Towards Machine Learning of Motor Skills for Robotics

From Simple Skills to Robot Table Tennis and Manipulation

Jan Peters

Technische Universität Darmstadt

Max Planck Institute for Intelligent Systems

Motivation

Can we create such robots?

Source: Movie iRobot

Motivation

Uncertainty in tasks
and environment

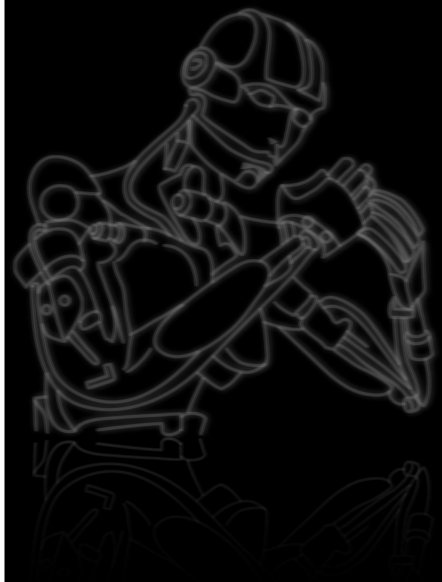
Adapt to humans

Programming complexity
beyond human imagination

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

- Smart Humans? Hand-coding 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? Can they scale?

➡ We need to develop **skill learning approaches** for autonomous robot systems!



Important Questions

I. How can we develop efficient motor learning methods?

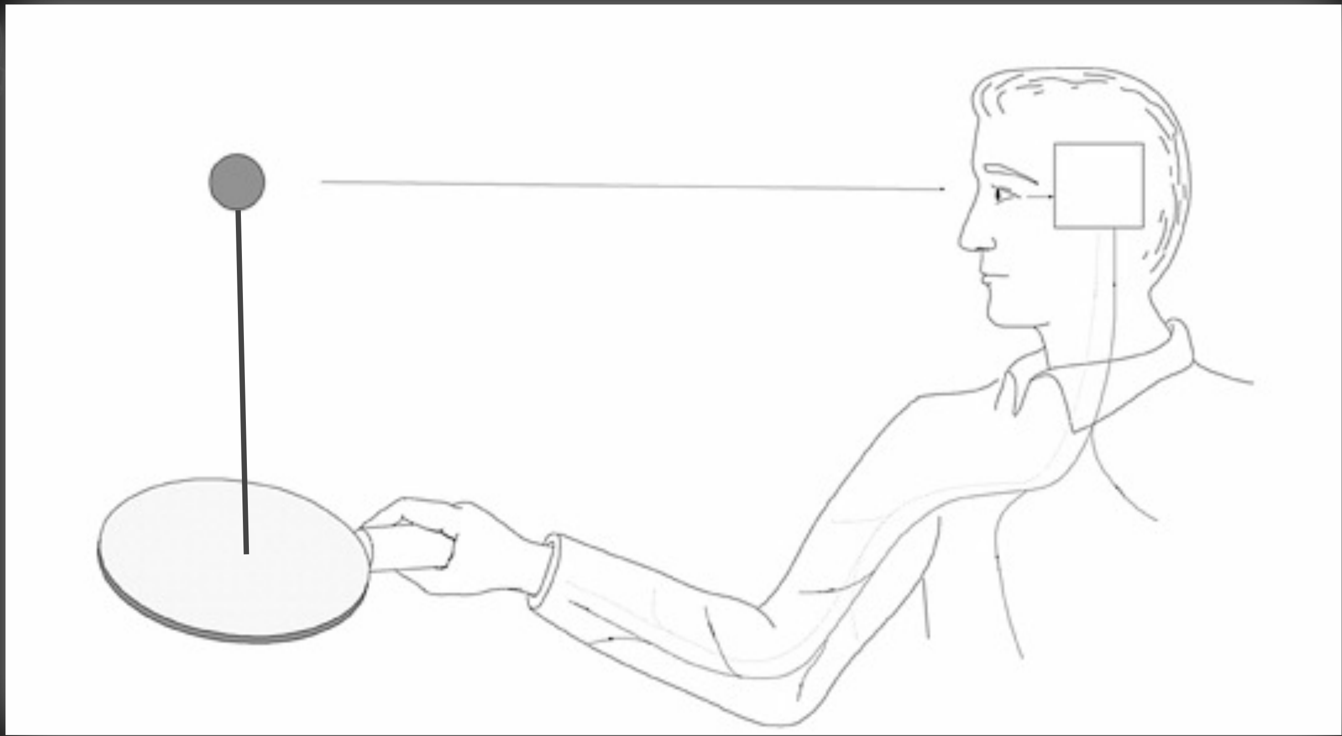
II. How can anthropomorphic robots learn basic skills similar to humans?

III. Can complex skills be composed with these elements?

Outline

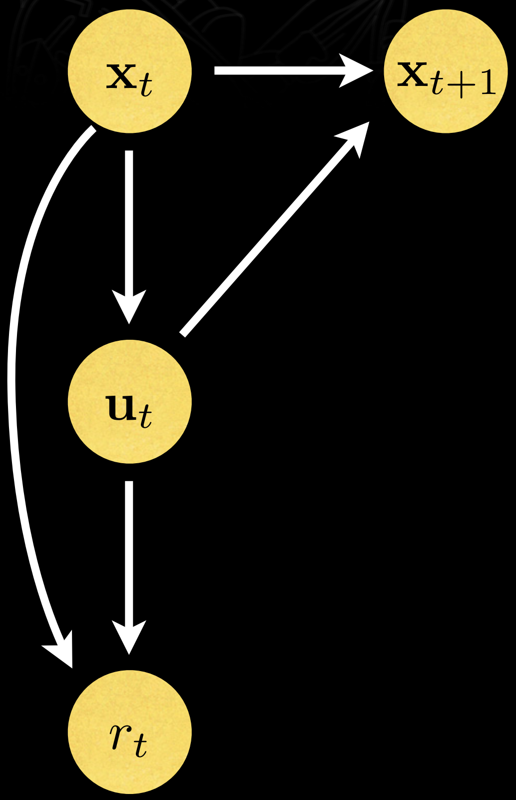
1. Introduction
2. How can we develop efficient motor learning methods?
3. How can anthropomorphic robots learn basic skills similar to humans?
4. Can complex skills be composed with these elements?
5. Outlook
6. Conclusion

Example:



Internal and external state $\mathbf{x}_t$, action $\mathbf{u}_t$.

Modeling Assumptions



Autonomous Learning System: Modeled by a policy that generates action $\mathbf{u}_t$ in state $\mathbf{x}_t$.

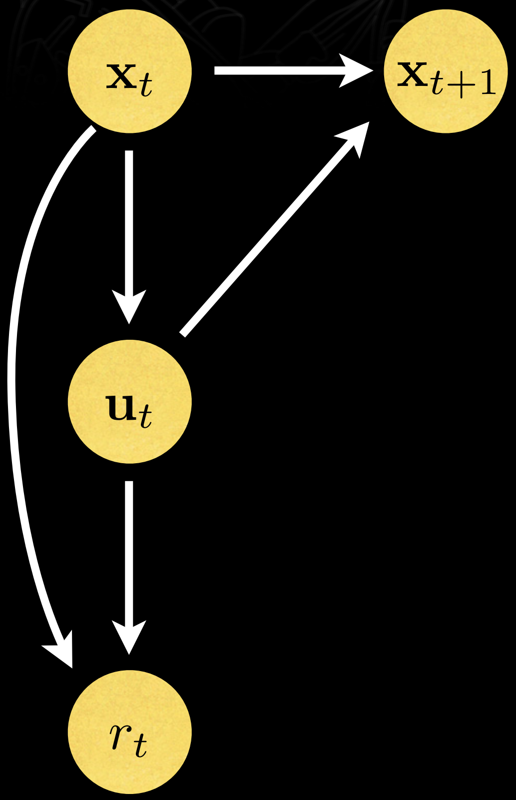
Teacher: Evaluates the performance and rates it with r_t.

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

Model in a perfect world: $\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t)$

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

Modeling Assumptions



Autonomous Learning System: Modeled by a policy that generates action $\mathbf{u}_t$ in state $\mathbf{x}_t$.

How can we model a behavior with "rules"?

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

NO! *Stochasticity* is important:

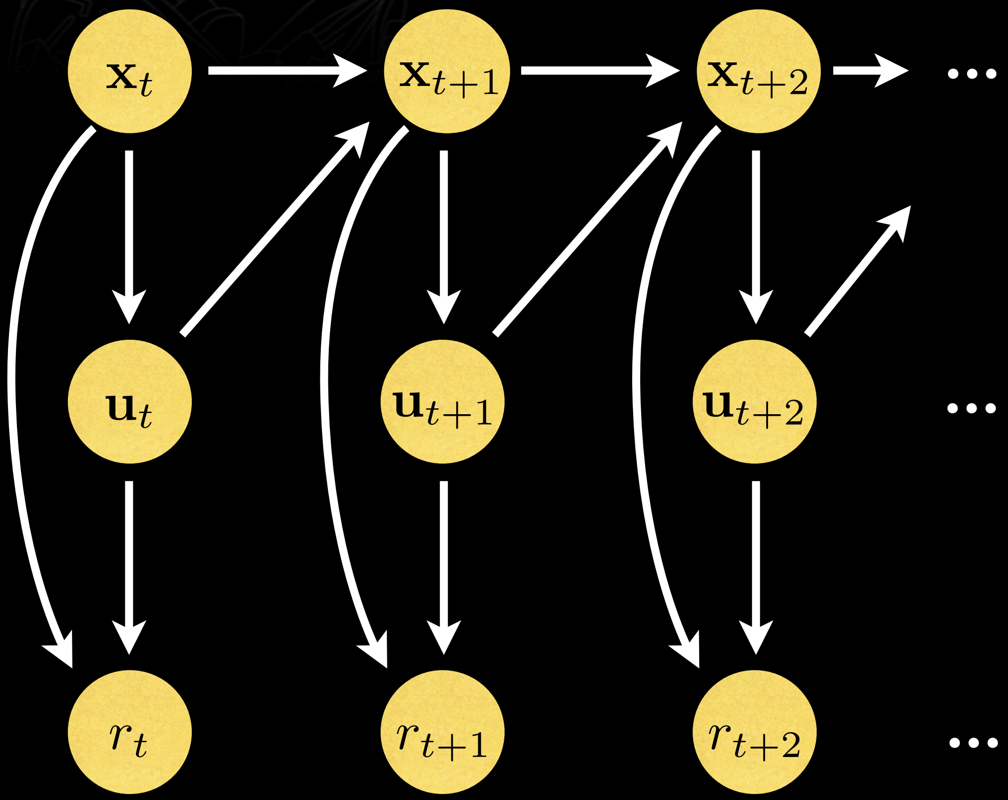
- needed for exploration
- eases algorithm design
- reduces the complexity
- optimal solution can be stochastic
- can model variance of the teacher

Hence, we use a stochastic policy:

Allow learning!

$$\mathbf{u}_t \sim \pi(\mathbf{u}_t|\mathbf{x}_t) = p(\mathbf{u}_t|\mathbf{x}_t, \theta)$$

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:

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

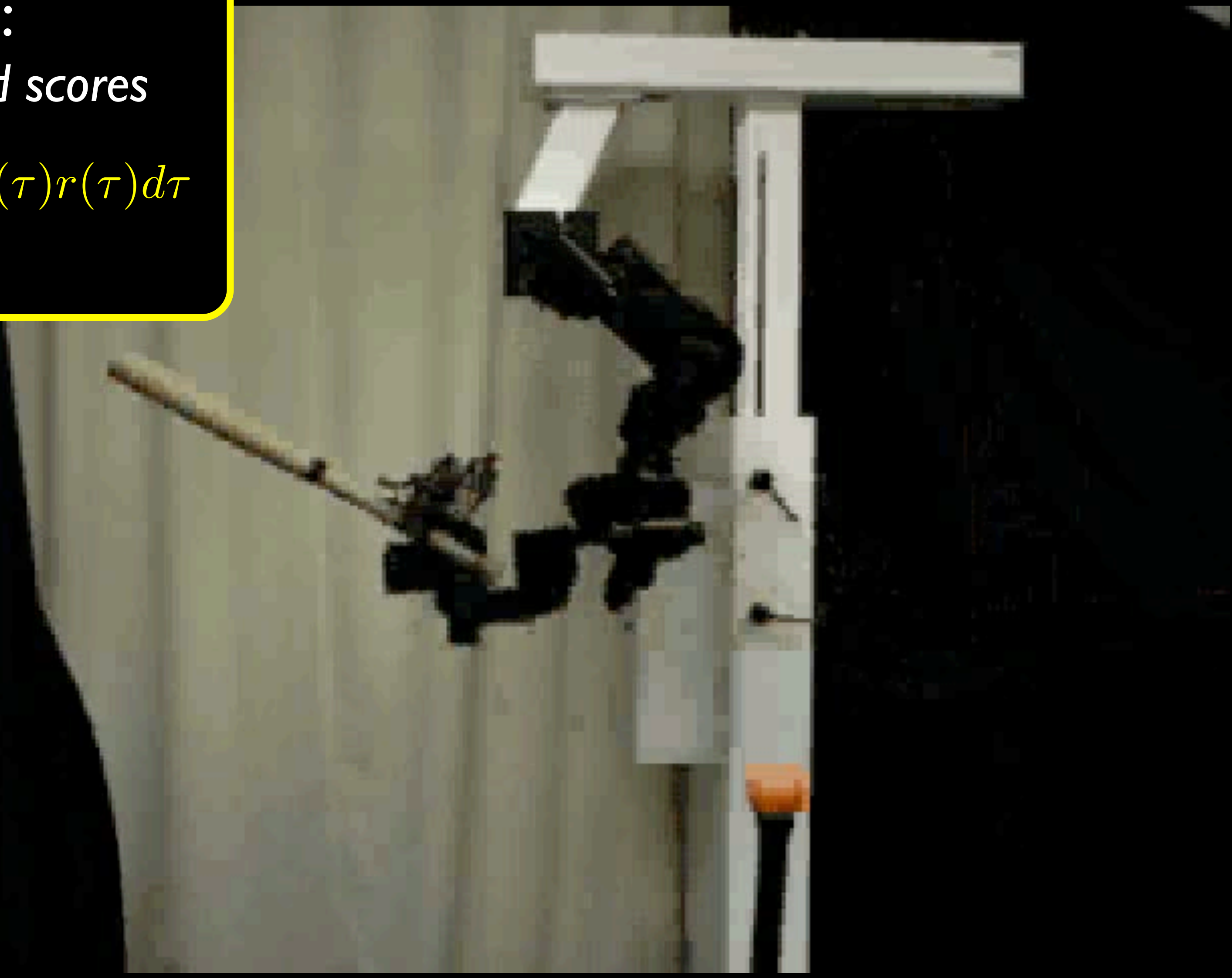
What is learning?

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 efficient motor learning methods?
3. How can anthropomorphic robots learn basic skills similar to humans?
4. Can complex skills be composed with these elements?
5. Outlook
6. Conclusion

Imitation Learning

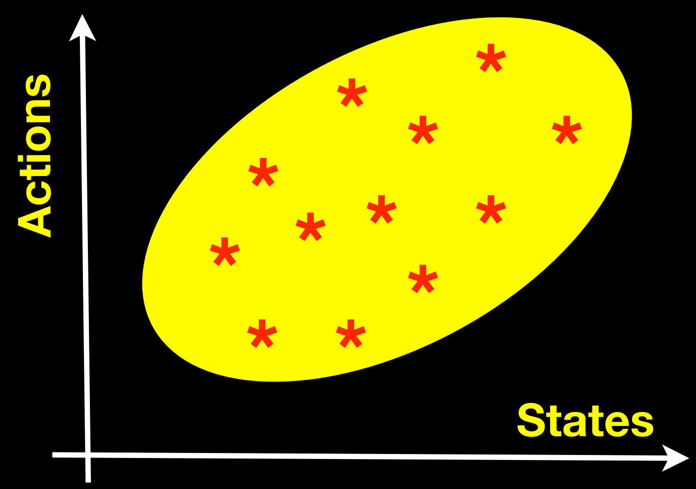
Given a path distribution, can we reproduce the policy?

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

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

- only adapt the policy parameters θ
- model-free, purely sample-based
- results in one-shot and expectation maximization algorithms

Imitation



Reinforcement Learning

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

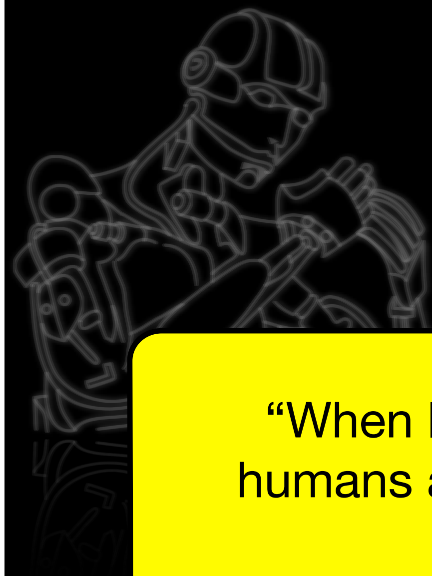
- *Goal*: maximize the return of the paths $r(\tau)$ generated by path distribution $p_\theta(\tau)$

- Optimization function is the 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...

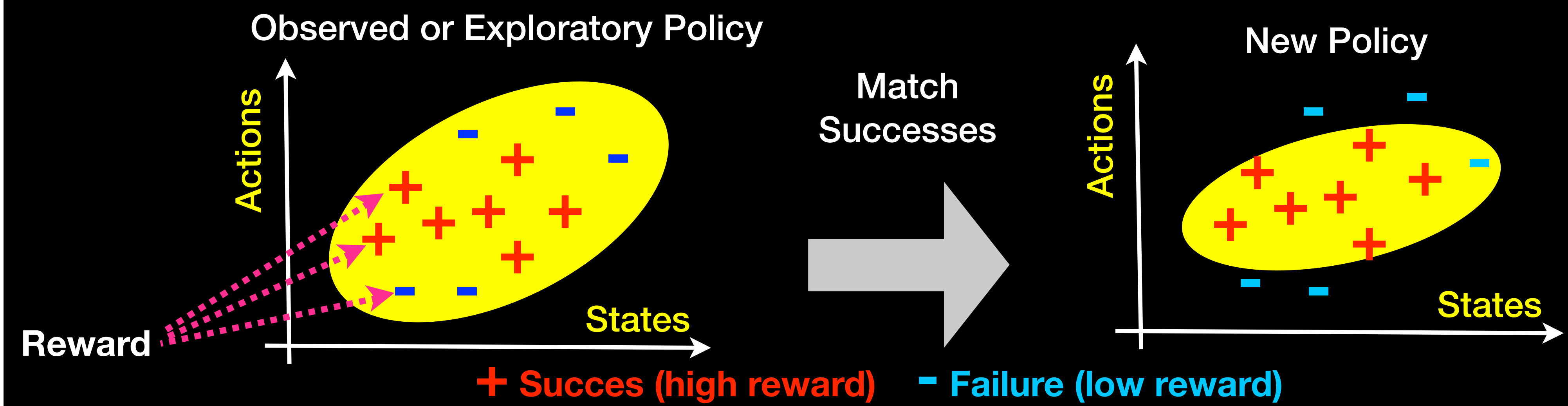
- *Problem*: Optimization Bias



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).

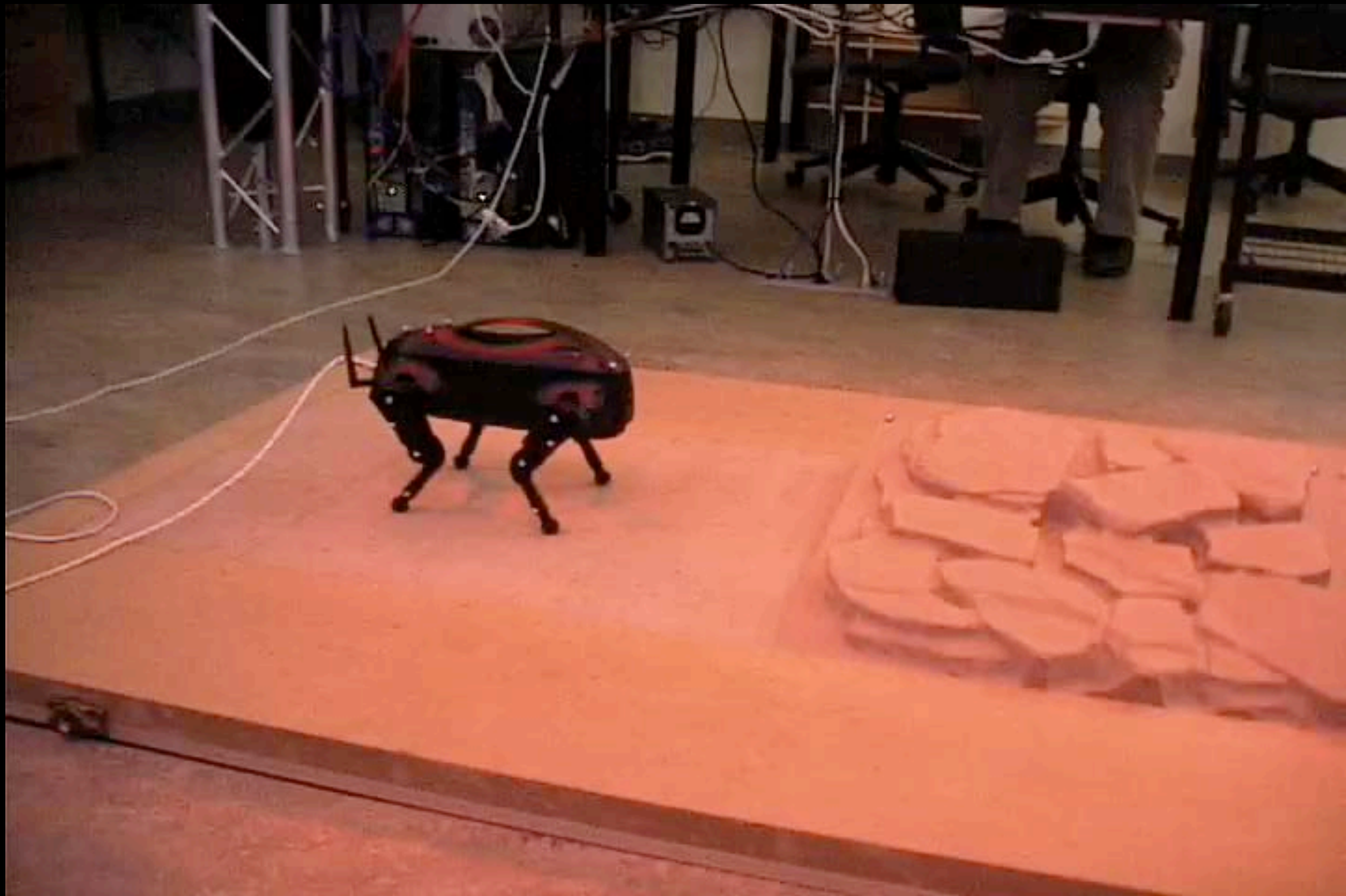
Can we create better policies by matching the reward-weighted previous policy ?



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 Reward-Weighted Imitation

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

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

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

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

New Policy Parameters   Features   Rewards   Actions

➡Reduces Reinforcement Learning onto Reward 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)

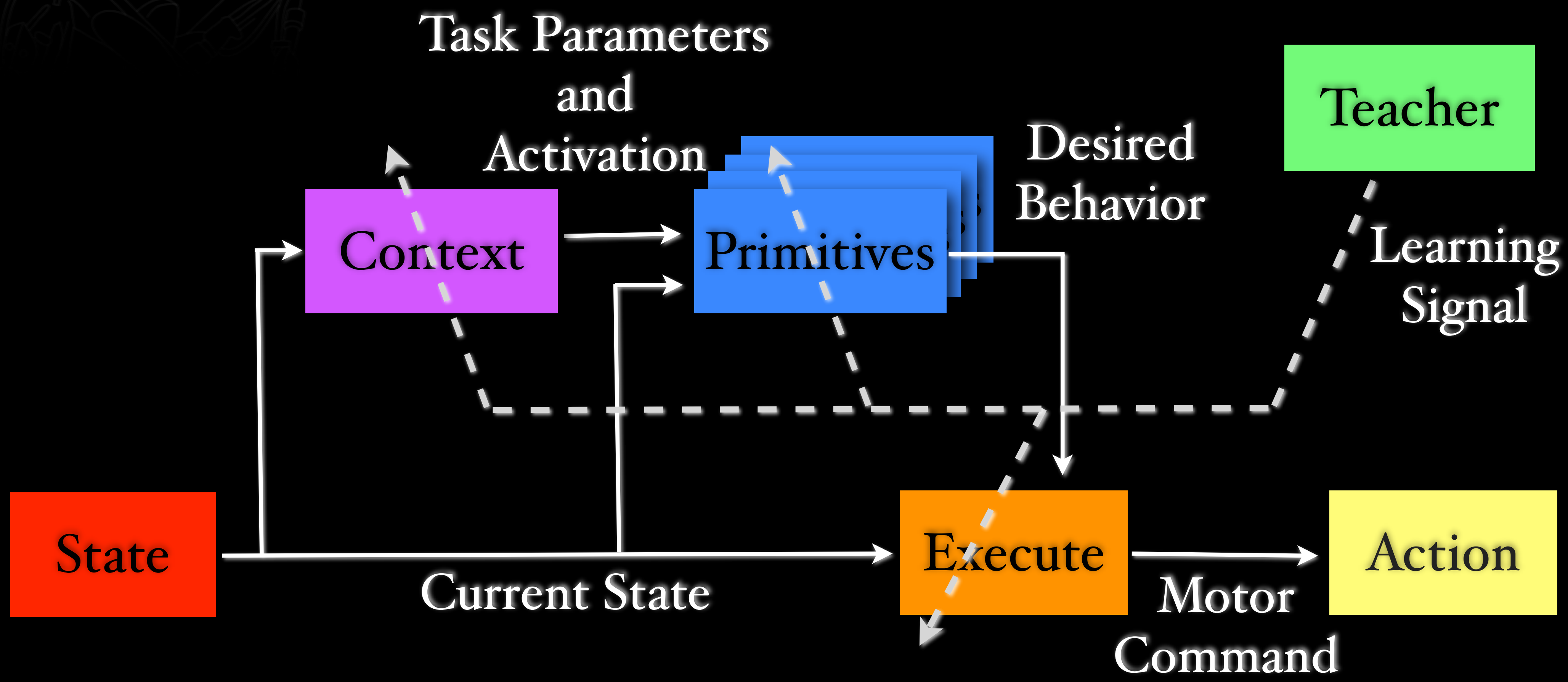
All of these approaches are extensions of this idea.

Our follow-up approach "Relative Entropy Policy Search" (Peters et al., AAAI, 2010; Daniel et al., AIStats 2012) also relies on most of these insights.

Outline

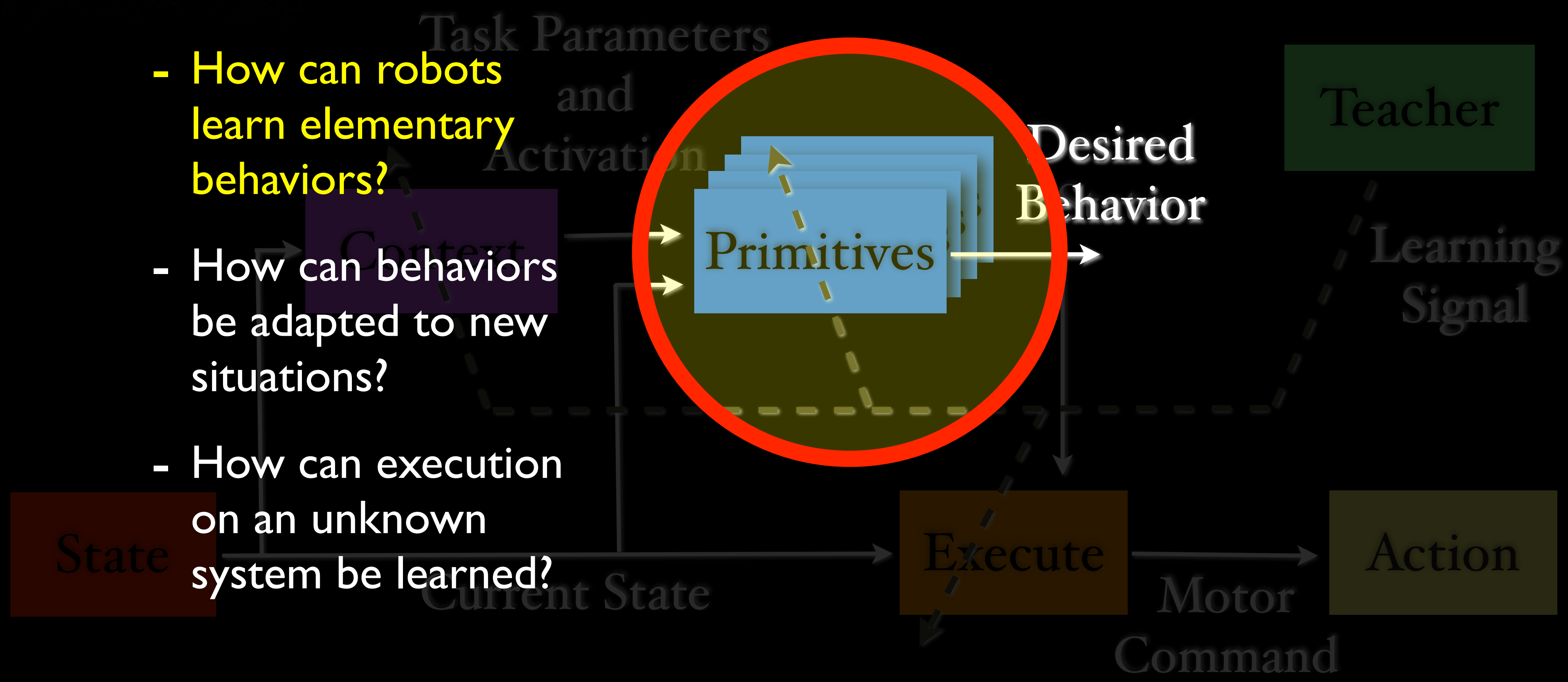
1. Introduction
2. How can we develop efficient motor learning methods?
3. How can anthropomorphic robots learn basic skills similar to humans?
4. Can complex skills be composed with these elements?
5. Outlook
6. Conclusion

A Blue Print for Skill Learning?

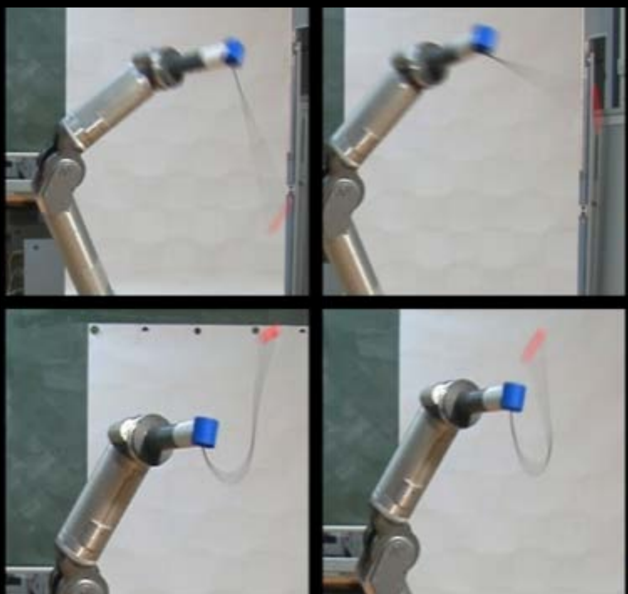


Outline

- How can robots learn elementary behaviors?
- How can behaviors be adapted to new situations?
- How can execution on an unknown system be learned?



Motor Primitives



How can we represent, acquire and refine elementary movements?

- Humans appear to rely on context-driven motor primitives (Flash & Hochner, TICS 2005)
- Many favorable properties:
  - Invariance under task parameters
  - Robust, superimposable, ...

➡ *Resulting approach:*
  - Use the dynamic system-based motor primitives (Ijspeert et al. NIPS2003; Schaal, Peters, Nakanishi, Ijspeert, ISRR2003).
  - Initialize by Imitation Learning.
  - Improve by trial and error on the real system with Reinforcement Learning.

Motor Primtives

Task/Hyperparameter

Trajectory Plan Dynamics

$$\dot{z} = \alpha_z\left(\beta_z\left(g - y\right) - z\right)$$
$$\dot{y} = \alpha_y\left(f\left(x,v\right) + z\right)$$

where

Canonical Dynamics

$$\dot{v} = \alpha_v\left(\beta_v\left(g - x\right) - v\right)$$
$$\dot{x} = \alpha_x v$$

Linear in learnable Policy Parameters

Local Linear Model Approx.

$$f\left(x,v\right) = \frac{\sum_{i=1}^{k} w_i b_i v}{\sum_{i=1}^{k} w_i}$$

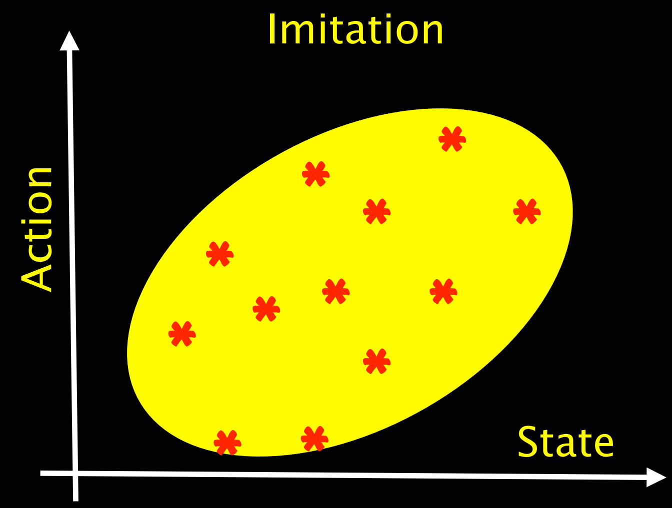
$$w_i = \exp\left(-\frac{1}{2} d_i \left(\bar{x} - c_i\right)^2\right) \text{ and } \bar{x} = \frac{x - x_0}{g - x_0}$$

(Ijspeert et al., NIPS 2003; Schaal, Peters, Nakanishi, Ijspeert, ISRR 2003)

Acquisition by Imitation

Teacher shows the task and the student reproduces it.

- maximize similarity

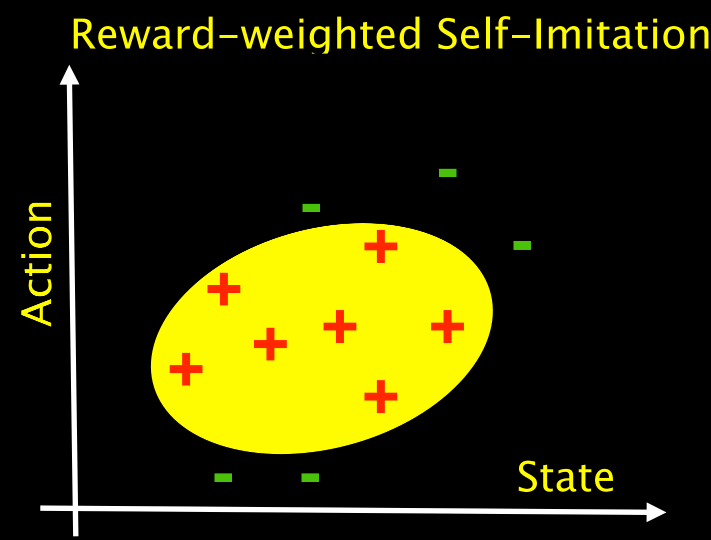


Kober & Peters (2009). Learning Motor Primitives, ICRA

Self-Improvement by Reinforcement Learning

Student improves by reproducing his successful trials.

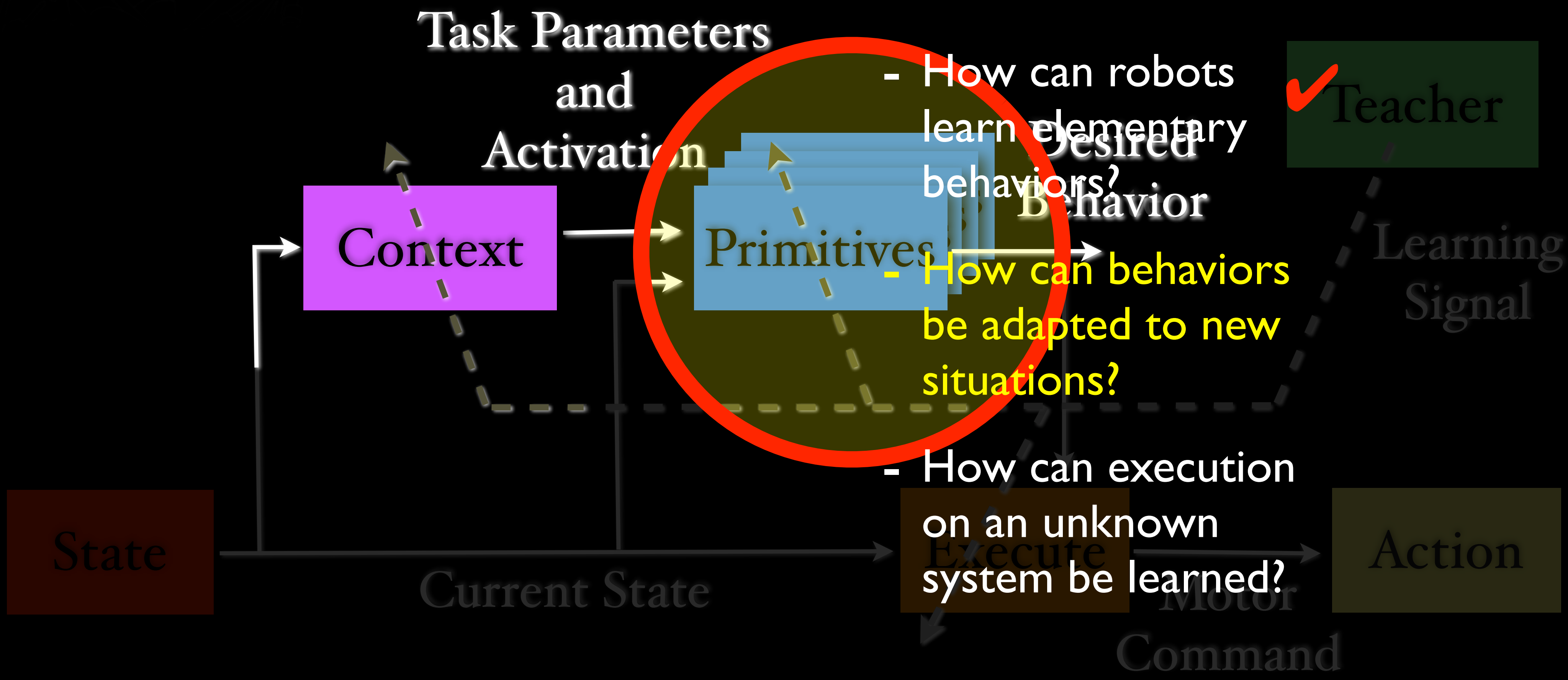
- maximize reward-weighted similarity



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?
- How can execution on an unknown system be learned?



Motor Primtives

Task/Hyperparameter

Trajectory Plan Dynamics

$$\begin{cases} \dot{z} = \alpha_z\left(\beta_z\left(g - y\right) - z\right) \\ \dot{y} = \alpha_y\left(f\left(x,v\right) + z\right) \end{cases}$$

where

Canonical Dynamics

$$\begin{cases} \dot{v} = \alpha_v\left(\beta_v\left(g - x\right) - v\right) \\ \dot{x} = \alpha_x v \end{cases}$$

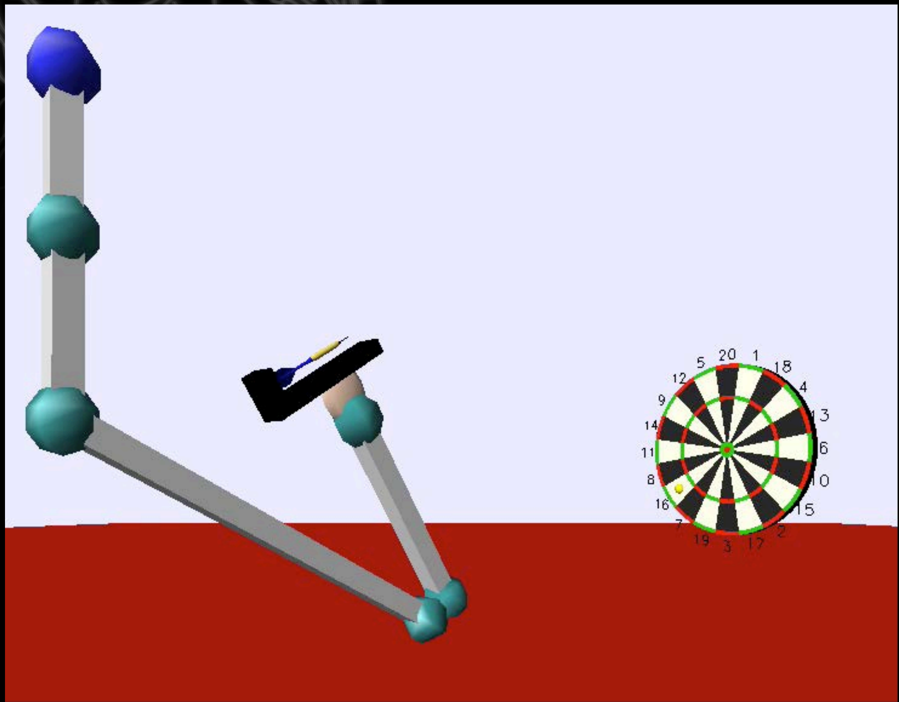
Linear in learnable Policy Parameters

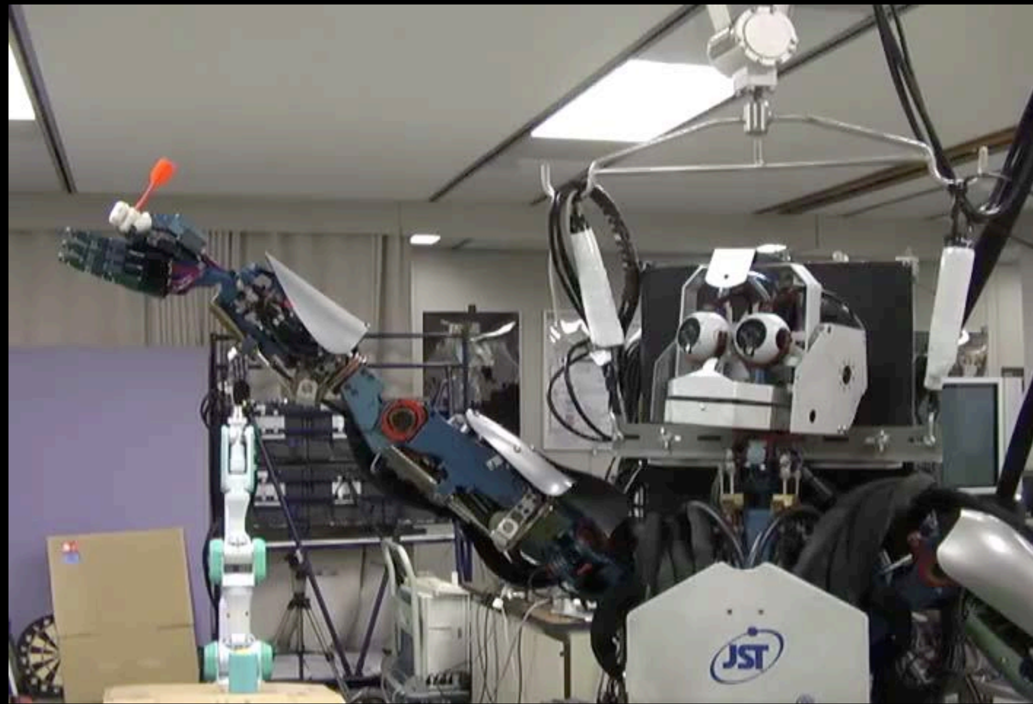
Local Linear Model Approx.

$$\begin{cases} f\left(x,v\right) = \dfrac{\sum_{i=1}^{k} w_i b_i v}{\sum_{i=1}^{k} w_i} \\ w_i = \exp\left(-\dfrac{1}{2} d_i \left(\bar{x} - c_i\right)^2\right) \text{ and } \bar{x} = \dfrac{x - x_0}{g - x_0} \end{cases}$$

(Ijspeert et al., NIPS 2003; Schaal, Peters, Nakanishi, Ijspeert, ISRR 2003)

Task Context: Goal Learning



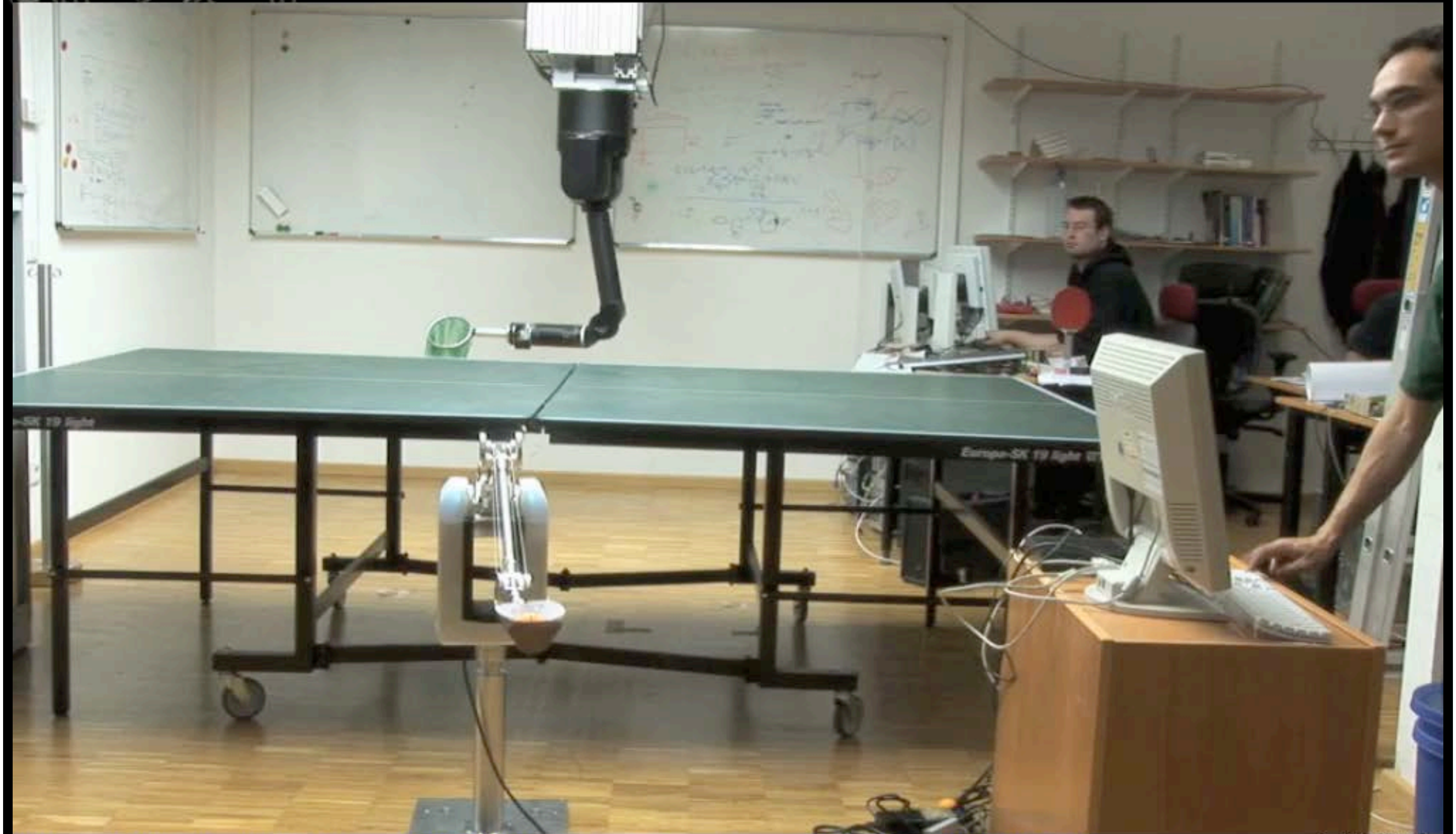


Adjusting Motor Primitives through their Hyperparameters:

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

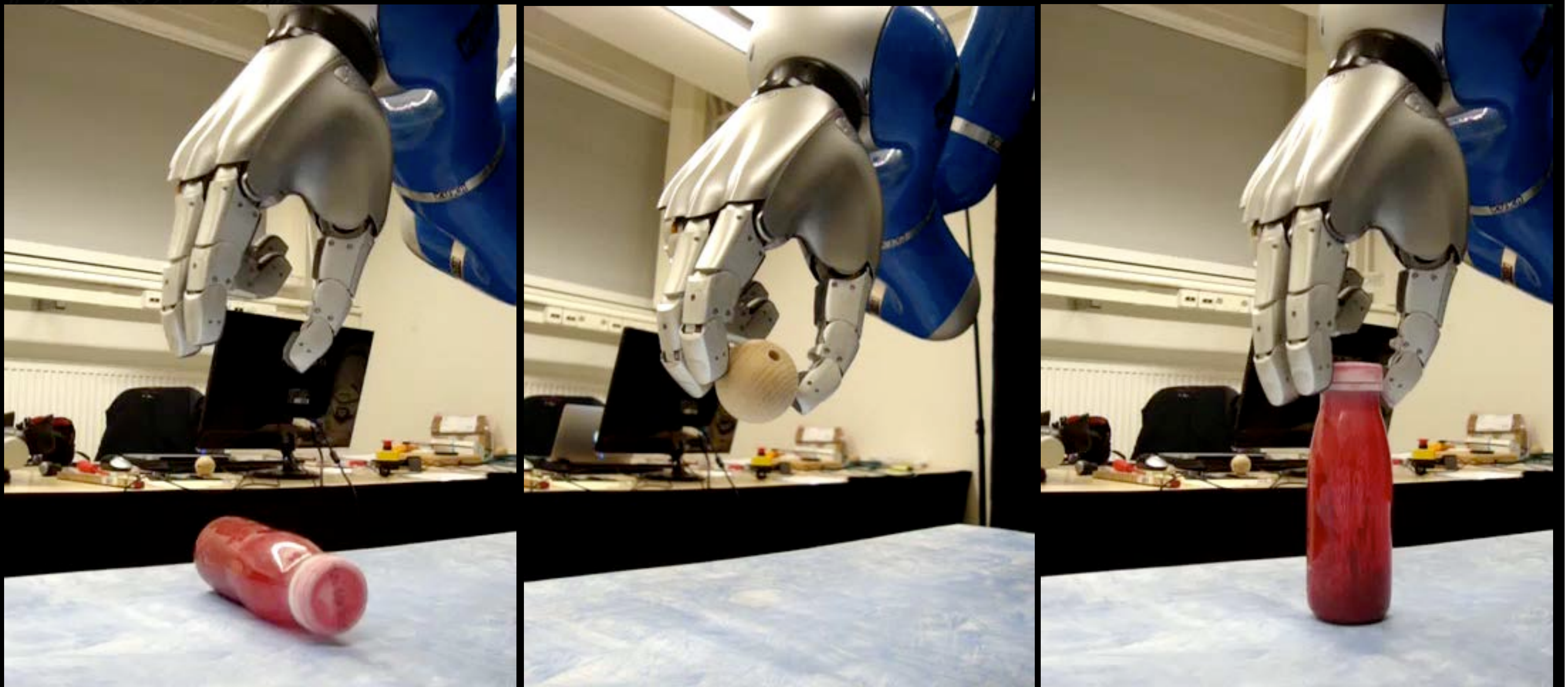
Kober, Oztop & Peters (2012). Goal Learning for Motor Primitives, Autonomous Robots

Throwing and Catching...



Kober, J; Muelling, K.; Peters, J. (2012). Learning Throwing and Catching Skills, IROS

Grasping and Manipulation

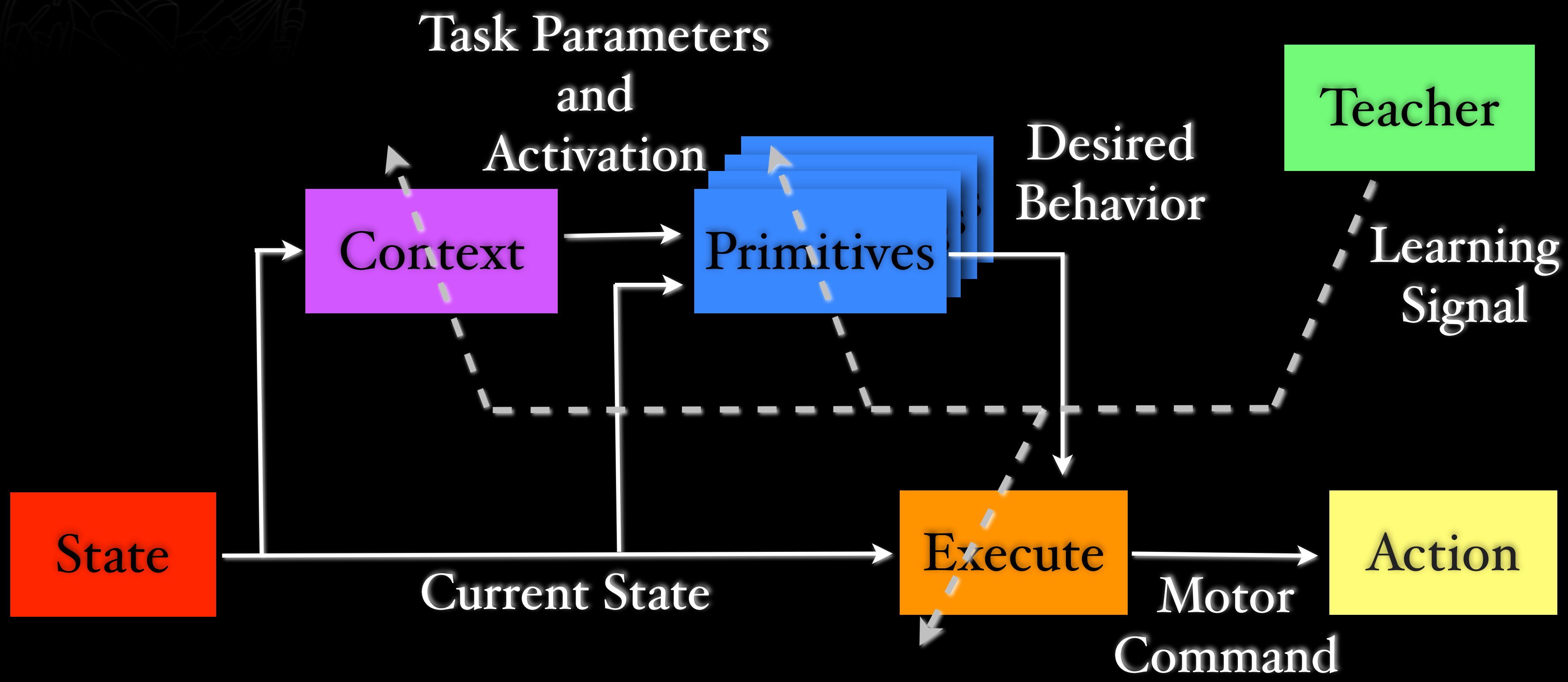


Kroemer, O.; van Hoof, H.; Neumann, G.; Peters, J. (2014). Learning to Predict Phases of Manipulation Tasks as Hidden States, Proceedings of 2014 IEEE International Conference on Robotics and Automation (ICRA).

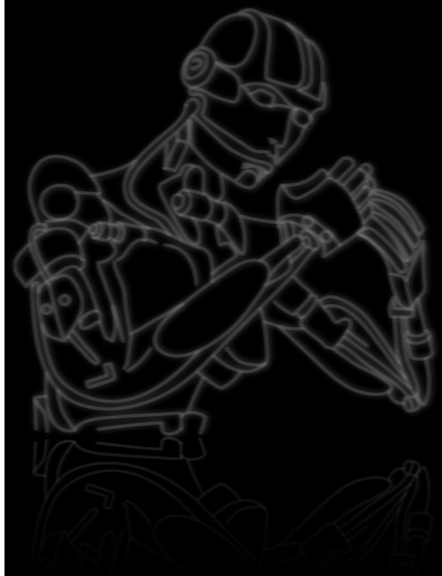
Outline

1. Introduction
2. How can we develop efficient motor learning methods?
3. How can anthropomorphic robots learn basic skills similar to humans?
4. Can complex skills be composed with these elements?
5. Outlook
6. Conclusion

Composition



Let us put all these elements together!



Applying the Whole Framework

Steps to Learned Table Tennis Player:

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

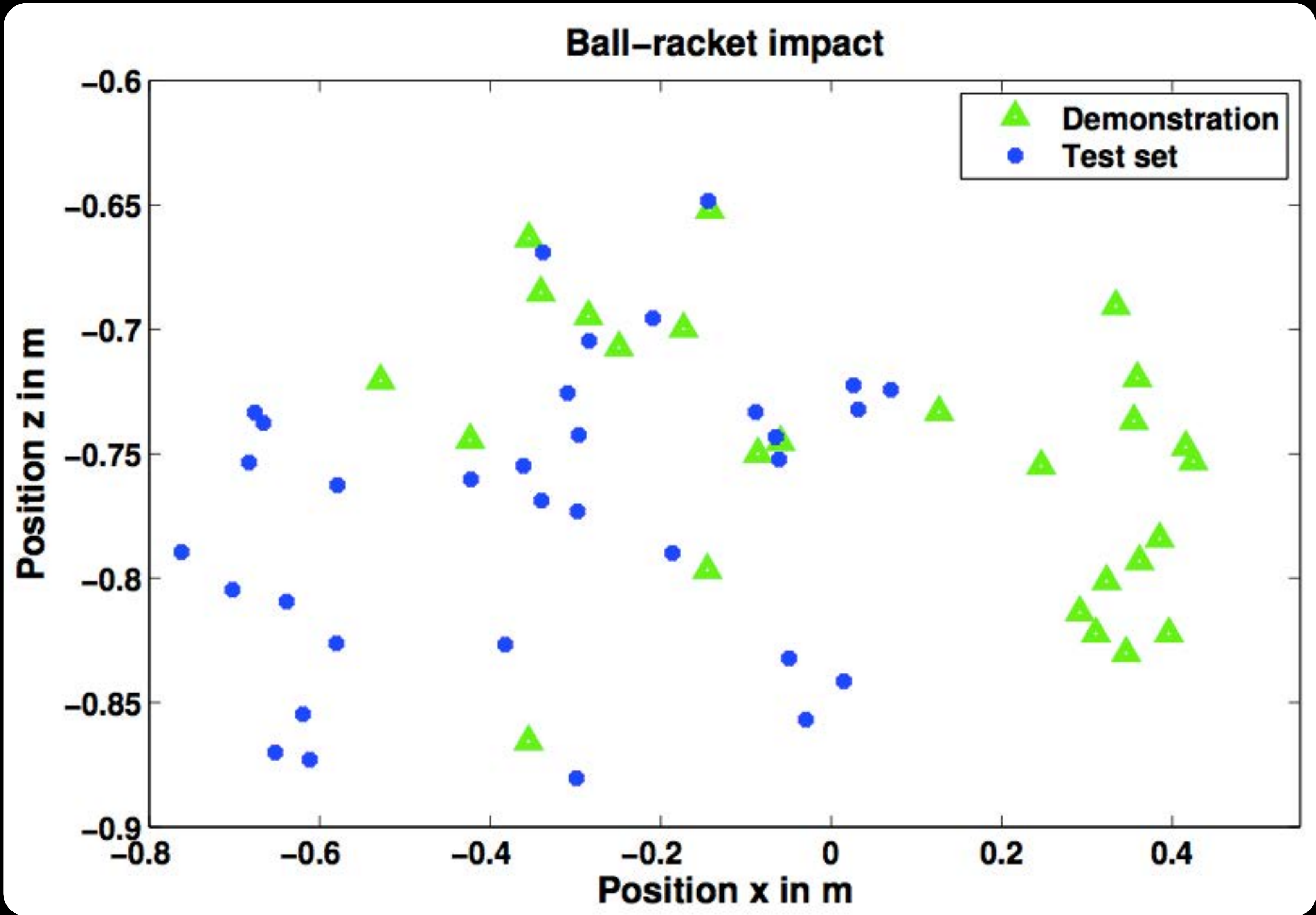
Mülling, K.; Kober, J.; Kroemer, O.; Peters, J. (2013). Learning to Select and Generalize Striking Movements in Robot Table Tennis, International Journal on Robotics Research.

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

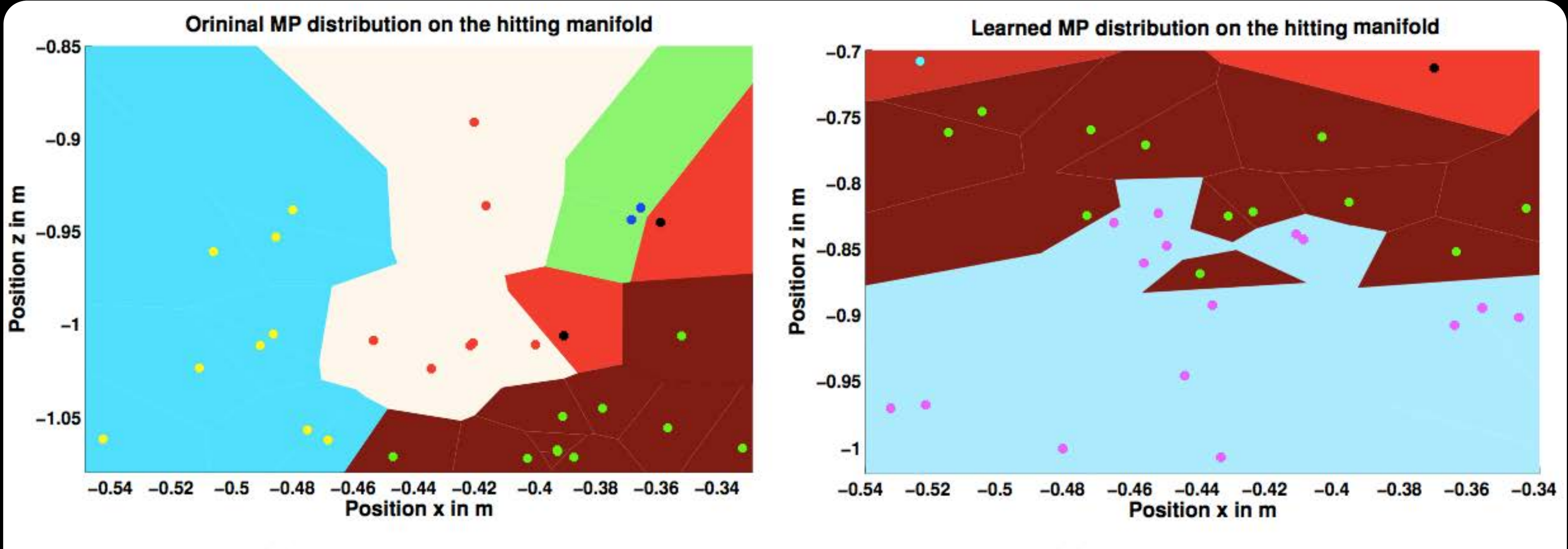


Mülling, K.; Kober, J.; Kroemer, O.; Peters, J. (2013). Learning to Select and Generalize Striking Movements in Robot Table Tennis, International Journal on Robotics Research.

Self-Improvement

Müllng, 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.

Current Problems

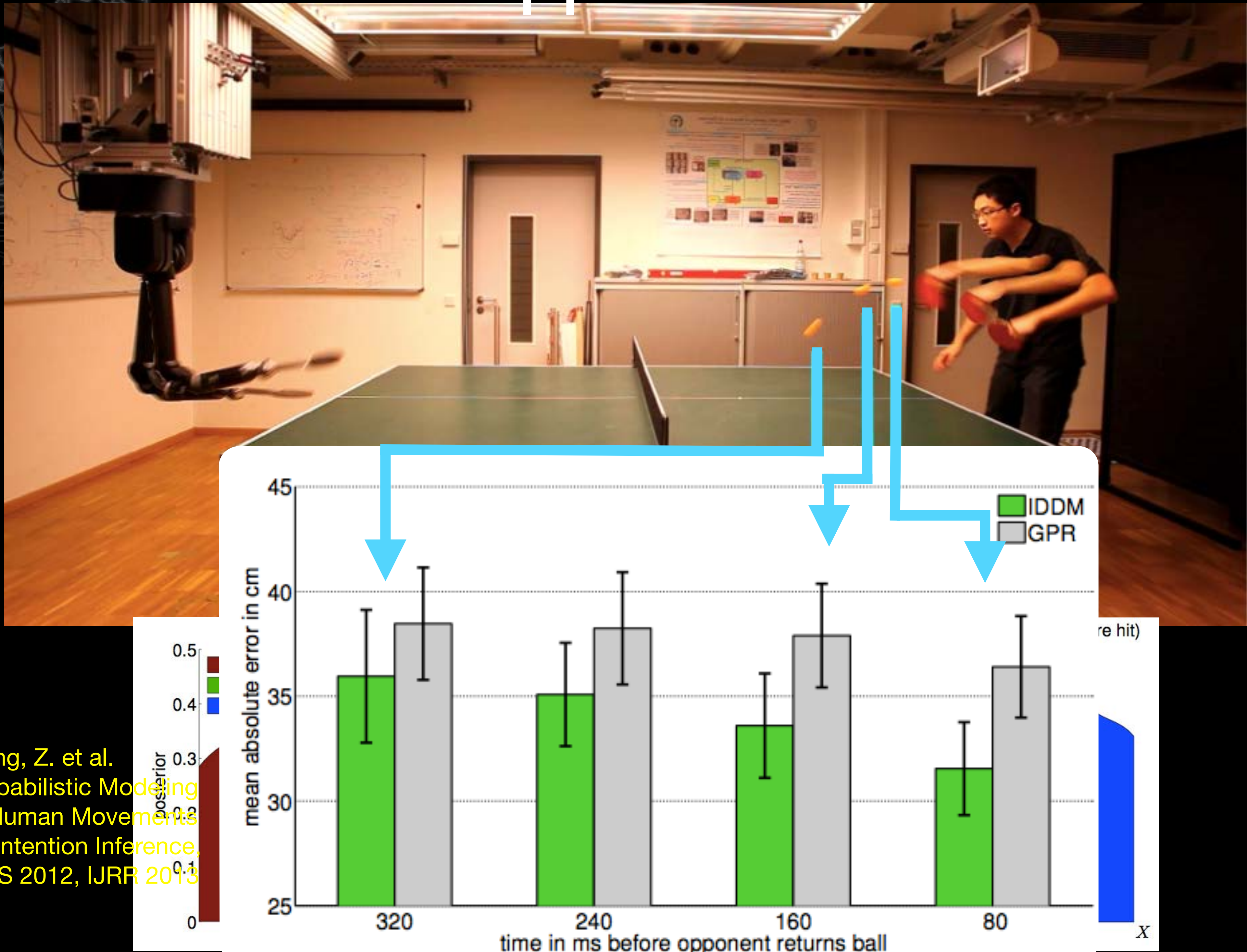
Problem I: Workspace is too limited.

Problem II: Arm accelerations are too low.

Problem III: Limited reaction time.

Problem III: Reaction Time

Reactive Opponent Prediction



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

Opponent Predictiom

Probabilistic Modeling of Human Movements for Intention Prediction

prototype system

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

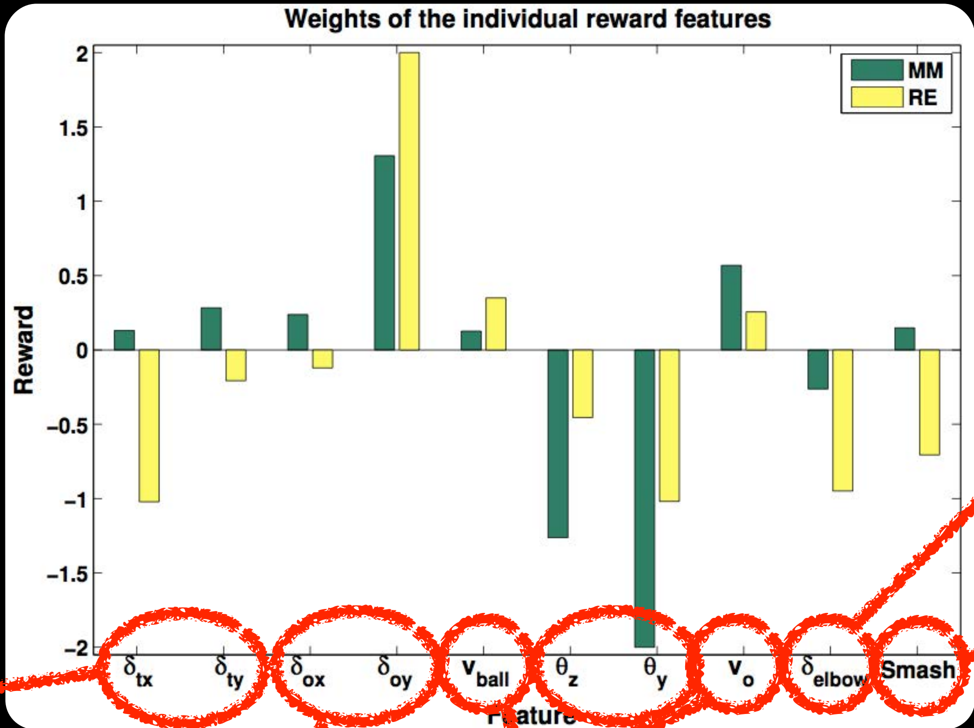
Wang, Z. et al. Probabilistic Modeling of Human Movements for Intention Inference, R:SS 2012, IJRR 2013

Extracting Strategies from Game Play

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

Extracting Strategies from Game Play

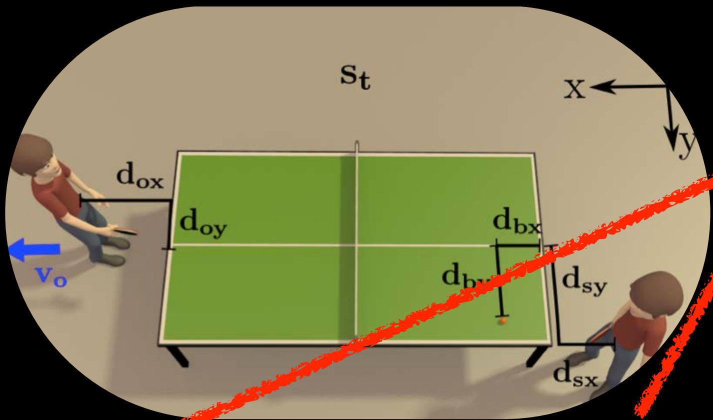
Weights of the most relevant features!



Opponent Elbow

Smash or not

Distance to the Edge of the Table



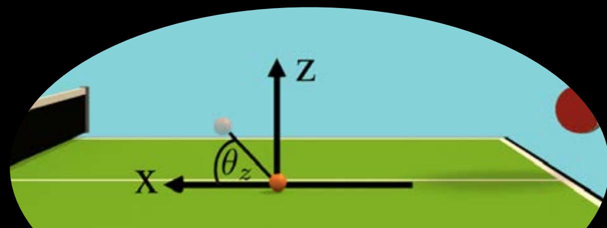
Angle of Incoming Bouncing Ball

Velocity of the Ball

Movement Direction of the Opponent

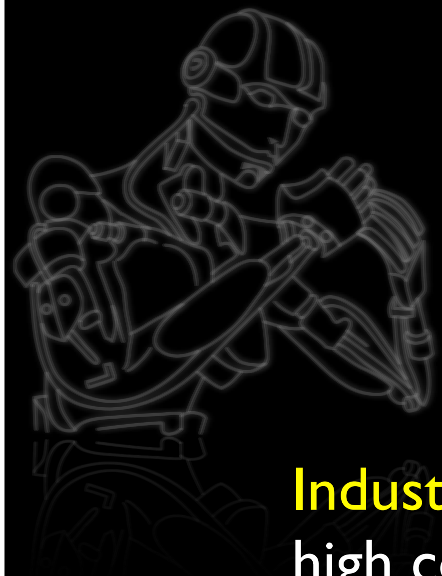
Distance to the Opponent

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



Outline

1. Introduction
2. How can we develop efficient motor learning methods?
3. How can anthropomorphic robots learn basic skills similar to humans?
4. Can complex skills be composed with these elements?
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.*

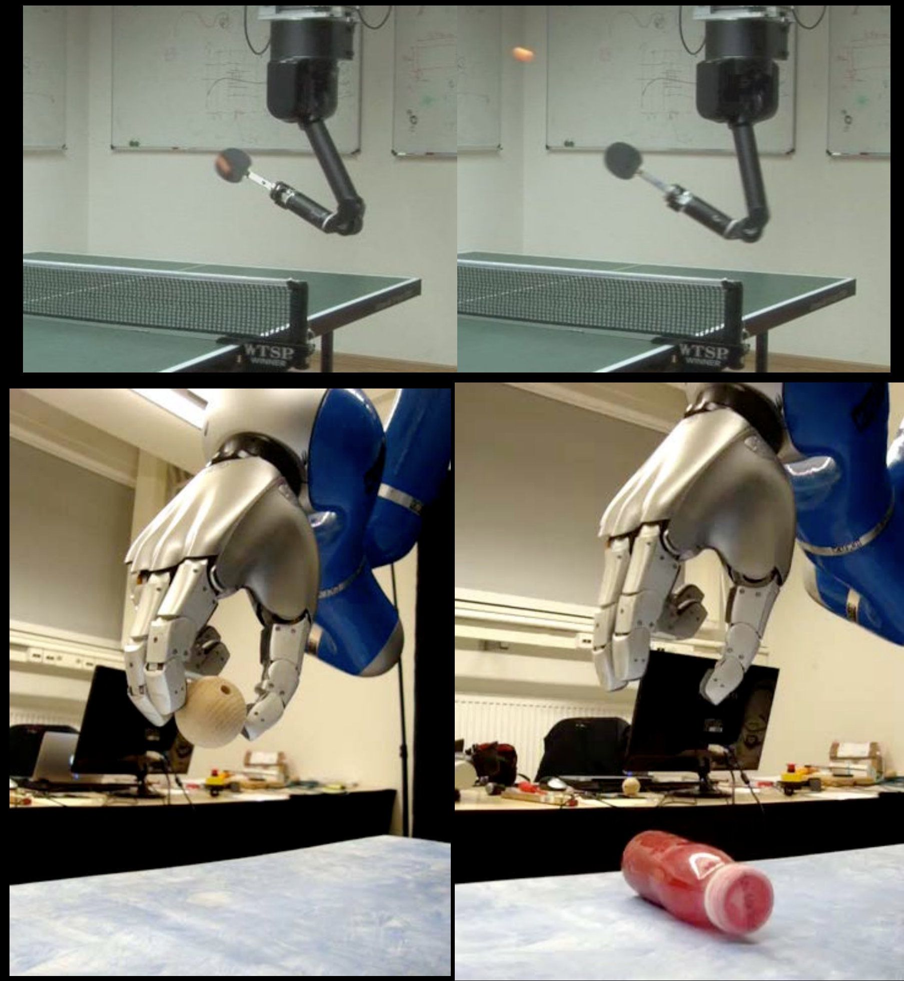
Assistive Robots: In hospital and rehablitation institutions, nurses need to “program” the robot – not computer scientists.

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

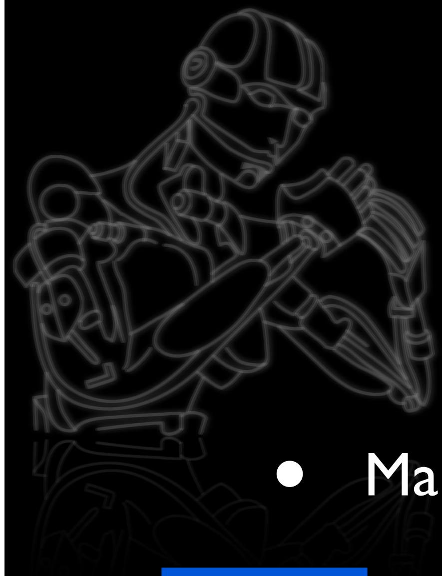
Transfer from Robot Table Tennis

Grasping with Dynamic Motor Primitives

- Hitting a ball: Velocity at hitting point
- Reaching and grasping
  - Avoiding obstacles
  - Approach direction
  - Adjusting fingers to object



Kroemer, O.; Detry, R.; Piater, J.; Peters, J. (2010). Grasping with Vision Descriptors and Motor Primitives, *(ICINCO)*.



Phases of Manipulation

- Manipulations consist of sequences of phases*

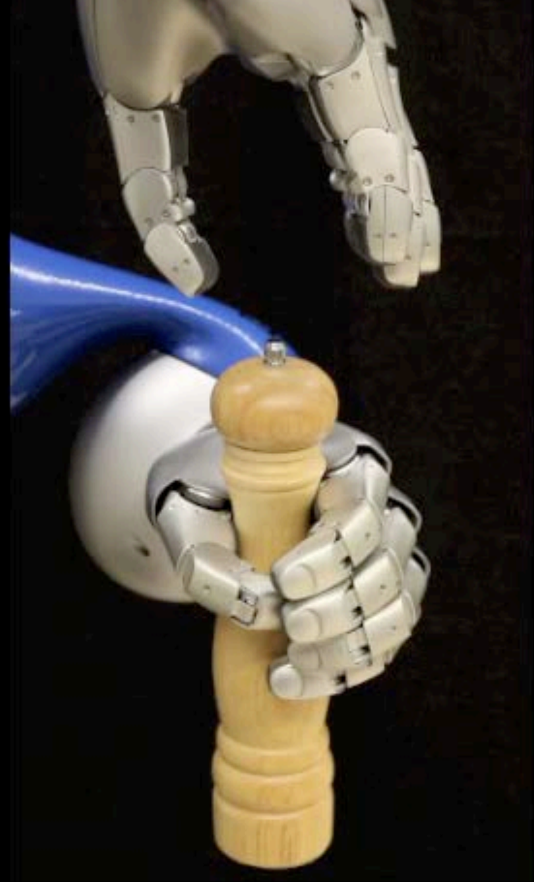
→ Reach → Load → Lift → Hold → Replace → Unload →

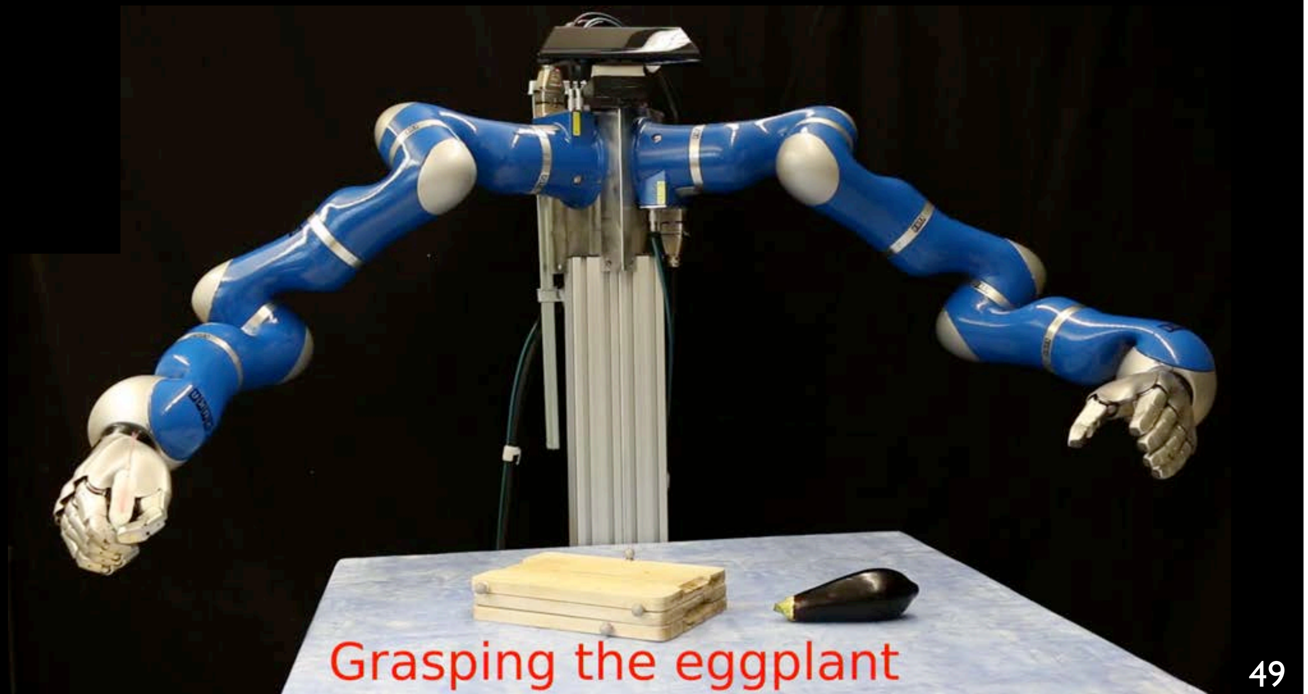
- Effects of actions depend on the current phase

- Phase transitions are constraints and subgoals of tasks

Kroemer, O.; van Hoof, H.; Neumann, G.; Peters, J. (2014). Learning to Predict Phases of Manipulation Tasks as Hidden States, ICRA

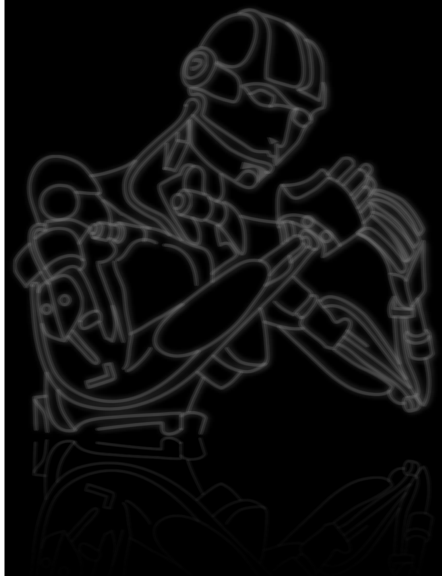
Transfer from Robot Table Tennis: First Examples



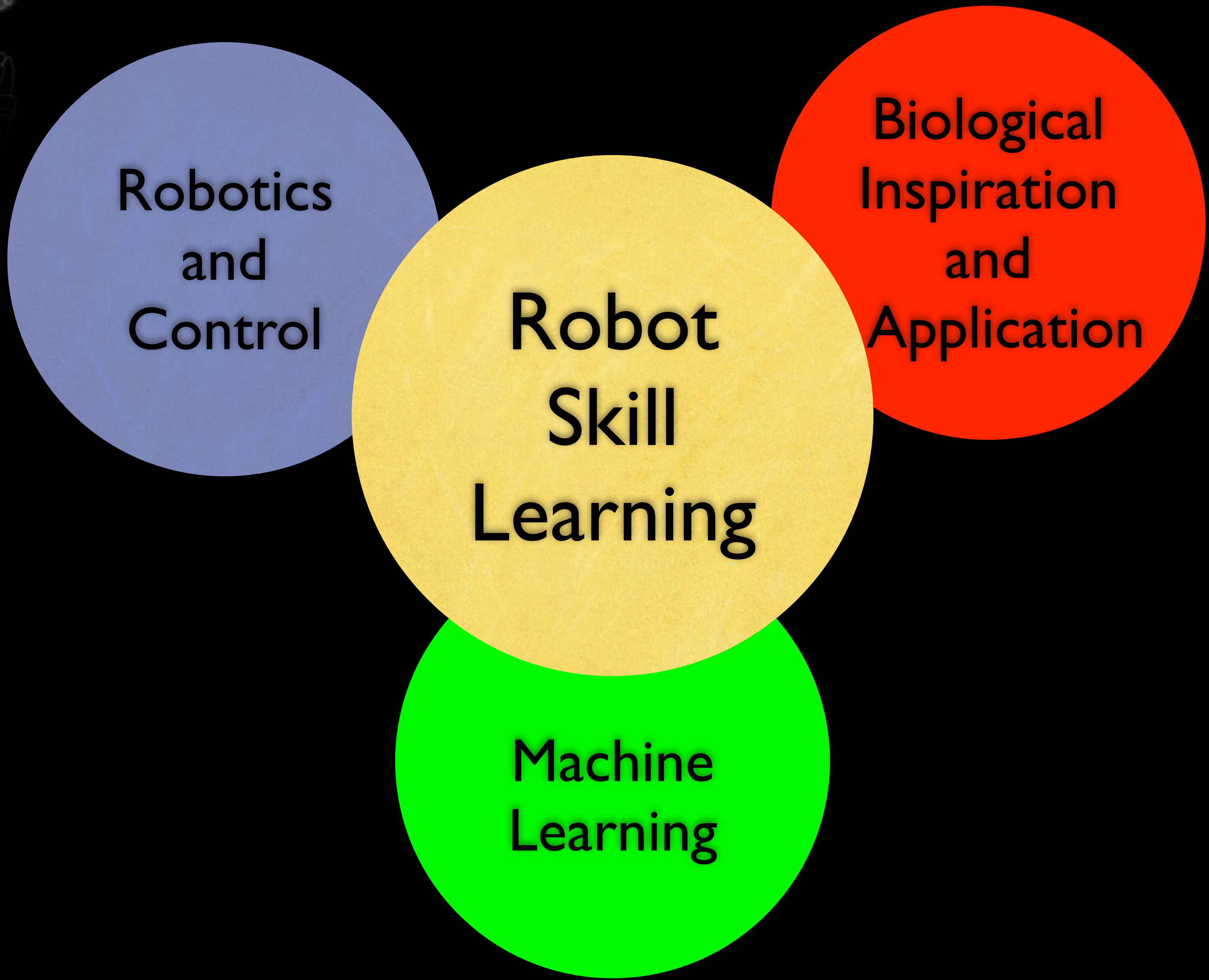


Kroemer, O.; van Hoof, H.; Neumann, G.; Peters, J. (2014). Learning to Predict Phases of Manipulation Tasks as Hidden States, Proceedings of 2014 IEEE International Conference on Robotics and Automation (ICRA).

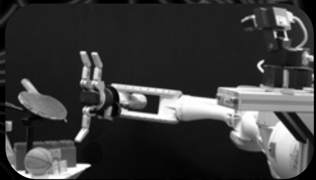
Lioutikov, R.; Kroemer, O.; Peters, J.; Maeda, G. (2014). Learning Manipulation by Sequencing Motor Primitives with a Two-Armed Robot, Proceedings of the 13th International Conference on Intelligent Autonomous Systems (IAS).



Outlook

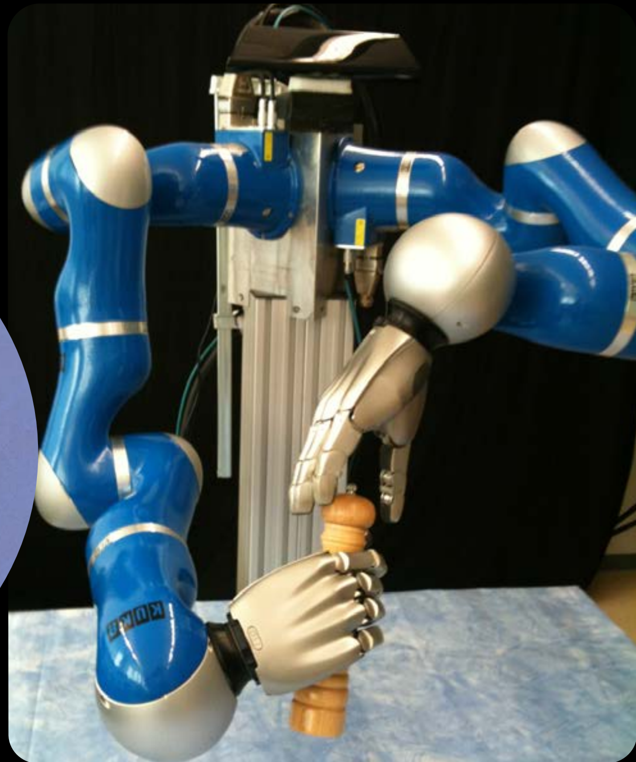

Robotics and Control
Robot Skill Learning
Biological Inspiration and Application
Machine Learning

Robotics & Control

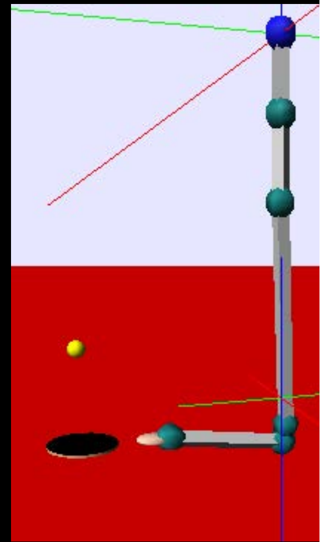


Robot Grasping
and Manipulation
(Krömer, Peters, Robotics & Autonomous Systems, 2010)

Robotics
and
Control

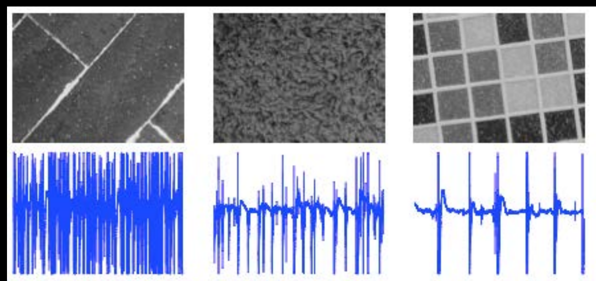


Real-Time Software &
Simulations for Robots



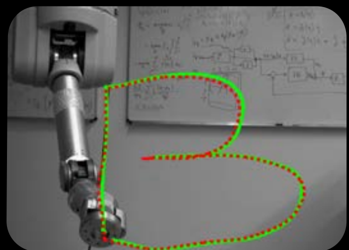
High-Speed
Real-Time Vision
(Lampert & Peters, Journal of Real-Time Vision)

Physics as prior
for Learning in
Planning & Control
(Nguyen-Tuong & Peters, ICRA 2010)



Tactile Sensing &
Sensory Integration
(Kroemer, Lampert & Peters, IEEE Trans. Robotics, 2011)

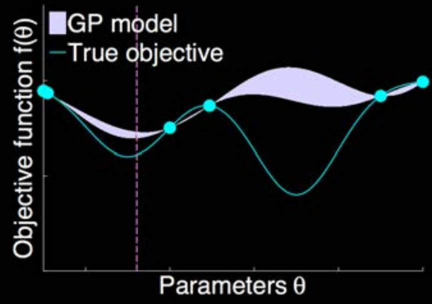
Optimal Control
(Kroemer & Peters, NIPS 2011)



Nonlinear Robot Control
(Peters et al, Autonomous Robots, 2008)

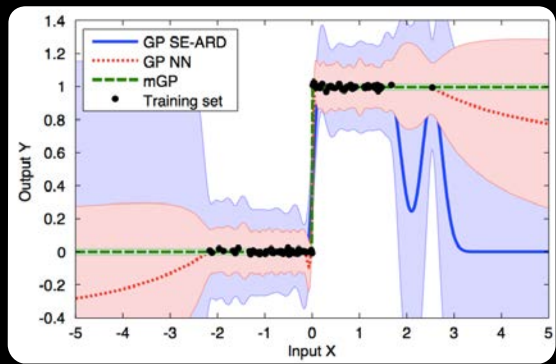
Machine Learning

Bayesian Optimization
(Calandra et al, 2014)



Much more Reinforcement Learning...
(Peters et al, Neural Networks 2008; Neurocomputing 2008)

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



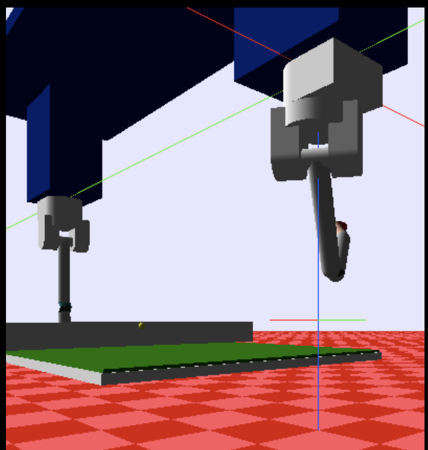
Maximum Entropy
(Peters et al., AAAI 2010; Daniel, Neumann & Peters, AIStats 2012)

Probabilistic Movement Representation
(Paraschos et al. NIPS 2013)

Manifold Gaussian Processes
(Calandra et al, 2014)

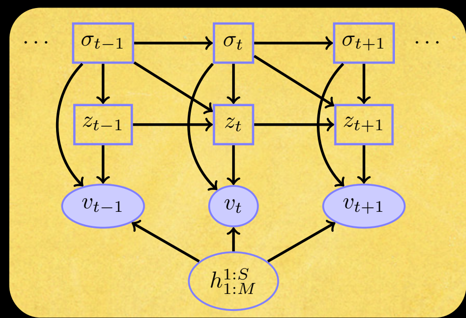
Policy Gradient Methods
(Peters et al, IROS 2006)

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



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

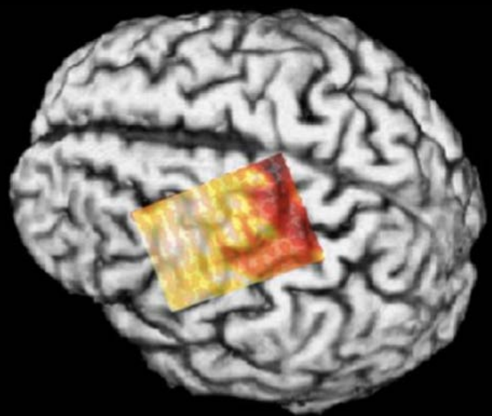
Machine Learning



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

Biological Inspiration and Application

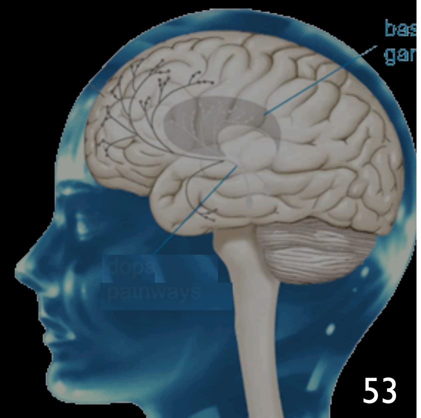
Brain-Computer Interfaces with ECoG for Stroke Patient Therapy
(Gomez, Peters & Grosse-Wentrup, Journal of Neuroengineering 2011)



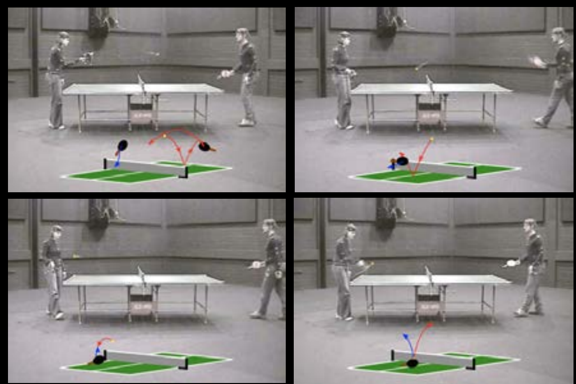
Brain Robot Interfaces
(Peters et al., Int. Conf. on Rehabilitation Robotics, 2011)

Biological Inspiration and Application

Computational Models of Motor Control & Learning

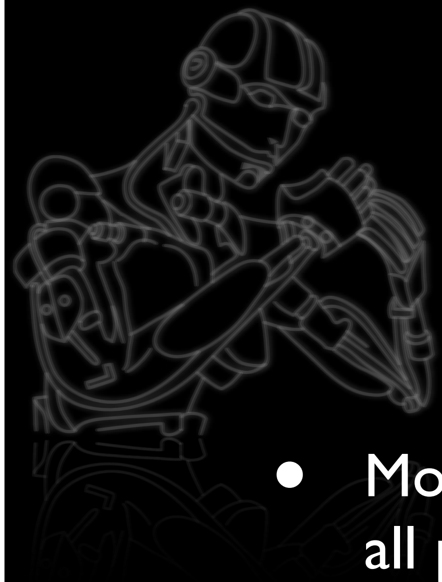


Understanding Human Movements
(Mülling, Kober & Peters, Adaptive Behavior 2011)



Outline

1. Introduction
2. How can we develop efficient motor learning methods?
3. How can anthropomorphic robots learn basic skills similar to humans?
4. Can complex skills be composed with these elements?
5. Outlook
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.
- Many interesting research topics benefit from this work!

Thanks for your Attention!