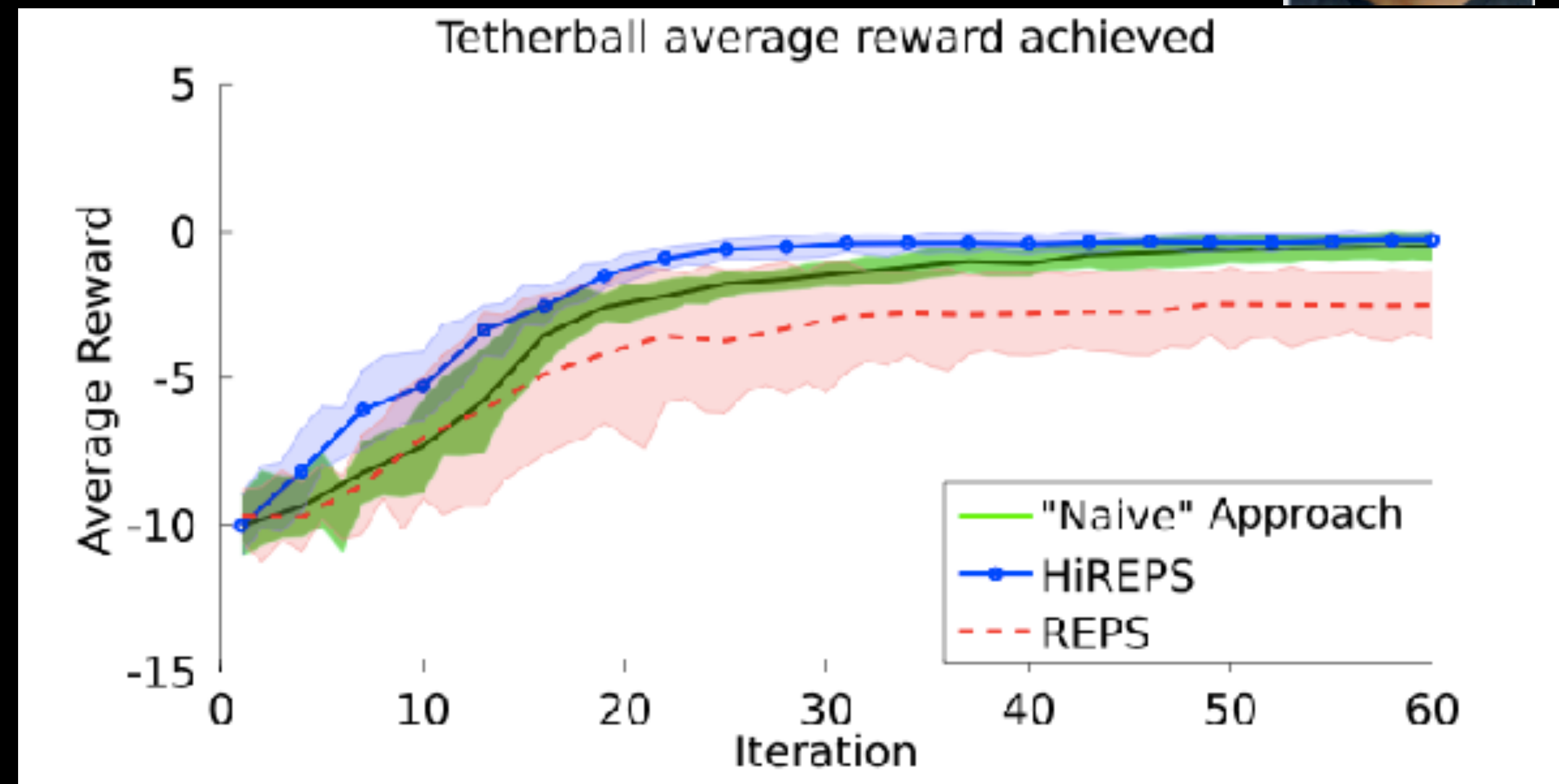
Force the primitives to limited responsibility

Daniel, Neumann & Peters. Hierarchical Relative Entropy Policy Search, JMLR

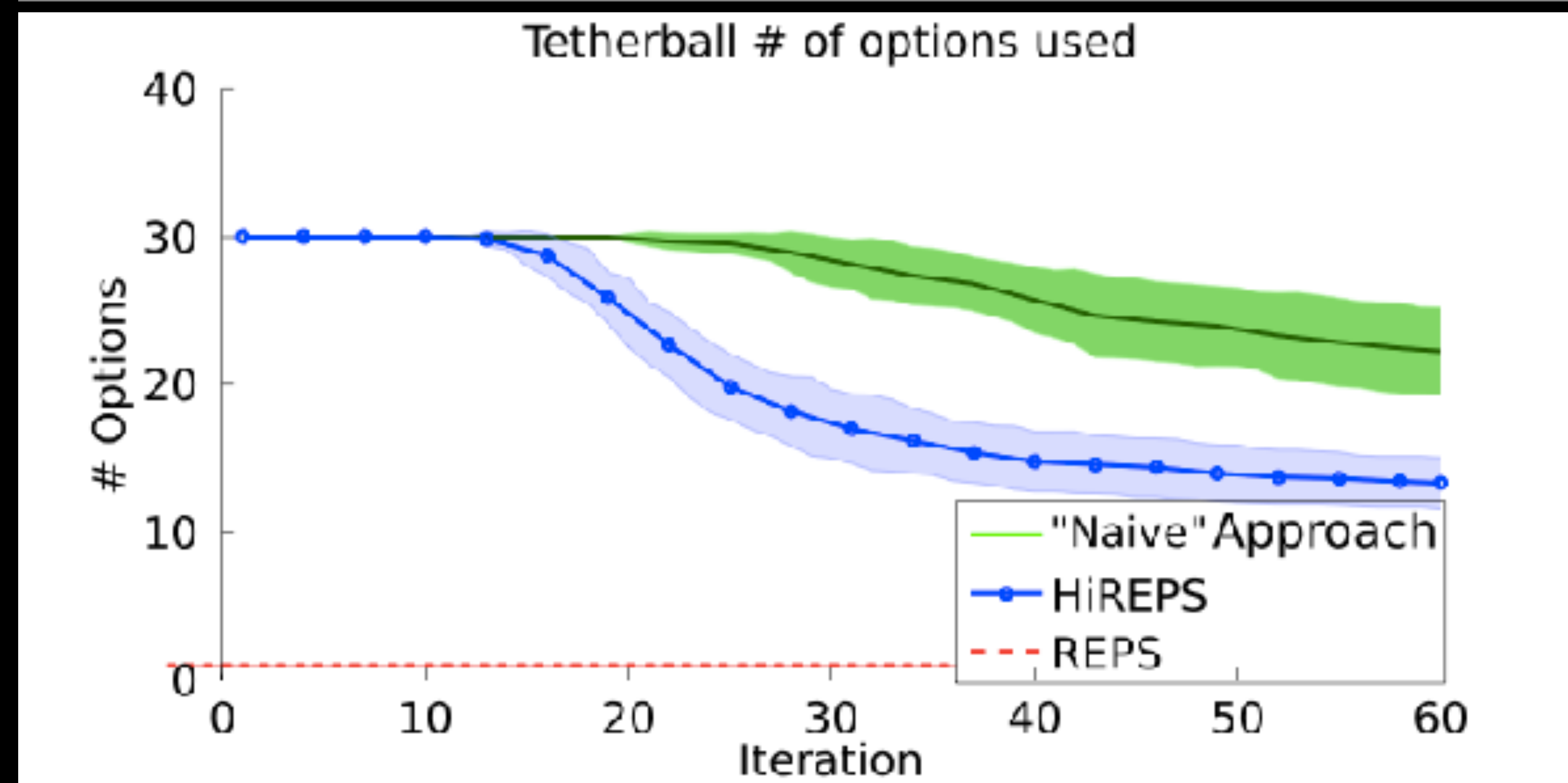
TECHNISCHE UNIVERSITÄT DARMSTADT

Localized behavior can be learned efficiently!

Good performance



Fast reduction in the number of primitives



Daniel, Neumann & Peters. Hierarchical Relative Entropy Policy Search, JMLR

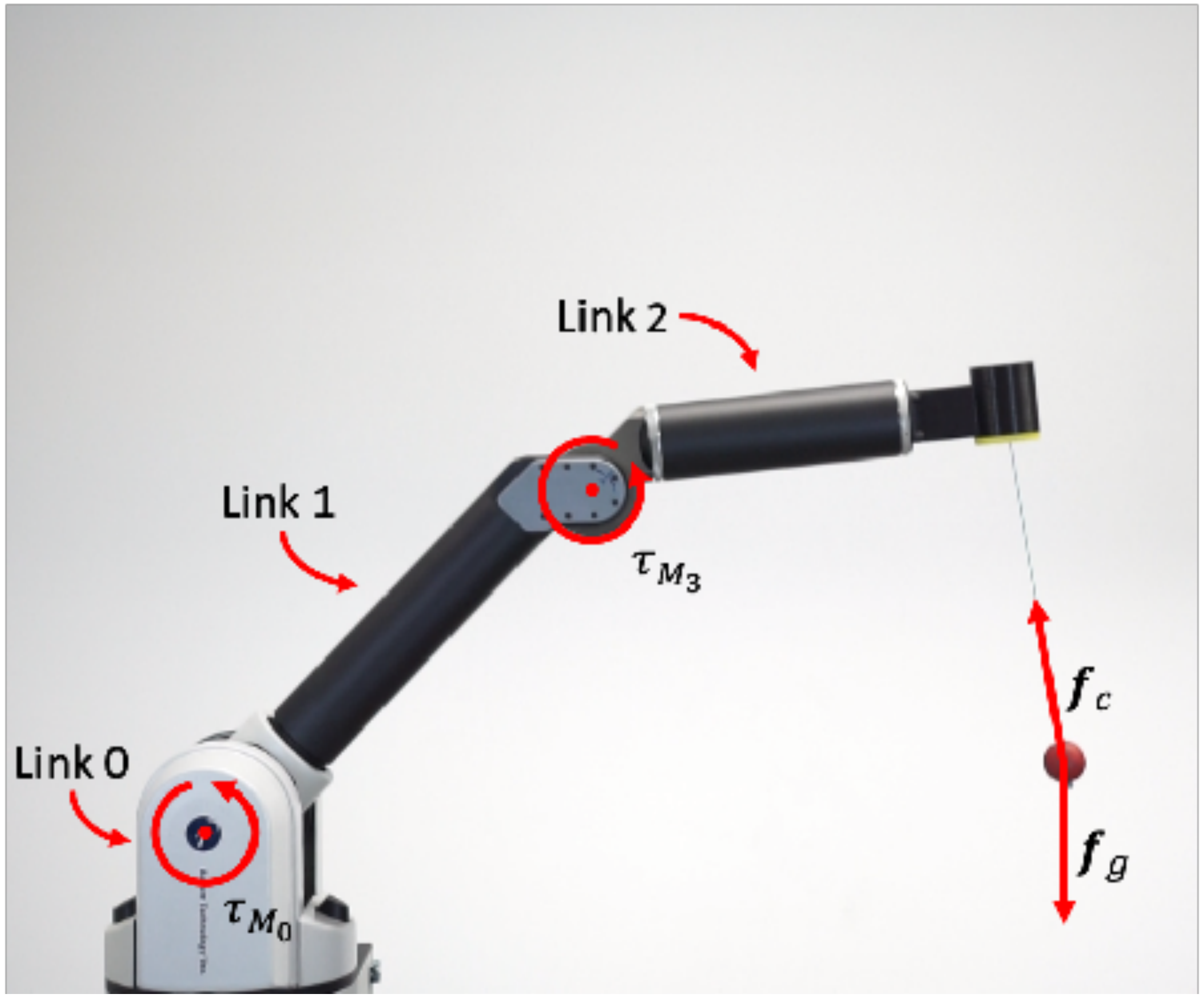
Outline

1. Inductive Bias: Stay close to your training data!
2. Inductive Bias: Use modular policy structure for composition!
3. Inductive Bias
4. Inductive Bias
5. Inductive Bias
6. Inductive Bias

Models can be very powerful, but errors in models are often exploited by RL algorithms…

Model-based RL with Differentiable Physics Models

Michael Lutter



Differentiable Newton Euler Algorithm:

$$\sum_{i=0}^{N} \boldsymbol{f}_i = \mathbf{I}_\theta \, \dot{\mathrm{V}} \quad \text{s.t.} \quad \mathrm{c}(\mathbf{q};\, \theta) = 0, \qquad c(\boldsymbol{x}; \theta) \leq 0$$

with the physics parameters θ consisting of inertia, mass, lengths, center of mass, string length, etc..

Learn Ball in a Cup via offline MBRL:

(1) Record dataset on the physical manipulator
(2) Learn model using the recorded data
(3) Learn trajectory with learned model & eREPS
(4) Evaluate learnt trajectory on the real WAM

Lutter, M.; Silberbauer, J.; Watson, J.; Peters, J. (Submitted). A Differentiable Newton-Euler Algorithm for Real-World Robotics, Submitted to the IEEE Transaction of Robotics (T-RO).

Performance based on 4min of Motor Babbling

Michael Lutter

Structured models enable out of distribution generalization.

- DiffNEA extrapolated to completely unseen states
- For MBRL generalization *might* be more important than perfect prediction

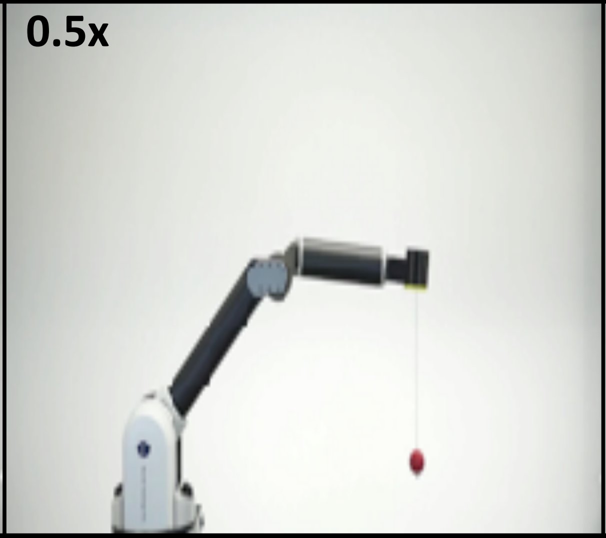
Nominal Model + eREPS

Transferability: 0%
Repeatability: 0%

DiffNEA Model + eREPS

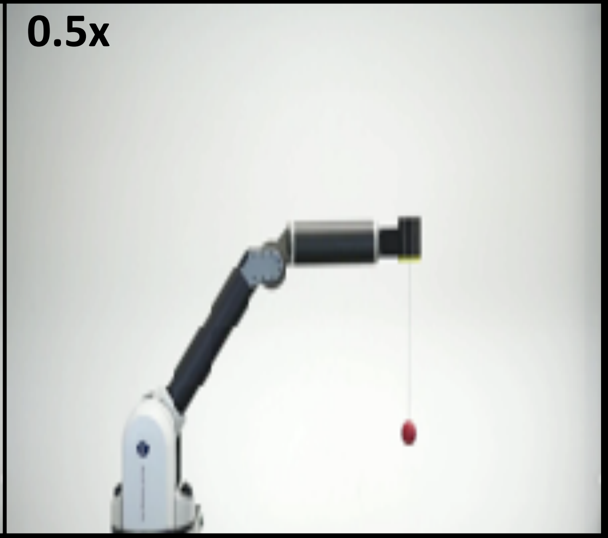
Transferability: 60%
Repeatability: 90%

FFNN Model + eREPS



Transferability: 0%
Repeatability: 0%

LSTM Model + eREPS



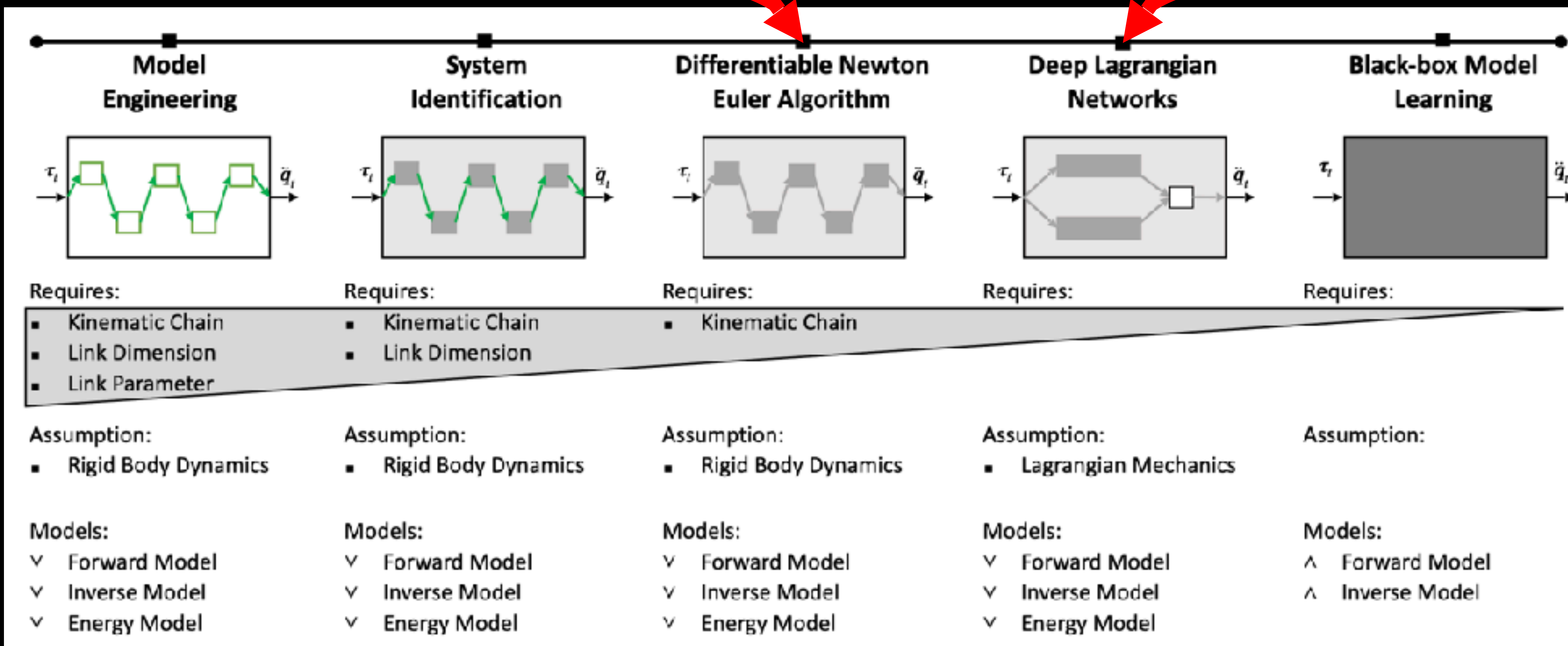
Transferability: 0%
Repeatability: 0%

Lutter, M.; Silberbauer, J.; Watson, J.; Peters, J. (Submitted). A Differentiable Newton-Euler Algorithm for Real-World Robotics, Submitted to the IEEE Transaction of Robotics (T-RO).

Lutter, M.; Silberbauer, J.; Watson, J.; Peters, J. (2021). Differentiable Physics Models for Real-world Offline Model-based Reinforcement Learning, ICRA.

Lutter, M.; Ritter, C.; Peters, J. (2019). Deep Lagrangian Networks: Using Physics as Model Prior for Deep Learning, International Conference on Learning Representations (ICLR).


Michael Lutter



3. Inductive Bias

Robotics
Inductive
Bias

Use physically consistent models!

TECHNISCHE UNIVERSITÄT DARMSTADT

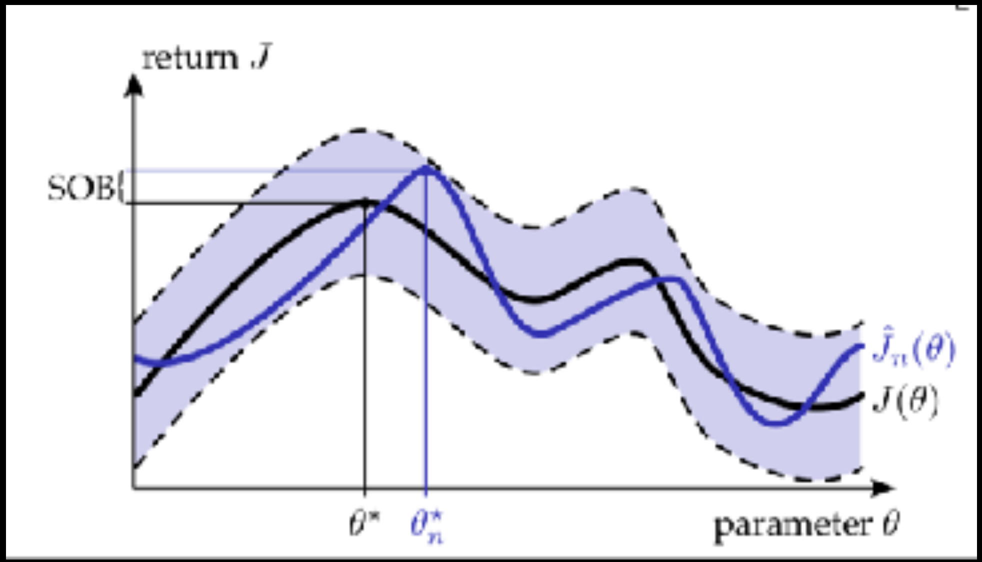
Simulation Optimization Bias (S.O.B.)

What's wrong about simulation?

$$G(\theta) = E_\xi \left\{ \max_{\hat\theta} \frac{1}{N} \sum_i^N J(\hat\theta, \xi) \right\} - \max_\theta E_\xi \left\{ \frac{1}{N} \sum_i^N J(\theta, \xi) \right\} \geq 0$$

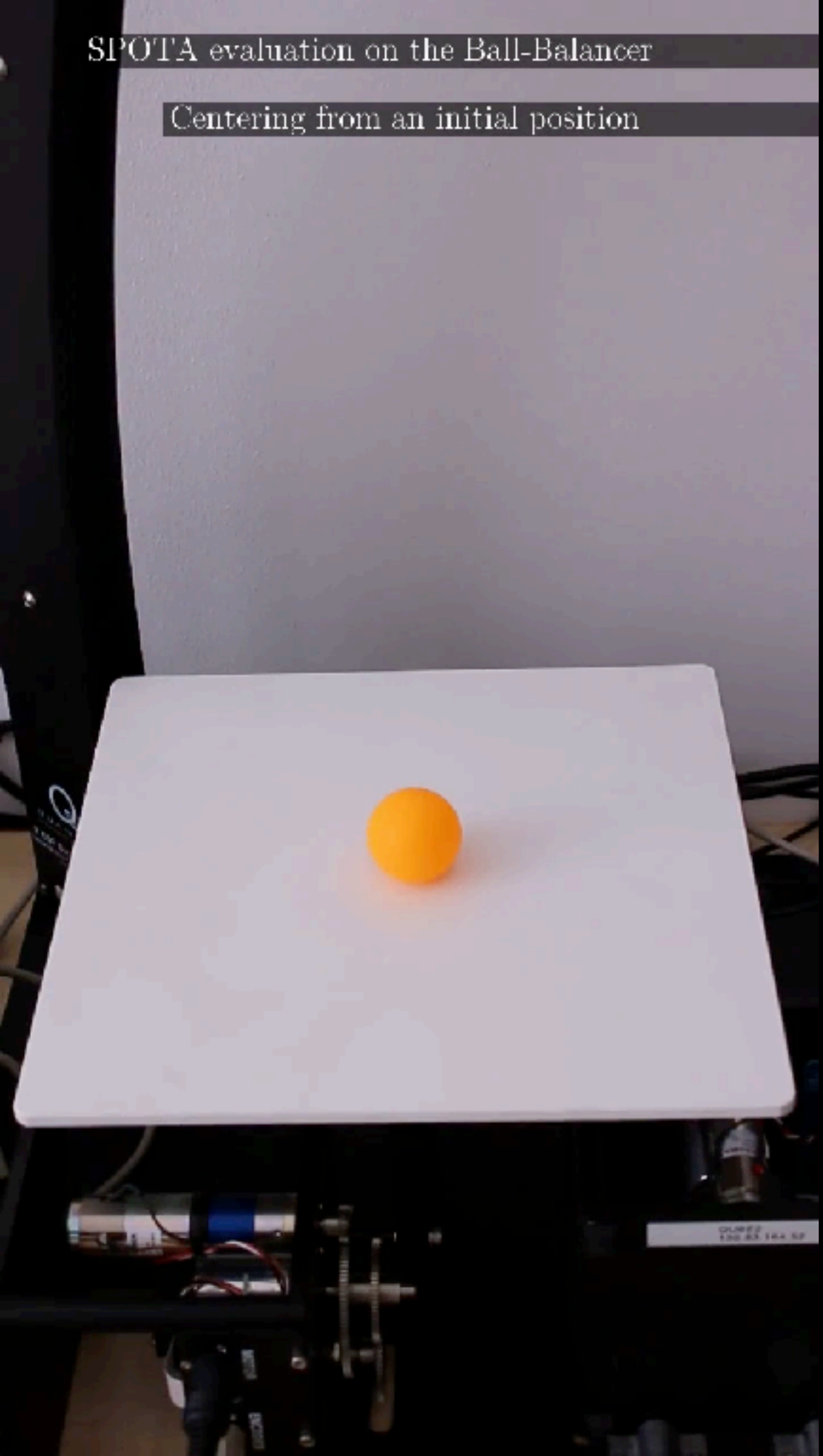
Optimal Solution for Samples

True Optimal Solution



We are guaranteed to be wrong!

Fabio Muratore



Fabio Muratore

SPOTA controls the S.O.B.

Muratore, F. et al. (2022). Assessing Transferability from Simulation to Reality for Reinforcement Learning, PAMI

4. Inductive Bias

Robotics
Inductive
Bias

Control your
optimization bias

Outline

1. Inductive Bias: Stay close to your training data!
2. Inductive Bias: Use modular policy structure for composition!
3. Inductive Bias: Use physically consistent models!
4. Inductive Bias: Control your optimization biases!
5. Inductive Bias: Use your constraints to direct your exploration!
6. Inductive Bias

The fastest way to destroy a robot system is by exploration…

Safe Exploration


Davide Tateo


Puze Liu

$$\max_\theta \ \mathbb{E}_{s_t, a_t} \left[\sum_{t=0}^{T} \gamma^t r(s_t, a_t) \right],$$

$$\text{s. t.} \quad f(q_t) = 0, \quad g(q_t) \leq 0$$

$$s_t = [q_t \ x_t]^{\mathrm{T}}$$

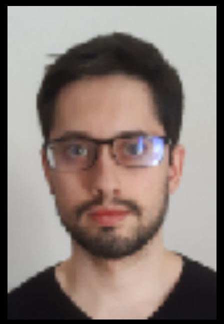
Robotics problems hide their difficulties in the constraints!

4. Inductive Bias

Robotics
Inductive
Bias

Use your constraints to direct your exploration!

Exploration on the Constraint Manifold



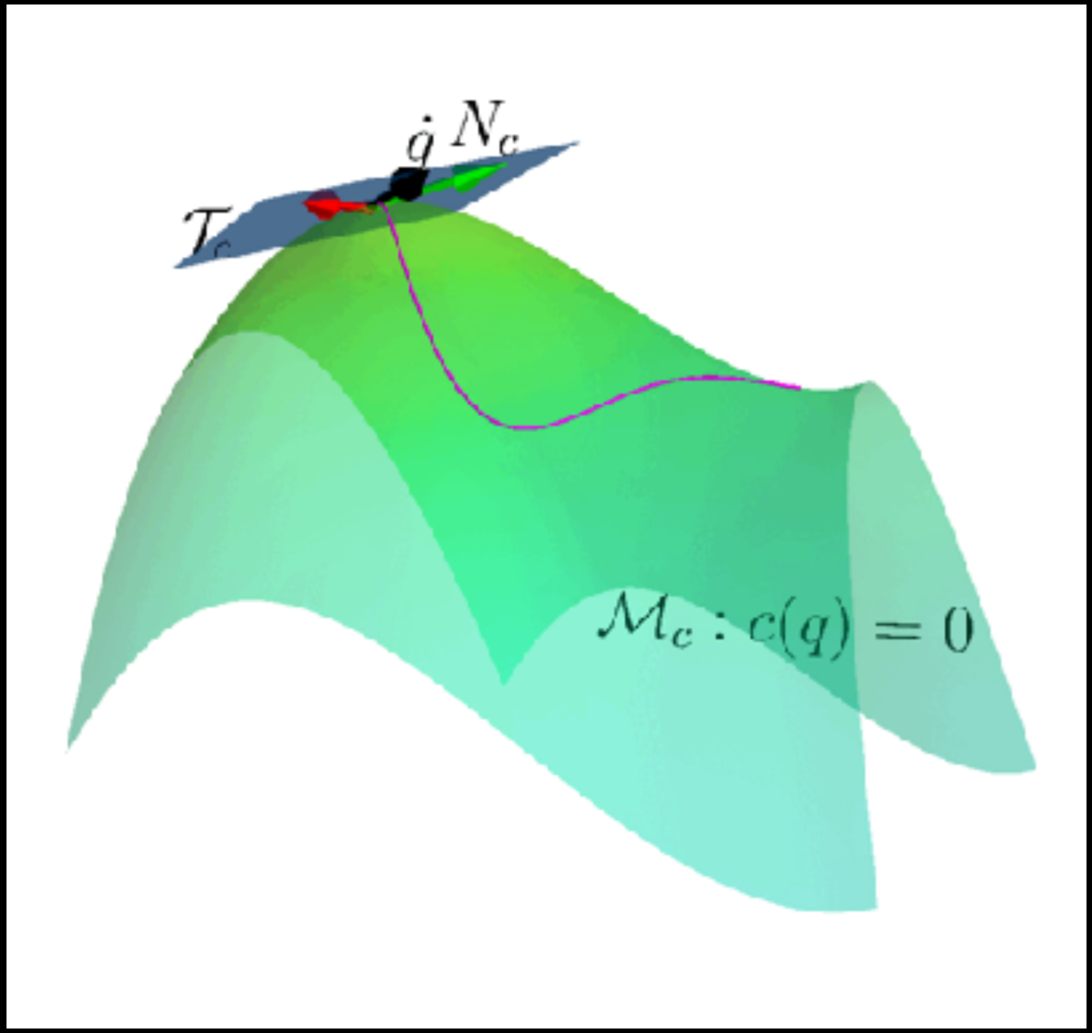
Davide Tateo

Puze Liu

1. Construct the constraint manifold

$$\mathcal{M}_c : c(q) = 0$$

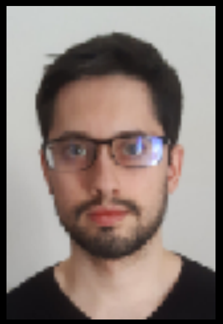
2. Determine the bases N_c of the tangent space $\mathcal{T}_c$
3. Sample state velocity in the tangent space



$$\begin{bmatrix} \ddot{q}_t \\ \dot{\mu}_t \end{bmatrix} = \underbrace{N_c(q_t, \mu_t)\alpha_t}_{\text{Tangent Space Action}} \overbrace{\underbrace{-J_c^{\dagger}(q_t, \mu_t)\psi(q_t, \dot{q}_t)}_{\text{Curvature Correction}} \underbrace{-J_c^{\dagger}(q_t, \mu_t)K_c c(q_t, \dot{q}_t, \mu_t)}_{\text{Error Correction}}}^{\text{Manifold Maintenance}}$$

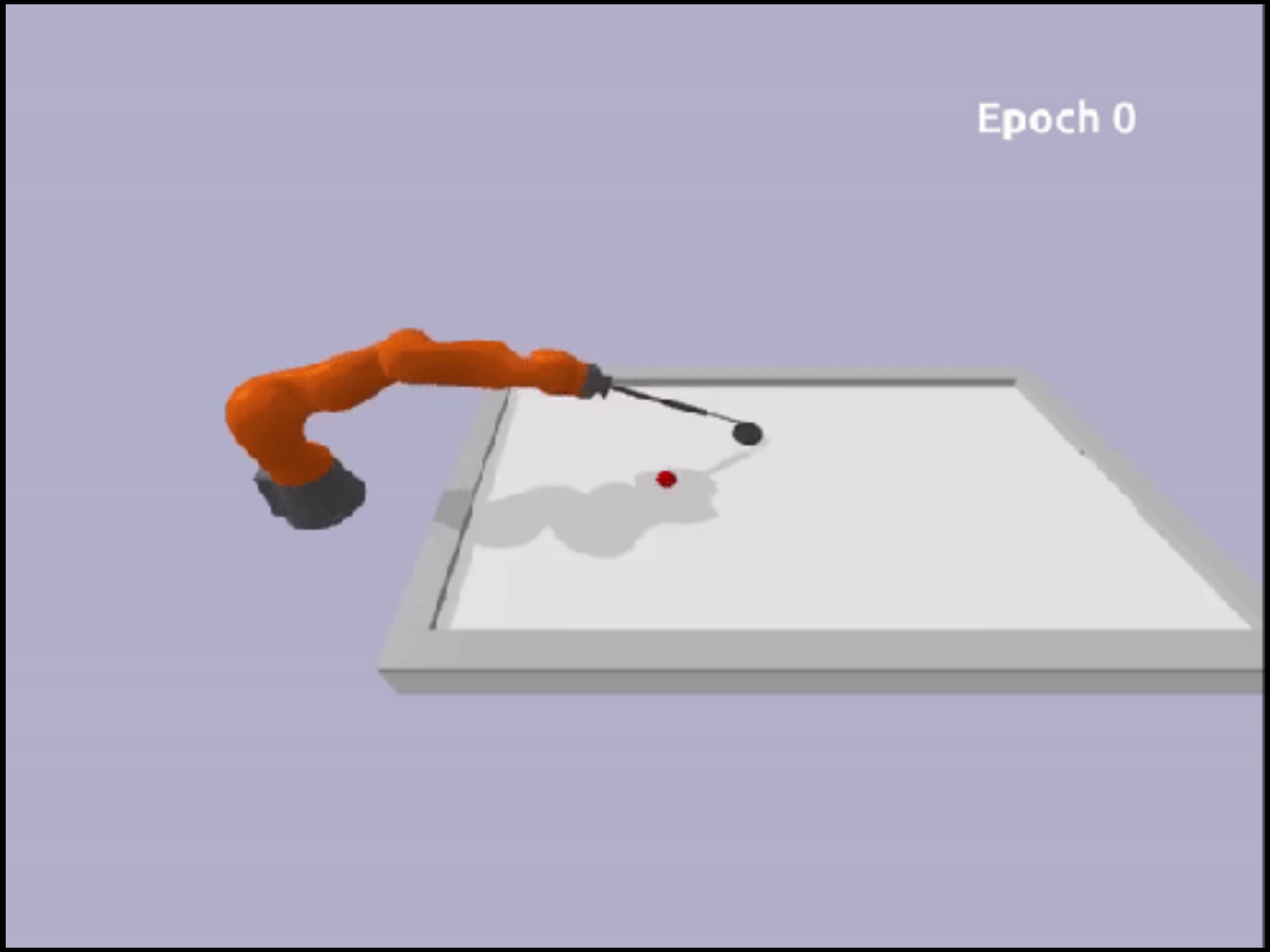
Learning Process

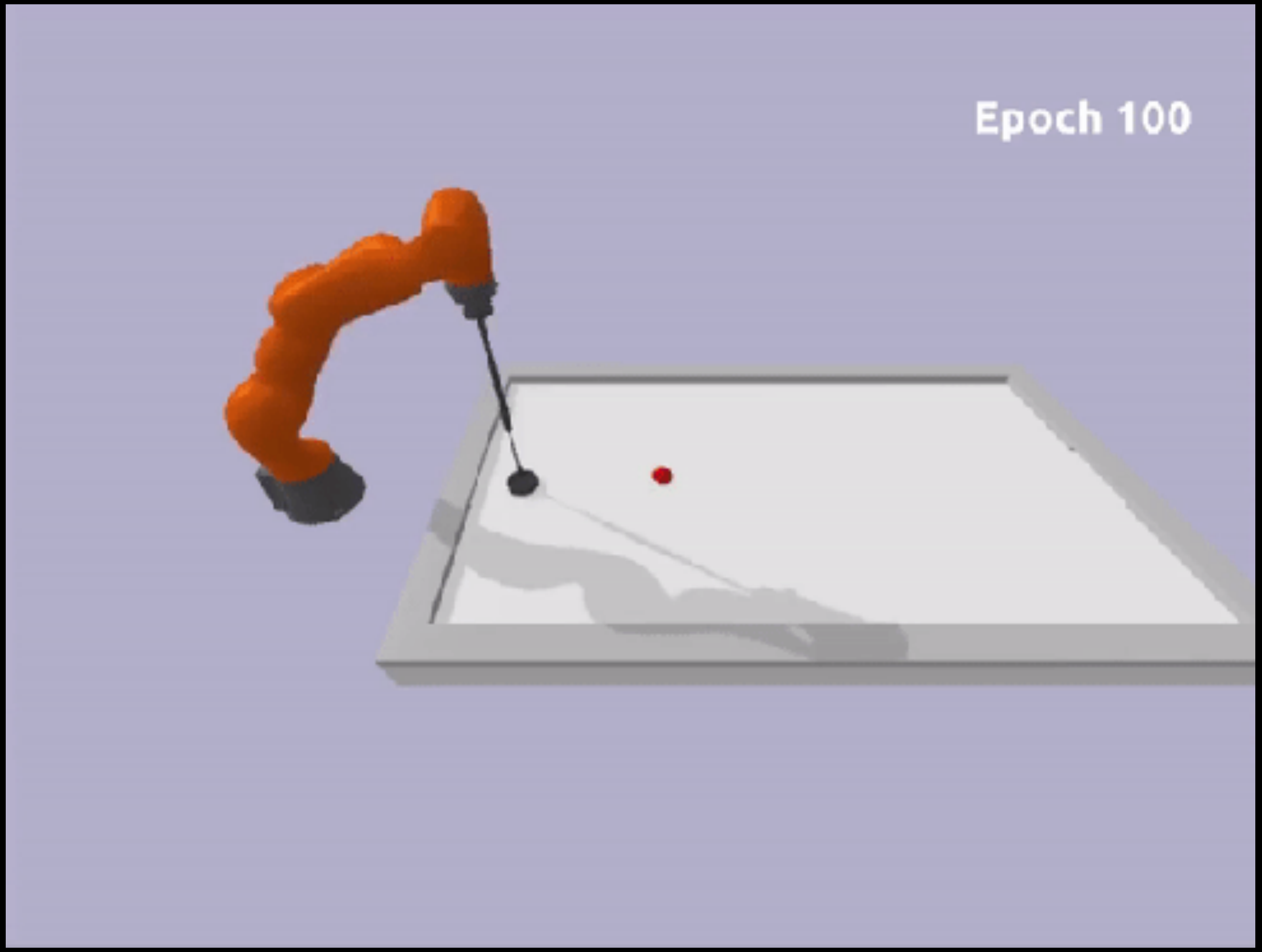
Piotr
Kicki

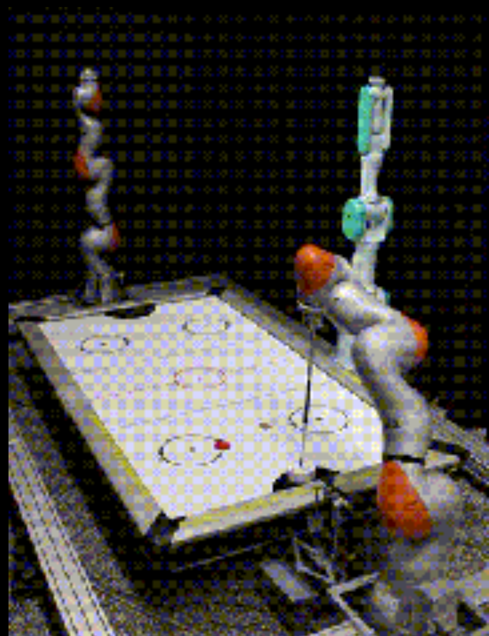


Davide
Tateo

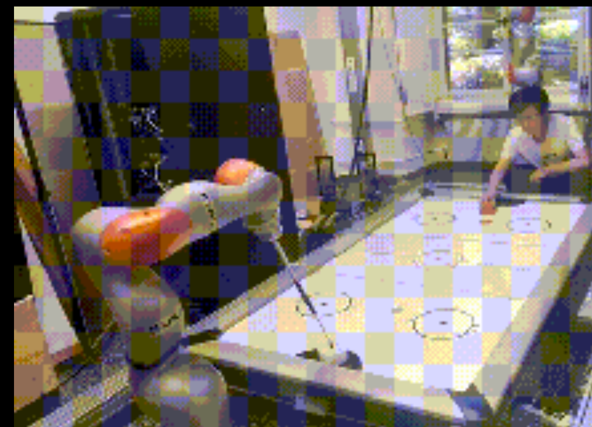
Puze
Liu



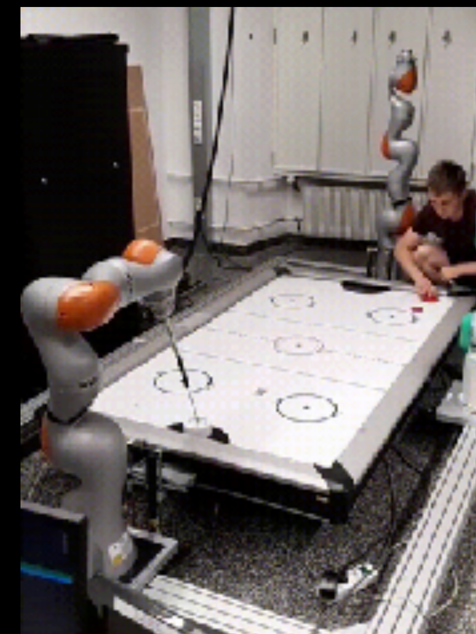




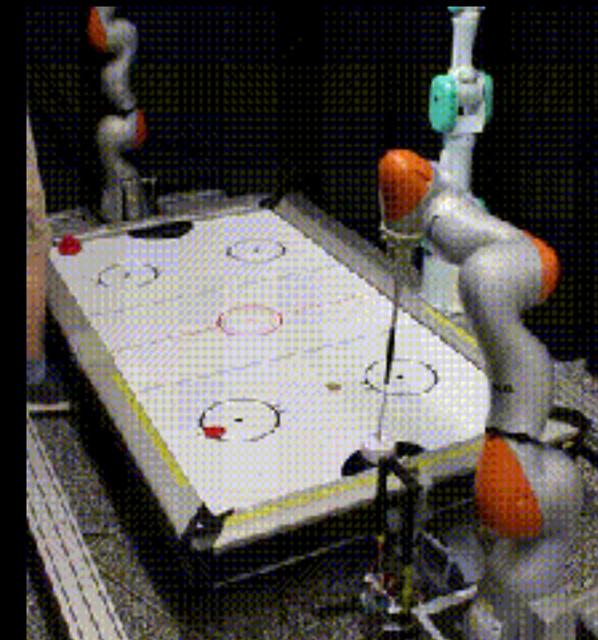
Smash



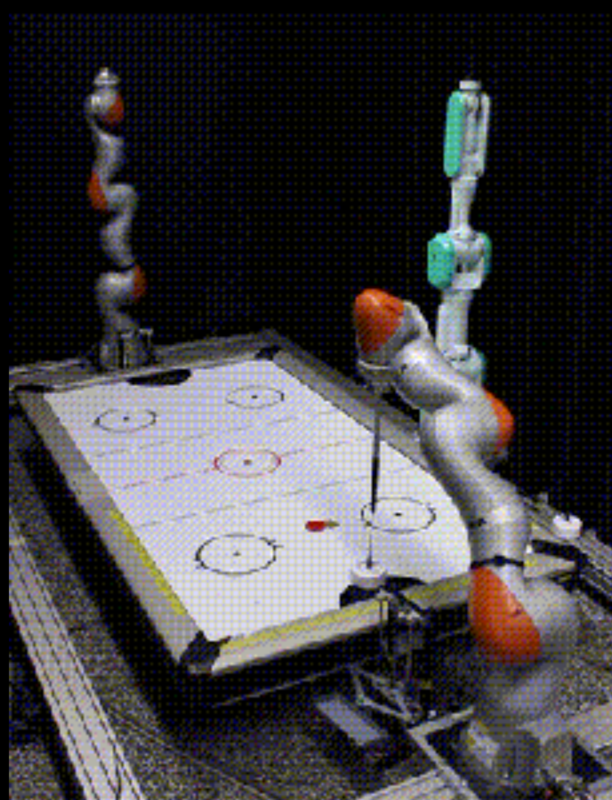
Cut



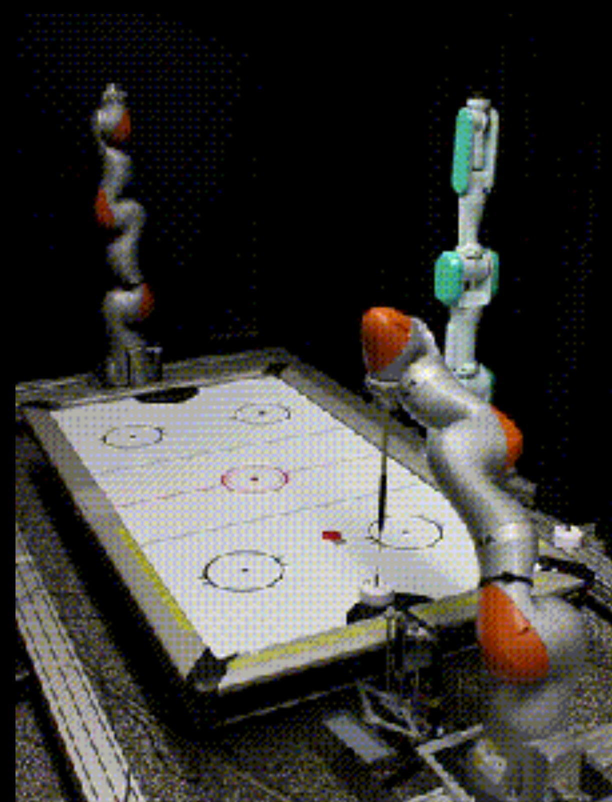
Repel



Prepare



Lissajous Hitting



Dynamic Hitting

Piotr Kicki

Haitham Bou Ammar

Davide Tateo

Puze Liu

Liu, P.; Tateo, D.; Bou-Ammar, H.; Peters, J. (2021). Robot Reinforcement Learning on the Constraint Manifold, Proceedings of the Conference on Robot Learning (CoRL).
Kicki, P.; Liu, P.; Tateo, D.; Bou Ammar, H.; Walas, K.; Skrzypczynski, P.; Peters, J. (2024). Fast Kinodynamic Planning on the Constraint Manifold with Deep Neural Networks, IEEE Trans. on Robotics

Outline

1. Inductive Bias: Stay close to your training data!
2. Inductive Bias: Use modular policy structure for composition!
3. Inductive Bias: Use physically consistent models!
4. Inductive Bias: Control your optimization biases!
5. Inductive Bias: Use your constraints to direct your exploration!
6. Inductive Bias

Let your body go with the flow…

Robot Bodies for Learning?

Dieter Büchler

Classical robotics builds the best body that can be controlled with classical approaches!

Human bodies would defy such an approach but generate high accelerations in order to

- reach high velocities
- perform skillful motions

Humans learn (typically) without breaking!

➡ Human performance robot learning needs better bodies!

Bodies for Learning

Dieter Büchler

High accelerations require

- strong actuators (pneumatic artificial muscles; 1,2kN)
- small moving masses (700g)

Antagonistic actuation

- prevents damages to the robot
- enables compliance

➡ Built for performance and learning *not* feedback control!

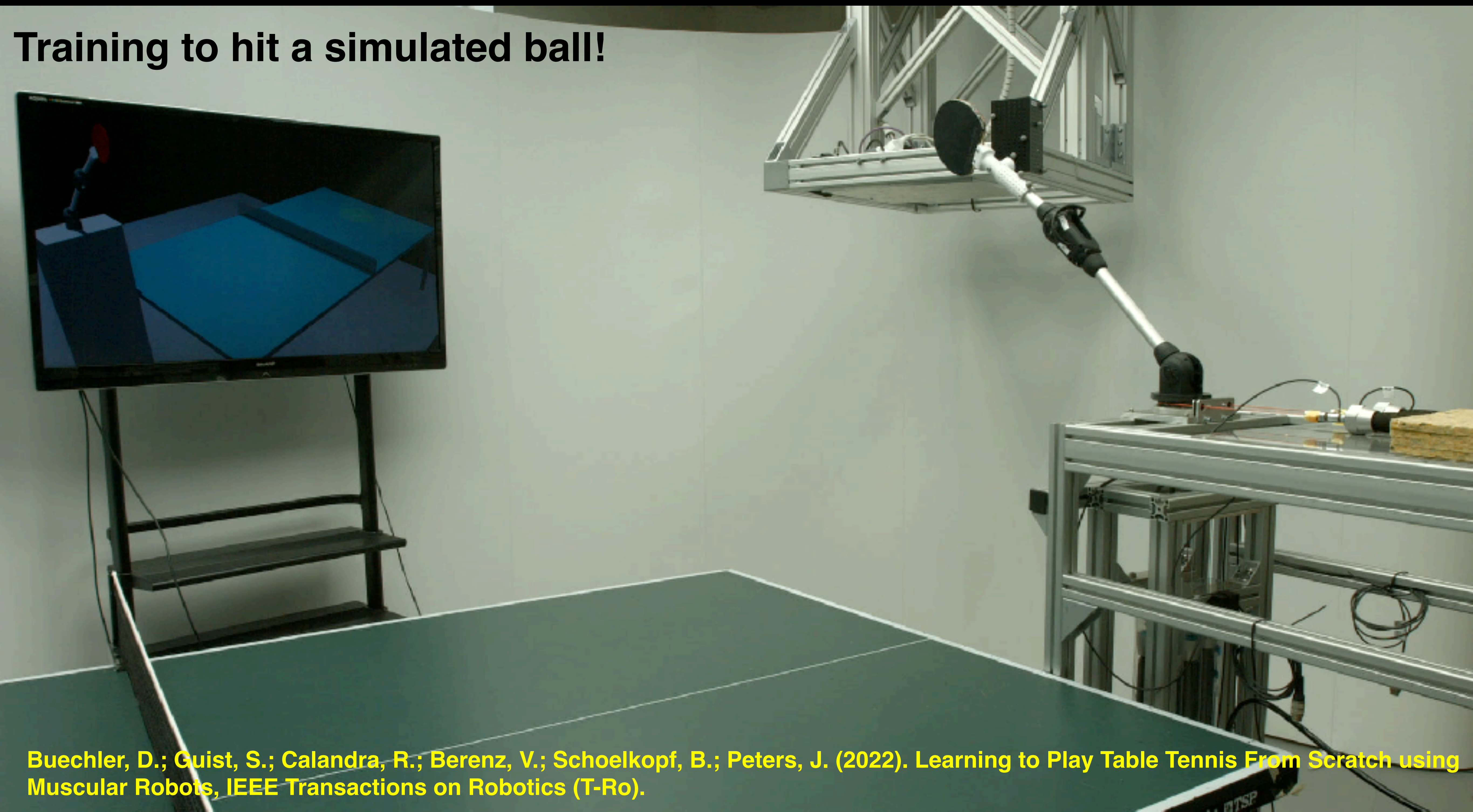
Buechler, D.; Calandra, R.; Peters, J. (2022). Learning to Control Highly Accelerated Ballistic Movements on Muscular Robots, Robotics and Autonomous Systems.

Learning Robot Table Tennis *from Scratch*

Dieter Büchler

Simon Guist

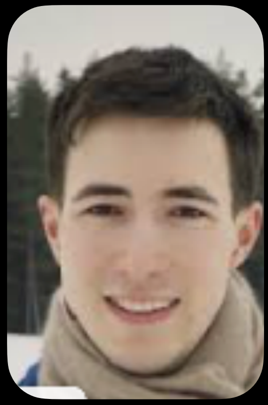
Training to hit a simulated ball!



Buechler, D.; Guist, S.; Calandra, R.; Berenz, V.; Schoelkopf, B.; Peters, J. (2022). Learning to Play Table Tennis From Scratch using Muscular Robots, IEEE Transactions on Robotics (T-Ro).

Learning Robot Table Tennis *from Scratch*


Dieter Büchler


Simon Guist

Training to hit a real ball!

Buechler, D.; Guist, S.; Calandra, R.; Berenz, V.; Schoelkopf, B.; Peters, J. (2022). Learning to Play Table Tennis From Scratch using Muscular Robots, IEEE Transactions on Robotics (T-Ro).

2. Inductive Bias

Robotics Inductive Bias

Let the natural robot dynamics guide your learning process!

Conclusion:
Use your inductive biases!

1. Inductive Bias: Stay close to your training data!
2. Inductive Bias: Use modular policy structure for composition!
3. Inductive Bias: Use physically consistent models!
4. Inductive Bias: Control your optimization biases!
5. Inductive Bias: Use your constraints to direct your exploration!
6. Inductive Bias: Let the natural robot dynamics guide your learning process!

Thanks for your attention!

TECHNISCHE UNIVERSITÄT DARMSTADT

Learning Robots
with inductive biases


Go2
Thanks for your attention!

Outline

1. Inductive Bias: Stay close to your training data!
2. Inductive Bias: Use modular policy structure for composition!
3. Inductive Bias: Use physically consistent models!
4. Inductive Bias: Control your optimization biases!
5. Inductive Bias: Use your constraints to direct your exploration!
6. Inductive Bias

Even if we had a perfect task model …

… we still need to find a solution to this task with limited resources

Pascal Klink

Pascal Klink

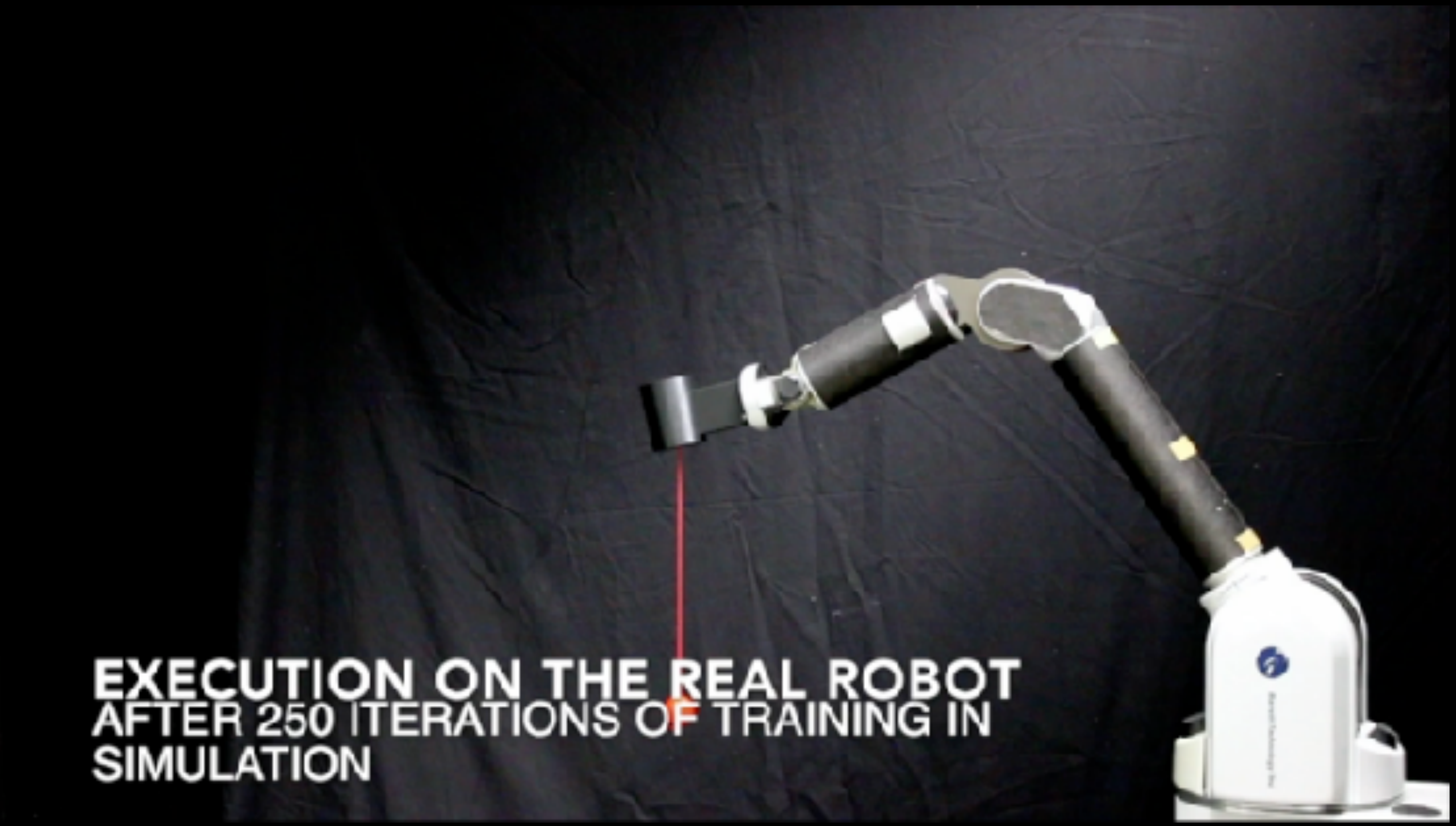
Curriculum Reinforcement Learning

Task Complexity

Klink, P. et al. (2021) A Probabilistic Interpretation of Self-Paced Learning with Applications to RL, JMLR
Klink, P. et al. (2022). Curriculum Reinforcement Learning via Constrained Optimal Transport, ICML

Curriculum Reinforcement Learning

Pascal Klink



Klink, P. et al. (2021) A Probabilistic Interpretation of Self-Paced Learning with Applications to RL, JMLR
Klink, P. et al. (2022). Curriculum Reinforcement Learning via Constrained Optimal Transport, ICML

N. Inductive Bias

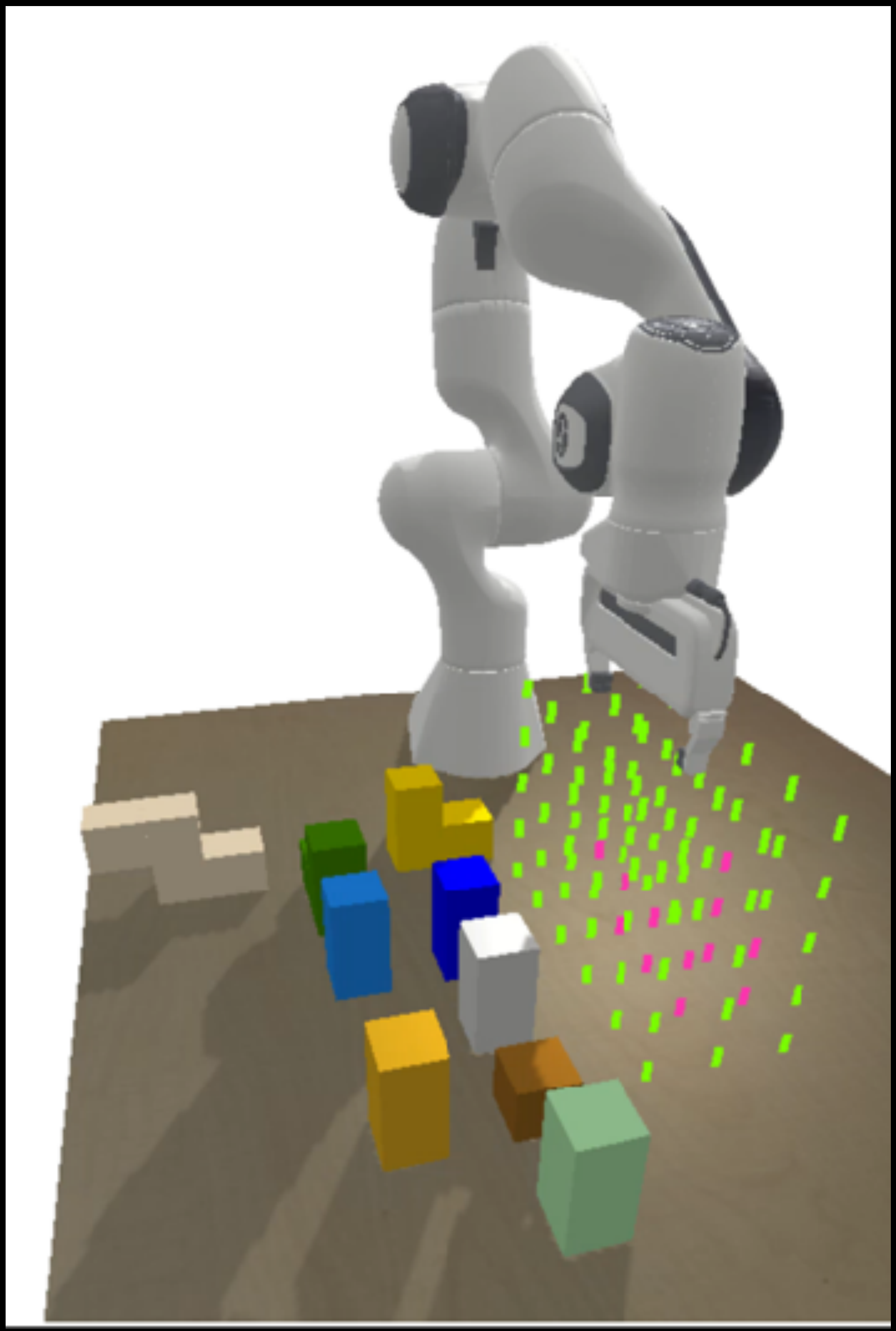
(Robo
tics)
Inducti
ve
Bias

Control your optimization complexity

Long horizon manipulation is challenging ...

Robot Assembly Discovery (RAD)

Niklas Funk



- Having to build arbitrary 3D structures given a set of building blocks
- Robot-in-the-loop
- Ensuring structural stability

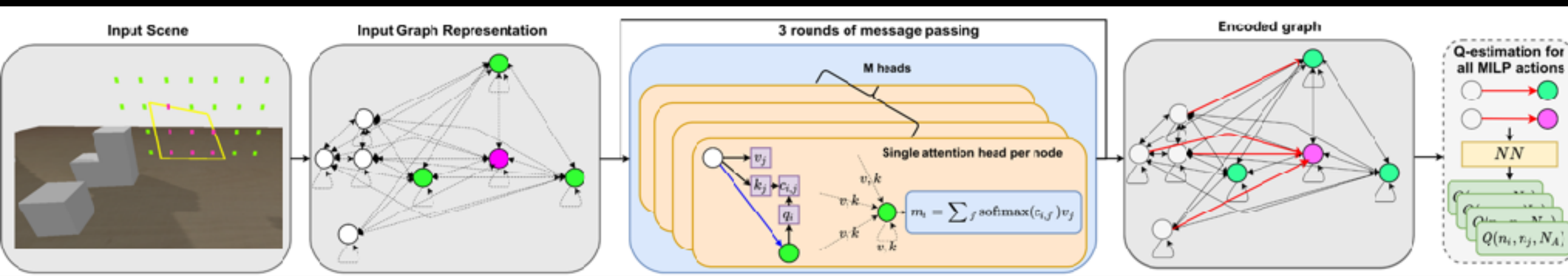
➡ 2-fold Combinatorial Complexity:

- Which parts to place where
- Execution sequence

Funk, N.; Belousov, B.; Chalvatzaki, G.; Peters, J. (2021). Learn2assemble with structured representations and search for robotic architectural construction.

Graph-based RL for RAD

Niklas Funk



Exploit graph-based representation

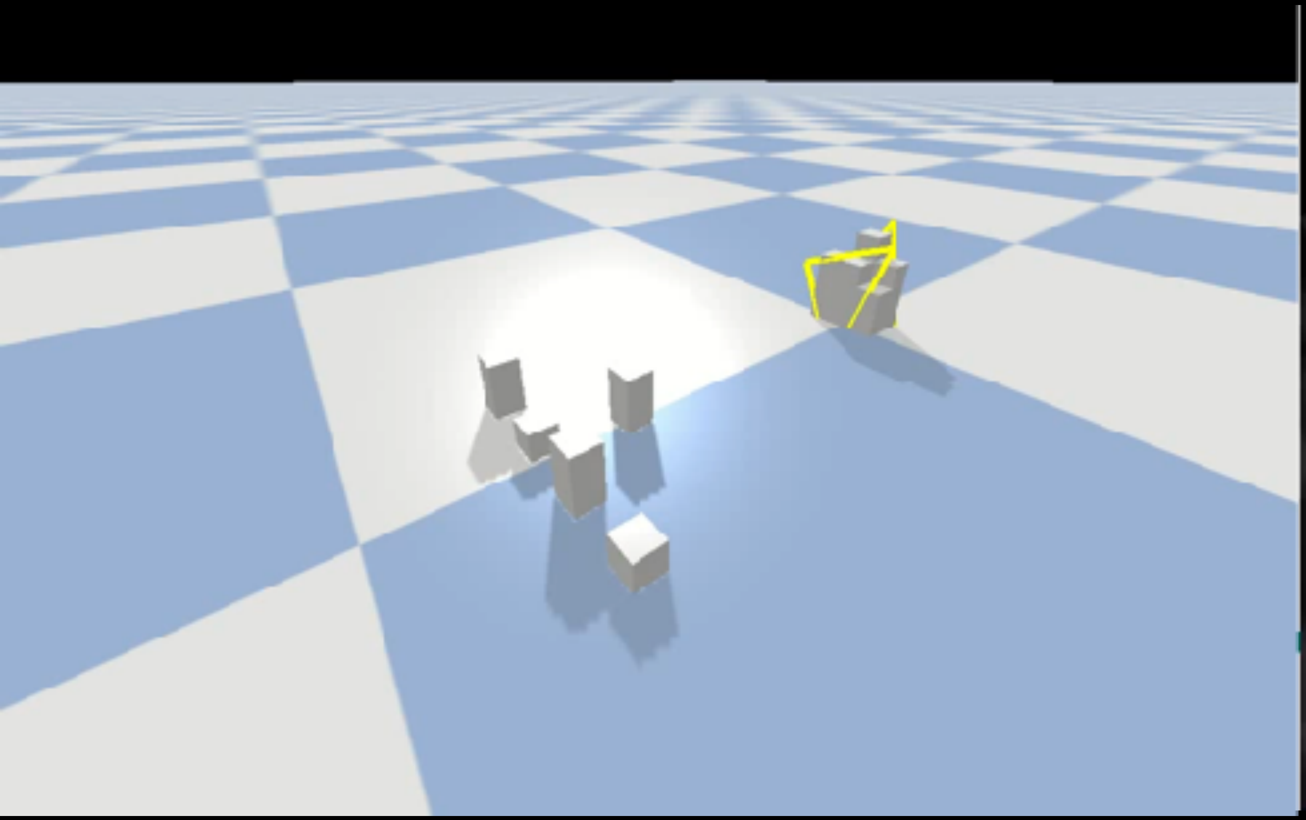
Eventually add prior knowledge from Mixed Integer Optimization

Funk, N.; Menzenbach, S.; Chalvatzaki, G.; Peters, J. (2022). Graph-based Reinforcement Learning meets Mixed Integer Programs: An application to 3D robot assembly discovery.

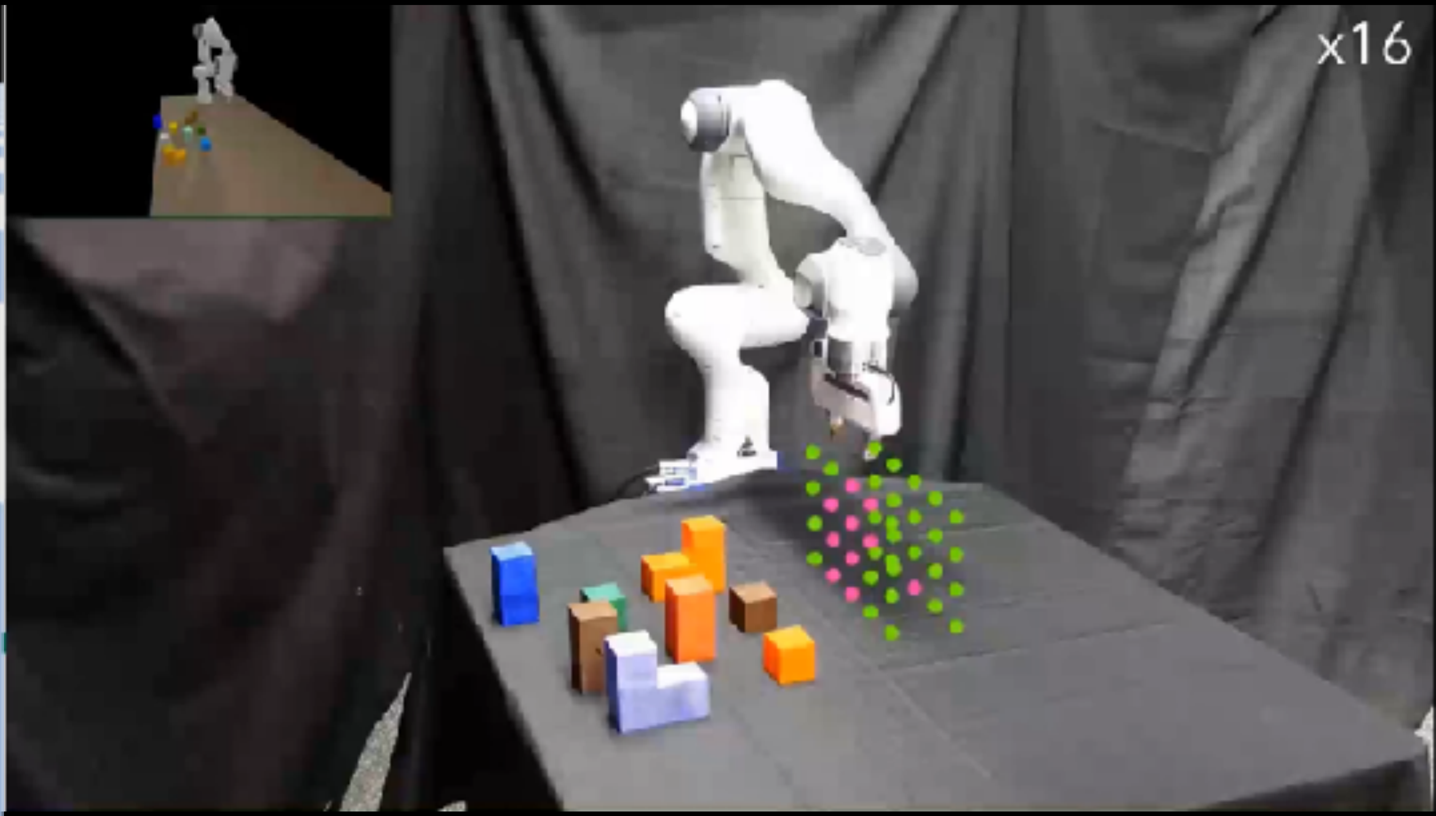
Funk, N.; Belousov, B.; Chalvatzaki, G.; Peters, J. (2021). Learn2assemble with structured representations and search for robotic architectural construction.

Resulting Assembly Policies

Niklas Funk



Simulation



Real-world transfer

➡ Graph-based representations allow generalization across scenes

Funk, N.; Belousov, B.; Chalvatzaki, G.; Peters, J. (2021). Learn2assemble with structured representations and search for robotic architectural construction.

2. Inductive Bias

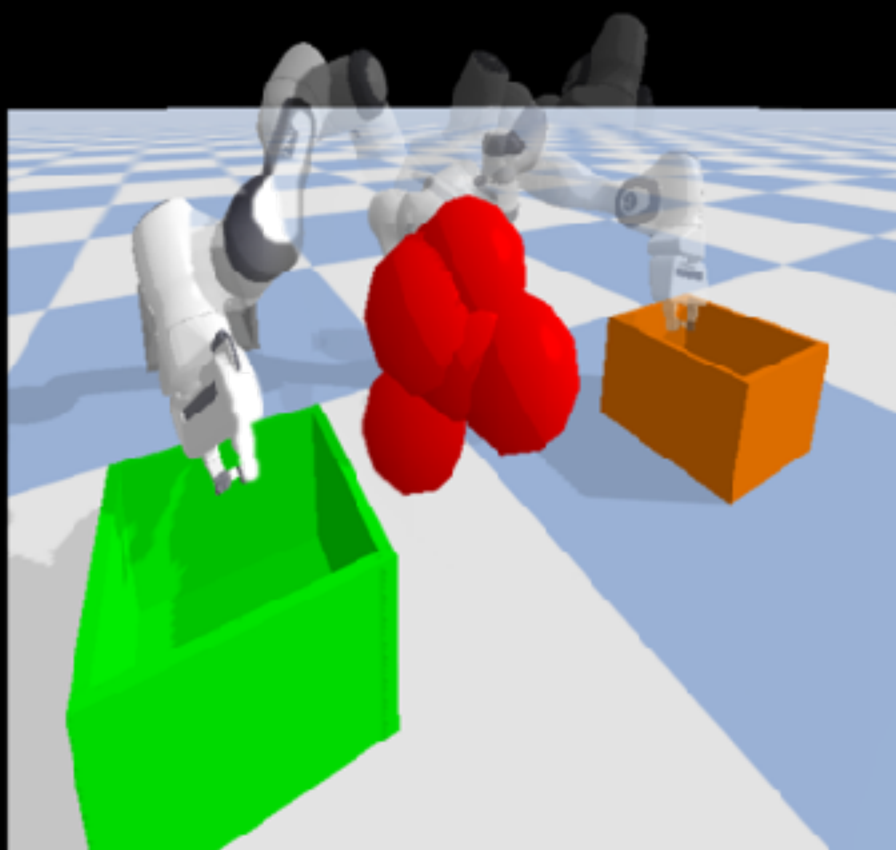
Robotics Inductive Bias

Use structured representations to enable generalization!

What about
dynamic and dense
environment?

Hierarchical Policy Blending



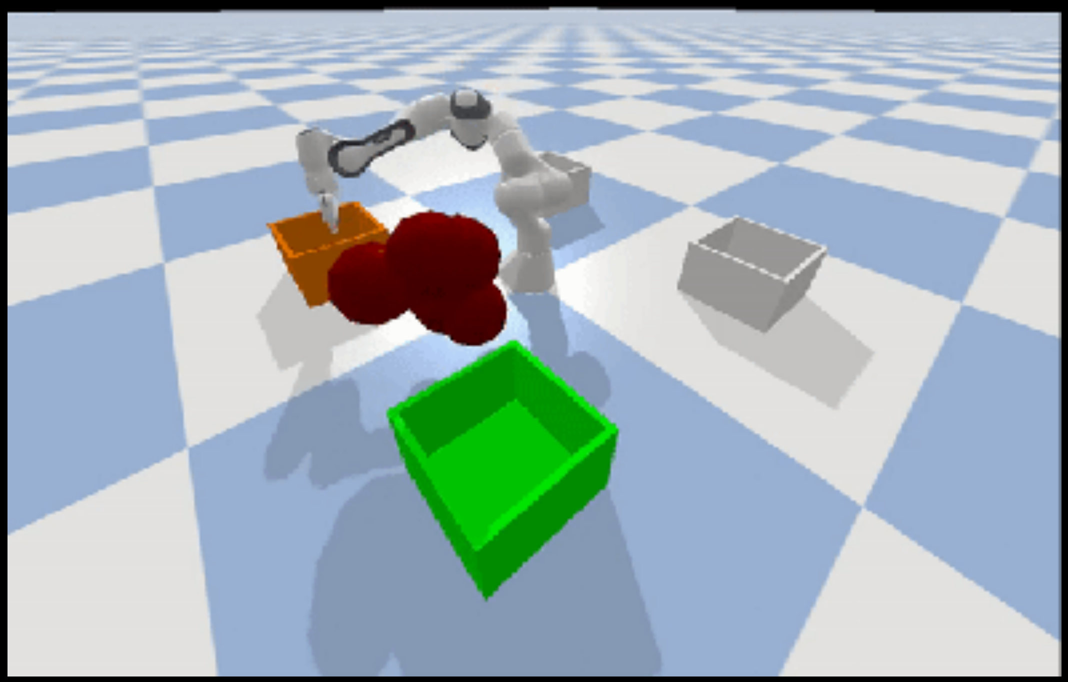


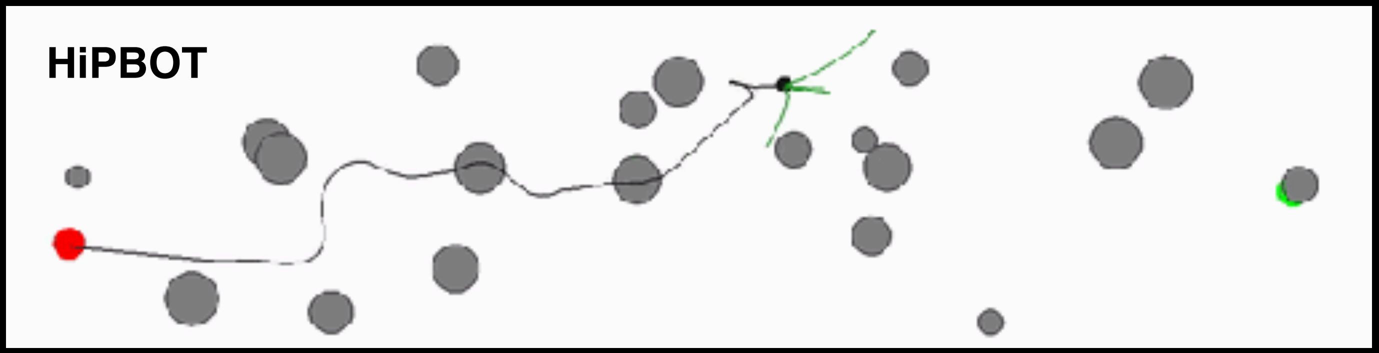
- Motion generation in dense and dynamic environments;
- Reactive policies:
  - Ensure fast response;
  - Risk of suboptimal behavior;
- Planning-based approaches:
  - Provide feasible trajectories;
  - High computational cost;
- Trade-off:
  - Safety vs Performance

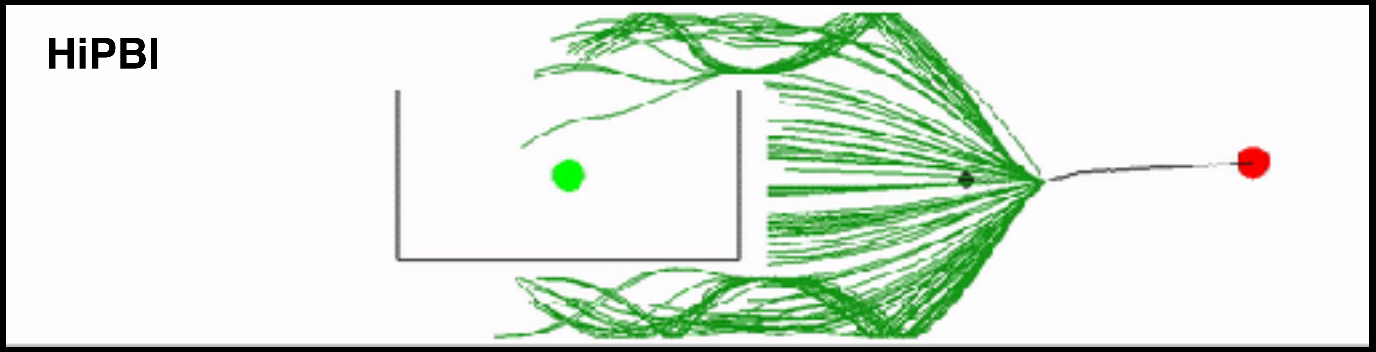
Hansel, K.; Urain, J.; Peters, J.; Chalvatzaki, G. (2022). Hierarchical Policy Blending as Inference for Reactive Robot Control.

Le, A. T.; Hansel, K.; Peters, J.; Chalvatzaki, G. (2022). Hierarchical Policy Blending As Optimal Transport

Hierarchical Policy Blending







➡ Adaptive Policy Blending enables safe and feasible robot motions across dense and dynamic environments.

HiPBI:

- Adopts Probabilistic Inference methods;
- Samping-Based Stochastic Optimization;

Hansel, K.; Urain, J.; Peters, J.; Chalvatzaki, G. (2022). Hierarchical Policy Blending as Inference for Reactive Robot Control.

HiPBOT:

- Leverages unbalanced optimal transport;
- Entropic-Regularized Linear Programming;

Le, A. T.; Hansel, K.; Peters, J.; Chalvatzaki, G. (2022). Hierarchical Policy Blending As Optimal Transport

IAS

TECHNISCHE UNIVERSITÄT DARMSTADT

2. Inductive Bias

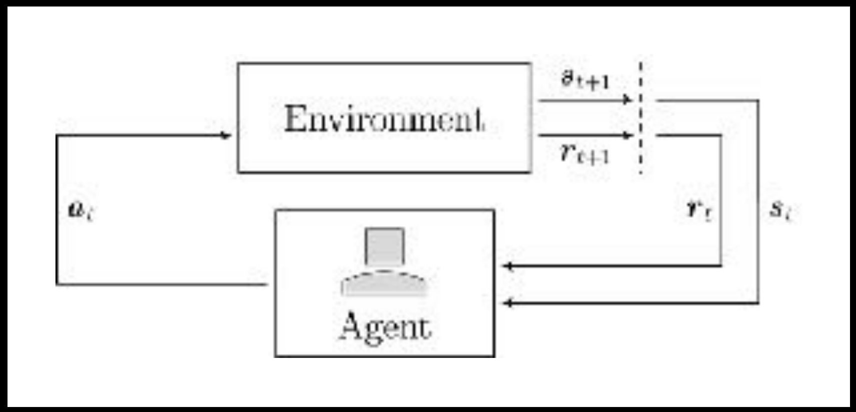
Robotics
Inductive
Bias

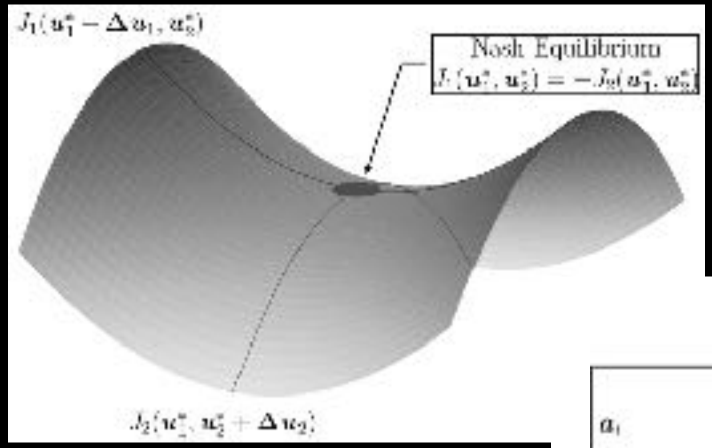
Inductive biases as hierarchical optimization for adaptive blending experts!

Learning to deal with environmental variations or uncertainties?

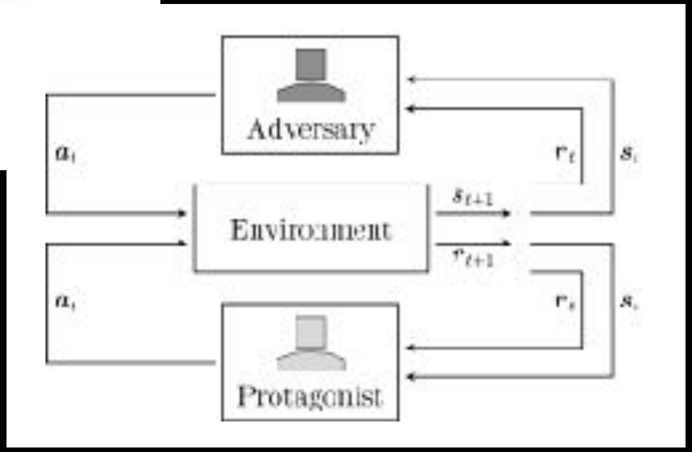
Robust Reinforcement Learning

Nominal MDP





The Nash Equilibrium



Adversarial MDP

- Reinforcement Learning:
  - Assumes underlying MDP
  - Struggles with:
    - Uncertainties;
    - Disturbances;
    - or structural changes in env
  - How to achieve "robustness"?

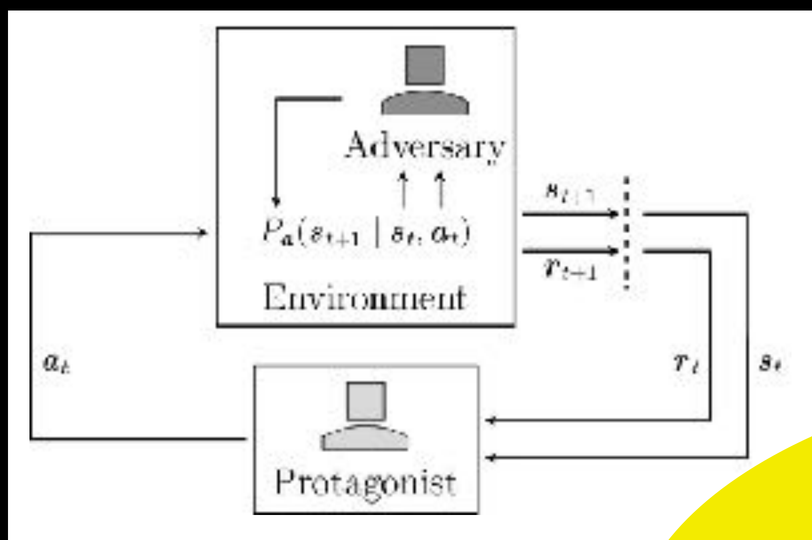
➡ Formalizing an Adversarial framework;

  - Worst-Case design
  - How to define the adversary?

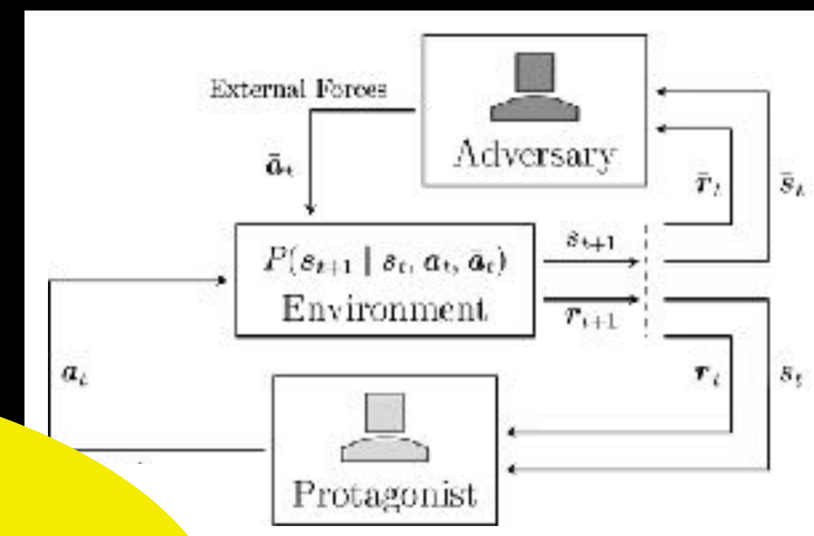
Hansel, K.; Moos, J.; Abdulsamad H.; Stark S.; Clever D.; Peters, J. (2022). Robust Reinforcement Learning: A Review of Foundations and Recent Advances

Robust Reinforcement Learning

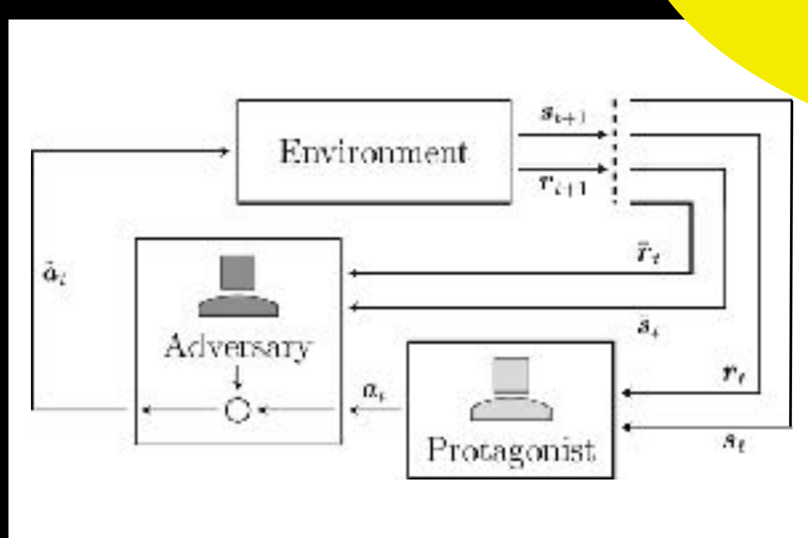
Transition-Robust-Design



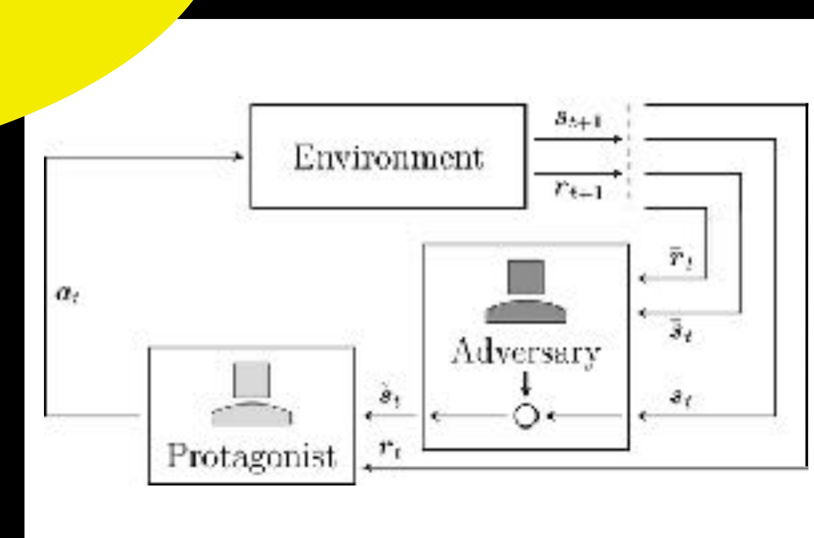
Disturbance-Robust-Design



Robustness in
Reinforcement Learning
via
Adversarial Design



Action-Robust-Design



Observation-Robust-Design

Hansel, K.; Moos, J.; Abdulsamad H.; Stark S.; Clever D.; Peters, J. (2022). Robust Reinforcement Learning: A Review of Foundations and Recent Advances

2. Inductive Bias

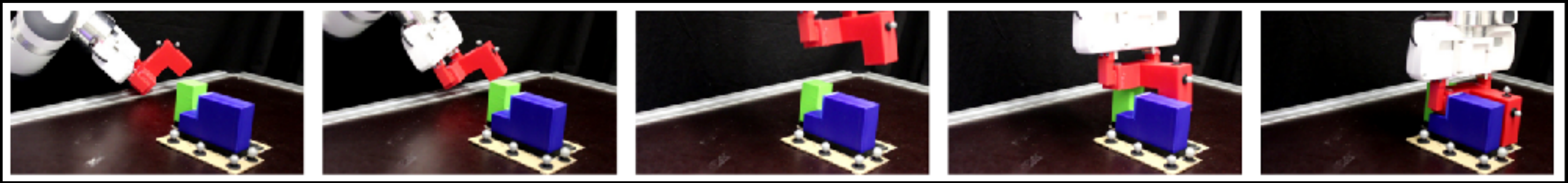
Robotics Inductive Bias

Use adversarial design as inductive bias to improve robustness!

Where to learn in contact-rich tasks?

Imitation and Residual Learning

Joao Carvalho



- Learning contact-rich assembly tasks in simulation is hard and does not transfer well
- We need methods to learn directly in the real system
- Adapt demonstrations from imitation learning with residual learning

Use the variability in human-demonstrations as an inductive bias for exploration

Carvalho, J.; Koert, D.; Daniv, M.; Peters, J. (2022). Adapting Object-Centric Probabilistic Movement Primitives with Residual Reinforcement Learning, 2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)

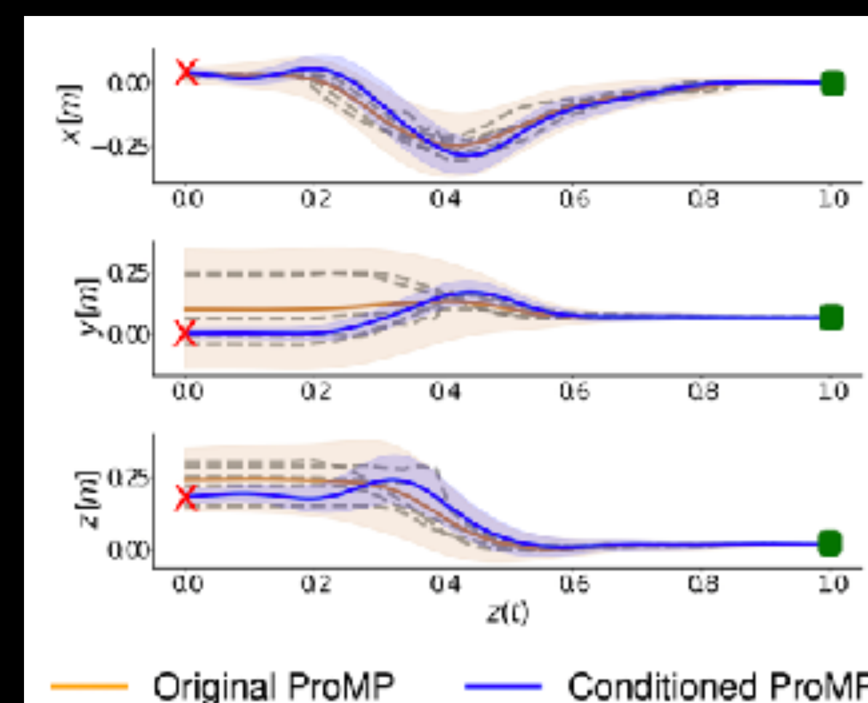
Imitation and Residual Learning

Joao Carvalho

- Residual learning combines a nominal policy π_{nom} and a learned policy π_θ

$$\begin{aligned}\pi(\boldsymbol{a}|\boldsymbol{s},t) &= \Psi\left(\pi_{\text{nom}}, \pi_\theta, \boldsymbol{s}, \boldsymbol{a}, t\right)\\ &= \alpha(\boldsymbol{s},t)\pi_{\text{nom}}(\boldsymbol{s},t) \oplus \beta(\boldsymbol{s},t)\pi_\theta(\boldsymbol{a}|\boldsymbol{s},t)\end{aligned}$$

- Given demonstrations of an insertion task, choose where to learn the residual based on the demonstrations' variance

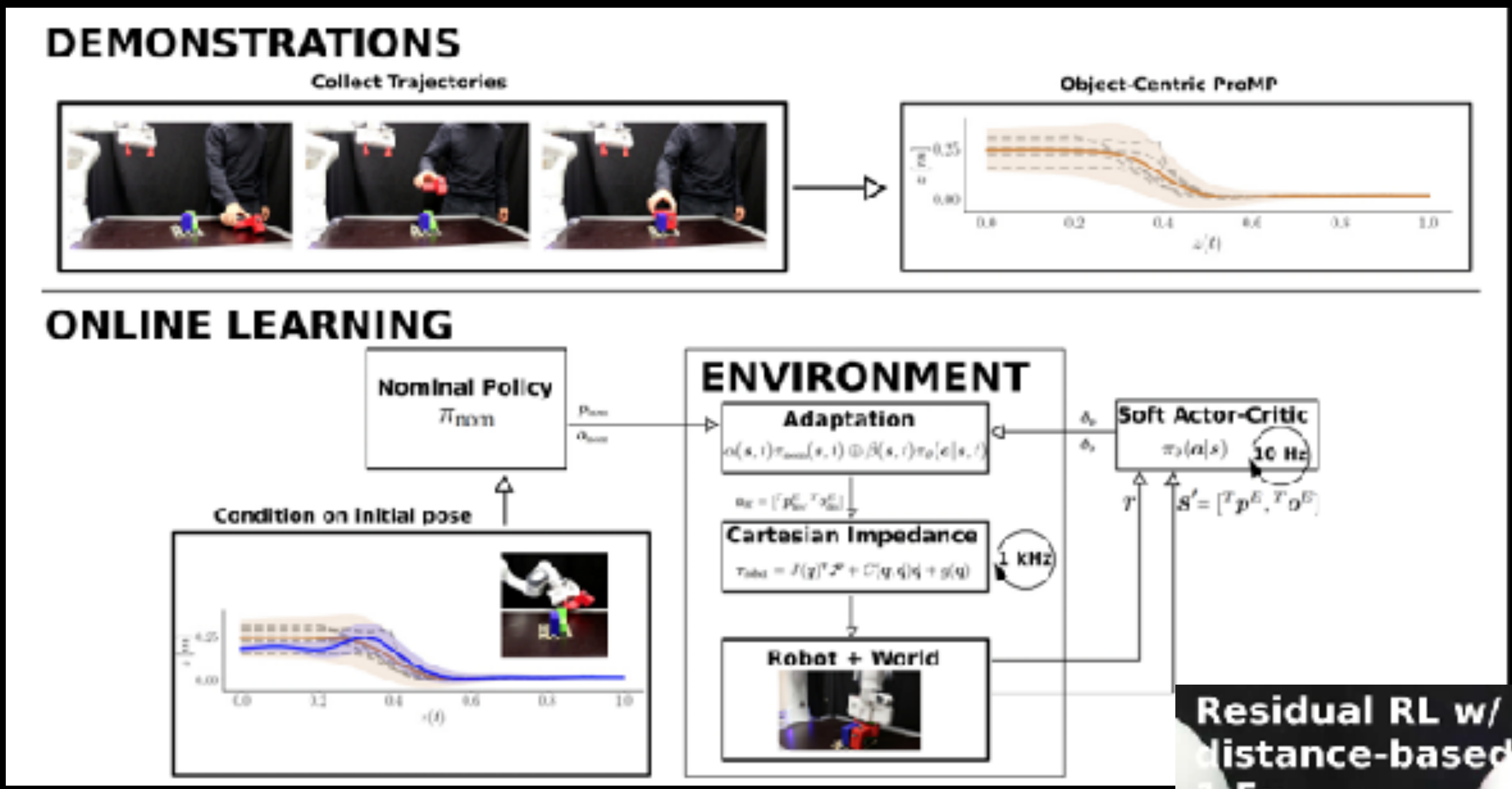


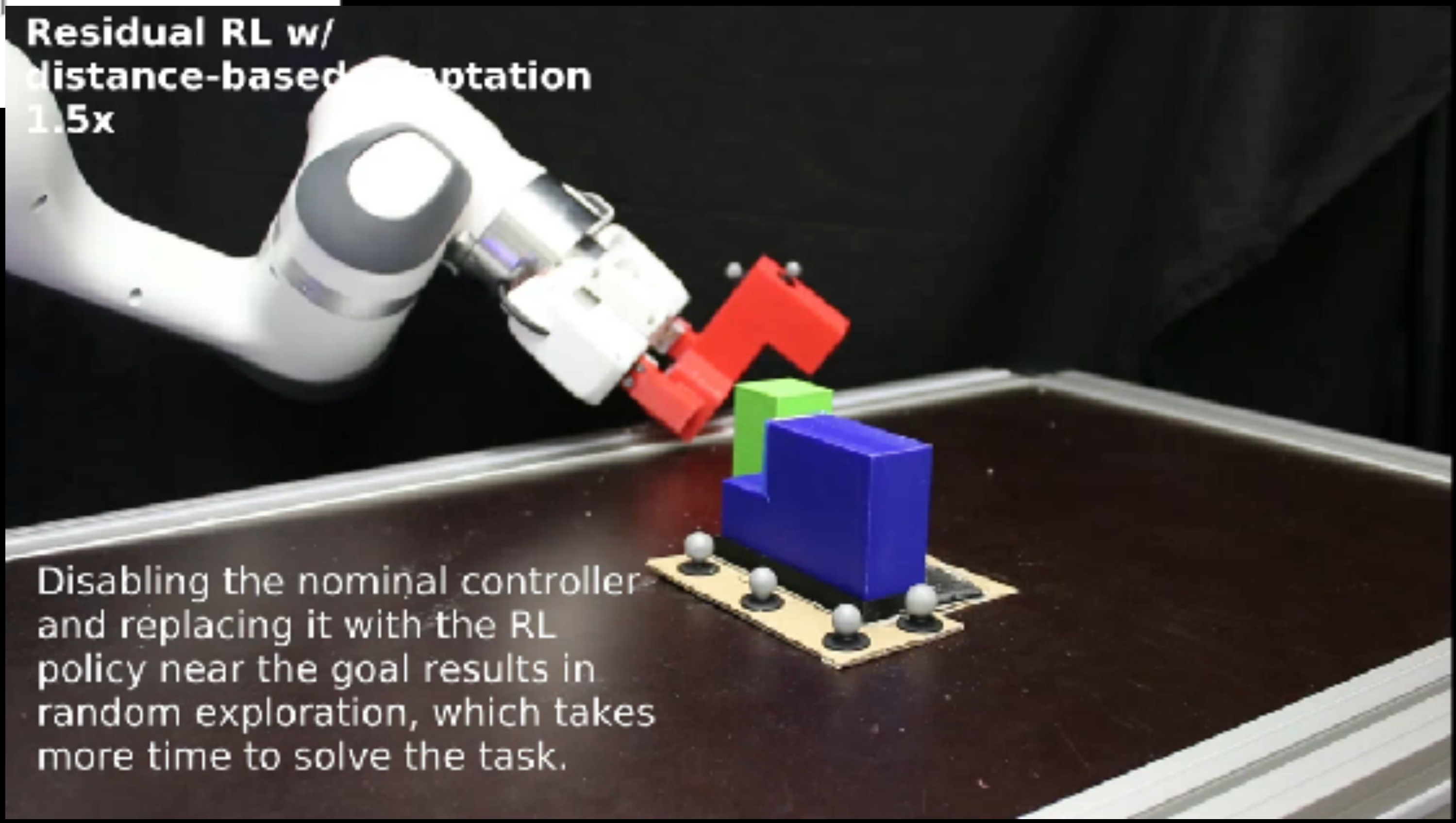
Carvalho, J.; Koert, D.; Daniv, M.; Peters, J. (2022). Adapting Object-Centric Probabilistic Movement Primitives with Residual Reinforcement Learning, 2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)

TECHNISCHE UNIVERSITÄT DARMSTADT

Imitation and Residual Learning

Joao Carvalho





Carvalho, J.; Koert, D.; Daniv, M.; Peters, J. (2022). Adapting Object-Centric Probabilistic Movement Primitives with Residual Reinforcement Learning, 2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)

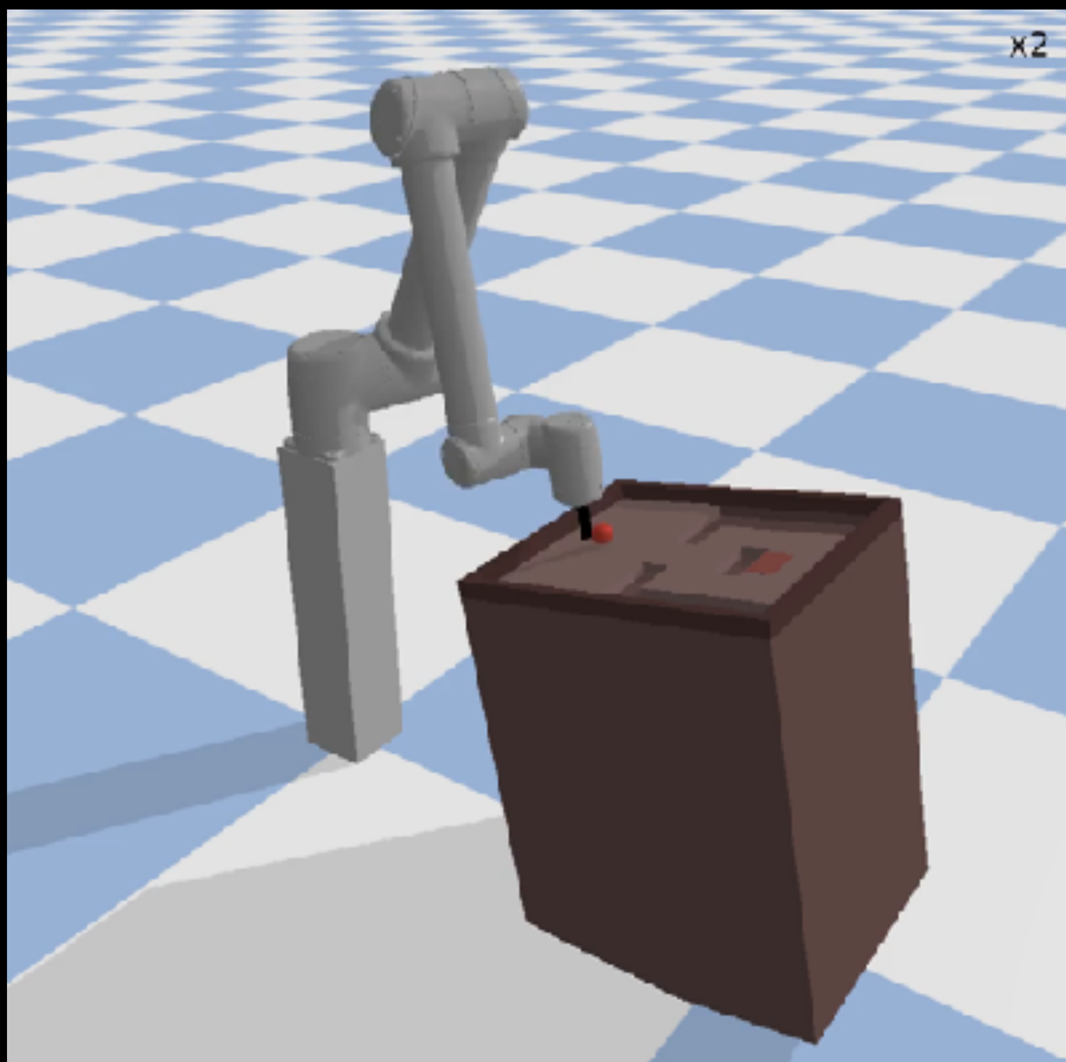
2. Inductive Bias

Robotics Inductive Bias

Use nominal policies and residual reinforcement learning to learn in the real system.

Data of real robotic systems is extremely scarce - we cannot afford to be wasteful!

Active Model Learning



The robot has to push the red ball into the target zone at the top of the table. The tilted table and a sparse reward make this task impossible to solve for classical RL approaches.

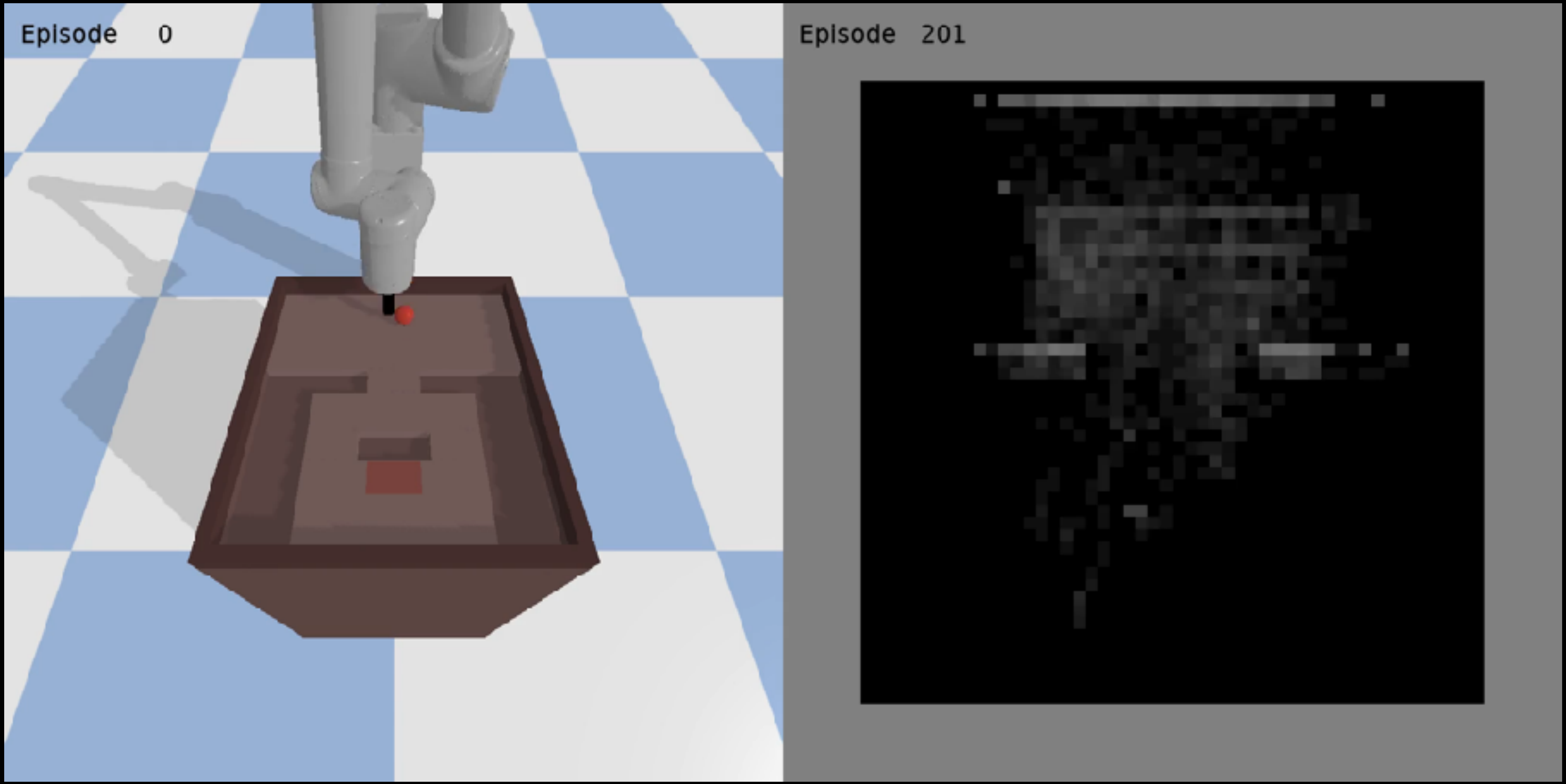
- How do we make sure the data we collect of the system is as useful for model-learning as possible?
- Maximize Information Gain w.r.t the model!

$$\max_{\pi} \underbrace{\mathrm{MI}\left(\theta, (\mathbf{x}, \mathbf{r}, \mathbf{a}) \mid \pi, x_t\right)}_{\text{Expected Information Gain}} + \beta \underbrace{\mathbb{E}_{P_\pi(r_{t+1:T})}\left[\sum_{\tau=t+1}^{T} r_\tau\right]}_{\text{Expected Reward}}$$

- Augment with expected reward to explore promising regions more thoroughly

Active Model Learning

Tim Schneider

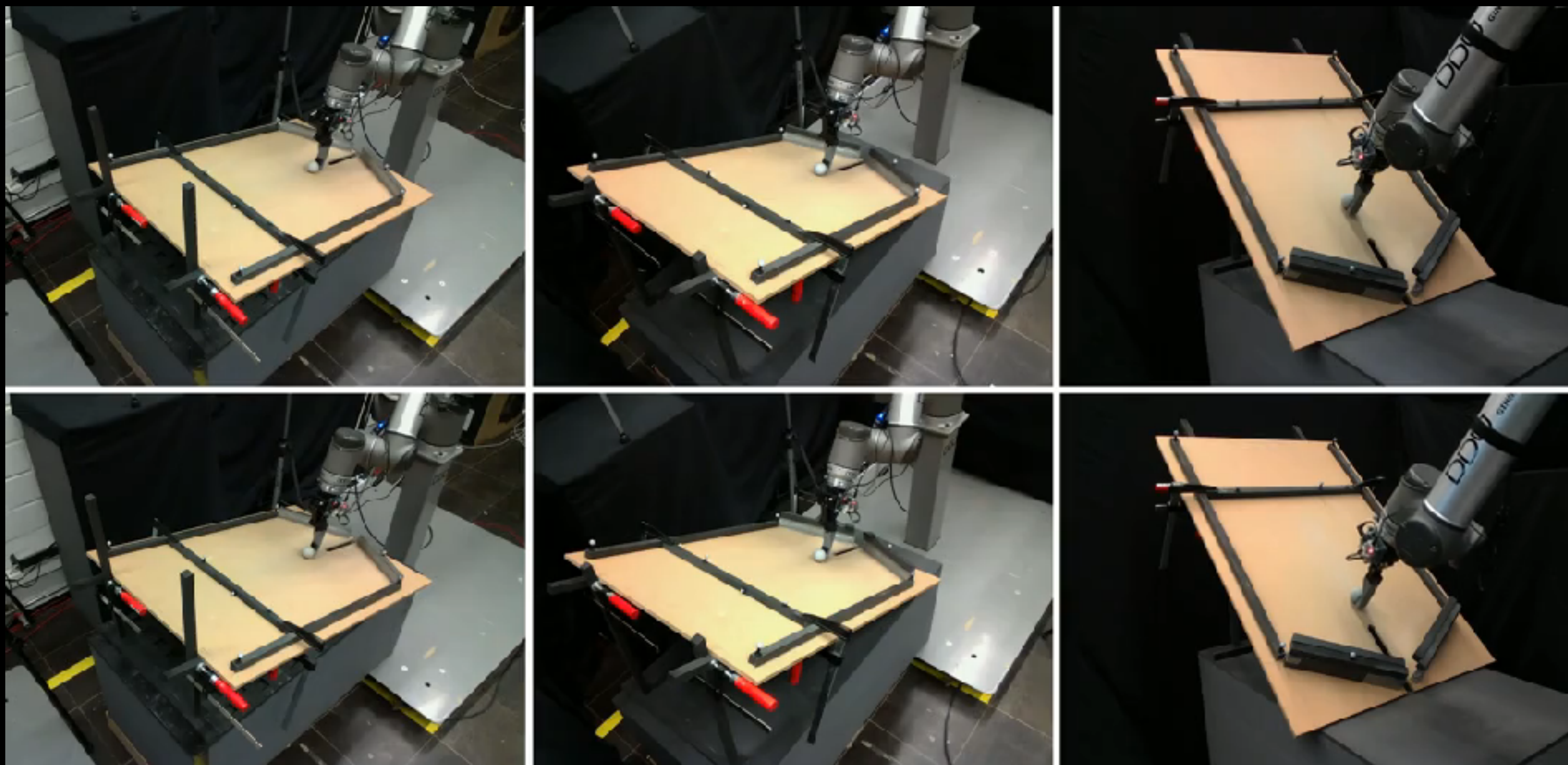


Our method explores this challenging environment efficiently and solves the task

Active Model Learning

Tim Schneider

fully automated training - **~25h** of training each - **6 configurations**



Our method solves this task by training on the real system from scratch for 5/6 configurations

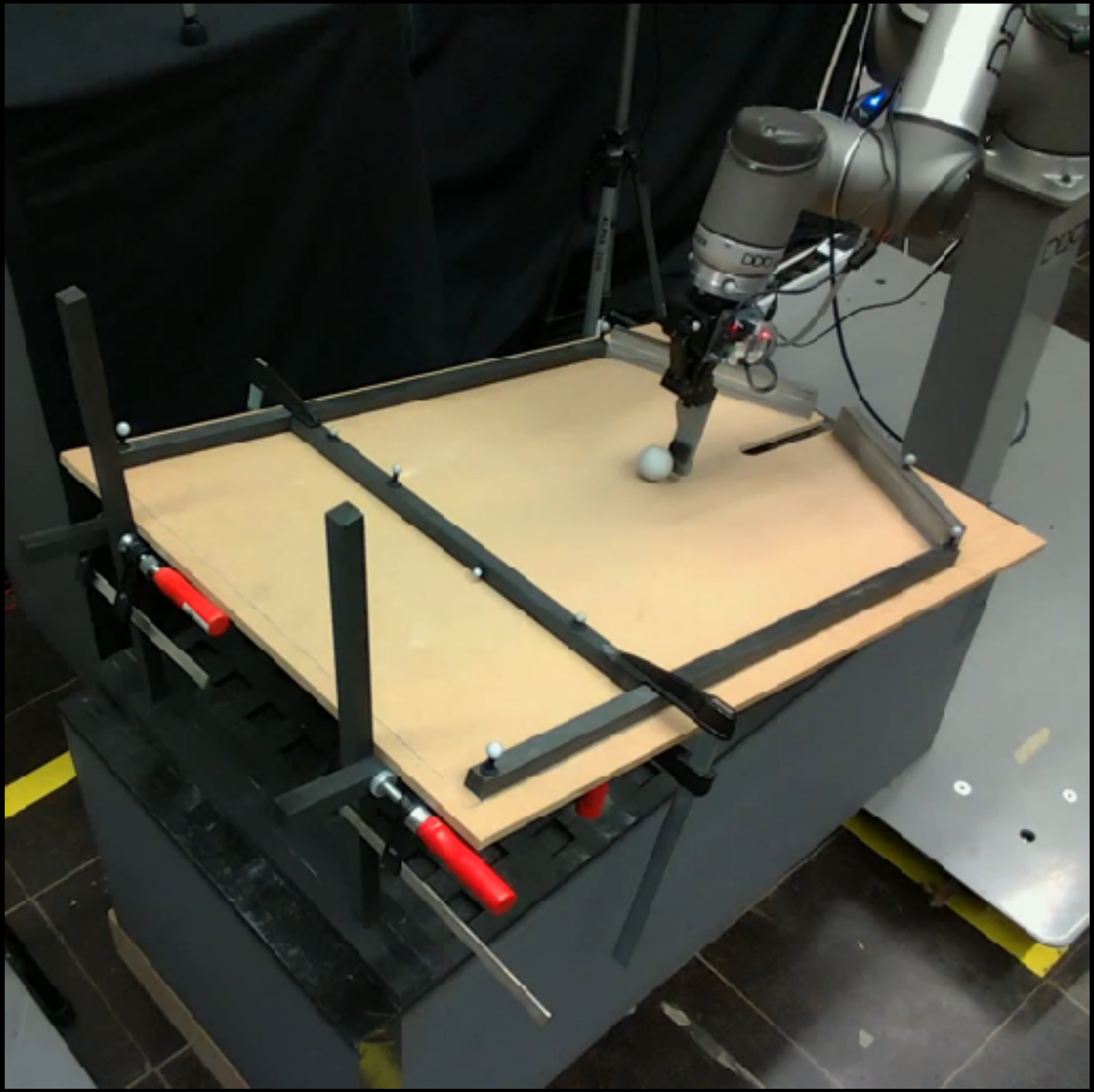
Schneider, T; Belousov, B; Chalvatzaki, G.; Romeres, D; Jha, D.K.; Peters, J. (2022).
Active exploration for robotic manipulation, International Conference on Intelligent Robots and Systems

TECHNISCHE UNIVERSITÄT DARMSTADT

Active Model Learning


Tim Schneider

How did we do it? Fully automated training!



Schneider, T; Belousov, B; Chalvatzaki, G.; Romeres, D; Jha, D.K.; Peters, J. (2022).
Active exploration for robotic manipulation, International Conference on Intelligent Robots and Systems

TECHNISCHE UNIVERSITÄT DARMSTADT

2. Inductive Bias

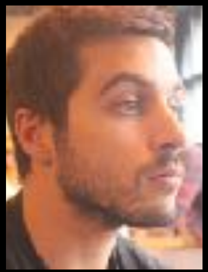
Robotics
Inductive
Bias

Control your system in a maximally informative way for model learning

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Stable Vector Fields for Goal-Conditioned Tasks

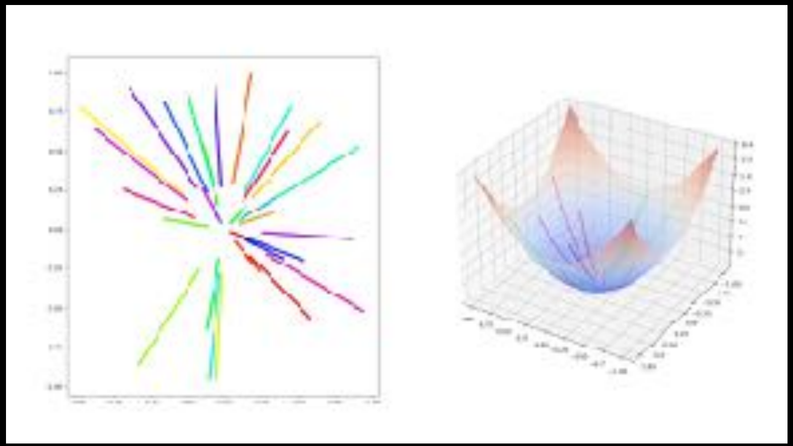
Davide Tateo



Julen Urain

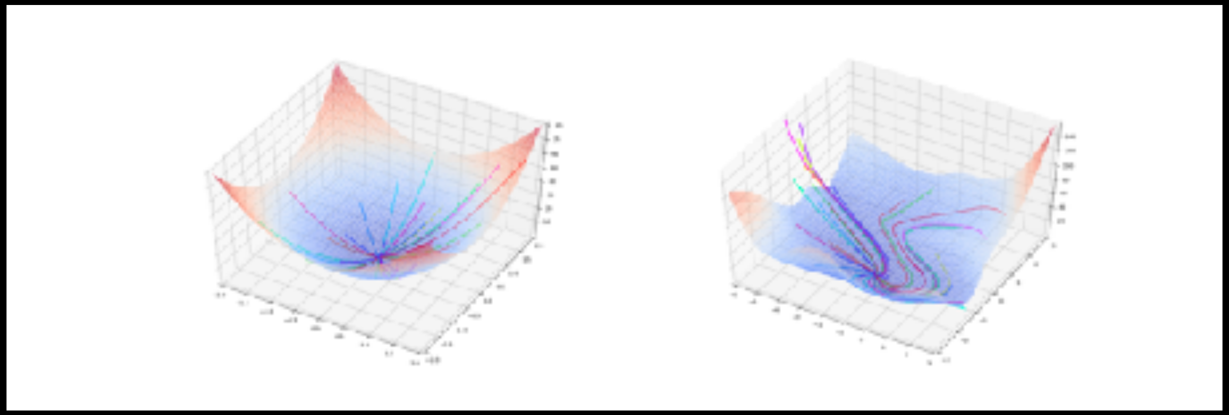
- Lot of desirable robot behaviours aims to arrive to a specific target.
- Simple solution: Linear attractor

$$\dot{\mathbf{x}} = -(\mathbf{x} - \mathbf{x}_{tar})$$



- How can we learn nonlinear attractors?
- Exploit the diffeomorphic function Φ of the Normalising Flows

$$\dot{\mathbf{x}} = -\mathbf{J}_{\Phi}(\Phi^{-1}(\mathbf{x}) - \Phi^{-1}(\mathbf{x}_{tar}))$$

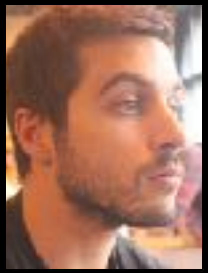


Urain, J.; Ginesi, M; Tateo, D.; Peters, J. (2020).
ImitationFlow: Learning Deep Stable Stochastic Dynamic Systems by Normalising Flows. IROS

TECHNISCHE UNIVERSITÄT DARMSTADT

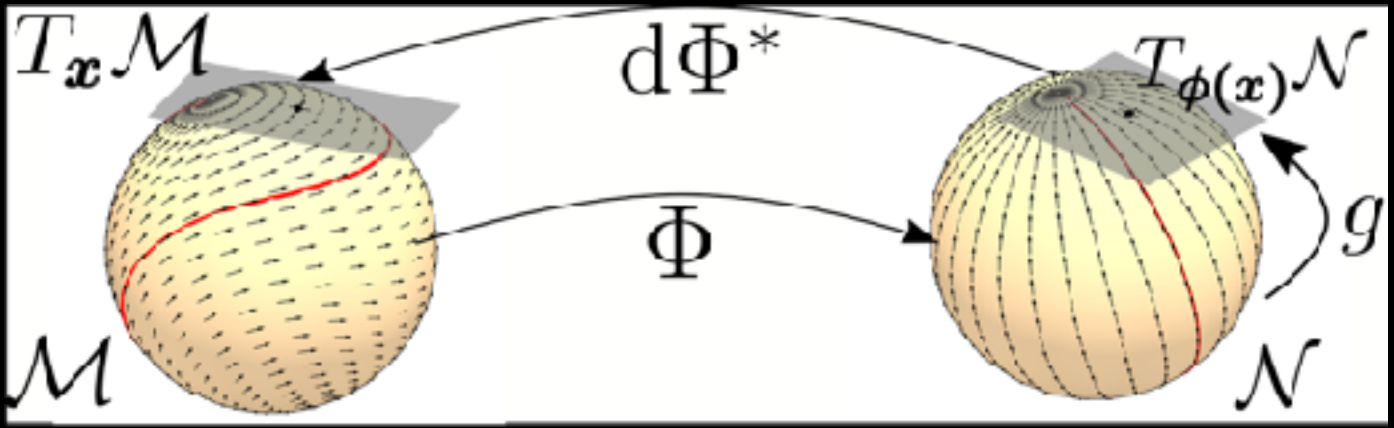
Stable Vector Fields in Manifolds

Davide Tateo

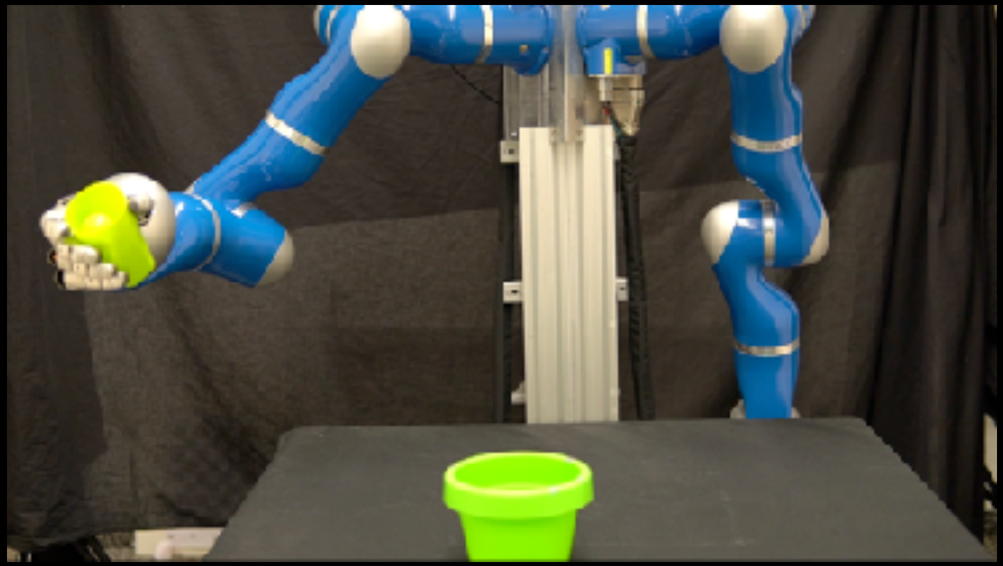


Julen Urain

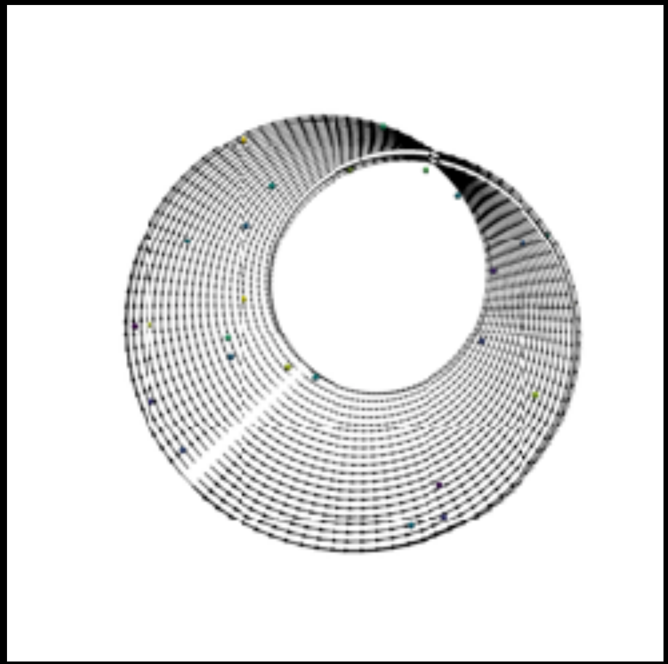
- To represent Stable Vector Fields for Orientations, we require to represent vector fields in non-Euclidean manifolds



Stable Vector Fields in SE(3) (Position + Orientation)



Stable Vector Fields in Mobius Strip

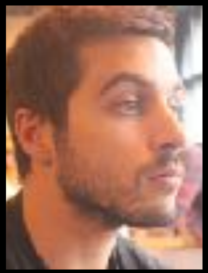


Urain, J., M; Tateo, D.; Peters, J. (2022).
Learning Stable Vector Fields on Lie groups. RA-L

Diffusion Models in Robotics


Niklas Funk


Georgia Chalvatzaki


Julen Urain

- Diffusion models are perfect candidates to represent dense and smooth cost functions in trajectory optimisation.

- Diffusion Models propose learning a function $\mathrm{s}_\theta(\tau)$ that represents the score function of a data distribution $\mathrm{s}_\theta(\tau) = \nabla_\tau \log q_{\mathcal{D}}(\tau)$

- Then, we can run an inverse diffusion process to generate samples from $q_{\mathcal{D}}(\tau)$

$$\tau_{k-1} = \tau_k + \frac{\alpha_k^2}{2} \nabla_\tau \log q_{\mathcal{D}}(\tau_k) + \alpha_k \epsilon \quad , \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

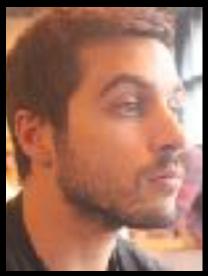
- Note that if we substitute $\log q(\tau_k) = \sum_k c_k(\tau)$ then, we have a gradient-based trajectory optimiser.

Urain, J.; Funk, N; Peters, J. ; Chalvatzaki G (2023).
SE(3)-DiffusionFields: Learning smooth cost functions for joint grasp and motion optimization through diffusion (ICRA).

6D-Grasp Diffusion Models
for joint grasp and motion optimization

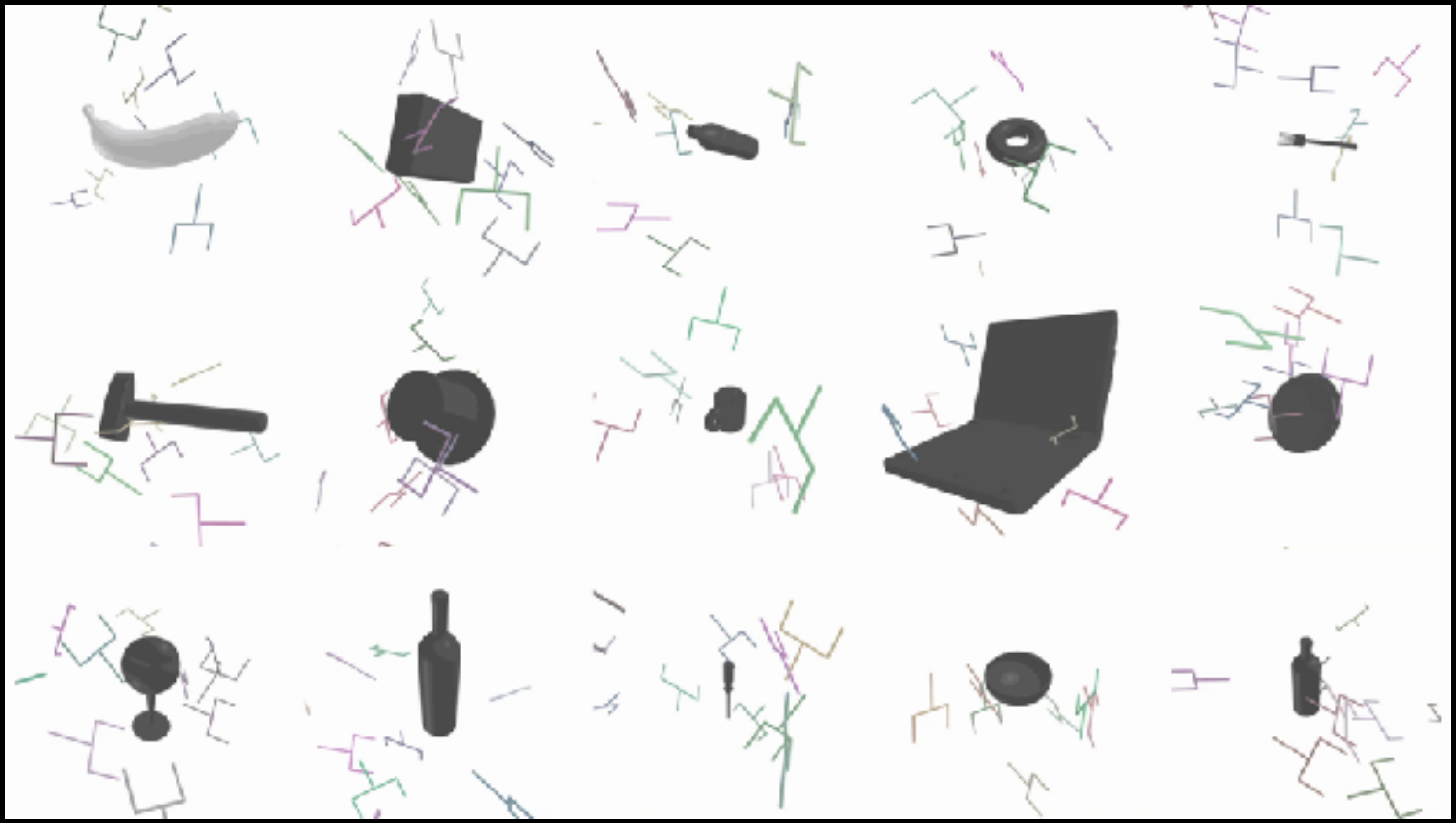

Niklas Funk


Georgia Chalvatzaki


Julen Urain

- We learn a diffusion model representing a distribution of SE(3) grasp poses for a variety of objects

$$f_\theta(\mathbf{H}) = \log q_{\mathcal{D}}(\mathbf{H}) \quad , \quad \mathbf{H} \in SE(3)$$

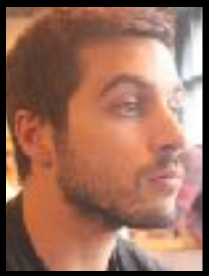


Urain, J.; Funk, N; Peters, J. ; Chalvatzaki G (2023).
SE(3)-DiffusionFields: Learning smooth cost functions for joint grasp and motion optimization through diffusion (ICRA).

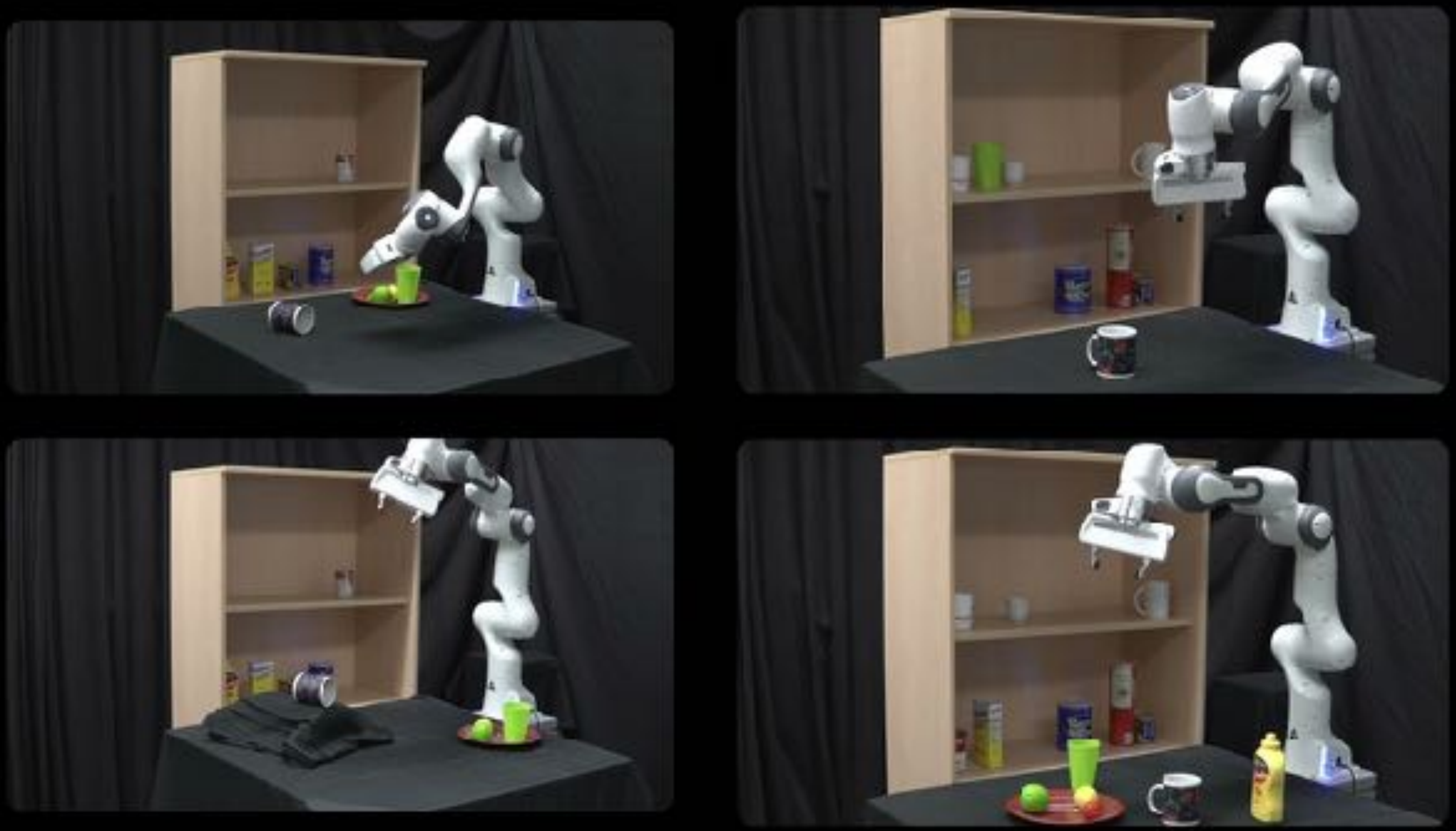
6D-Grasp Diffusion Models
for joint grasp and motion optimization


Niklas Funk


Georgia Chalvatzaki

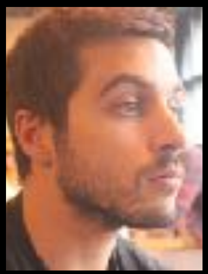

Julen Urain

- We combine the learned grasp diffusion model with heuristic costs(obstacle avoidance, joint limits…) and generate trajectories to solve complex pick and place tasks



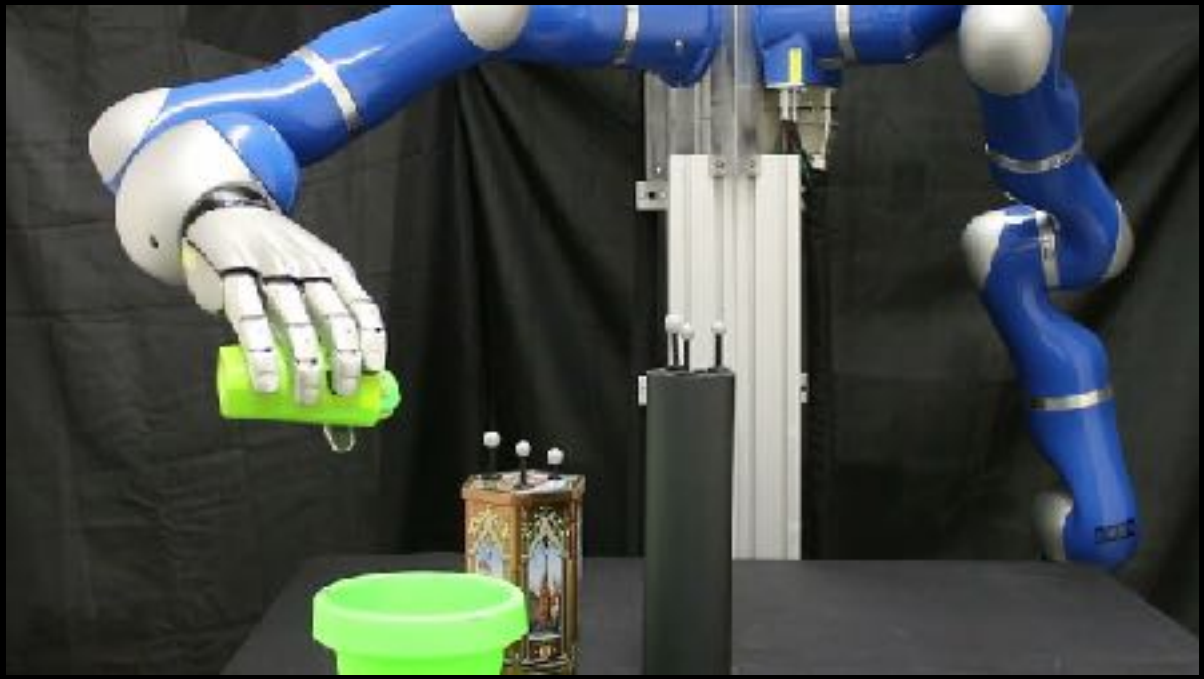
Urain, J.; Funk, N; Peters, J. ; Chalvatzaki G (2023).
SE(3)-DiffusionFields: Learning smooth cost functions for joint grasp and motion optimization through diffusion (ICRA).

Learning smooth cost functions for complex tasks



An Thai Lee Georgia Chalvatzaki Julen Urain

- We apply similar approaches to learn more complex behaviours such as pouring and combine it with additional cost functions



Urain, J.; Lee, A.T.; Lambert, A.; Chalvatzaki G.; Peters, J. (2022). Learning Implicit priors for motion optimisation (IROS)

2. Inductive Bias

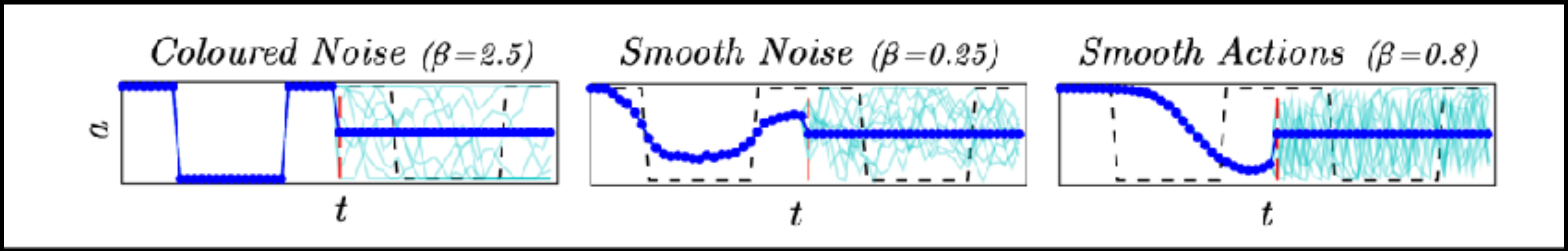
Robotics Inductive Bias

Temporal smoothness of action sequences

Inferring smooth control

Joe Watson

- Every good roboticist knows that robots like **smooth actions**, but this is hard to achieve in settings such as **sample-based model predictive control.**
- How can we achieve this?
  - **Filter** the actions — introduces delay!
  - Sample **smoothed** noise — won't preserve smoothness!
  - Movement **primitives** — requires handcrafted features!

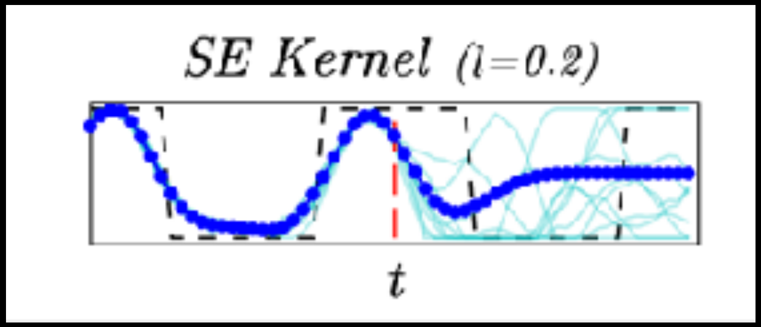


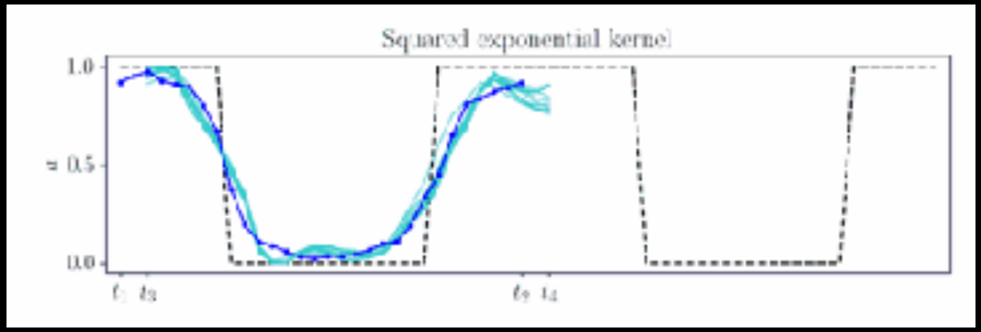
Watson, J; Peters, J. (2022).
Inferring smooth control: Monte Carlo Posterior Policy Iteration with Gaussian Processes (CoRL) [oral].

Inferring smooth control

Joe Watson

- When using REPS, our update looks like a Bayesian posterior $q(\theta) \propto \exp(\alpha R(\theta)\, p(\theta)$, which we call a 'pseudo'-posterior
- In the context of MPC, we are optimising an open-loop action sequence $A = [a_1, a_2, a_3, \ldots]$
- To encode smoothness we can design a continuous-time Gaussian process (GP) prior, $p(a \mid t) = \mathrm{GP}(\mu(t), \Sigma(t))$
- REPS-style optimization can be implemented as inference on the GP, with Kalman filter-like updates each time step



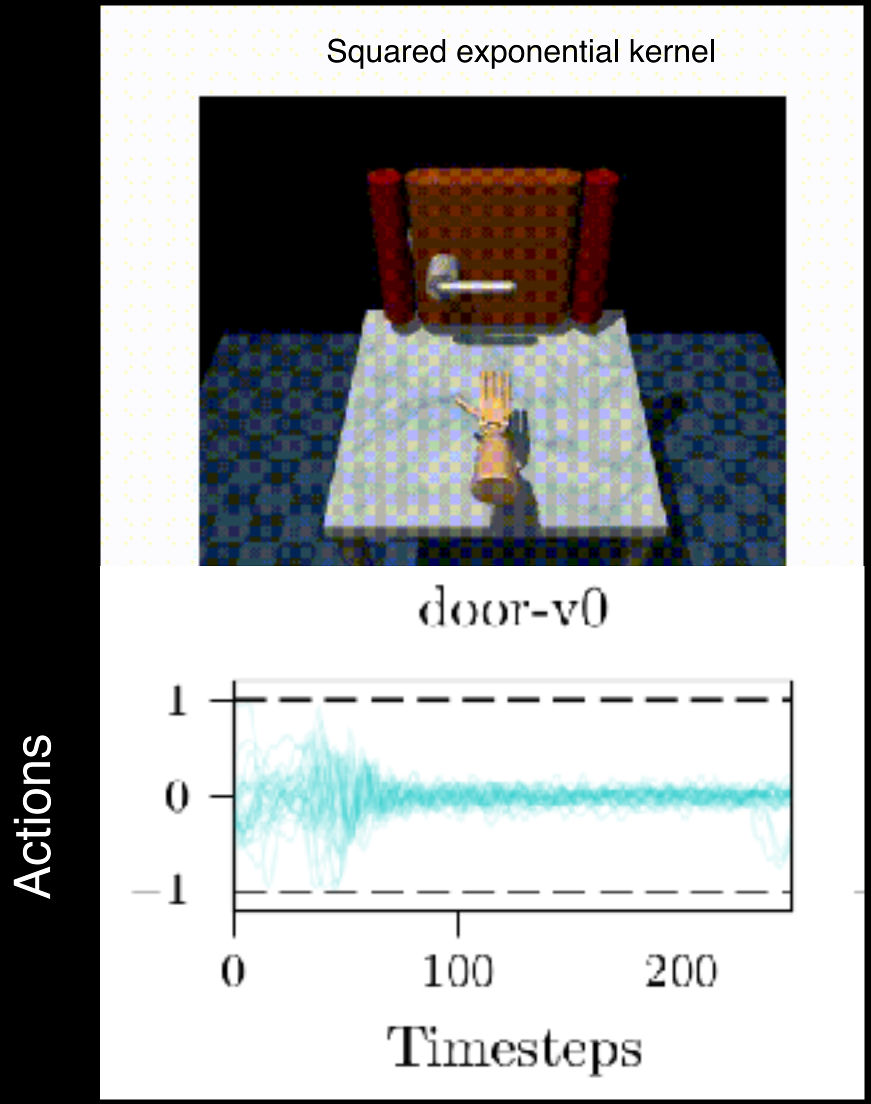


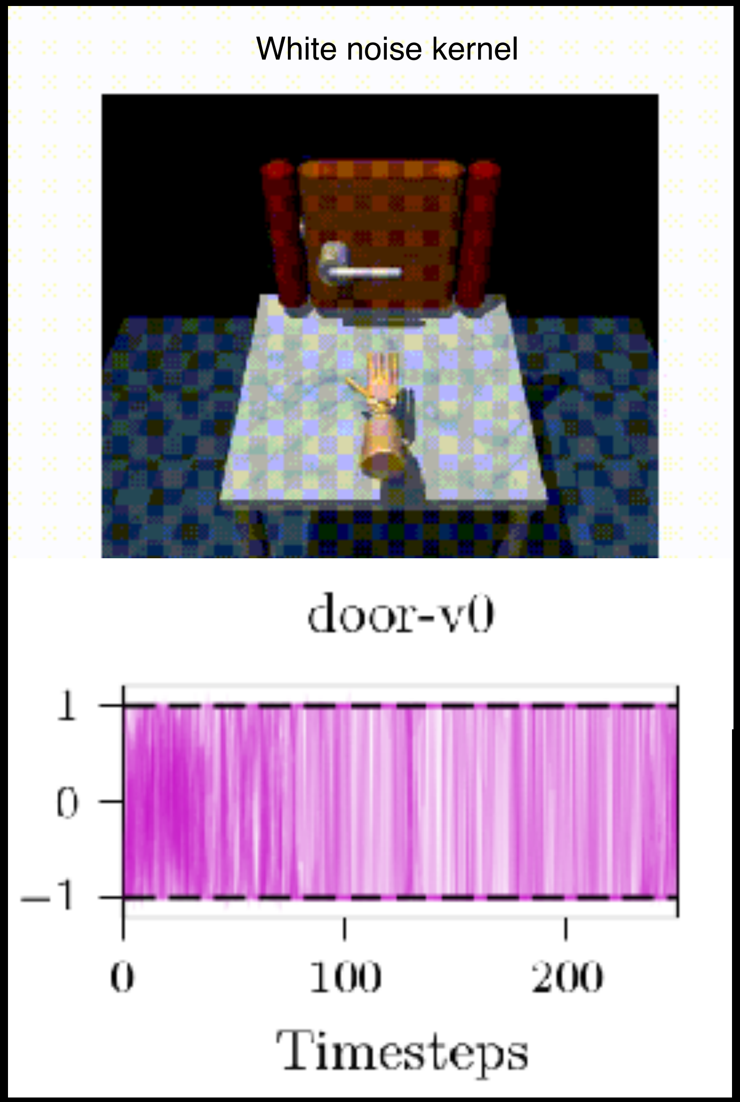
Watson, J; Peters, J. (2022).
Inferring smooth control: Monte Carlo Posterior Policy Iteration with Gaussian Processes (CoRL) [oral].

Inferring smooth control

Joe Watson

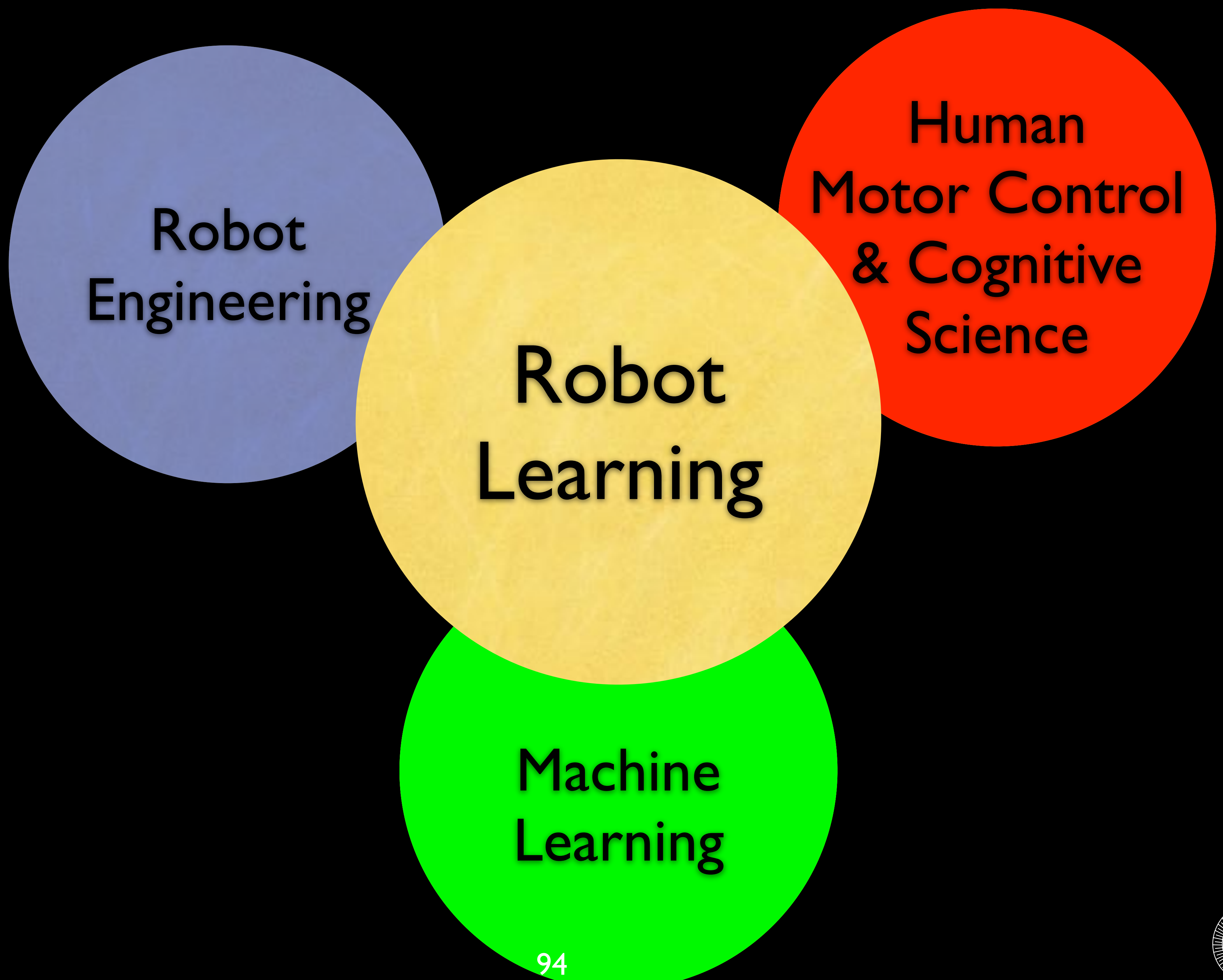
- This approach scales to high-dimensional MPC and dramatically changes the solution



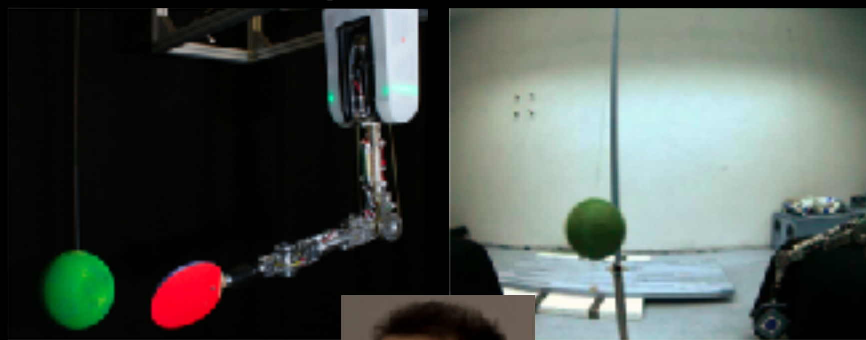


Watson, J; Peters, J. (2022).
Inferring smooth control: Monte Carlo Posterior Policy Iteration with Gaussian Processes (CoRL) [oral].

Outlook

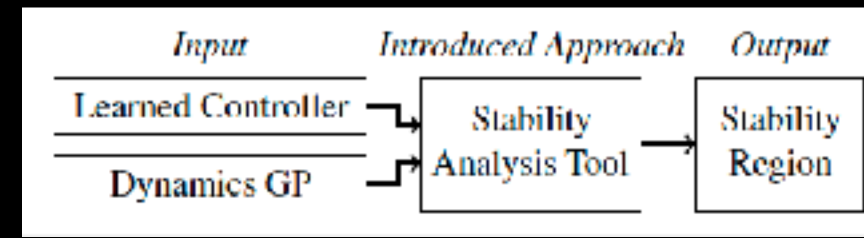

Robot Engineering
Robot Learning
Human Motor Control & Cognitive Science
Machine Learning

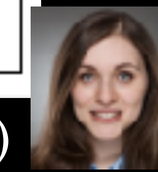
Learning State Representations for Robotics



(Simone Parisi @ MLJ)

Automated Stability Proofs





(Julia Vinogradska @ ICML+JMLR)

Robot Engineering

Self-Paced Robot Reinforcement Learning

SELF-PACED CONTEXTUAL REINFORCEMENT LEARNING (SPRL)

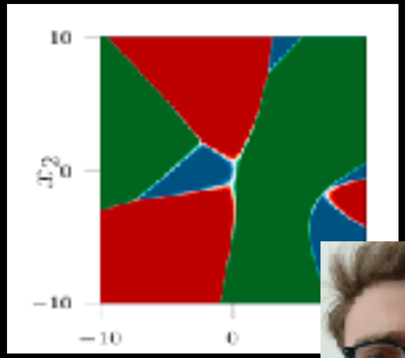
Pascal Klink, Hany Abdulsamad, Boris Belousov, Jan Peters

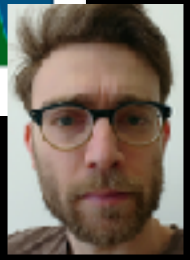
Intelligent Autonomous Systems, TU Darmstadt

SPARSE BALL-IN-A-CUP TASK

(Pascal Klink @ CoRL+NeurIPS)

Inferring Hybrid Control From Data



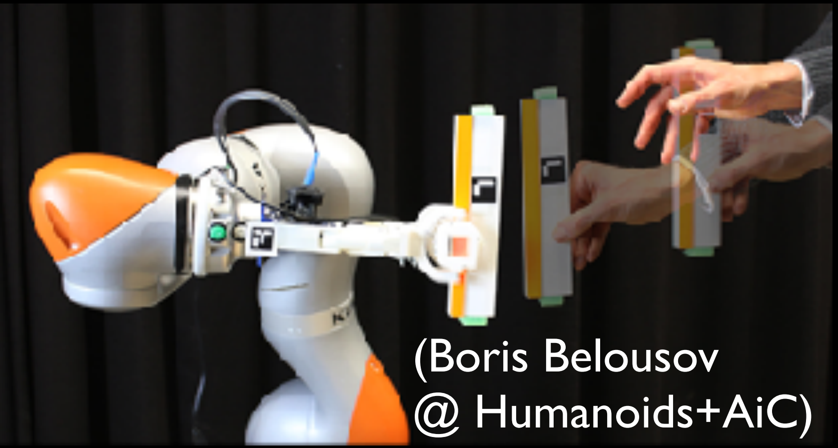


(Hany Abdulsamad @ LCNS)

(Learning) Control for Table Tennis

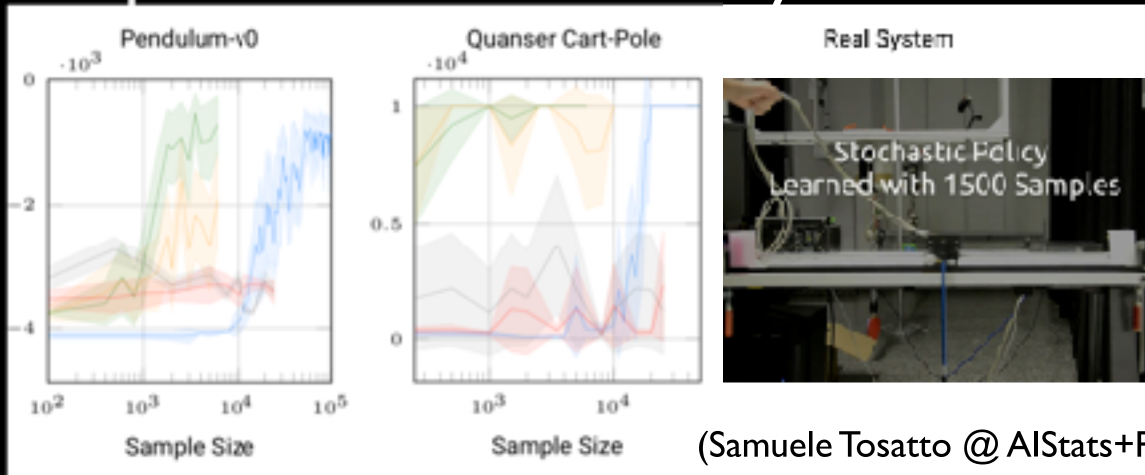
(Okan Koç @ R-AL/ ICRA)

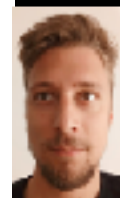
Tactile Skill Libraries



(Boris Belousov @ Humanoids+AiC)

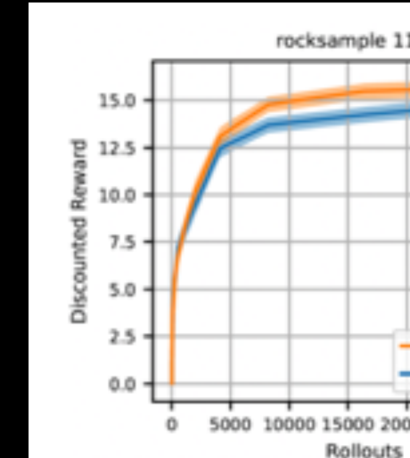
Sample Efficient Off-Policy Gradients





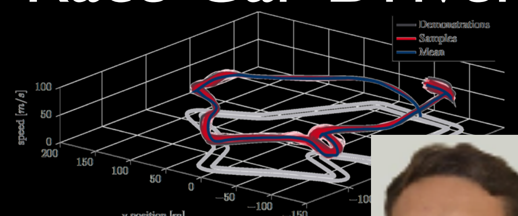
(Samuele Tosatto @ AIStats+PAMI)

Generalized Mean Estimation with MCTS



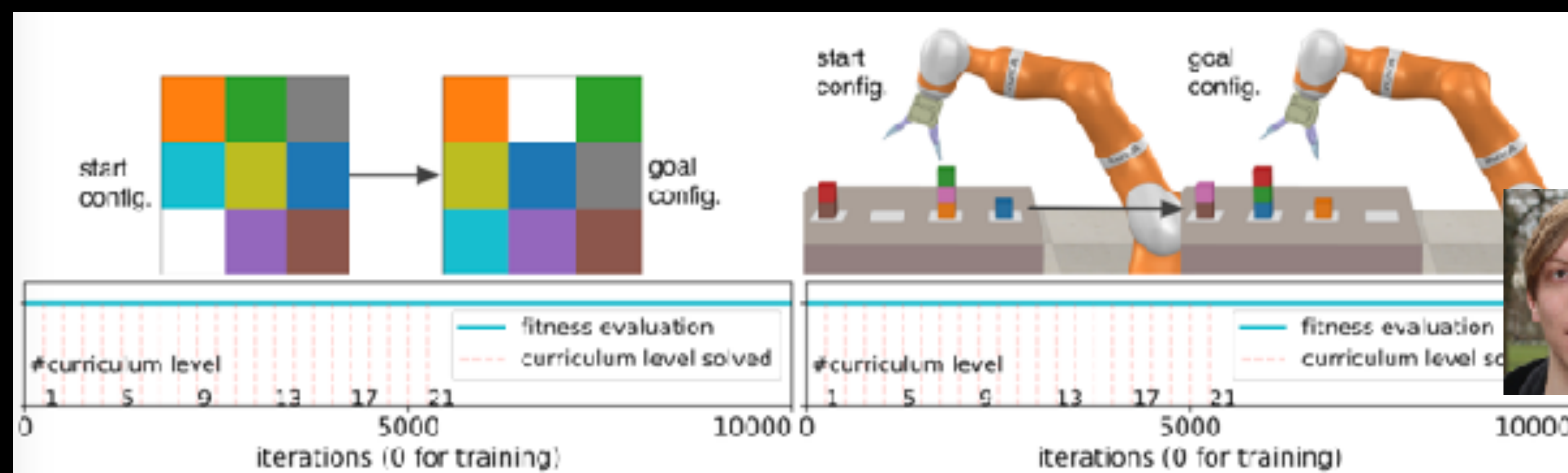
(Tuan Dam @ IJCAI)

Imitation of Race Car Drivers



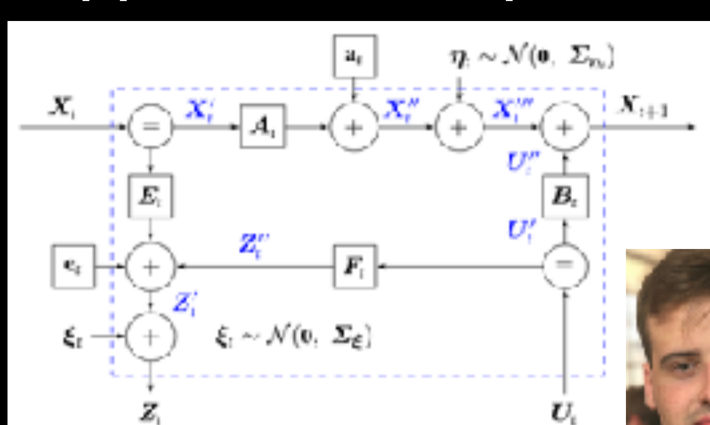
(Stefan Löckel @ IEEE R-AL)

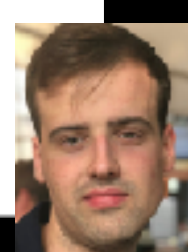
Learning Abstract Strategies independent of the Task Domain



(Daniel Tanneberg @ Nature MI)

Stochastic Optimal Control by Approximate Input Inference

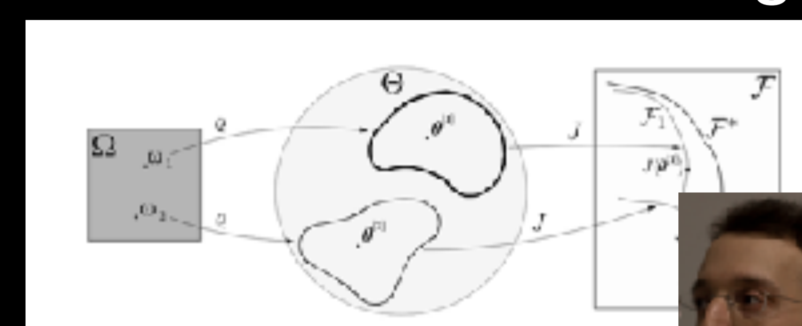




(Joe Watson @ CoRL 2019)

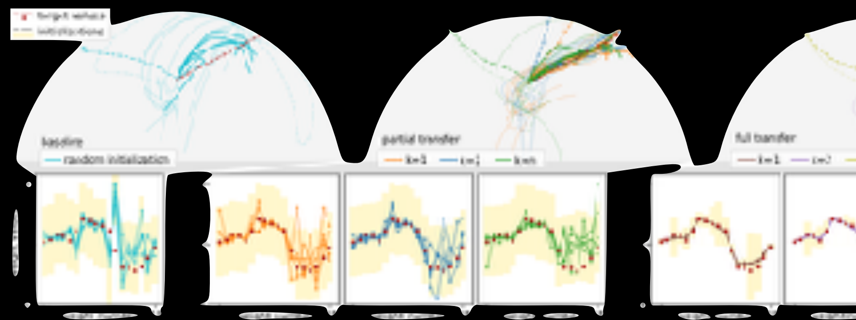
Machine Learning

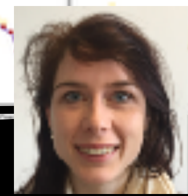
Multi-Objective Reinforcement Learning



(Simone Parisi @ NECO)

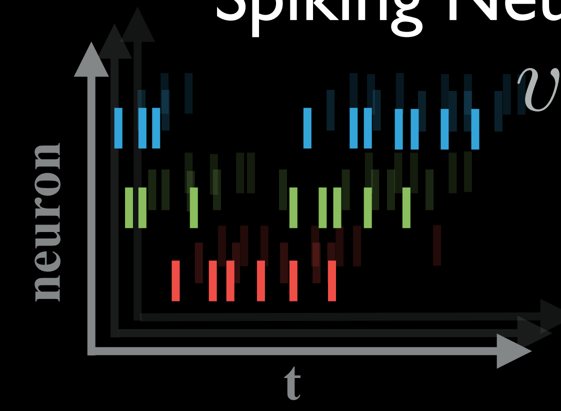
Human-like Experience Reuse

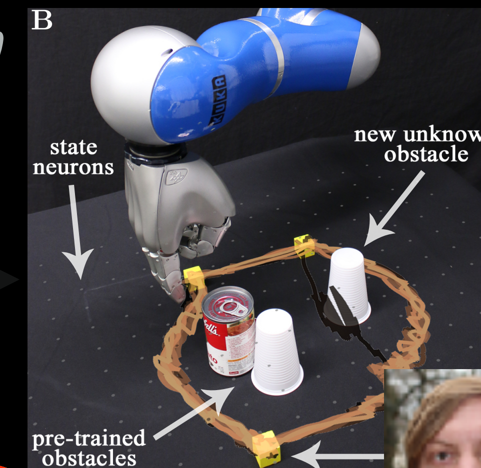


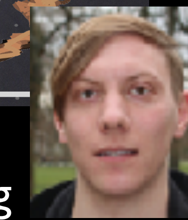


(Svenja Stark @ IROS)

Spiking Neural Models

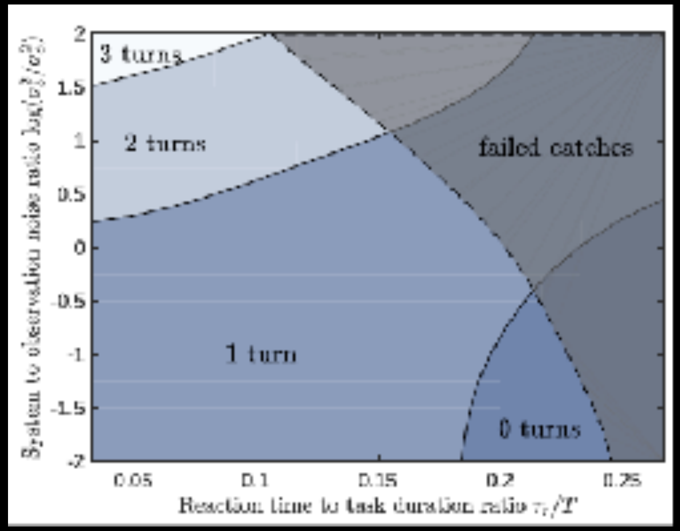






(Daniel Tanneberg @ Neural Networks)

Human Ball Catching



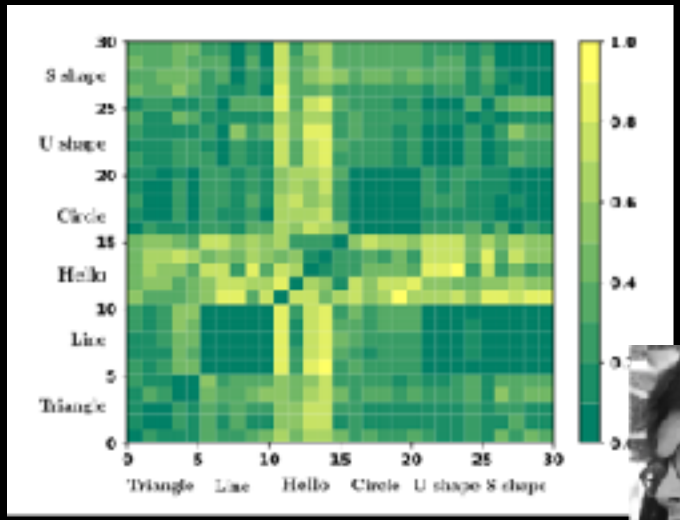
(Boris Belousov @ NeurIPS)

Human Motor Control & Cognitive Science

Human Intent Prediction

(Dorothea Koert @ R-AL/IROS)

Trajectory Similarity Measures



(Julen Urain @ IROS)


Robot Juggling
Thanks for your Attention!


Thanks for your Attention!
Robot Beer Pong

Thanks for your Attention!

Robot Coffee Making

Thanks for your Attention!

Robot Dart Games


Changing a Light-Bulb
Thanks for your Attention!

Cooperative Throwing and Catching

Thanks for your Attention!

Thanks for your Attention!

Demonstration of Pouring

Robot Pouring


Robot Juggling
Thanks for your Attention!

Thanks for your Attention!

Robot Beer Pong

Thanks for your Attention!

Robot Coffee Making

Thanks for your Attention!

Robot Dart Games

Changing a Light-Bulb

Thanks for your Attention!

Cooperative Throwing and Catching

Thanks for your Attention!

Thanks for your Attention!

Demonstration of Pouring

Robot Pouring