Towards Machine Learning of Motor Skills for Robotics

Jan Peters
Technische Universität Darmstadt

Max Planck Institute for Intelligent Systems

TECHNISCHE UNIVERSITÄT DARMSTADT

Motivation

How can we create all of these behaviors?

Source: Movie iRobot

Motivation

Uncertainty in tasks
and environment

Adapt to humans
and interact

Programming complexity
beyond human imagination

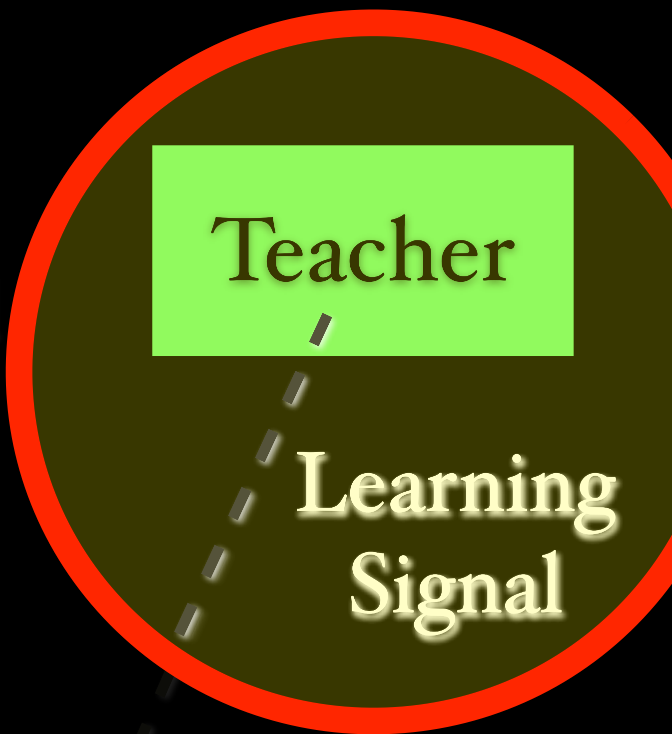
How can we fulfill Hollywood's vision of future robots?

- Smart Humans? Hand-engineering of behaviors has allowed us to go *very far*!
- Maybe we should allow the robot to learn new tricks, adapt to situations, refine skills?
- "Off-the-shelf" machine learning approaches for regression/classification?

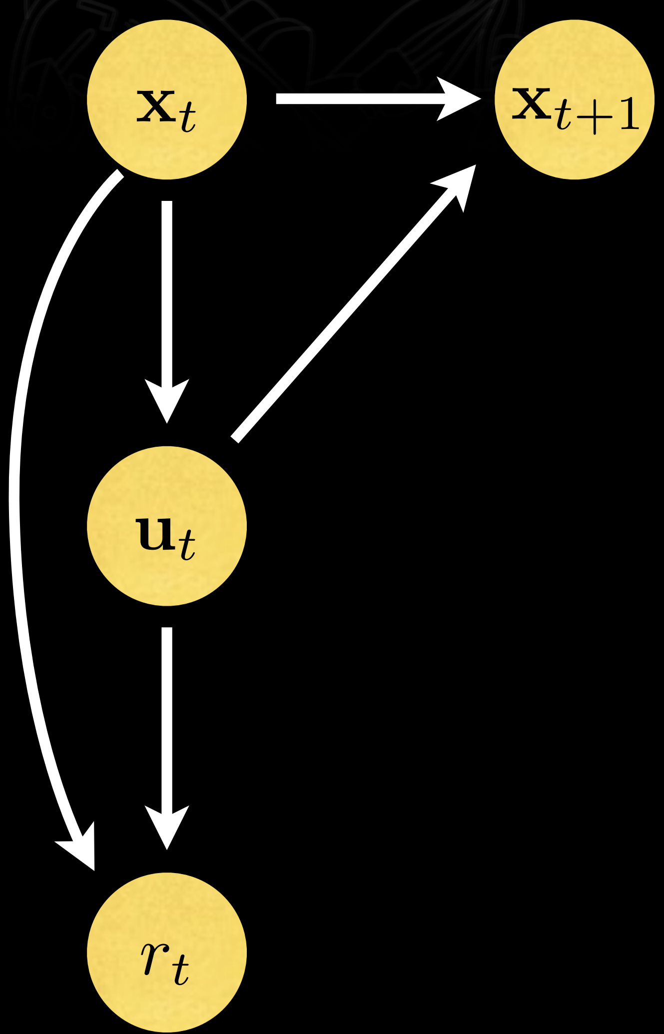
➡ We need to develop **learning approaches** suitable for **robotics**!

Outline

1. Introduction
2. How can we develop suitable machine learning methods?
3. How can elementary behavior be learned with such machine learning methods?
4. Can complex skills be learned leveraging on elementary behaviors?
5. How can we adapt to humans and learn interaction?
6. Conclusion



Modeling Assumptions



Policy: Generates action $\mathbf{u}_t$ in state $\mathbf{x}_t$.

Should we use a deterministic policy $\mathbf{u}_t = \pi(\mathbf{x}_t)$?

NO! Stochasticity is important:

- needed for exploration
- breaks "curse of dimensionality"
- optimal solution can be stochastic

Hence, we use a stochastic policy: $\mathbf{u}_t \sim \pi(\mathbf{u}_t|\mathbf{x}_t)$

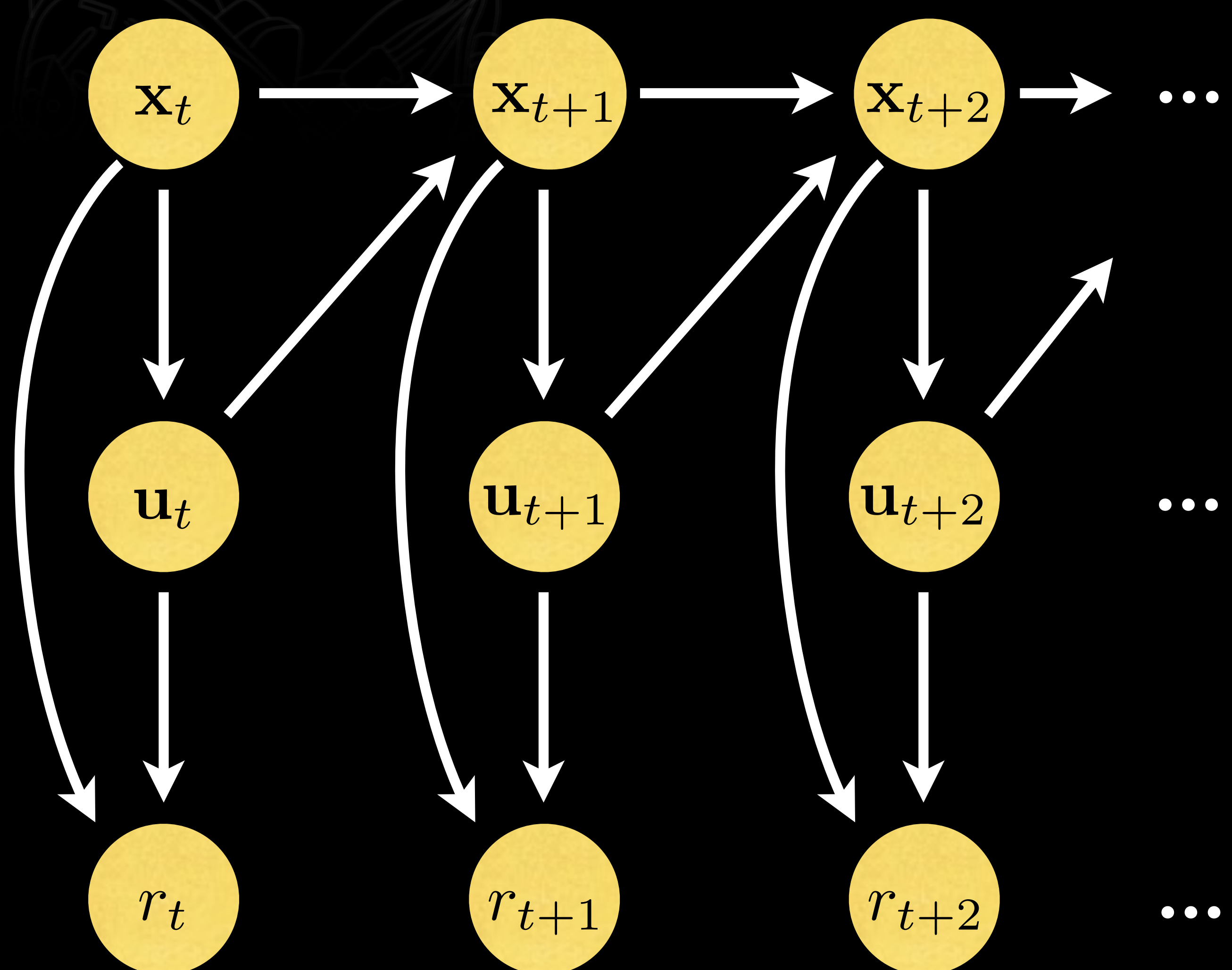
Robot learning implies "policy optimization"!

Teacher: Evaluates the performance and rates it with r_t.

Environment: An action $\mathbf{u}_t$ causes the system to change state from $\mathbf{x}_t$ to $\mathbf{x}_{t+1}$.

Model in the real world: $\mathbf{x}_{t+1} \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{u}_t)$

Let the loop roll out!



Trajectories

$$\boldsymbol{\tau} = [\mathbf{x}_0, \mathbf{u}_0, \mathbf{x}_1, \mathbf{u}_1 \ldots, \mathbf{x}_{T-1}, \mathbf{u}_{T-1}, \mathbf{x}_T]$$

Path distributions

$$p(\tau) = p(\mathbf{x}_0) \prod_{t=0}^{T-1} p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{u}_t)\pi(\mathbf{u}_t|\mathbf{x}_t)$$

Path rewards:

$$r(\boldsymbol{\tau}) = \sum_{t=0}^{T} \alpha_t r(\mathbf{x}_t, \mathbf{u}_t)$$

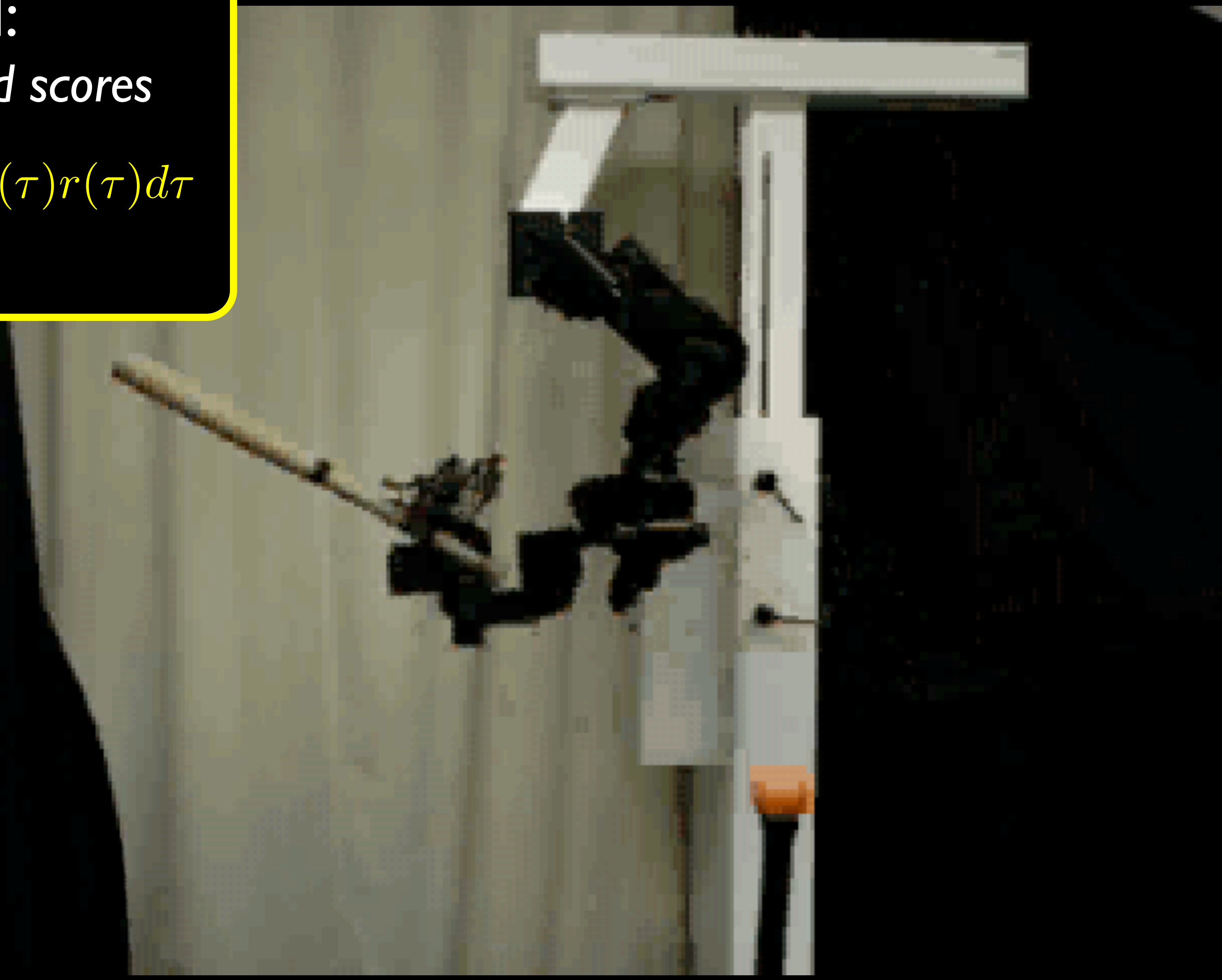
What is learning?

In our model:

Optimize the *expected scores*

$$J(\theta) = E_\tau\{r(\tau)\} = \int_{\mathbb{T}} p_\theta(\tau) r(\tau) d\tau$$

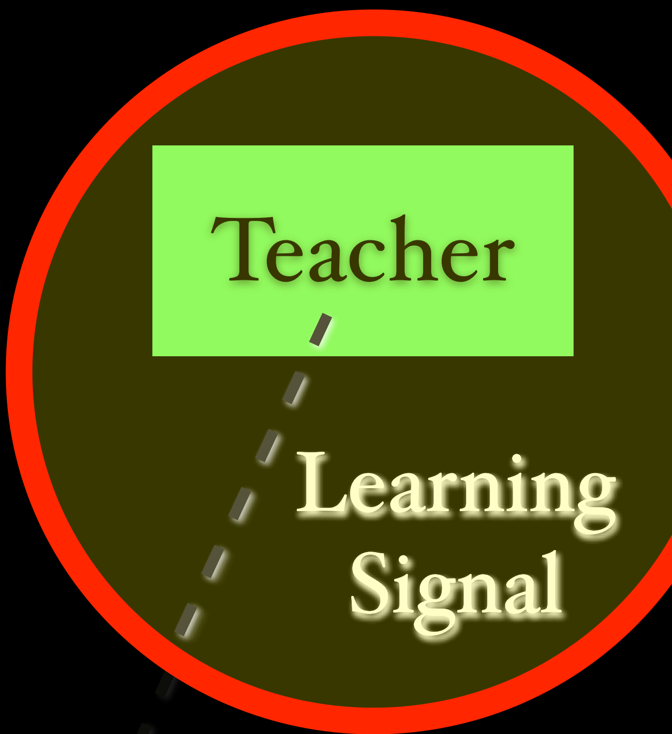
of the teacher.



Peters & Schaal (2003). Reinforcement Learning for Humanoid Robotics, HUMANOIDS

Outline

1. Introduction
2. How can we develop suitable machine learning methods?
3. How can elementary behavior be learned with such machine learning methods?
4. Can complex skills be learned leveraging on elementary behaviors?
5. How can we adapt to humans and learn interaction?
6. Conclusion



Imitation Learning

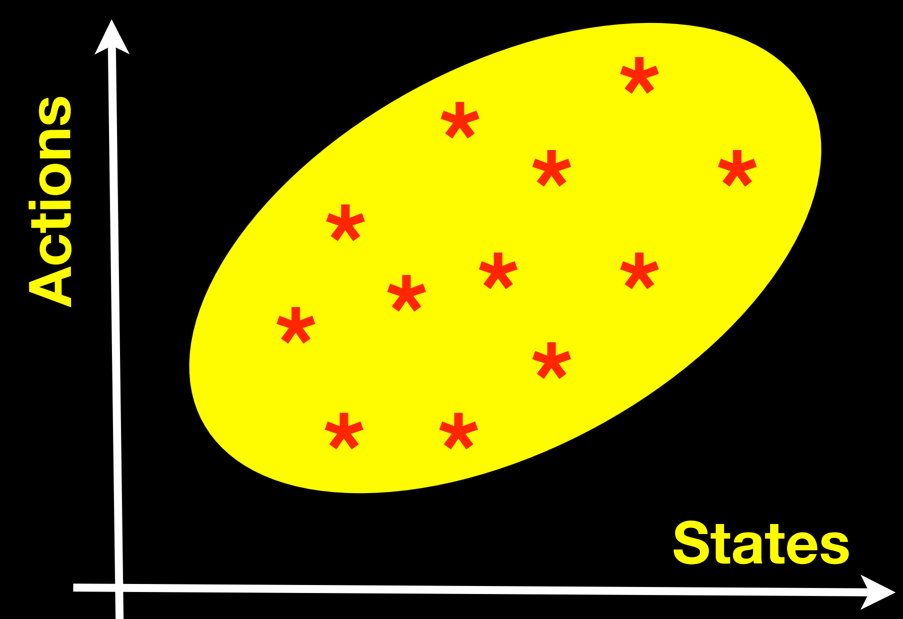
Given a path distribution, can we reproduce the policy?

- We need to measure similarity between distributions, e.g., using an *f*-measure as reward

$$r(\tau) = f\left(p_\theta(\tau), p(\tau)\right).$$

- Using $f(p,q) = \log(p/q)$ as *f*-measure, we obtain

Imitation



$$J(\pi) = \int_{\mathbb{T}} p_\theta(\tau) \log \frac{p_\theta(\tau)}{p(\tau)} d\tau = -D(p_\theta(\tau)||p(\tau))$$

Boularias, A. et al. (2011). Relative Entropy Inverse Reinforcement Learning, AISTATS 2011
Englert, P. et al. (2013). Probabilistic Model-based Imitation Learning, Adaptive Behavior

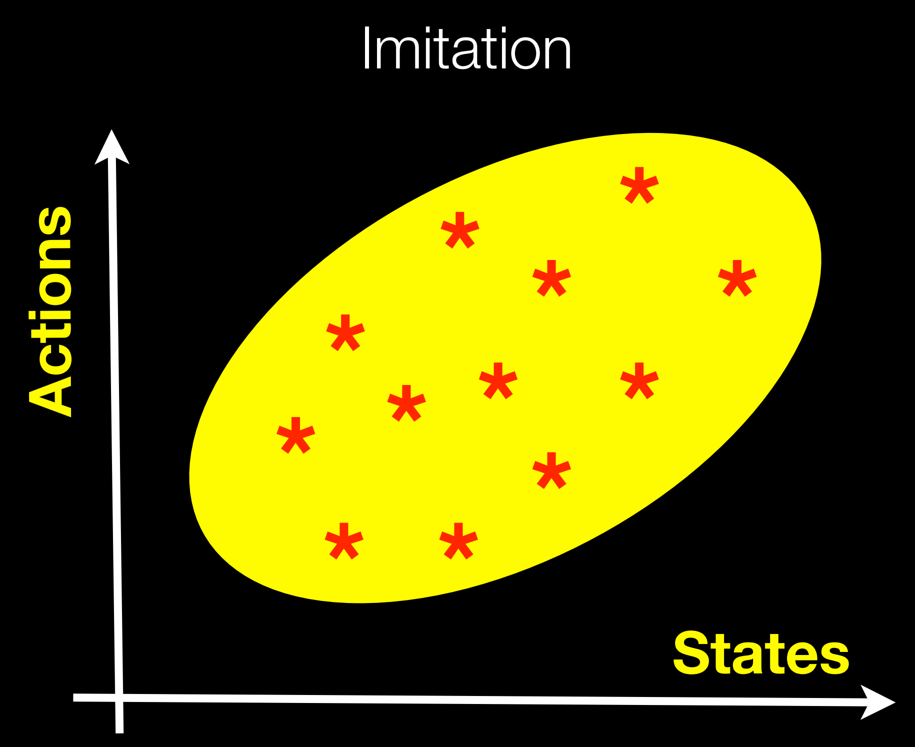
Imitation Learning

Given a path distribution, can we reproduce the policy?

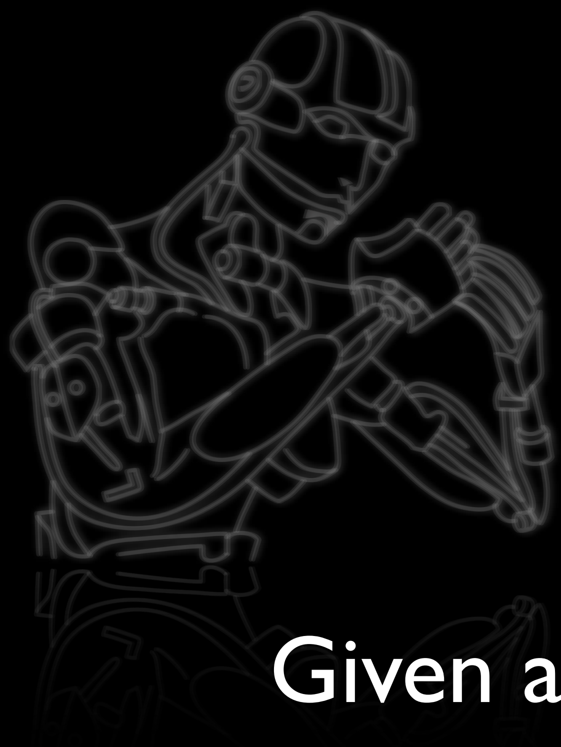
- match given path distribution $p(\tau)$ with a new one $p_\theta(\tau)$, i.e.,

$$D(p_{\boldsymbol{\theta}}(\boldsymbol{\tau})||p(\boldsymbol{\tau})) \to \min$$

- adapt the policy parameters θ
- possible model-free, purely sample-based (Boularias et al., 2011) and model-based (Englert et al.,2013)
- results in one-shot and expectation maximization algorithms



Boularias, A. et al. (2011). Relative Entropy Inverse Reinforcement Learning, AISTATS 2011
Englert, P. et al. (2013). Probabilistic Model-based Imitation Learning, Adaptive Behavior



Reinforcement Learning

Given a path distribution, can we find the optimal policy?

- *Goal*: maximize the return of the paths $r(\tau)$ generated by path distribution $p_\theta(\tau)$
- Optimization function is an ***arbitrary*** expected reward

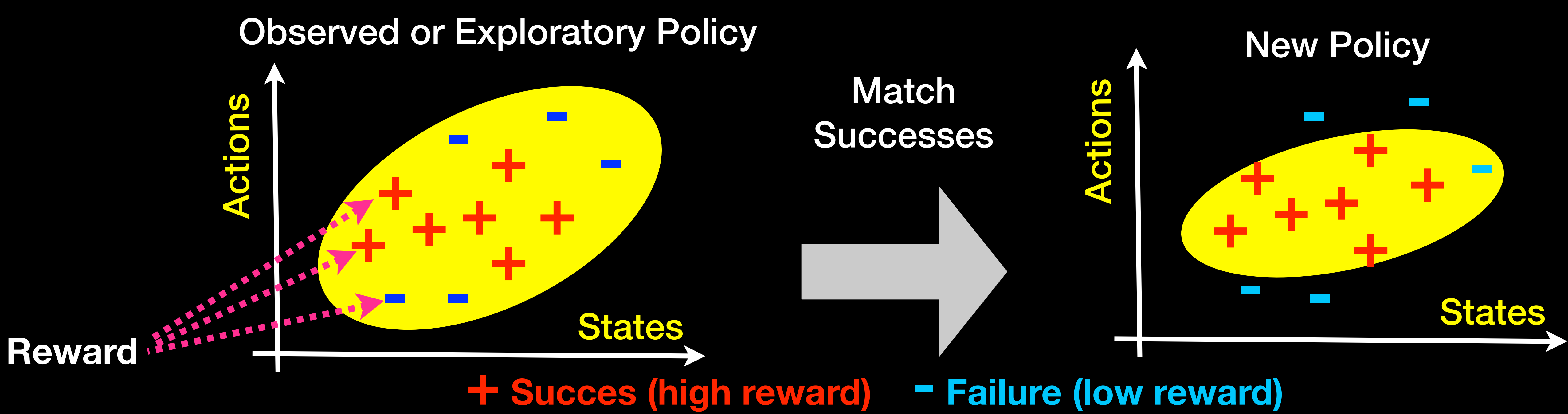
$$J(\boldsymbol{\theta}) = \int_{\mathbb{T}} p_{\boldsymbol{\theta}}(\boldsymbol{\tau}) r(\boldsymbol{\tau}) d\boldsymbol{\tau}$$

- This part usually results into a greedy, softmax updates or a \`vanilla' policy gradient algorithm...
- *Problem*: Small steps, optimization bias, results 'fragile'.

Success Matching

"When learning from a set of their own trials in iterated decision problems, humans attempt to match not the best taken action but the reward-weighted frequency of their actions and outcomes" (Arrow, 1958).

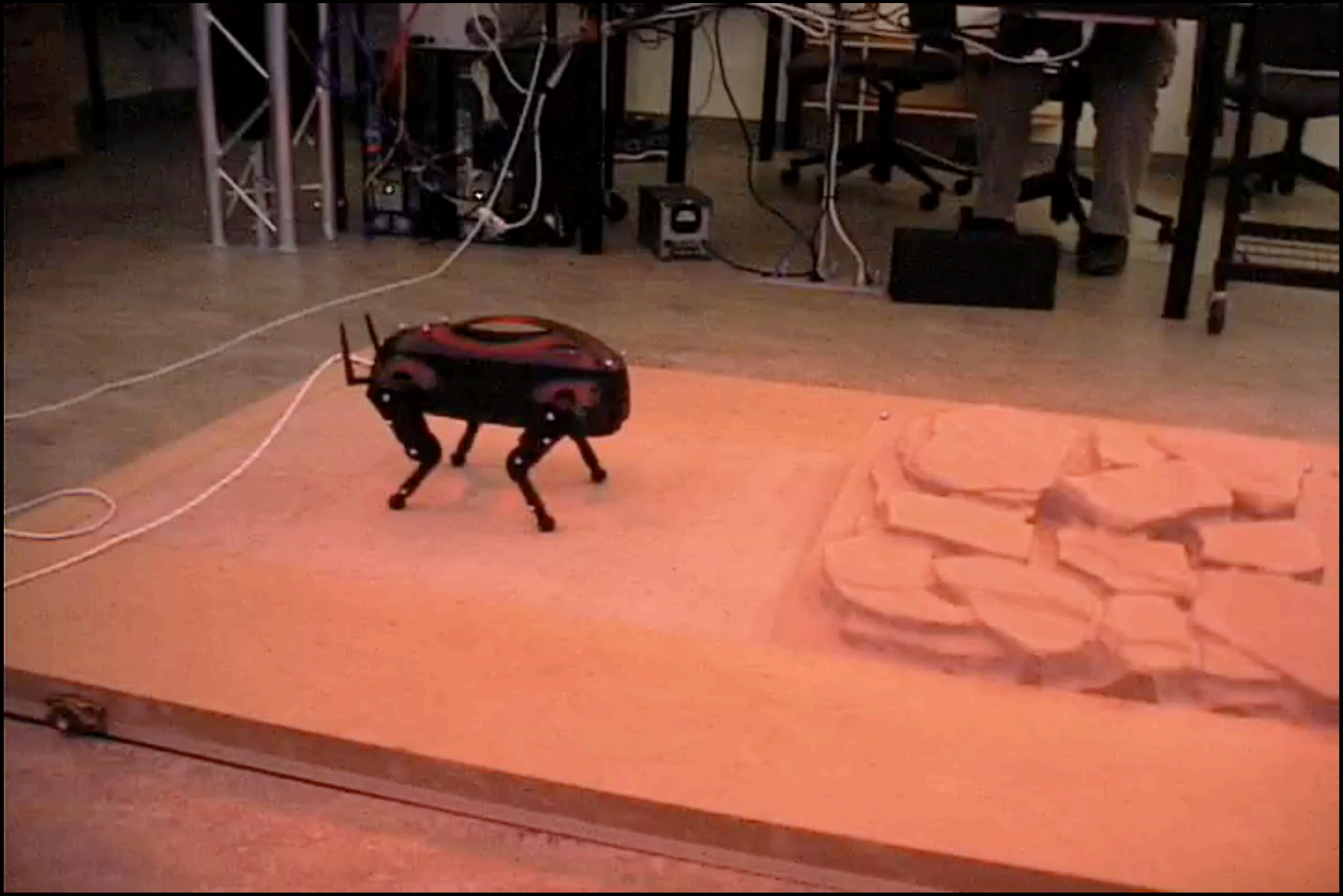
Can we create better policies by matching the reward-weighted previous policy ?



Many related frameworks, e.g., (Dayan&Hinton 1992;Andrews,'03;Attias,'04;Bagnell,'03;Toussaint,'06;...).

Illustrative Example

Foothold Selection



Match successful footholds!

Reinforcement Learning by Return-Weighted Imitation

Matching successful actions corresponds to minimizing the Kullback-Leibler 'distance'

$$D(p_{\boldsymbol{\theta}}(\boldsymbol{\tau})||r(\boldsymbol{\tau})p(\boldsymbol{\tau})) \rightarrow \min$$

For a Gaussian policy $\pi(\mathbf{u}|\mathbf{x}) = \mathcal{N}(\mathbf{u}|\phi(\mathbf{x})^T\boldsymbol{\theta}, \sigma^2\mathbf{I})$, we get the update rule

$$\boxed{\theta_{k+1} = (\boldsymbol{\Phi}^T\mathbf{R}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\mathbf{R}\mathbf{U}}$$

New Policy Parameters   Features   Returns   Actions

➡Reduces Reinforcement Learning onto Return-Weighted Regression!

Peters & Schaal (2007). Policy Learning for Motor Skills, International Conference on Machine Learning (ICML)
Kober & Peters (2009). Policy Search for Motor Primitives in Robotics, Advances in Neural Information Processing Systems (NIPS)

Resulting EM-like Policy Search Methods

This insight has allowed us to derive a series of new reinforcement learning methods:

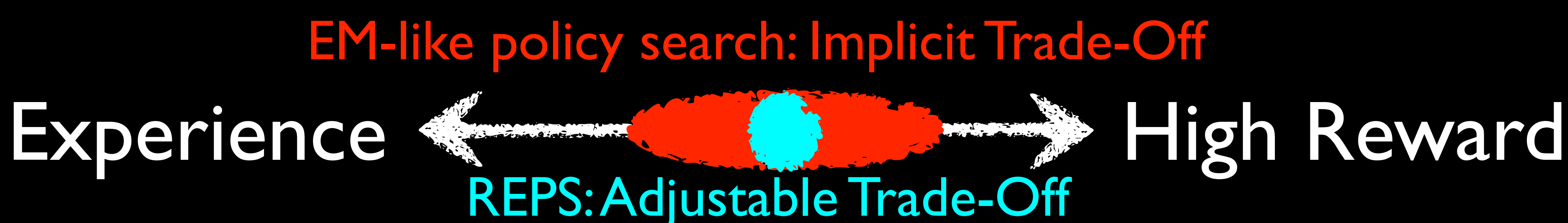
- Reward-Weighted Regression (Peters & Schaal, ICML 2007)
- PoWER (Kober & Peters, NIPS 2009)
- LaWER (Neumann & Peters, NIPS 2009+ICML 2009)
- CrKR (Kober, Oztop & Peters, R:SS 2010; IJCAI 2011)

All of these approaches are extensions of this idea.

Experience vs Reward Trade-Off

Requirements:

- Uses experience and initial demonstrations
- Aims at high reward but only "updates to a safe distance"
- EM-like policy search does this only implicitly



Peters, Muelling, Altun (2010). Relative Entropy Policy Search, AAAI
Lioutikov et al. (2014). Generalizing Movements with Information Theoretic Stochastic Optimal Control, JAIS
Daniel, Neumann & Peters (conditionally accepted). Hierarchical Relative Entropy Policy Search, JMLR

More focussed trade-off?

Relative Entropy Policy Search (REPS)

I. Maximize expect reward

$$\max_\theta J(\boldsymbol{\theta}) = \int_{\mathbb{T}} p_{\boldsymbol{\theta}}(\boldsymbol{\tau}) r(\boldsymbol{\tau}) d\boldsymbol{\tau}$$

II. Ensure path distribution remains a probability distribution

$$s.t. \quad \int_{\mathbb{T}} p_\theta(\tau) d\tau = 1 \qquad\qquad p_\theta(\tau) \geq 0$$

III. Trade off/limit **information loss** to past trial or trials

$$\epsilon \geq \int_{\mathbb{T}} p_\theta(\tau) \log \frac{p_\theta(\tau)}{p(\tau)} d\tau$$

Variations of this program yield analytic solutions for the policy!

Peters, Muelling, Altun (2010). Relative Entropy Policy Search, AAAI
Lioutikov et al. (2014). Generalizing Movements with Information Theoretic Stochastic Optimal Control, JAIS
Daniel, Neumann & Peters (conditionally accepted). Hierarchical Relative Entropy Policy Search, JMLR

Relative Entropy Policy Search

... is currently our favorite policy search method!

... results in an analytic solution which resembles a reward-weighted method with a reward transformation.

... explicitly trades experience against reward maximization.

... results in very efficient exploration.

... can be kernelized well (van Hoof et al. 2015, Learning of Non-Parametric Control Policies with High-Dimensional State Features, AISTATS)

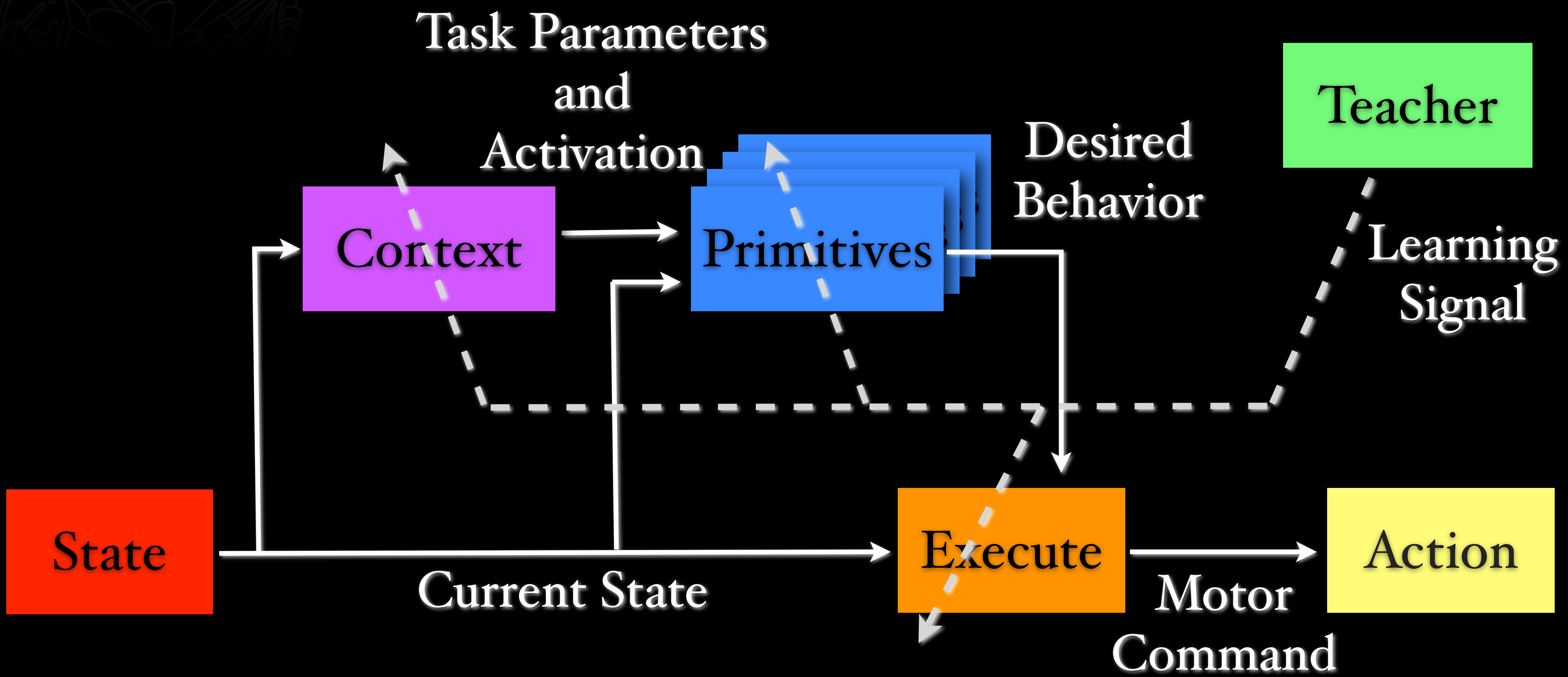
... has been extended with quite some success by Levine & Abbeel (NIPS 2013/4, ICML 2014).

Peters, Muelling, Altun (2010). Relative Entropy Policy Search, AAAI
Lioutikov et al. (2014). Generalizing Movements with Information Theoretic Stochastic Optimal Control, JAIS
Daniel, Neumann & Peters (conditionally accepted). Hierarchical Relative Entropy Policy Search, JMLR

Outline

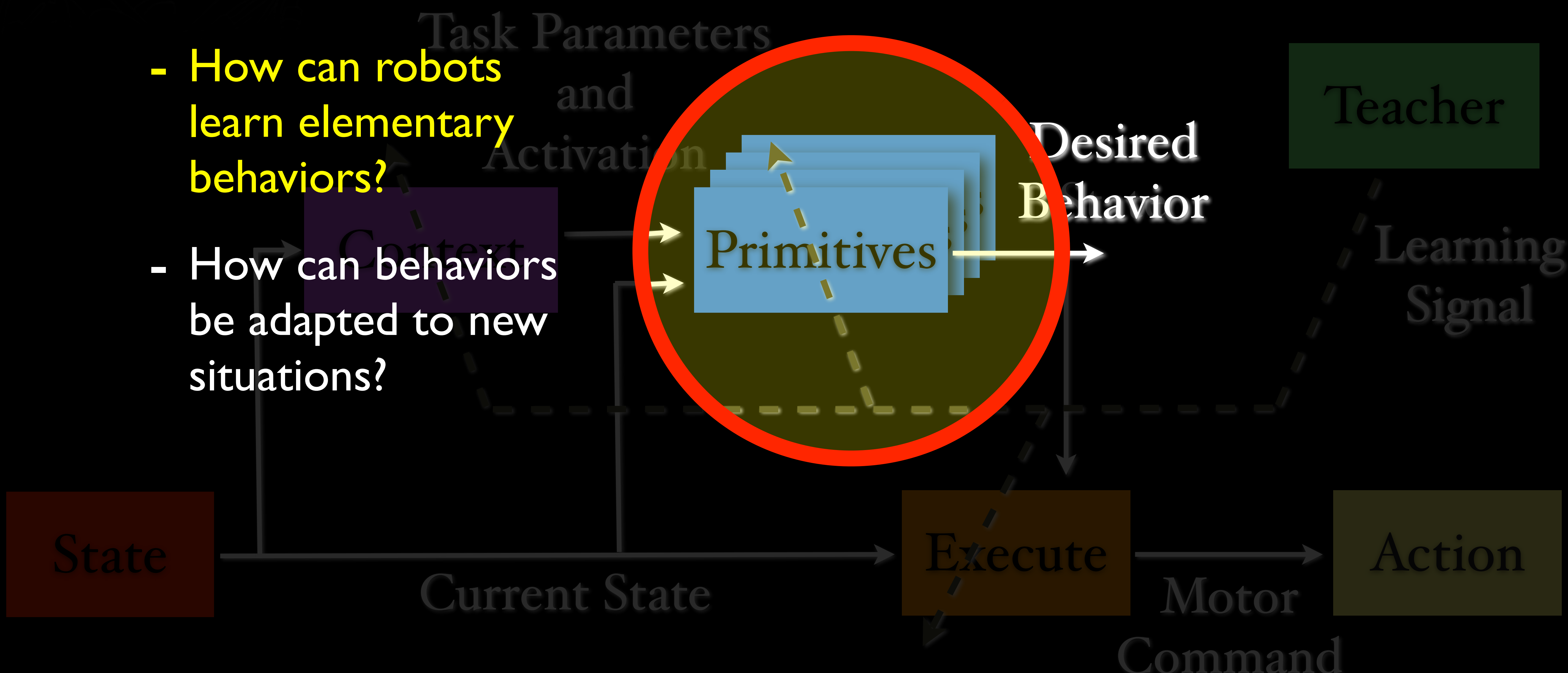
1. Introduction
2. How can we develop suitable machine learning methods?
3. How can elementary behavior be learned with such machine learning methods?
4. Can complex skills be learned leveraging on elementary behaviors?
5. How can we adapt to humans and learn interaction?
6. Conclusion

A Blue Print for Skill Learning?



Outline

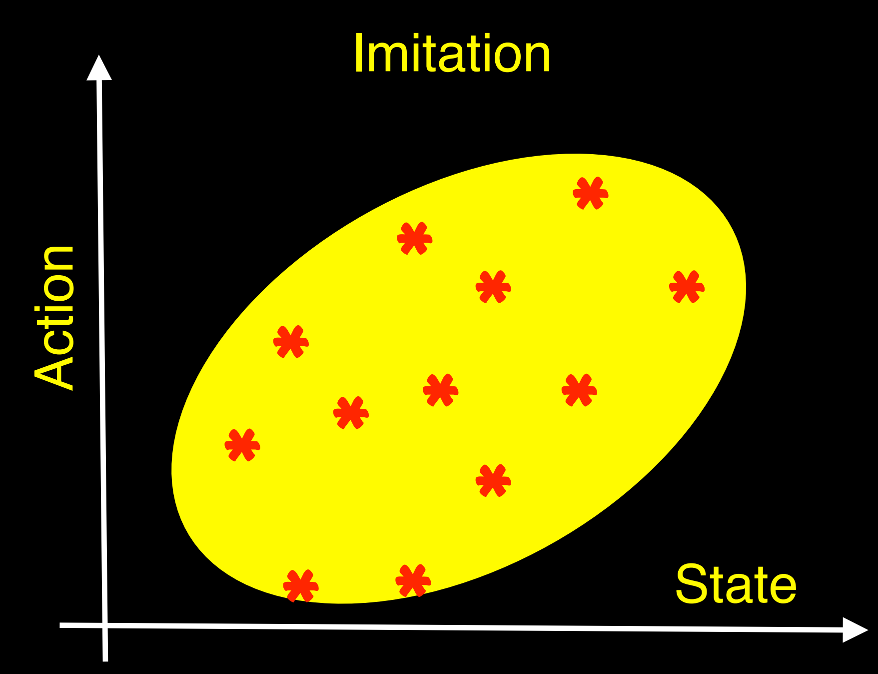
- How can robots learn elementary behaviors?
- How can behaviors be adapted to new situations?



Acquisition by Imitation

Teacher shows the task and the student reproduces it.

- maximize similarity

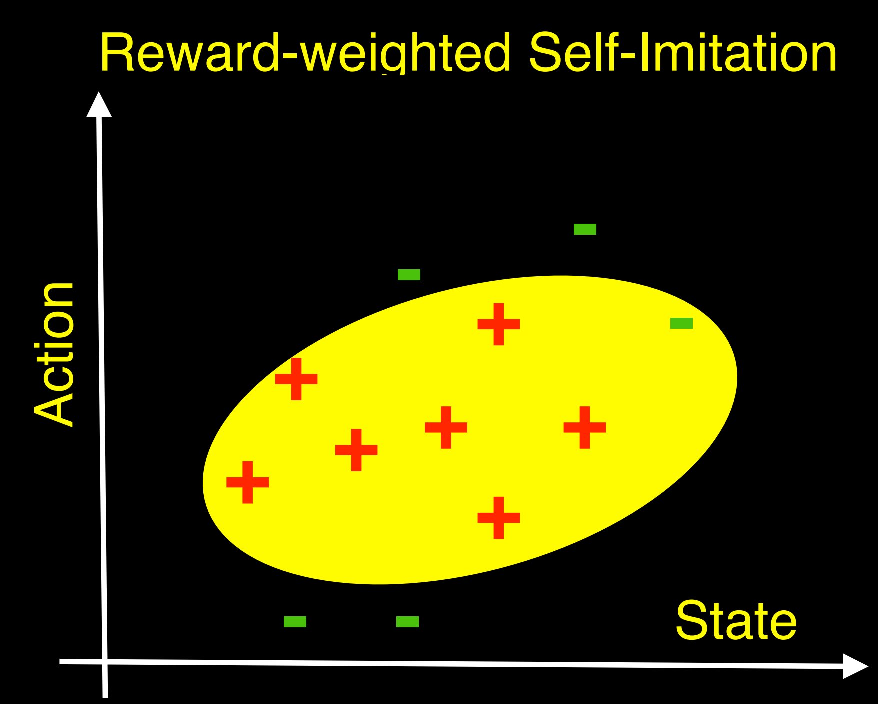


Kober & Peters (2009). Learning Motor Primitives, ICRA

Self-Improvement by Reinforcement Learning

Student improves by reproducing his successful trials.

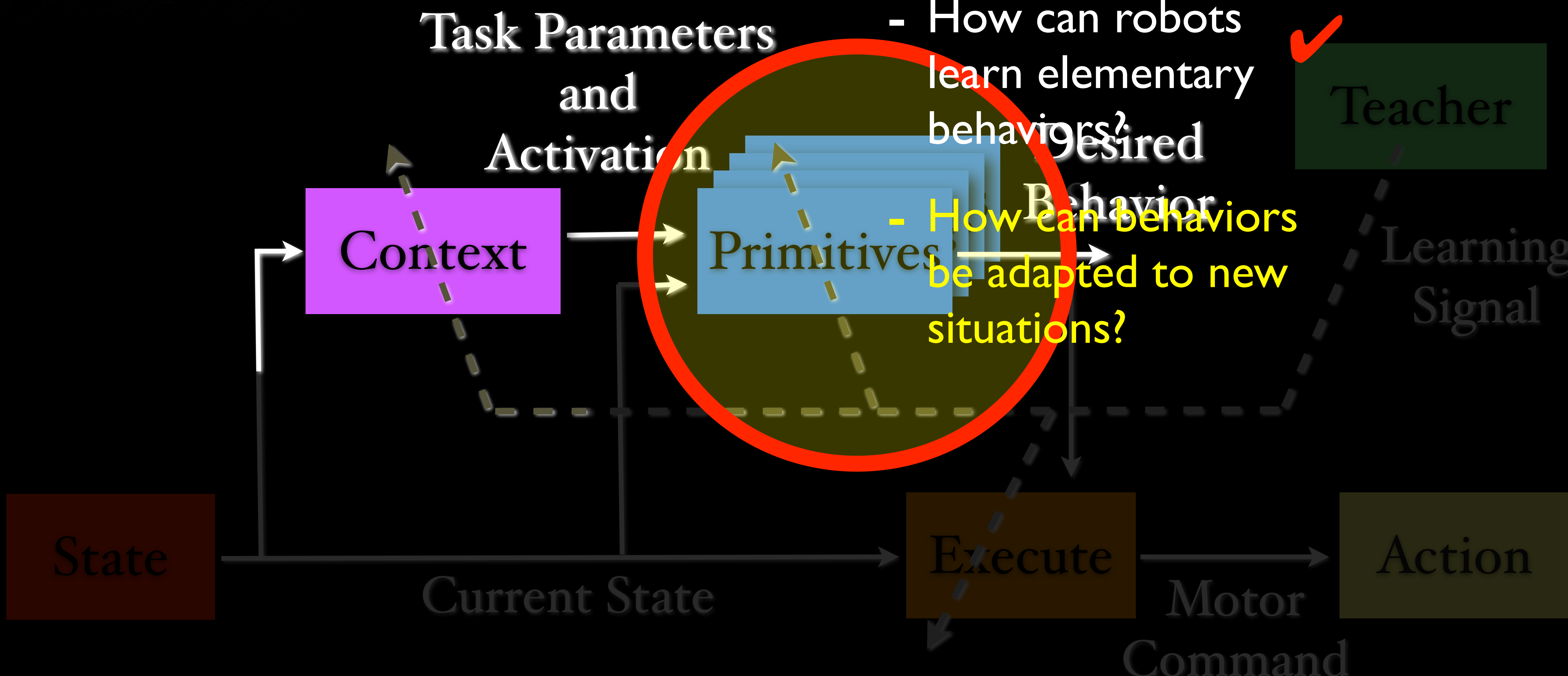
- maximize reward-weighted similarity



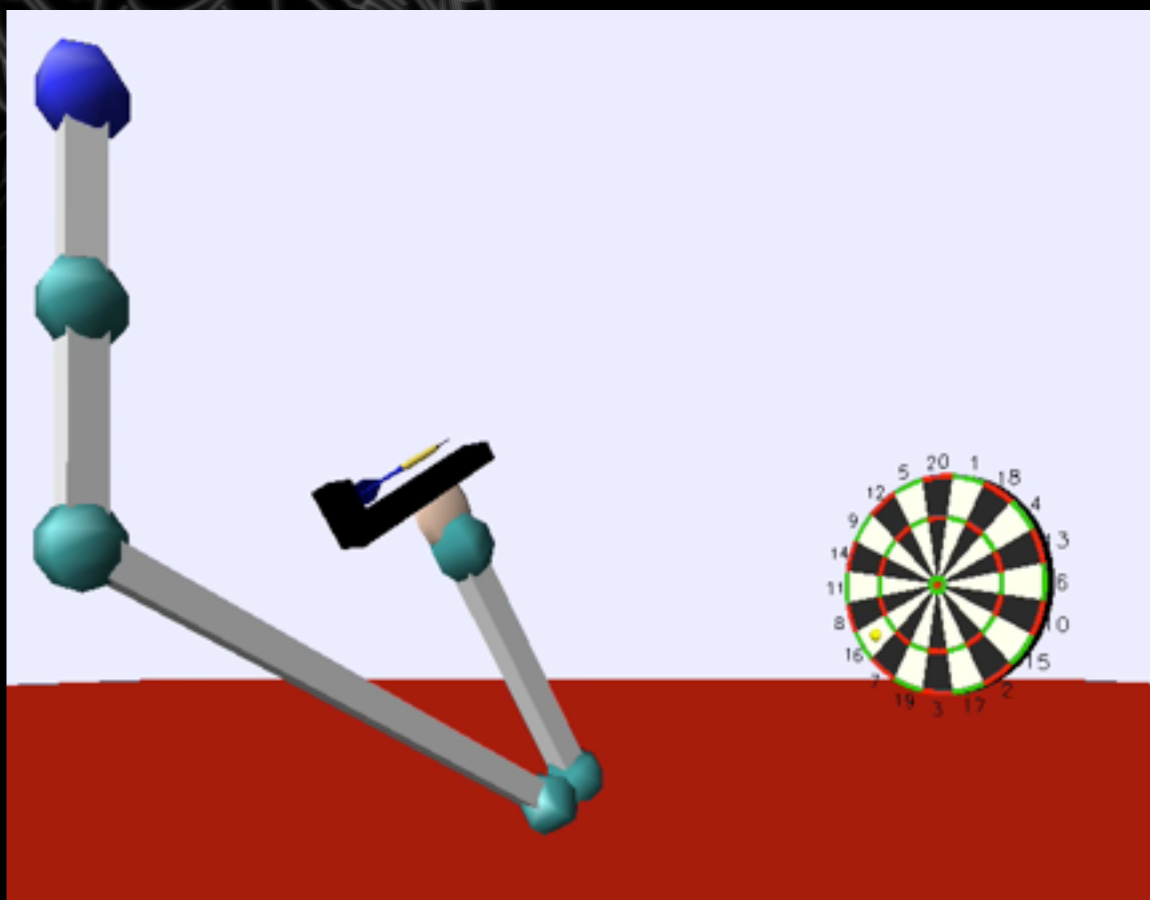
Kober & Peters (2009). Policy Search for Motor Primitives in Robotics, NIPS

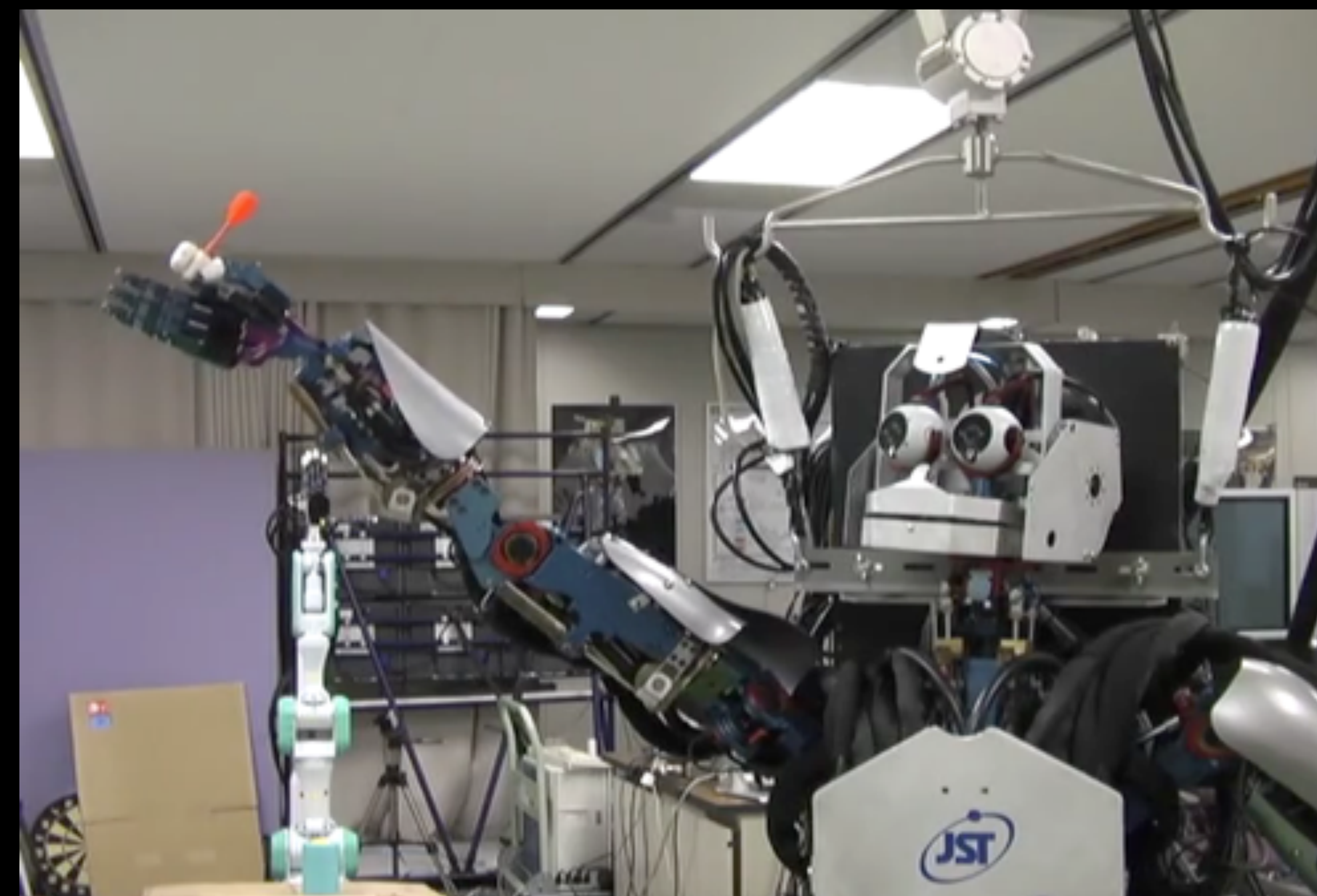
Outline

- How can robots learn elementary behaviors? ✔
- How can behaviors be adapted to new situations?



Task Context: Goal Learning





Adjusting Motor Primitives through their Hyperparameters:

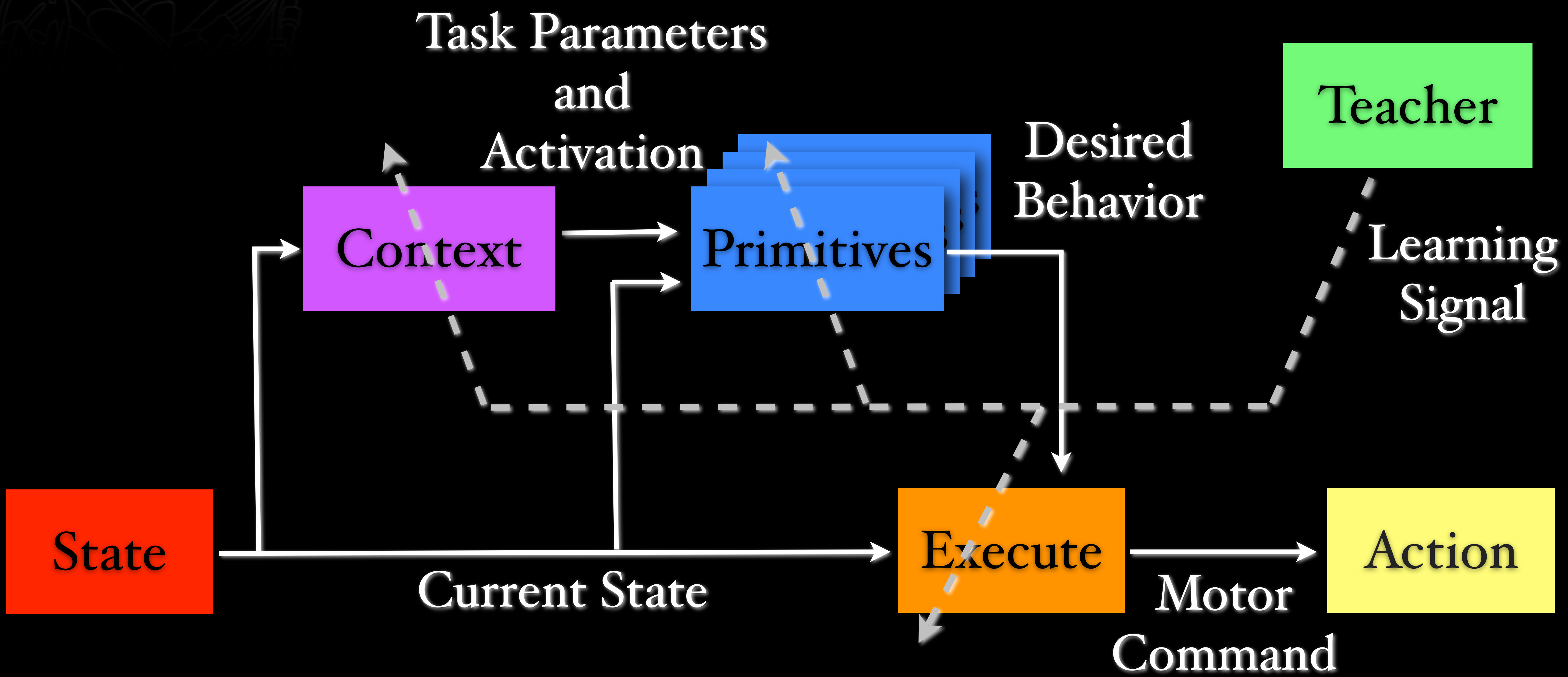
1. learn a single motor primitive using imitation and reinforcement learning
2. learn policies for the goal parameter and timing parameters by reinforcement learning

Kober, Oztop & Peters (2012). Goal Learning for Motor Primitives, Autonomous Robots

Outline

1. Introduction
2. How can we develop suitable machine learning methods?
3. How can elementary behavior be learned with such machine learning methods?
4. Can complex skills be learned leveraging on elementary behaviors?
5. How can we adapt to humans and learn interaction?
6. Conclusion

Composition by
Selection, Superposition & Sequencing



Let us put all these elements together!

Selection and Superposition of Motor Primitives

"Naïve" Approach:

1. Learn several motor primitives by imitation.
2. Self-Improvement on repetitive targets by reinforcement learning.
3. Generalize among targets and hitting points.

Mülling, K.; Kober, J.; Kroemer, O.; Peters, J. (2013). Learning to Select and Generalize Striking Movements in Robot Table Tennis, International Journal on Robotics Research.

Demonstrations

Demonstrations with Kinesthetic Teach-In

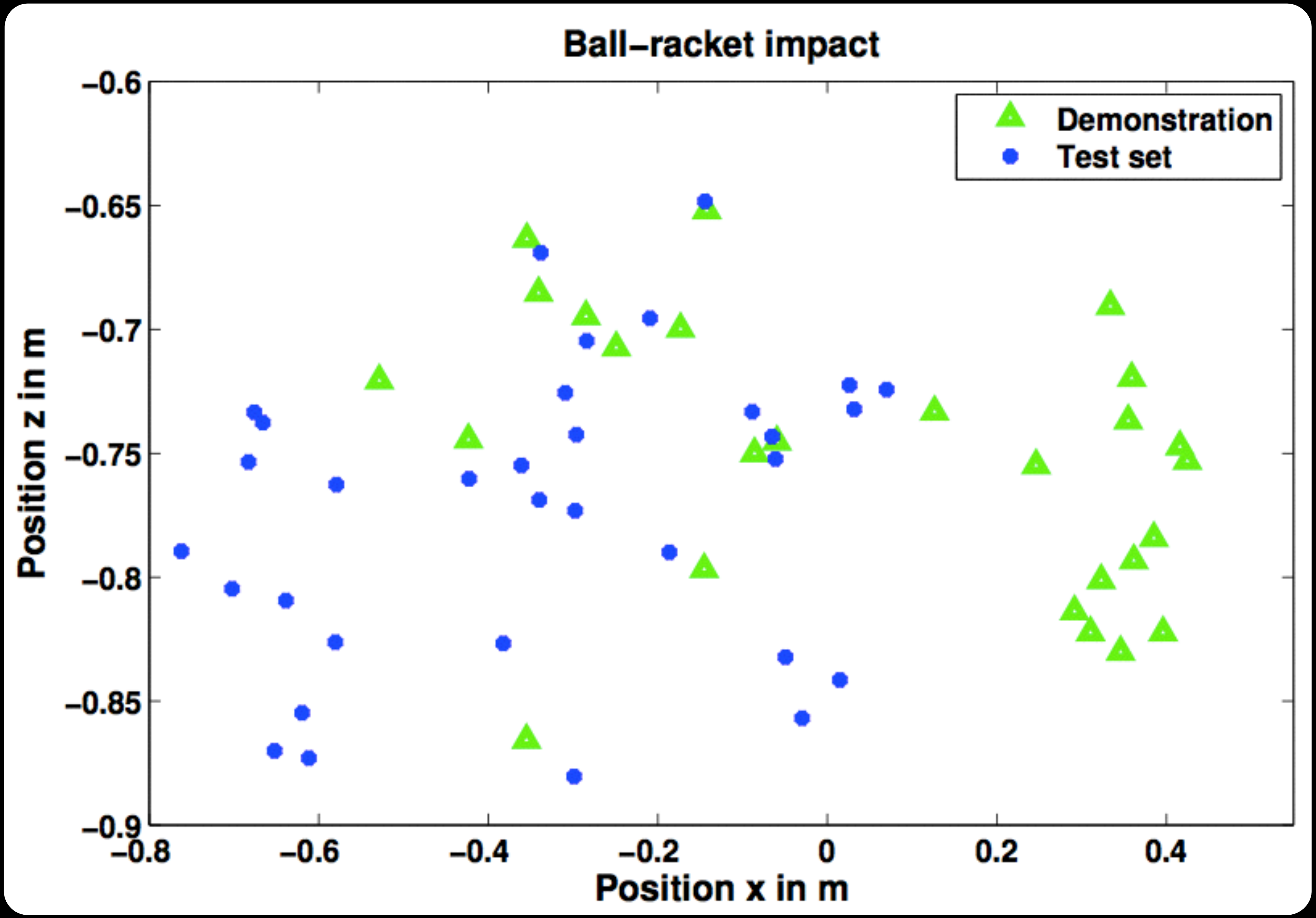
Mülling, K.; Kober, J.; Kroemer, O.; Peters, J. (2013). Learning to Select and Generalize Striking Movements in Robot Table Tennis, International Journal on Robotics Research.

Select & Generalize

From Imitation Learning we obtain 25 Movement Primitives

Mülling, K.; Kober, J.; Kroemer, O.; Peters, J. (2013). Learning to Select and Generalize Striking Movements in Robot Table Tennis, International Journal on Robotics Research.

Covered Situations



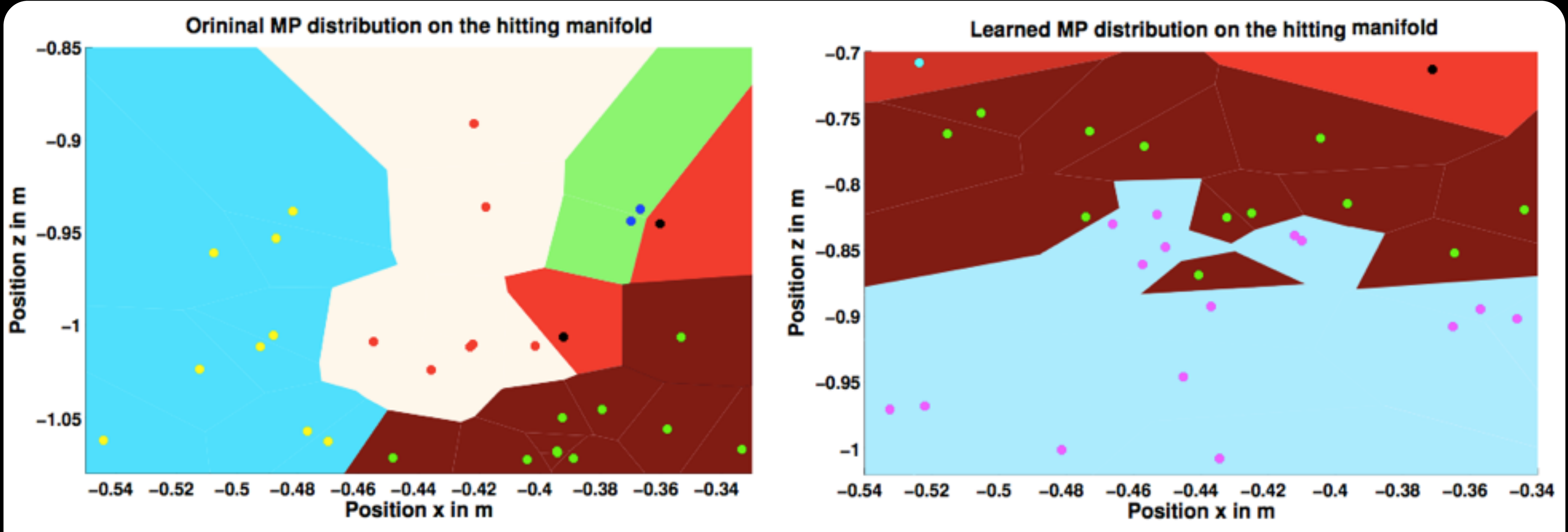
Mülling, K.; Kober, J.; Kroemer, O.; Peters, J. (2013). Learning to Select and Generalize Striking Movements in Robot Table Tennis, International Journal on Robotics Research.

Self-Improvement

Training a Hitting Region
with an Initial Success Rate
of 0%

Mülling, K.; Kober, J.; Kroemer, O.; Peters, J. (2013). Learning to Select and Generalize Striking Movements in Robot Table Tennis, International Journal on Robotics Research.

Changed Primitive Activation



(a) Before training. (b) After training.

Mülling, K.; Kober, J.; Kroemer, O.; Peters, J. (2013). Learning to Select and Generalize Striking Movements in Robot Table Tennis, International Journal on Robotics Research.

Current Gameplay

Final Challenge:
Match against a Human

Mülling, K.; Kober, J.; Kroemer, O.; Peters, J. (2013). Learning to Select and Generalize Striking Movements in Robot Table Tennis, International Journal on Robotics Research.

Selection and Superposition of Motor Primitives

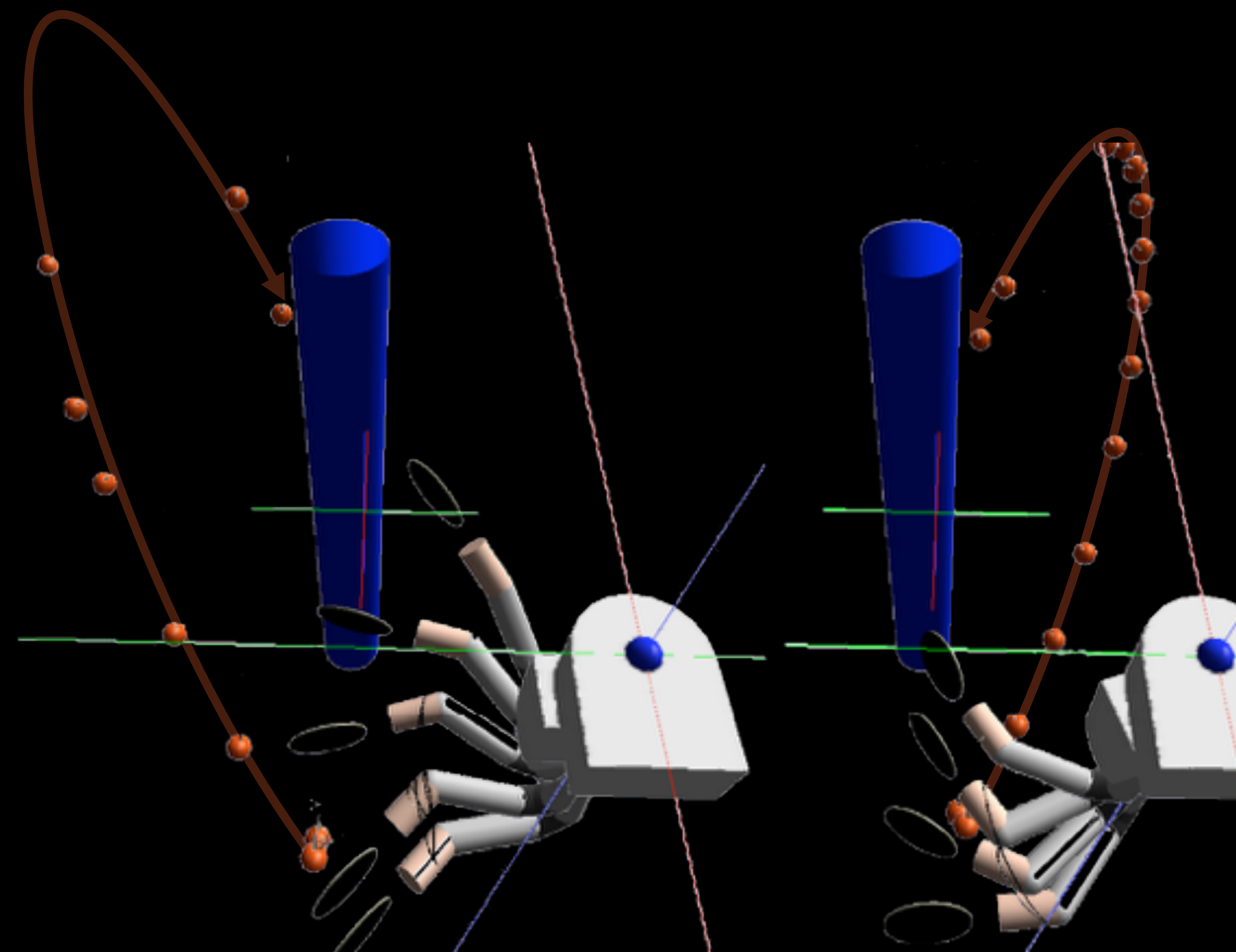
Problems with the "Naïve" Approach?

1. Weighted superposition works well in Robot Table Tennis:

   - convex combinations possible

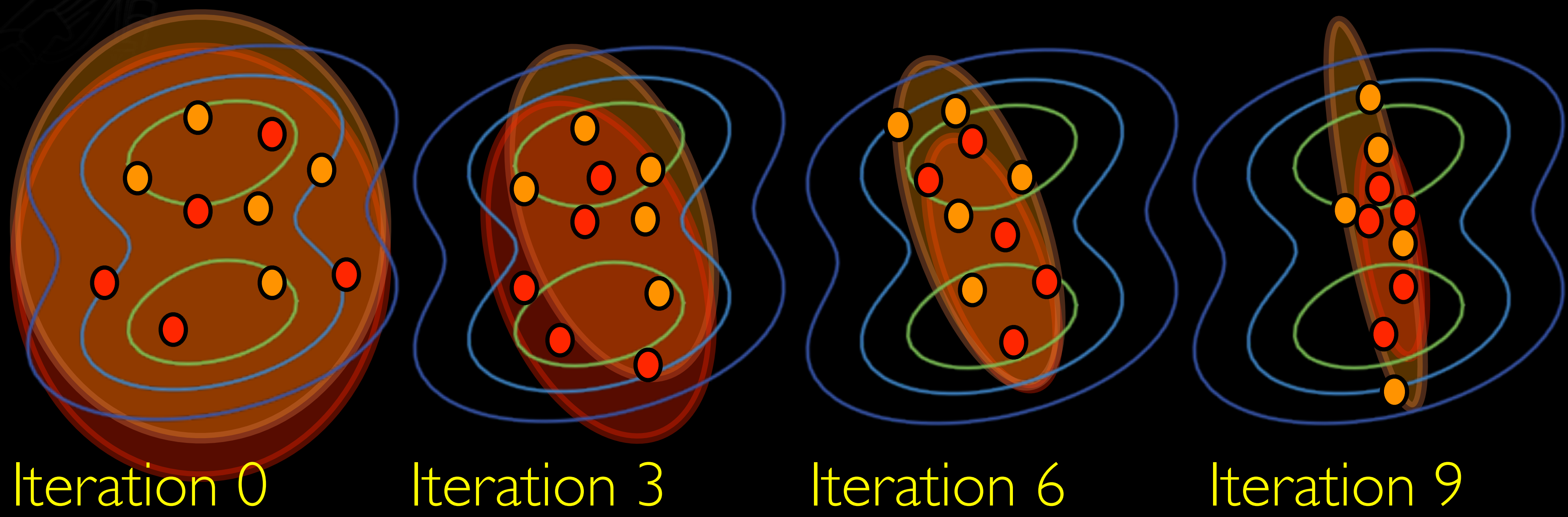
   - few primitives are equally responsible for an incoming ball

2. It fails if **selection** is needed!



Daniel, Neumann & Peters (conditionally accepted). Hierarchical Relative Entropy Policy Search, JMLR

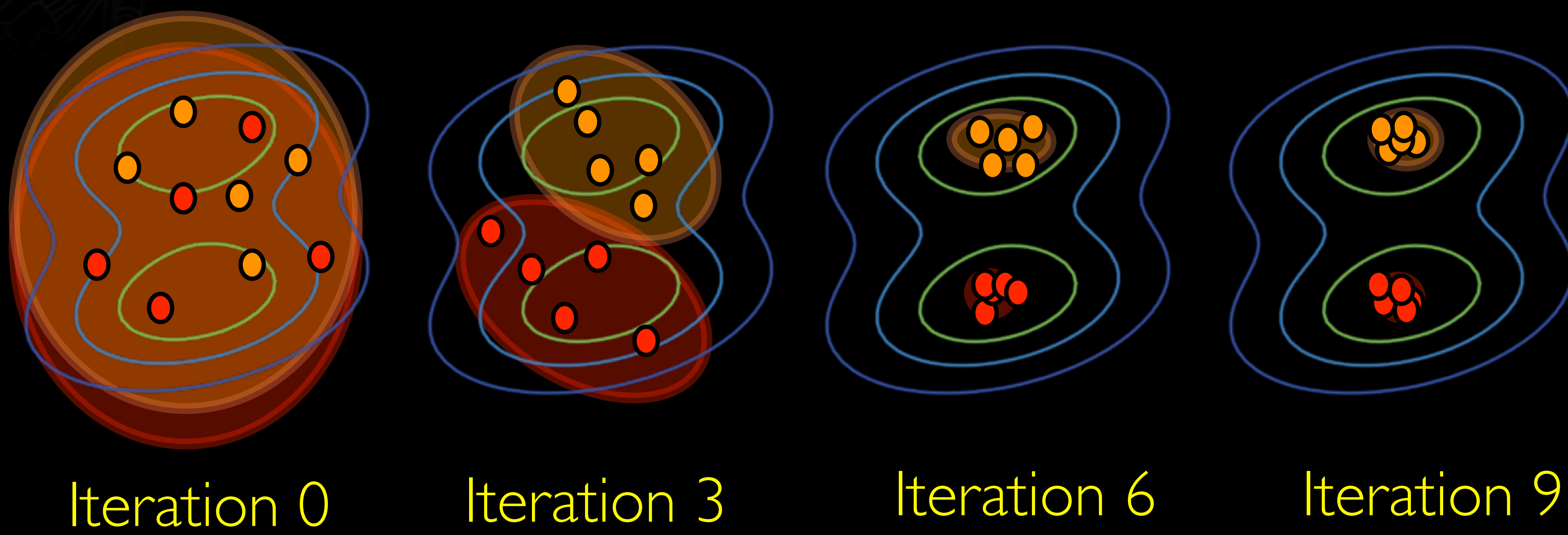
Problems with the Naïve Approach



If all primitives are equally responsible, we can represent versatile behavior but it will never be parsimonious.

Daniel, Neumann & Peters (conditionally accepted). Hierarchical Relative Entropy Policy Search, JMLR

Localized behavior can be learned efficiently!



We can reduce to the number of needed primitives!

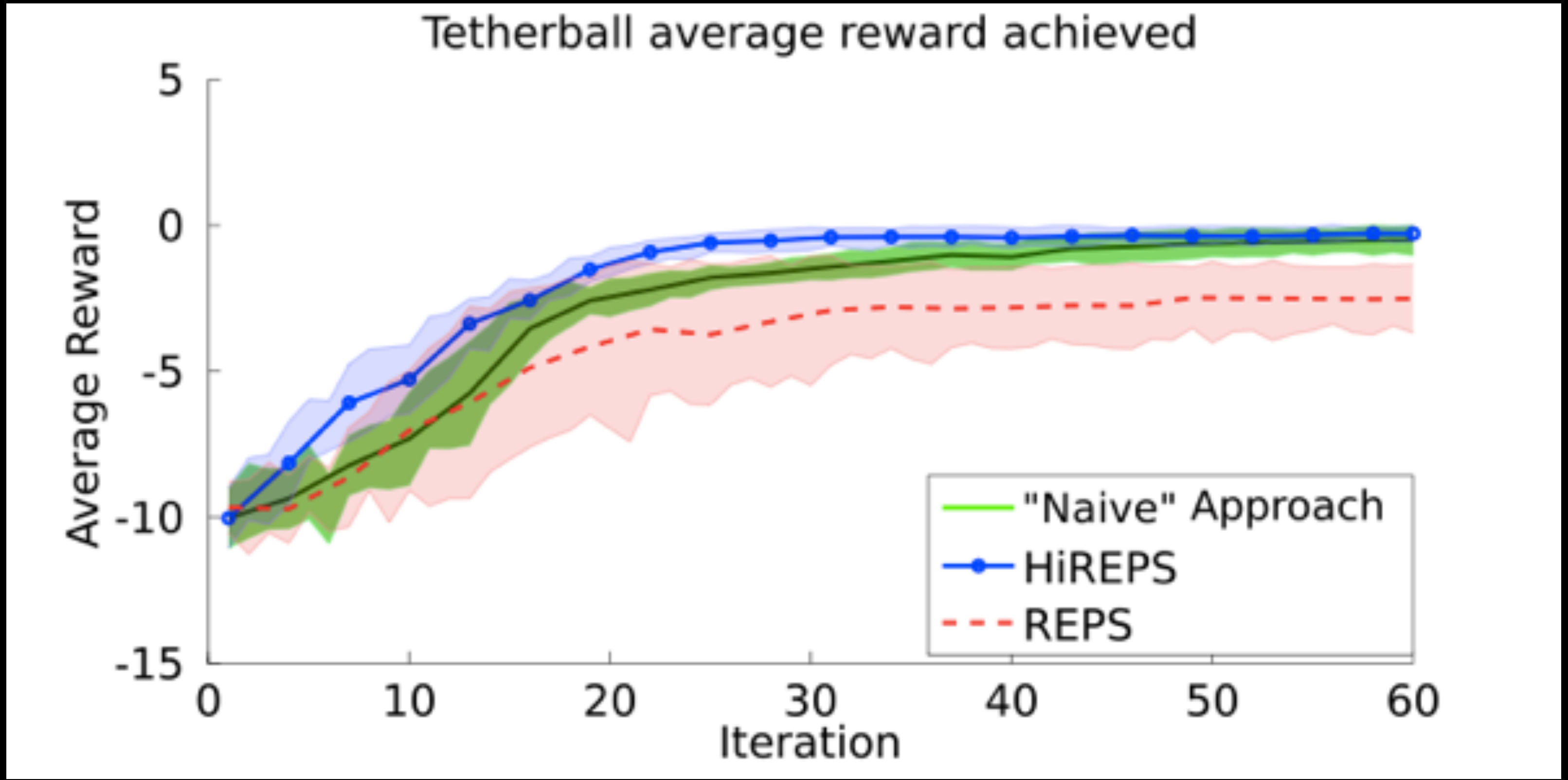
$$\kappa \geq \mathbb{E}_{\boldsymbol{s},\boldsymbol{a}}\left[\sum_{o} -p(o|\boldsymbol{s},\boldsymbol{a}) \log p(o|\boldsymbol{s},\boldsymbol{a})\right]$$

Force the primitives to limited responsibility

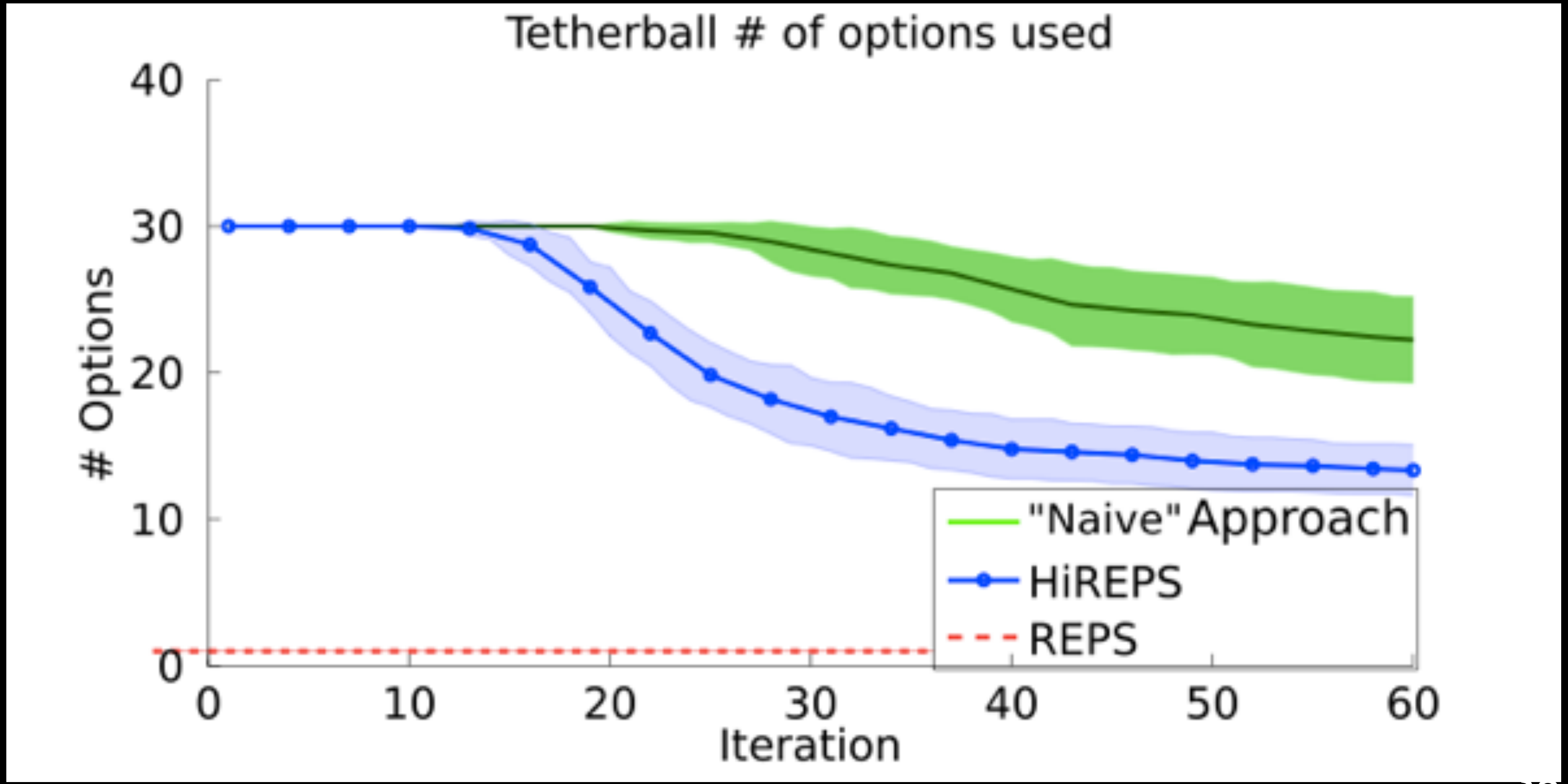
Daniel, Neumann & Peters (conditionally accepted). Hierarchical Relative Entropy Policy Search, JMLR

Localized behavior can be learned efficiently!

Good performance

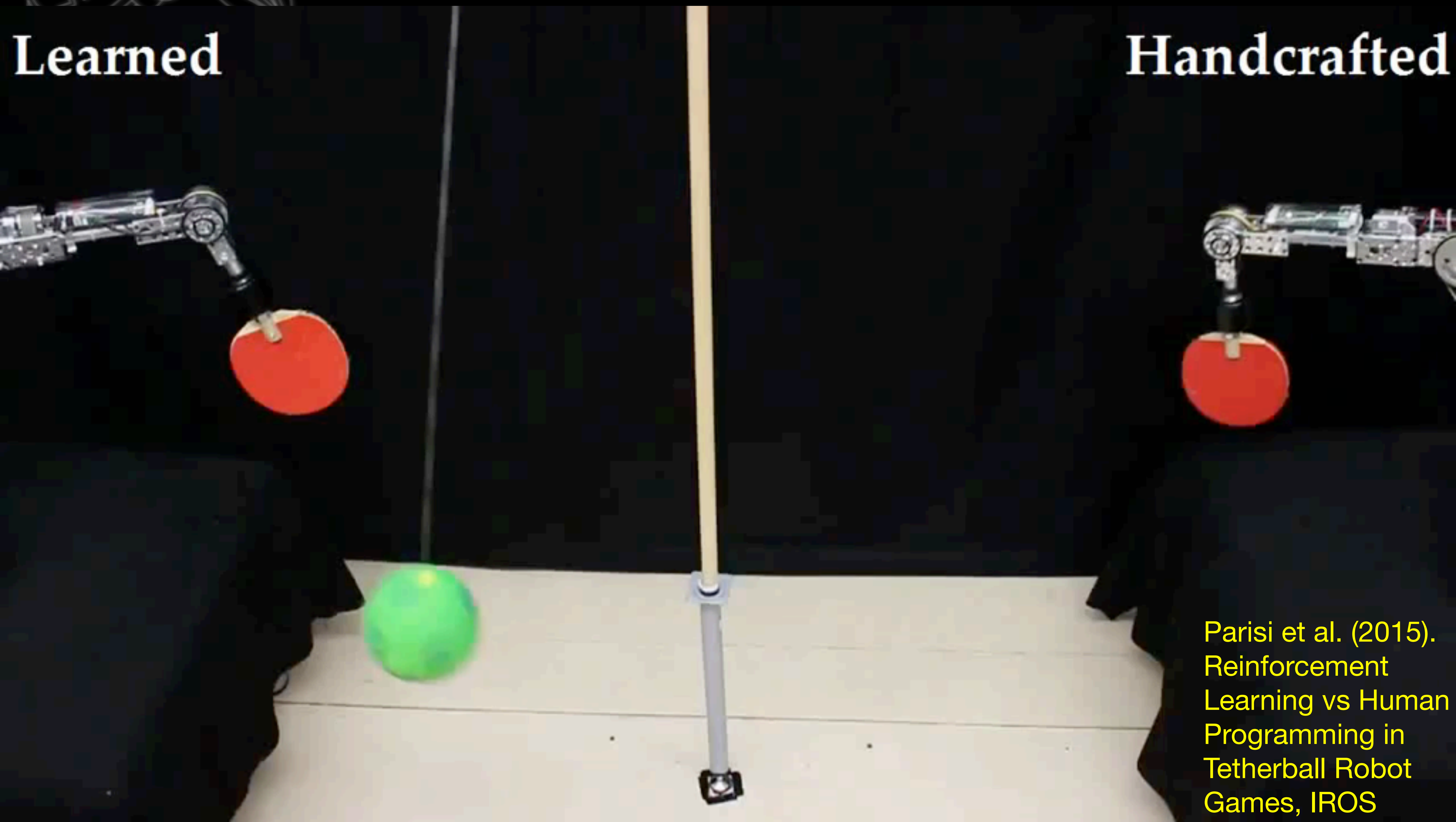


Fast reduction in the number of primitives



Daniel, Neumann & Peters (conditionally accepted). Hierarchical Relative Entropy Policy Search, JMLR

What's next? The Reinforcement Learning Games!



Parisi et al. (2015). Reinforcement Learning vs Human Programming in Tetherball Robot Games, IROS

Outline

1. Introduction
2. How can we develop suitable machine learning methods?
3. How can elementary behavior be learned with such machine learning methods?
4. Can complex skills be learned leveraging on elementary behaviors?
5. How can we adapt to humans and learn interaction?
6. Conclusion

Problems in Robot Table Tennis

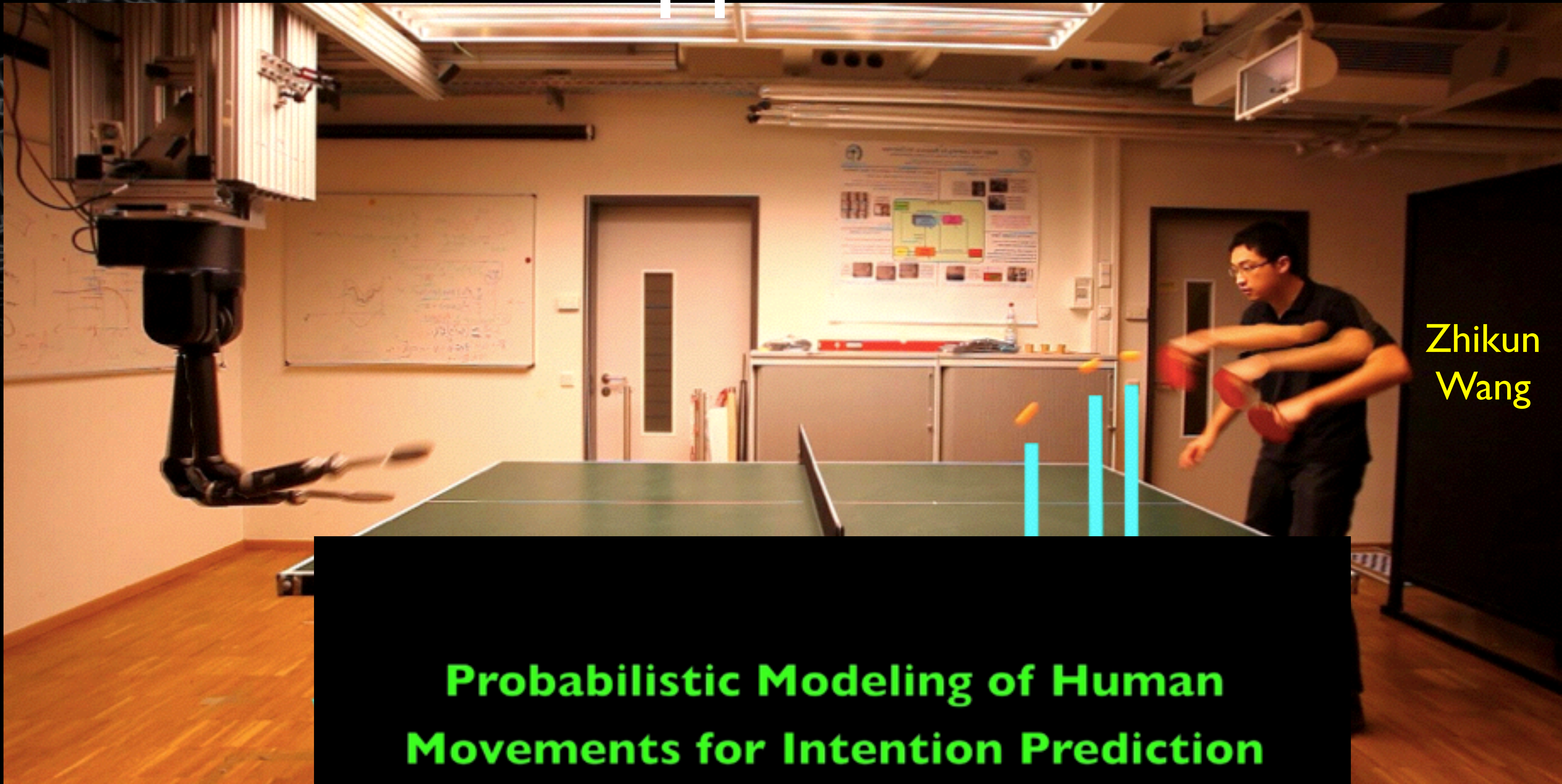
Problem I: Workspace is too limited.

Problem II: Arm accelerations are too low.

Problem III: Limited reaction time.

Problem III: Reaction Time

Reactive Opponent Prediction



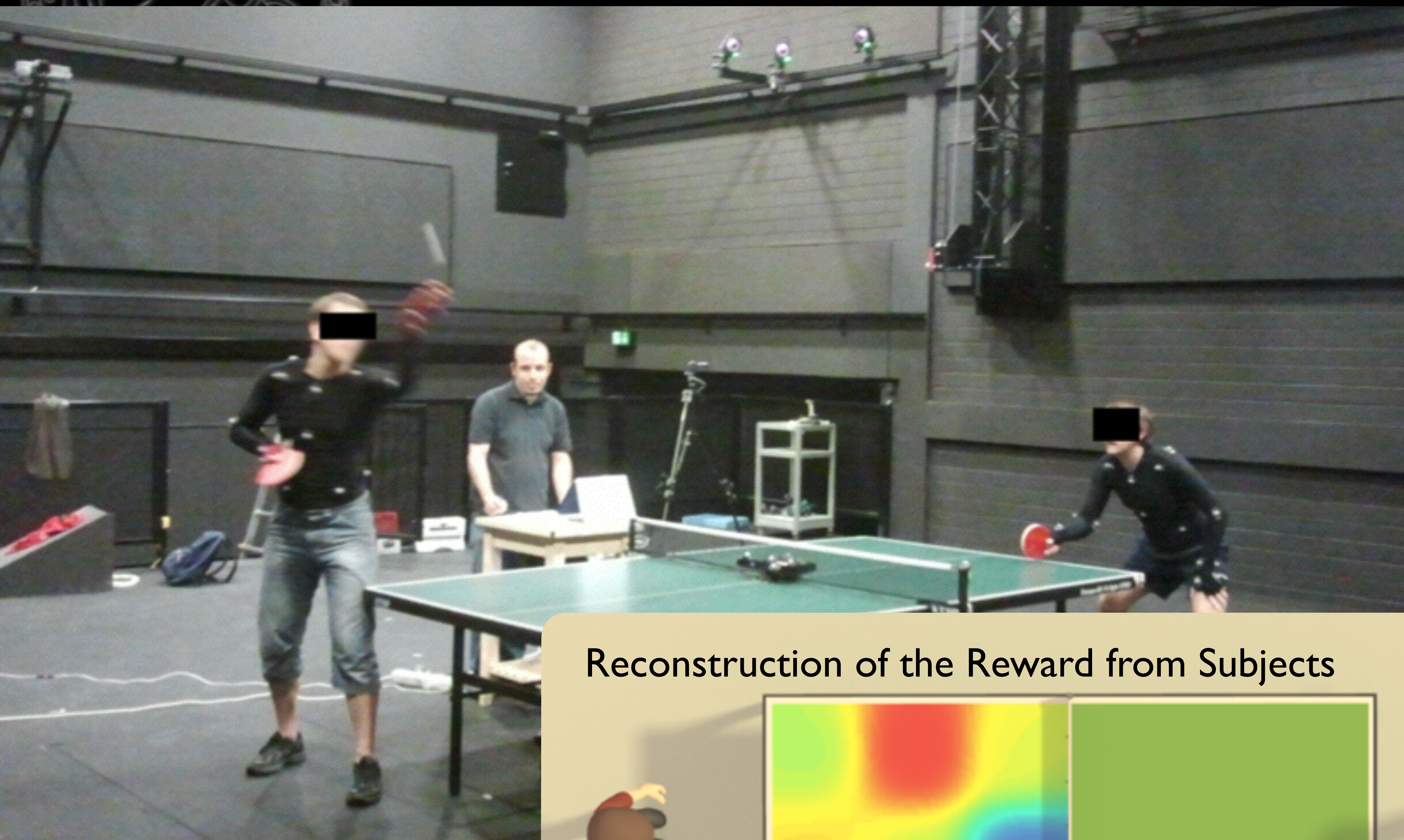
Probabilistic Modeling of Human Movements for Intention Prediction

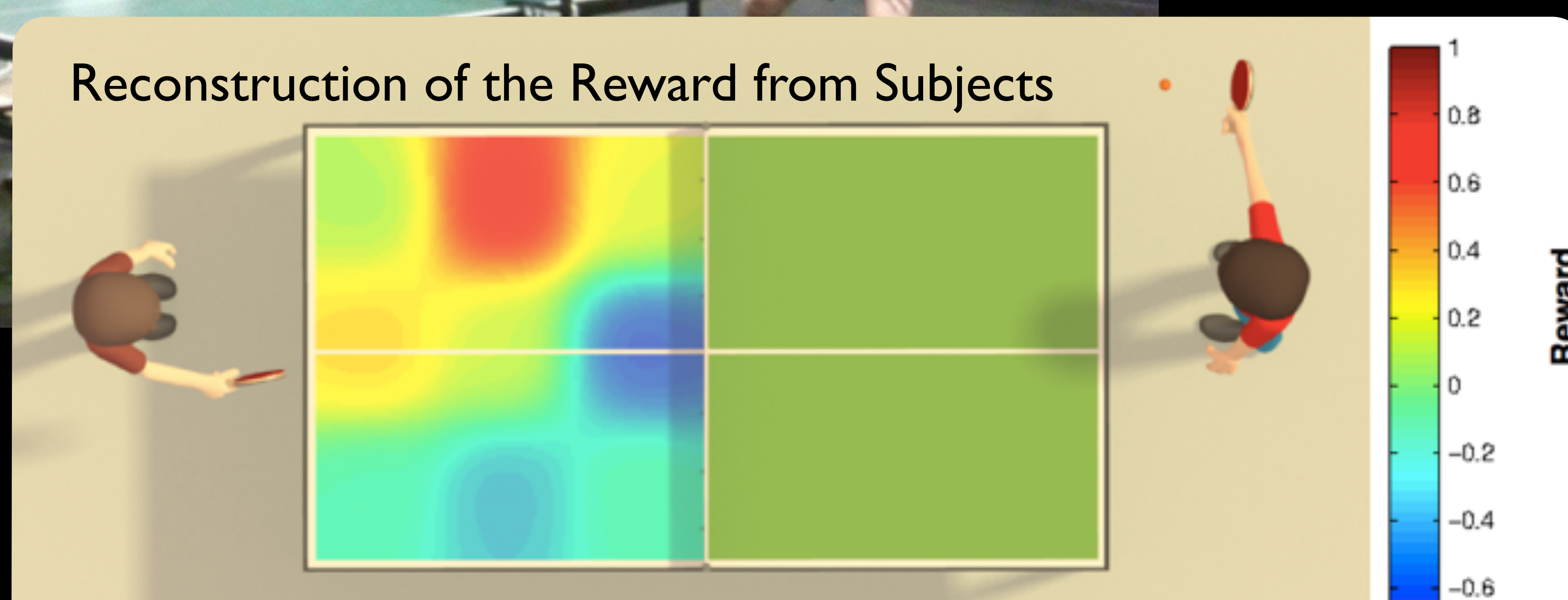
prototype system

Z. Wang, K. Muelling, M. Deisenroth, B. Schoelkopf, and J. Peters

Wang, Z. et al. Probabilistic Modeling of Human Movements for Intention Inference, R:SS 2012, IJRR 2013

Extracting Strategies from Game Play

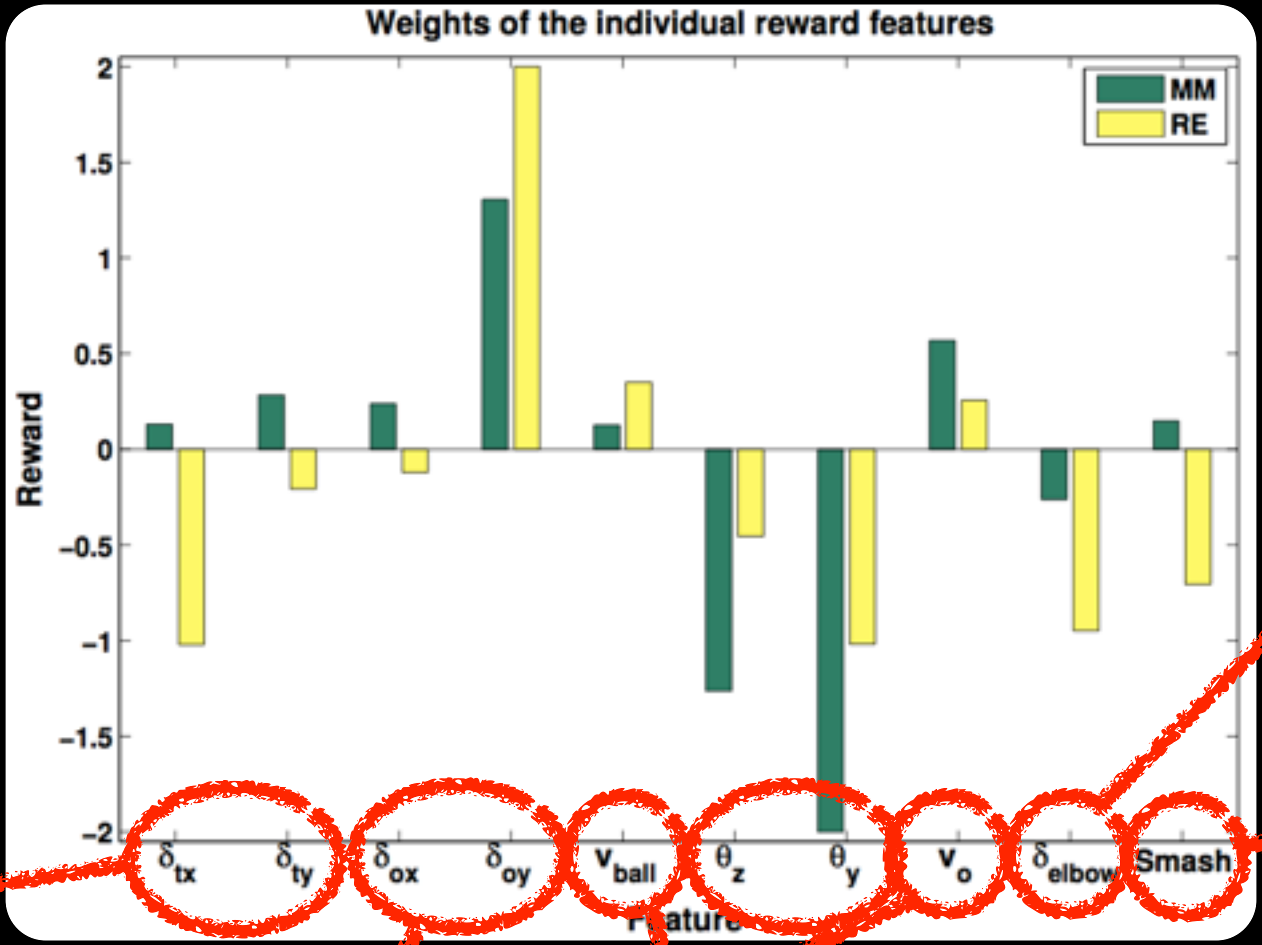




Mülling, K. et al. (2014). Biological Cybernetics.

Extracting Strategies from Game Play

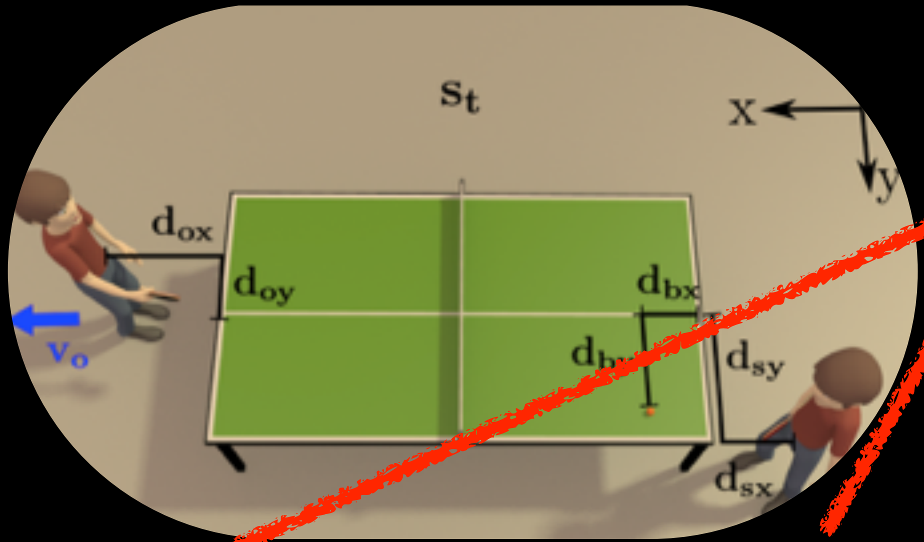
Weights of the most relevant features!



Opponent Elbow

Smash or not

Distance to the Edge of the Table



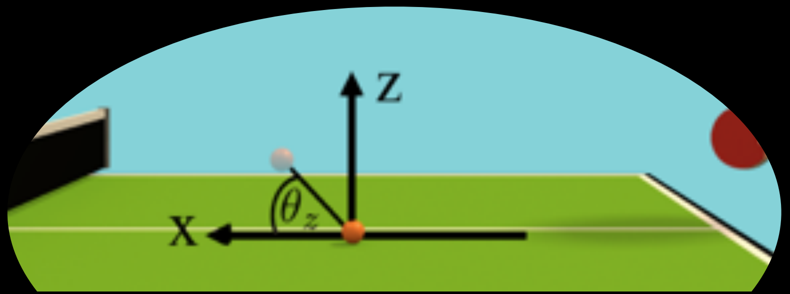
Angle of Incoming Bouncing Ball

Velocity of the Ball

Movement Direction of the Opponent

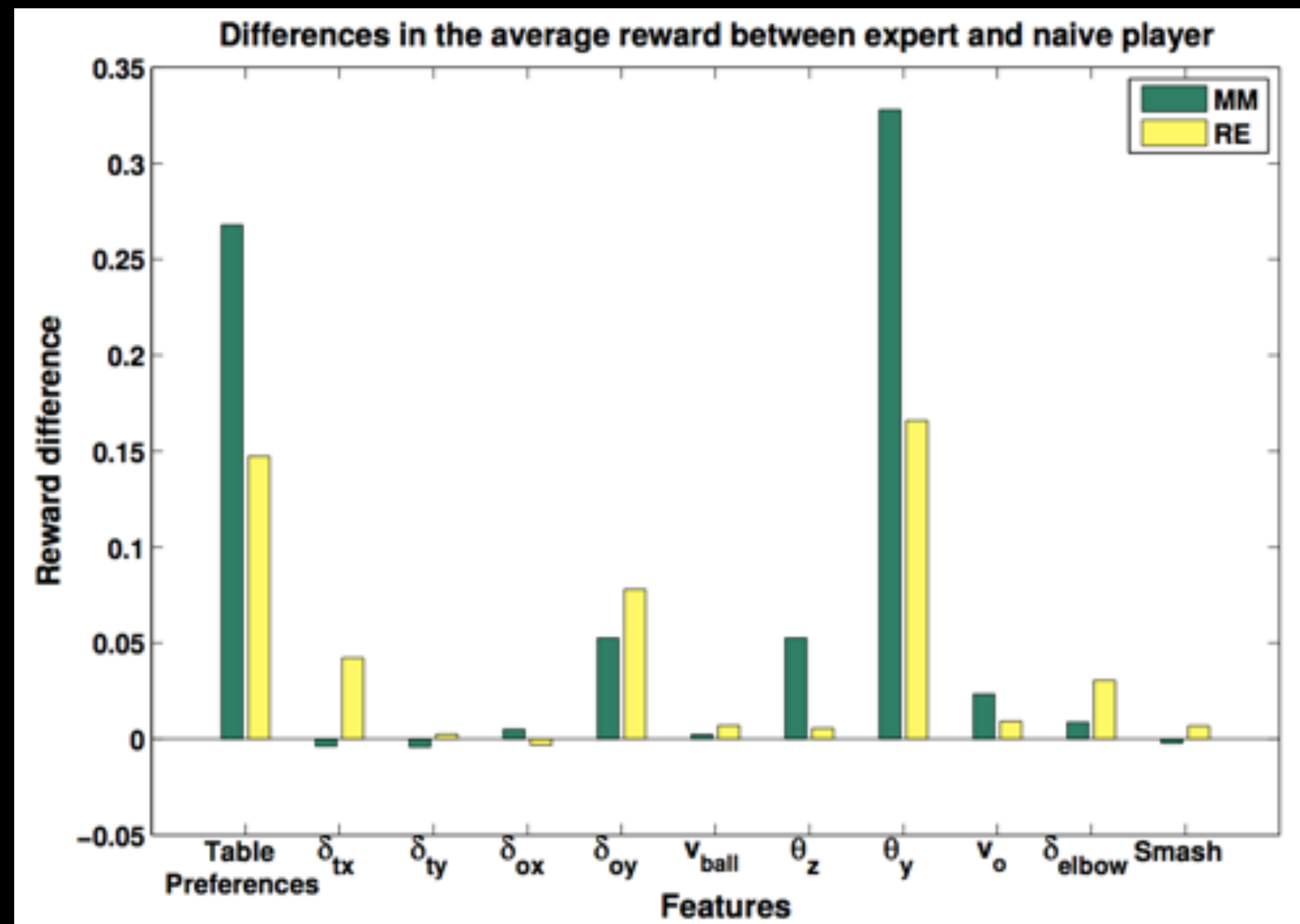
Distance to the Opponent

Mülling, K. et al. (2014) Biological Cybernetics.



Extracting Strategies from Game Play

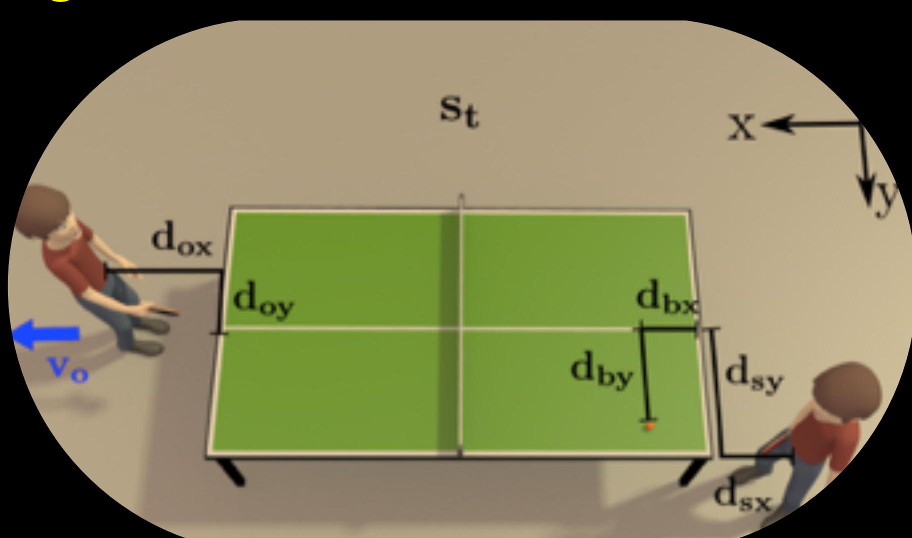
Differences between Experts and Naive Player only in few features!



Opponent Elbow

Smash or not

Distance to the Edge of the Table

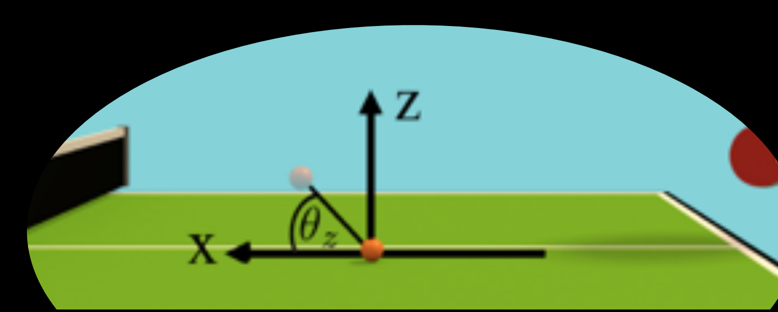


Movement Direction of the Opponent

Distance to the Opponent

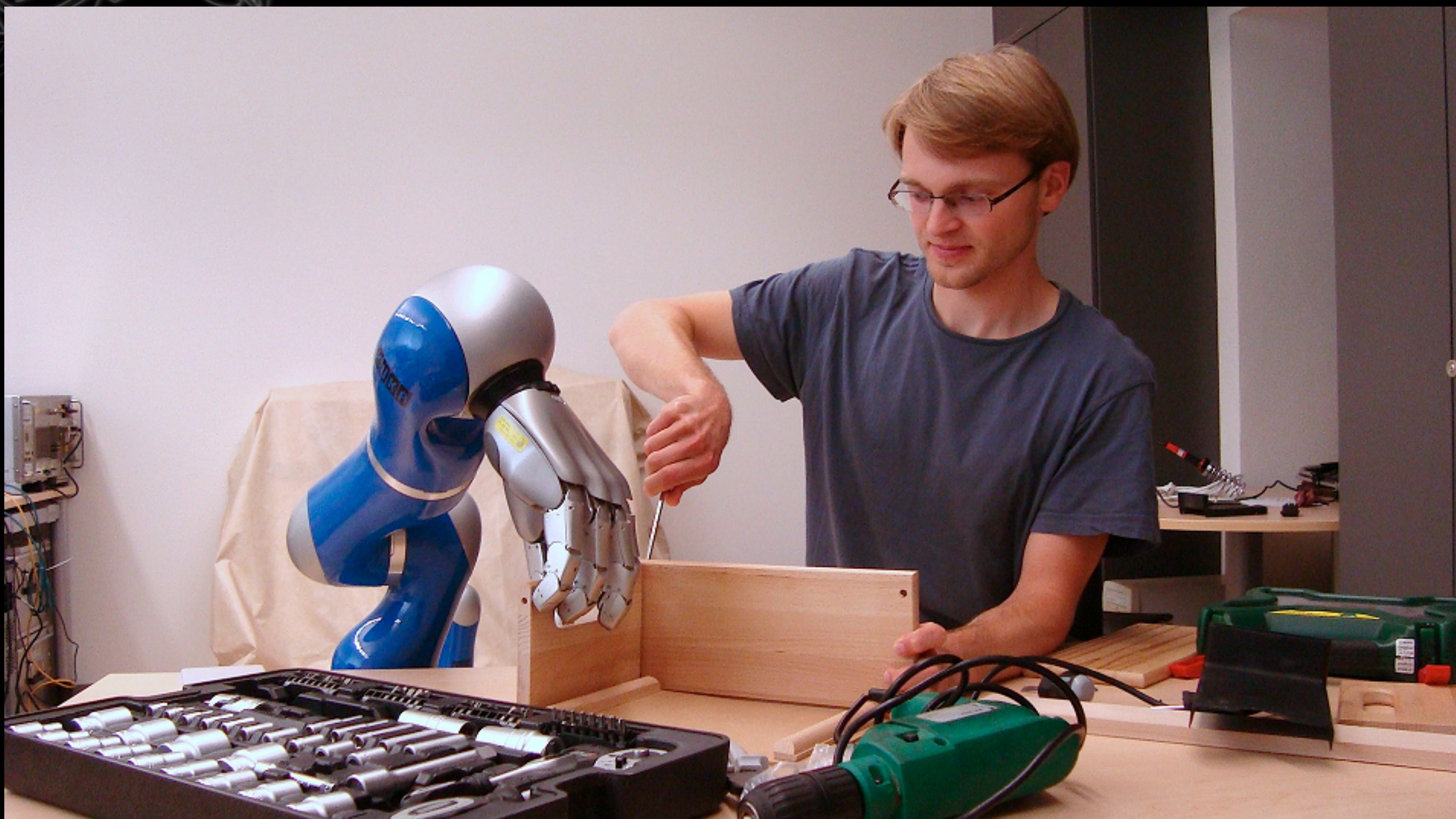
Angle of Incoming Bouncing Ball

Velocity of the Ball



Mülling, K. et al. (2014) Biological Cybernetics.

Interaction Primitives for a Semi-Autonomous 3rd Hand?

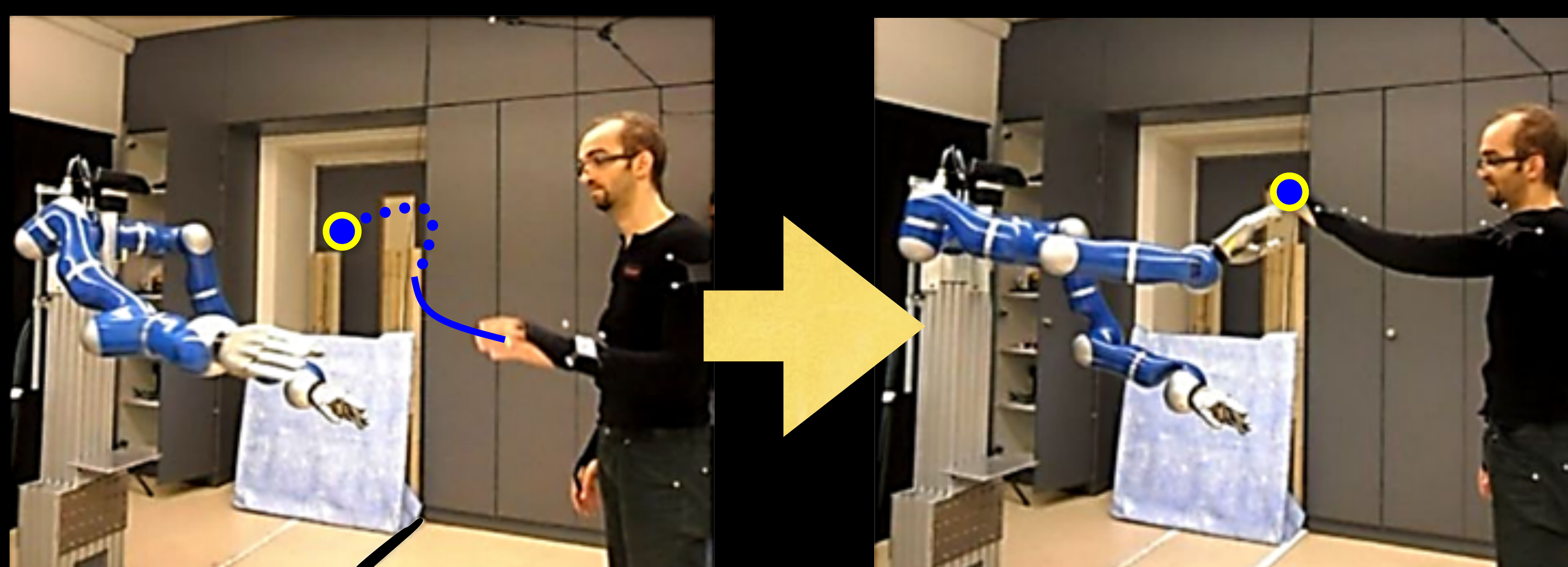


Interaction Primitives

The High-Five Task

- **Infer the task (aka primitive)**
- **Infer the human trajectory**

Generate the appropriate robot trajectory



Observed trajectory   Predicted trajectory   Predicted goal

Interaction Primitives

known agent **unknown agent**

Agent 1 (M joints) Agent 2 (N joints)

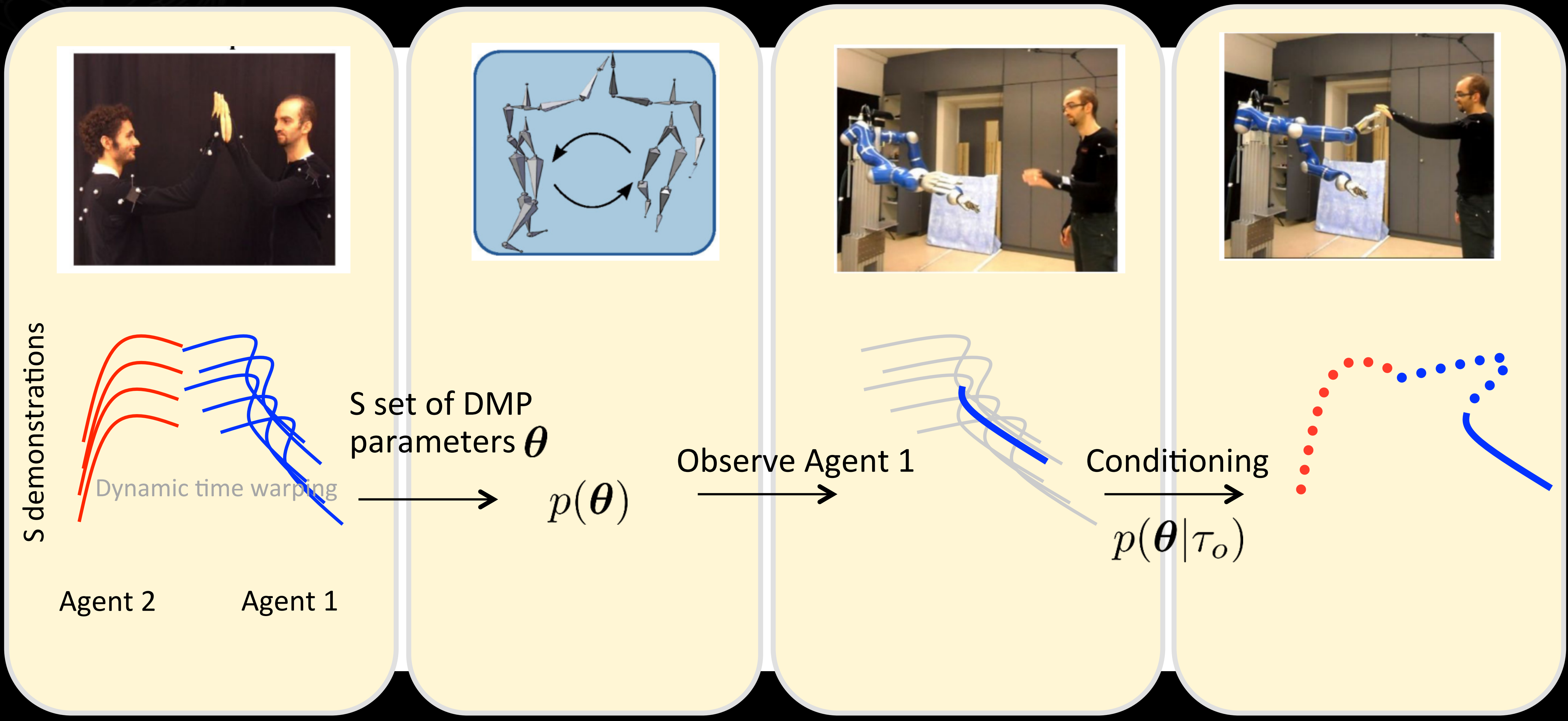
$$\boldsymbol{\theta}^{[1]} = [\ \boldsymbol{w}_1^T\ g_1\ \ldots\ \boldsymbol{w}_M^T\ g_M\ \ \boldsymbol{w}_1^T\ g_1\ \ldots\ \boldsymbol{w}_N^T\ g_N\ ]$$

Goal

$$\mathbf{w}_1 = [w_{1,1}\ \ldots\ w_{B,1}]^T$$

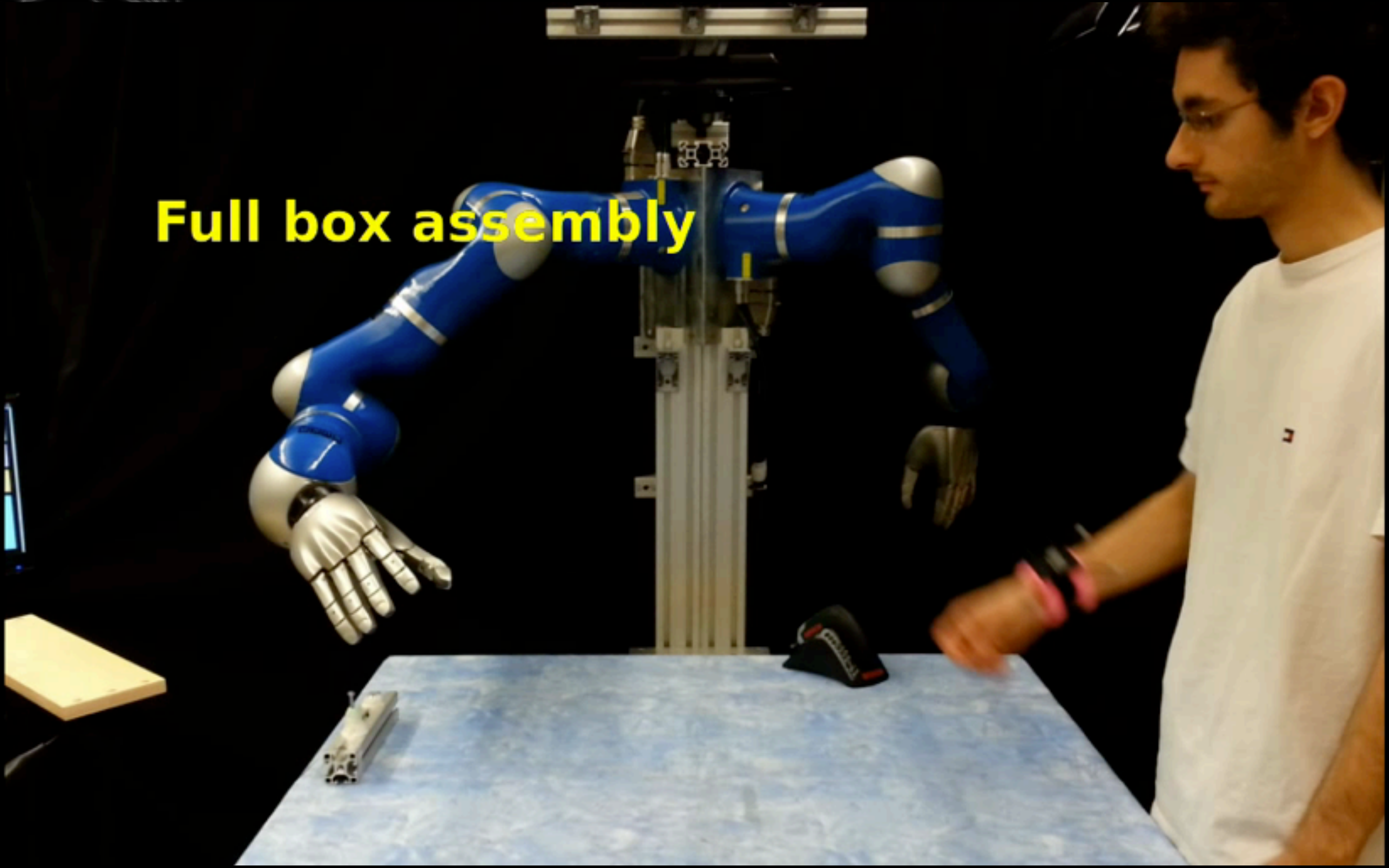
An Interaction primitive can simply be a motor primitive that includes both the known agent and the unknown agent.

Interaction Primitives for a Semi-Autonomous 3rd Hand



Ben Amor, H.; Neumann, G.; Kamthe, S.; Kroemer, O.; Peters, J. (2014). Interaction Primitives for Human-Robot Cooperation Tasks , Proceedings of 2014 IEEE International Conference on Robotics and Automation (ICRA).

Interaction Primitives for a Semi-Autonomous 3rd Hand



Ewerton, M.; Neumann, G.; Lioutikov, R.; Ben Amor, H.; Peters, J.; Maeda, G. (2015). Learning Multiple Collaborative Tasks with a Mixture of Interaction Primitives, International Conference on Robotics and Automation (ICRA).

Outline

1. Introduction
2. How can we develop suitable machine learning methods?
3. How can elementary behavior be learned with such machine learning methods?
4. Can complex skills be learned leveraging on elementary behaviors?
5. Outlook
6. Conclusion

It's not all Table Tennis...

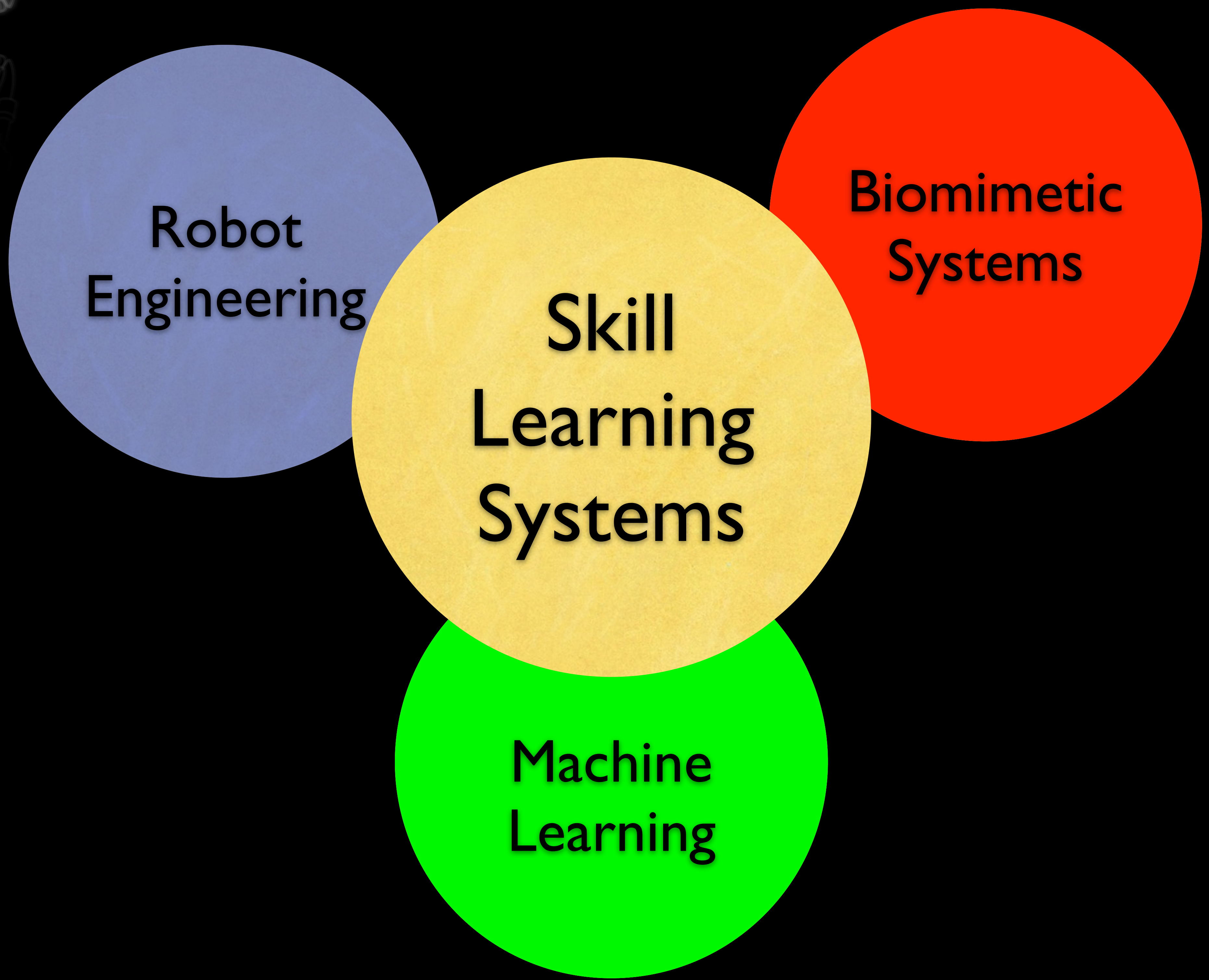
Industrial Application: Key bottleneck in manufacturing is the high cost of robot programming and slow implementation.

> **Bosch**: *If a product costs less than 50€ or is produced less than 10.000 times, it is not competitive with manual labor.*

Assistive Robots & Companion Technologies: In hospital and rehablitation institutions, nurses need to "program" the robot – not computer scientists.

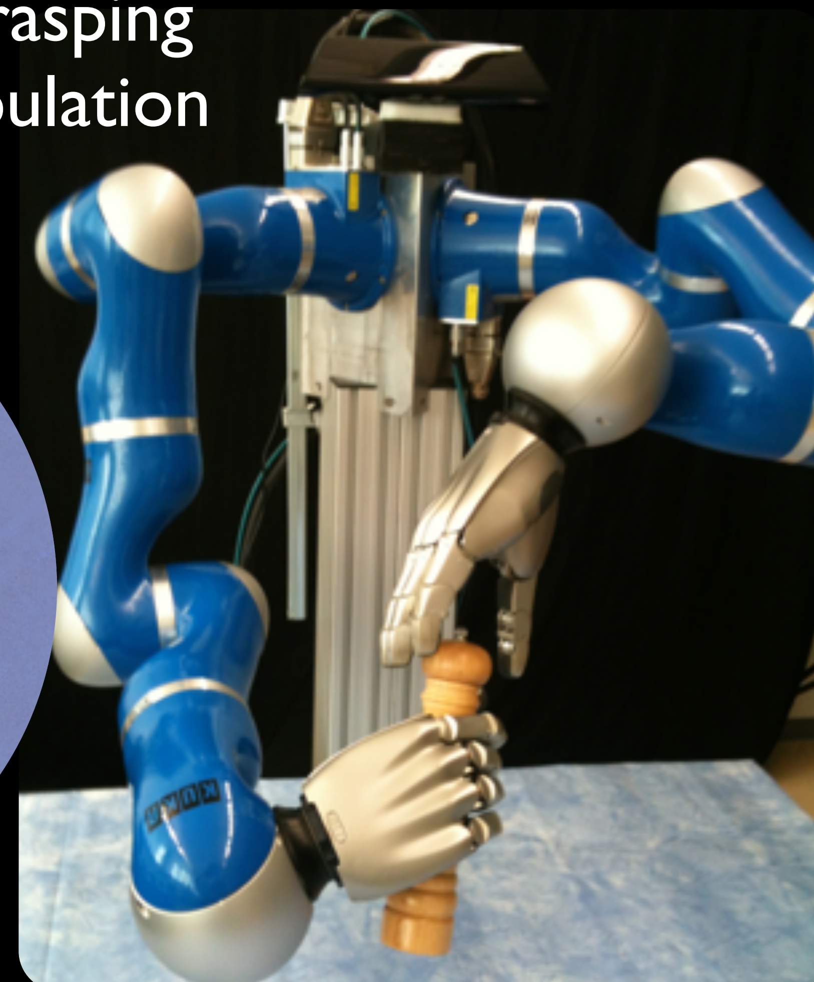
Robots@Home: Robots need to adapt to the human and "blend into the kitchen".

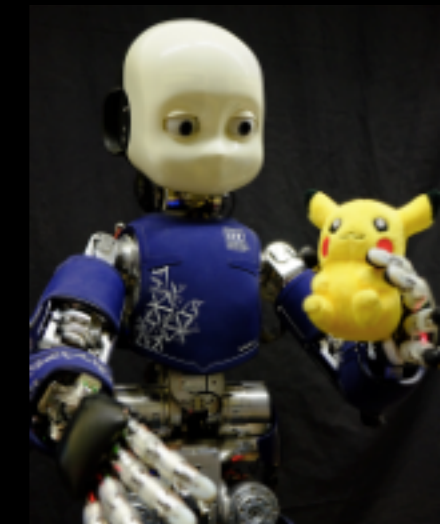
Outlook


Robot Engineering
Skill Learning Systems
Biomimetic Systems
Machine Learning

Robot Systems

Robot Grasping
and Manipulation

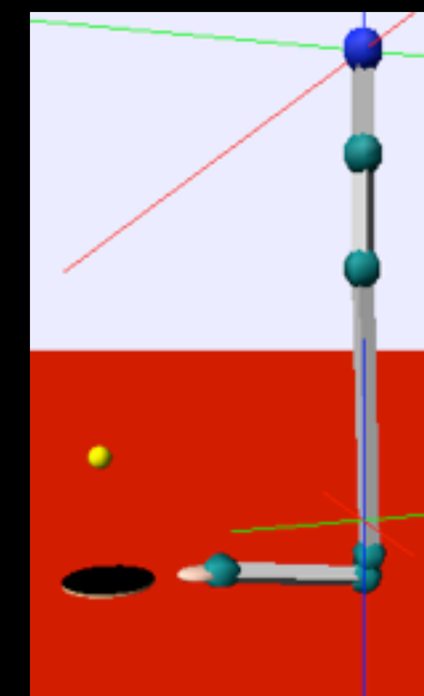




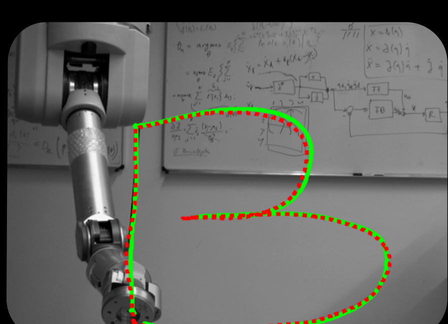
Humanoid Robotics

Robot
Engineering

Real-Time Software &
Simulations for Robots



High-Speed
Real-Time Vision



Nonlinear Robot Control

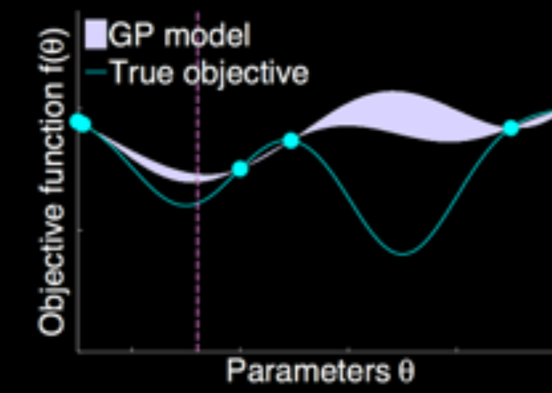
Tactile Perception &
Sensory Integration

Industrial
Partnership with
Honda, ABB and
Bosch.

Machine Learning

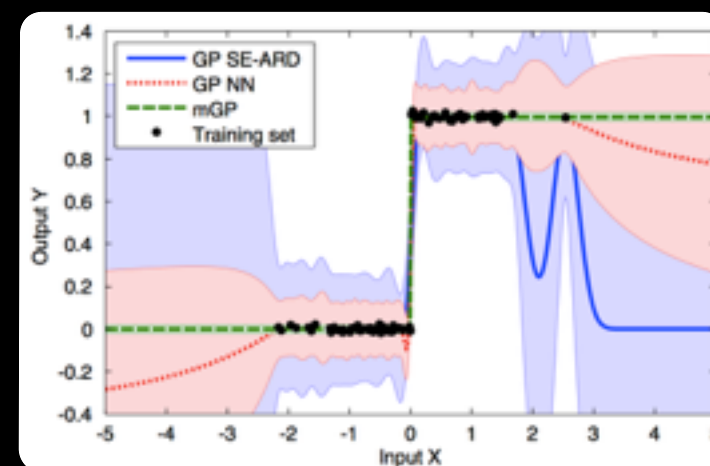
Real-Time Regression
(Nguyen-Tuong & Peters, Neurocomputing 2011)

Bayesian Optimization
(Calandra et al, 2014)



Much more Reinforcement Learning...

Model Learning
(Nguyen-Tuong & Peters, Advanced Robotics 2010)



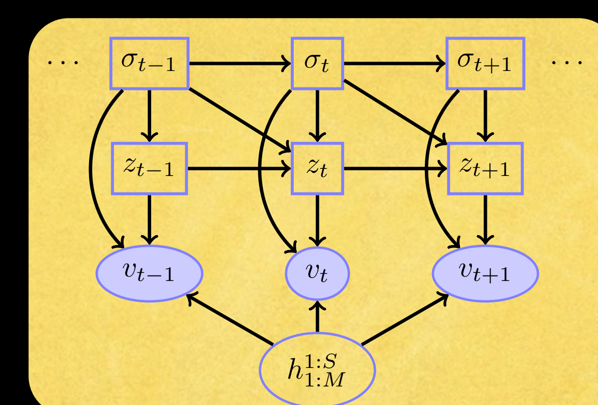
Maximum Entropy
(Peters et al., AAAI 2010; Daniel, Neumann & Peters, AIStats 2012)

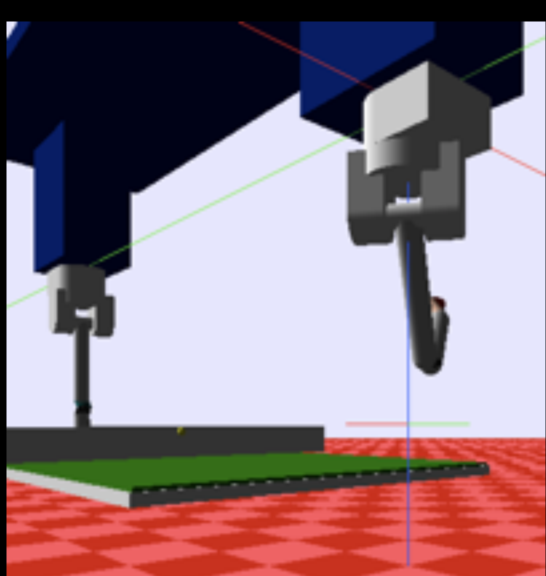
Probabilistic Movement Representation
(Paraschos et al. NIPS 2013)

Partnership with the Max Planck Institute for Intelligent Systems.

Manifold Gaussian Processes
(Calandra et al, 2014)

Policy Gradient Methods
(Peters et al, IROS 2006)





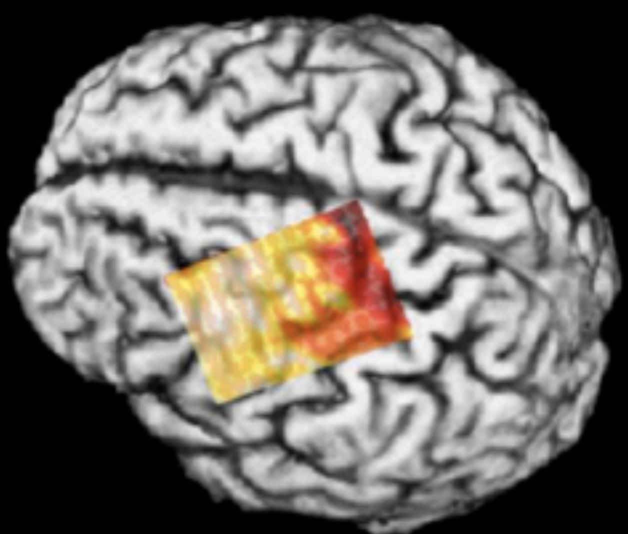
Machine Learning for Motor Games
(Wang, Boularias & Peters, AAAI 2011)

Machine Learning

Pattern Recognition in Time Series
(Alvarez, Peters et al., NIPS 2010a; Chiappa & Peters, NIPS 2010b)

Biological Inspiration and Application

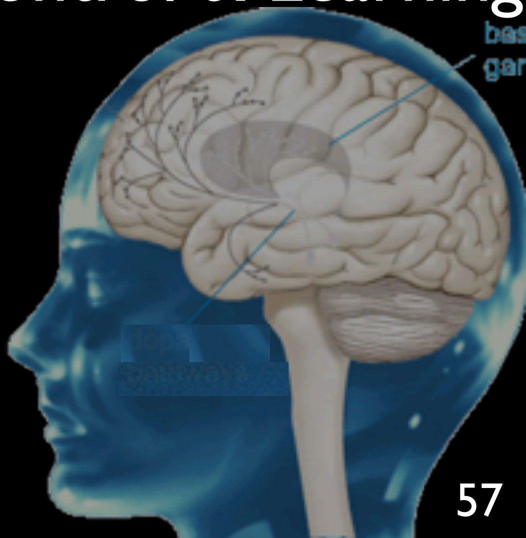
Brain-Computer Interfaces with ECoG for Stroke Patient Therapy
(Gomez, Peters & Grosse-Wentrup, Journal of Neuroengineering 2011)



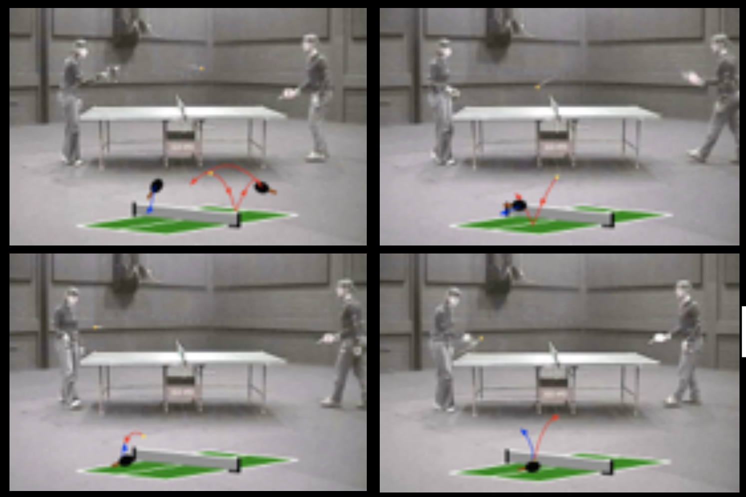
Brain Robot Interfaces
(Peters et al., Int. Conf. on Rehabilitation Robotics, 2011)

Biomimetic Systems

Computational Models of Motor Control & Learning



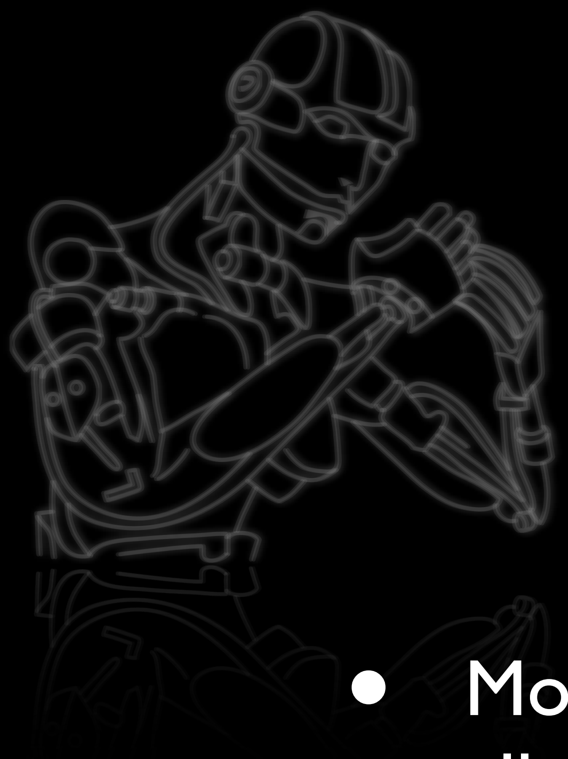
Collaboration with the Max Planck Institute for Intelligent Systems and the Tübingen University Hospital.



Understanding Human Movements
(Mülling, Kober & Peters, Adaptive Behavior 2011)

Outline

1. Introduction
2. How can we develop suitable machine learning methods?
3. How can elementary behavior be learned with such machine learning methods?
4. Can complex skills be learned leveraging on elementary behaviors?
5. How can we adapt to humans and learn interaction?
6. Conclusion



Conclusion

- Motor skill learning is a promising way to avoid programming all possible scenarios and continuously adapt to the environment.
- We have efficient Imitation and Reinforcement Learning Methods which scale to anthropomorphic robots.
- Basic skill learning capabilities of humans can be produced in artificial skill learning systems.
- We are working towards learning of complex tasks such as table tennis and a semi-autonomous 3rd hand.

Thanks for your Attention!