

# Microsoft Research

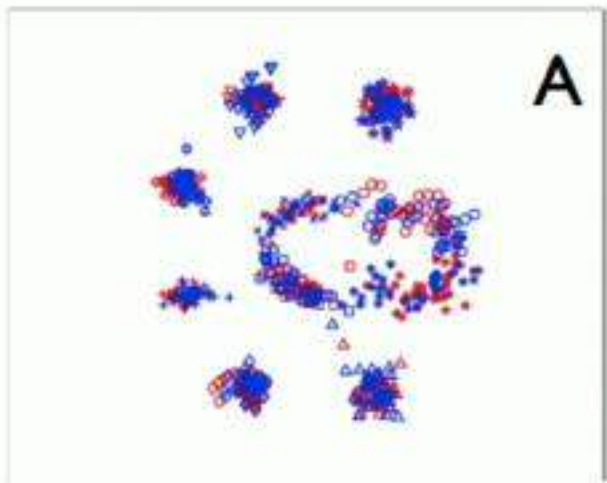
Each year Microsoft Research hosts hundreds of influential speakers from around the world including leading scientists, renowned experts in technology, book authors, and leading academics, and makes videos of these lectures freely available.  
2016 © Microsoft Corporation. All rights reserved.

# New Frontiers in Imitation Learning

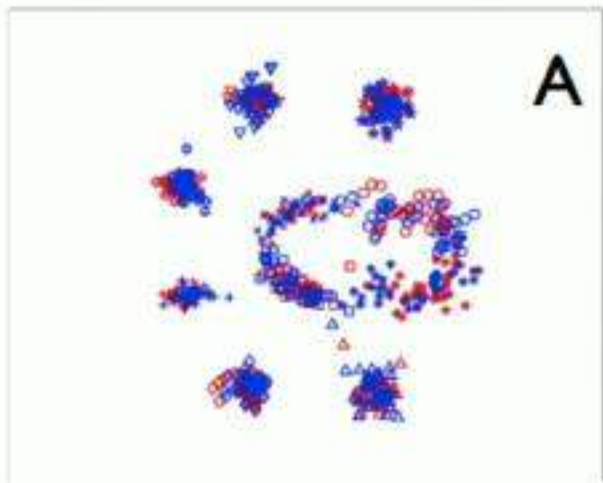
Yisong Yue

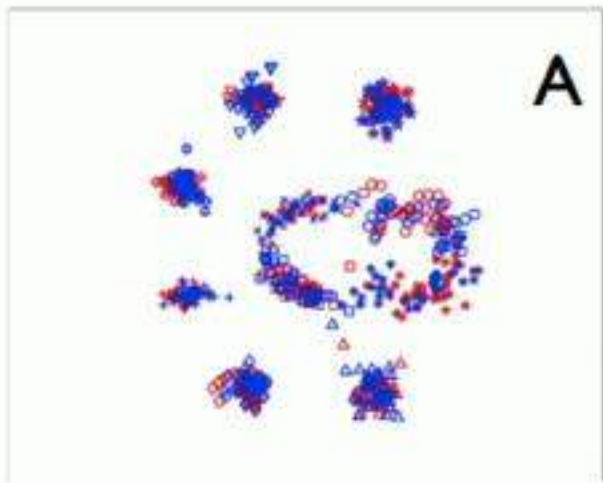
# Behavioral Modeling



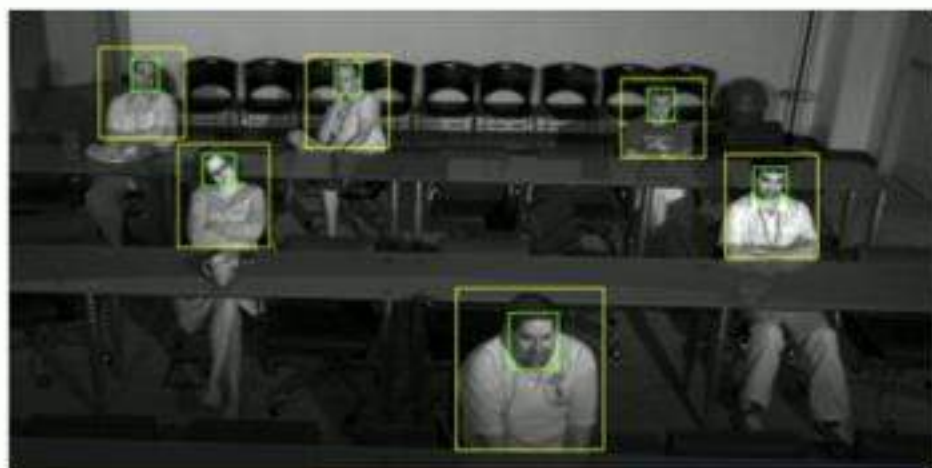
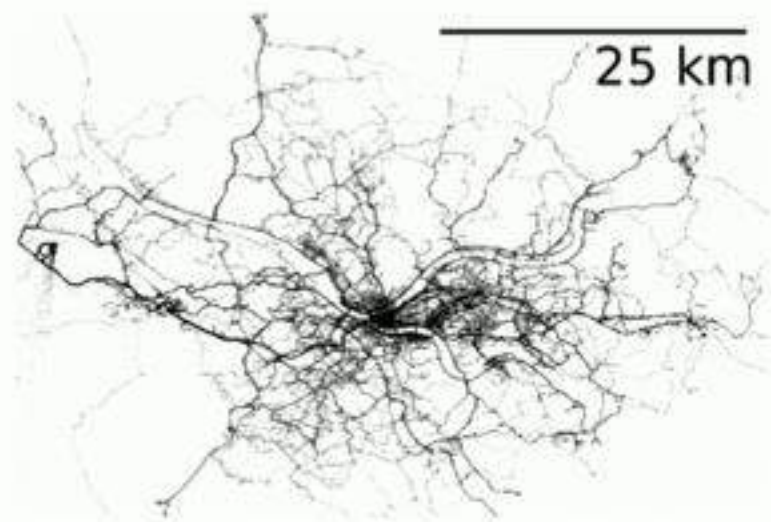
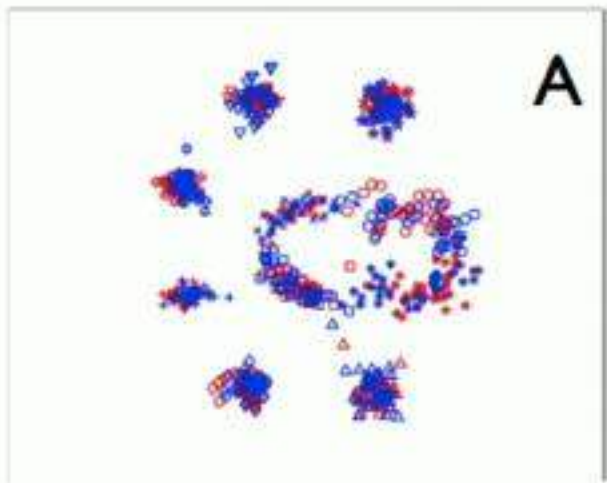




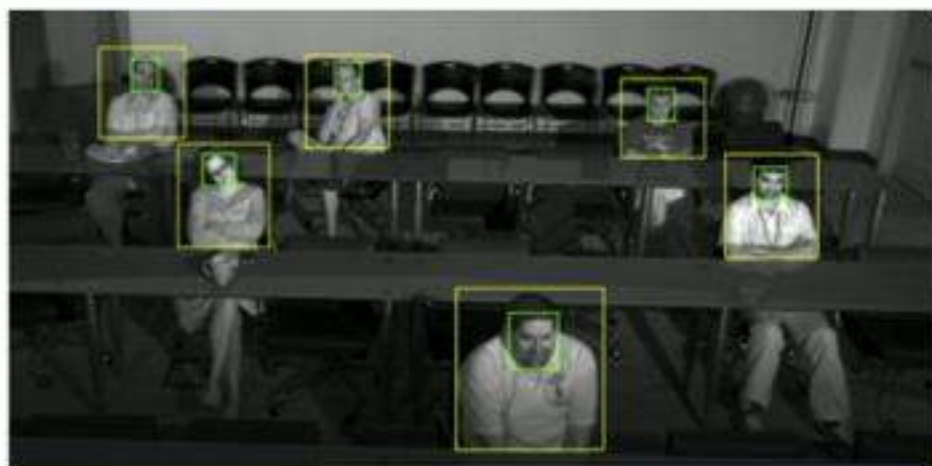
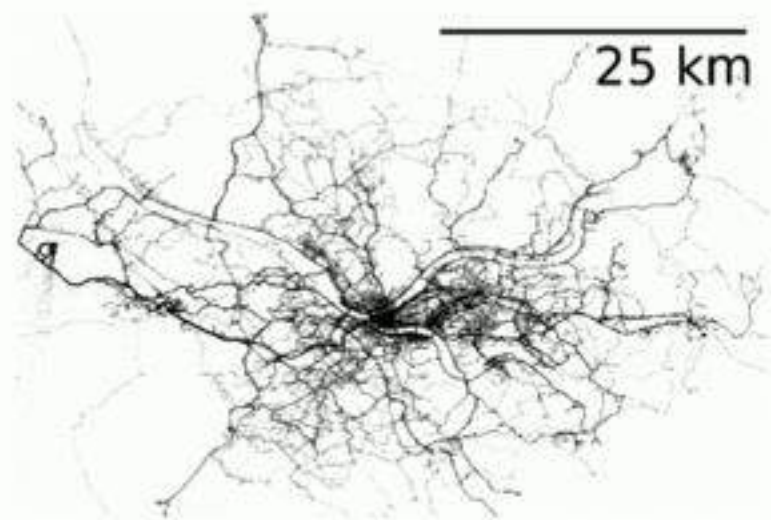
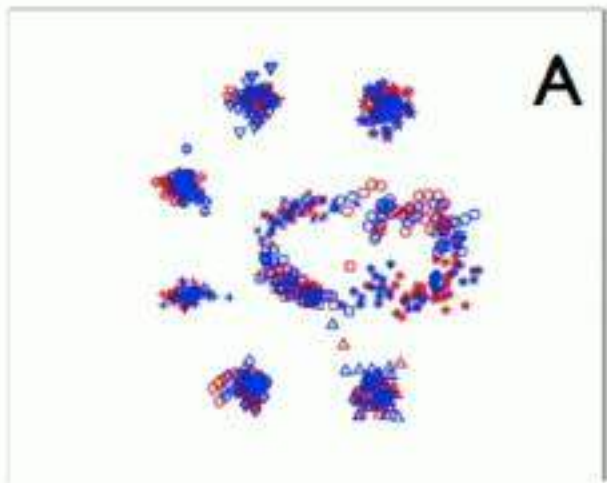




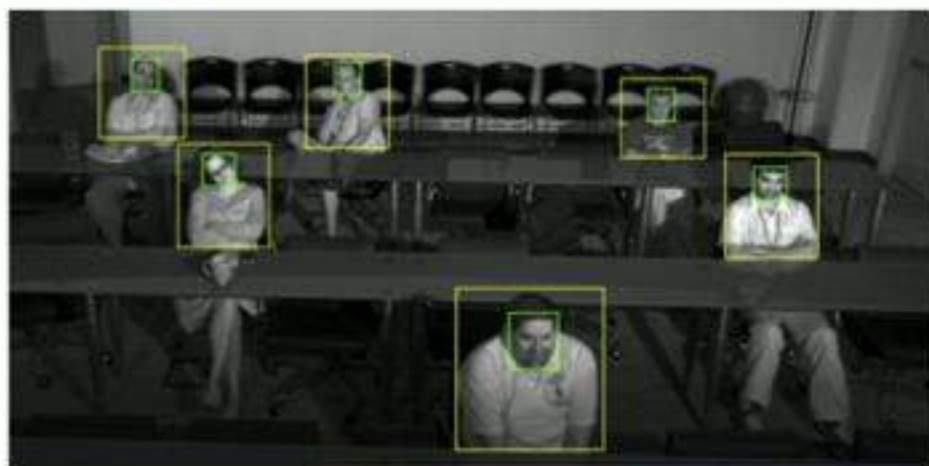
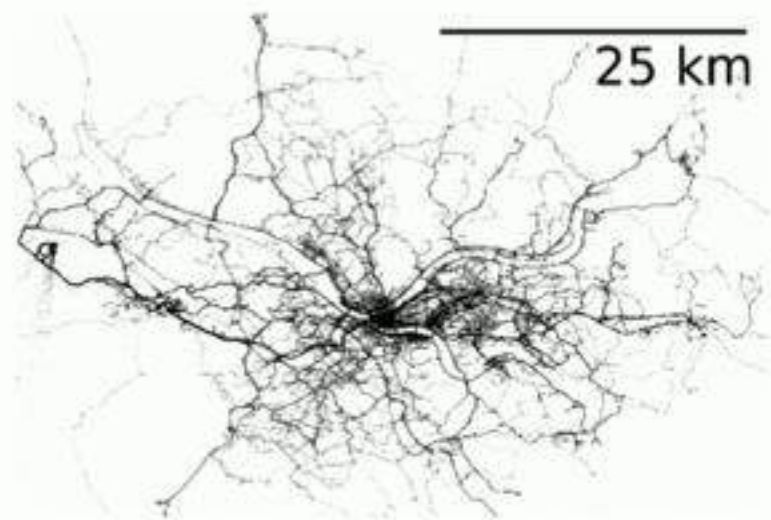
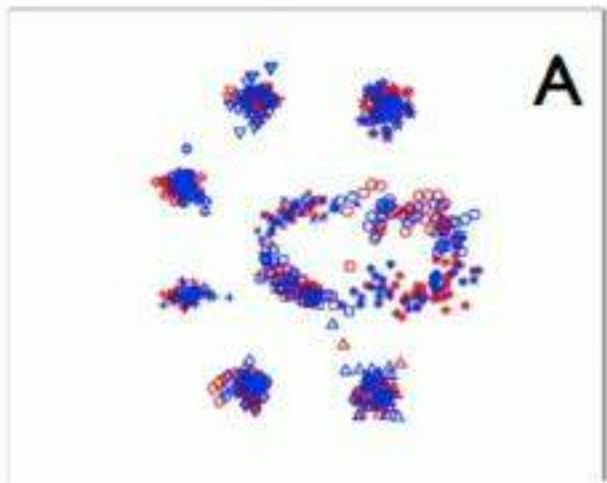




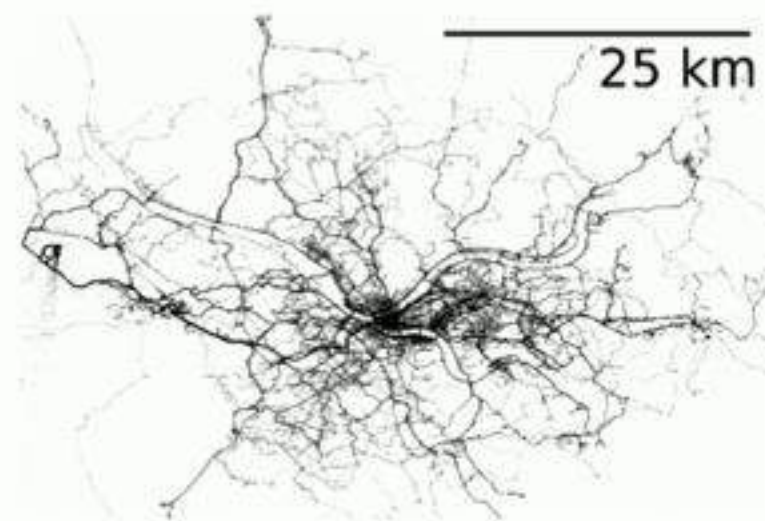
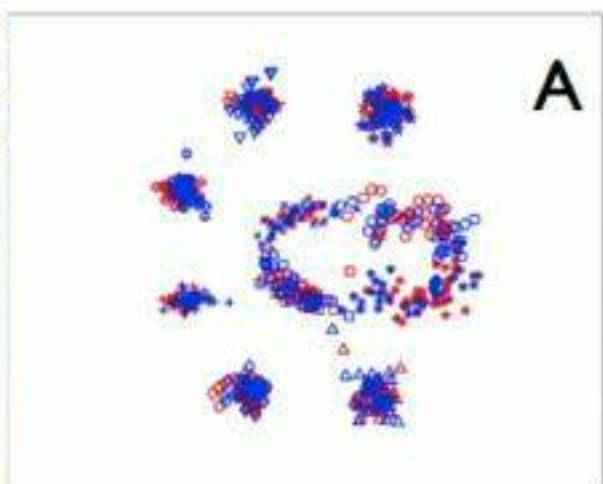














# Warm Up: Supervised Learning

- Find function from input space  $X$  to output space  $Y$

$$h : X \longrightarrow Y$$

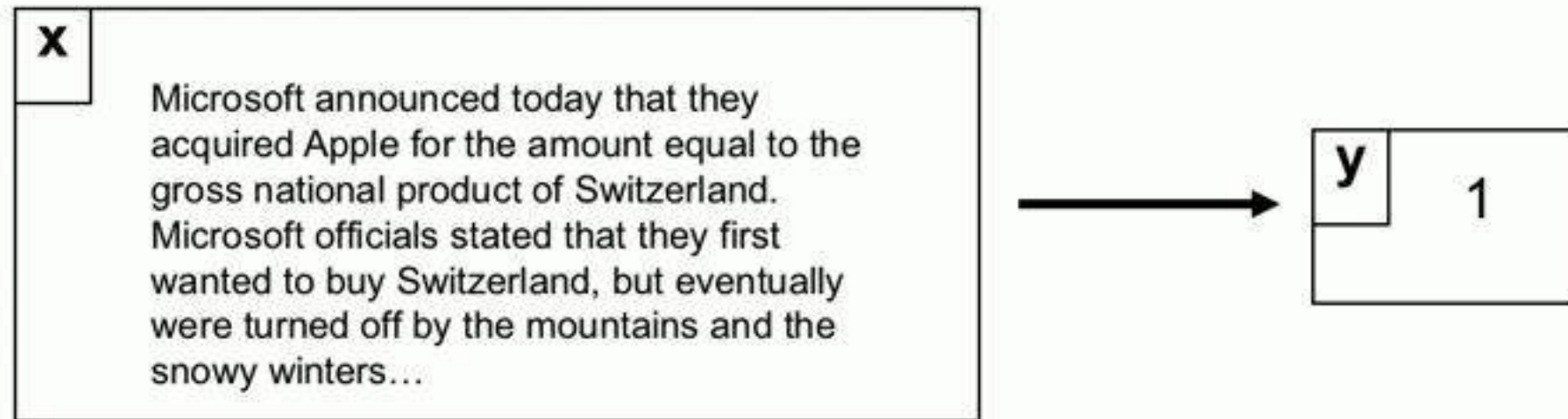
such that the prediction error is low.

# Warm Up: Supervised Learning

- Find function from input space  $X$  to output space  $Y$

$$h : X \longrightarrow Y$$

such that the prediction error is low.

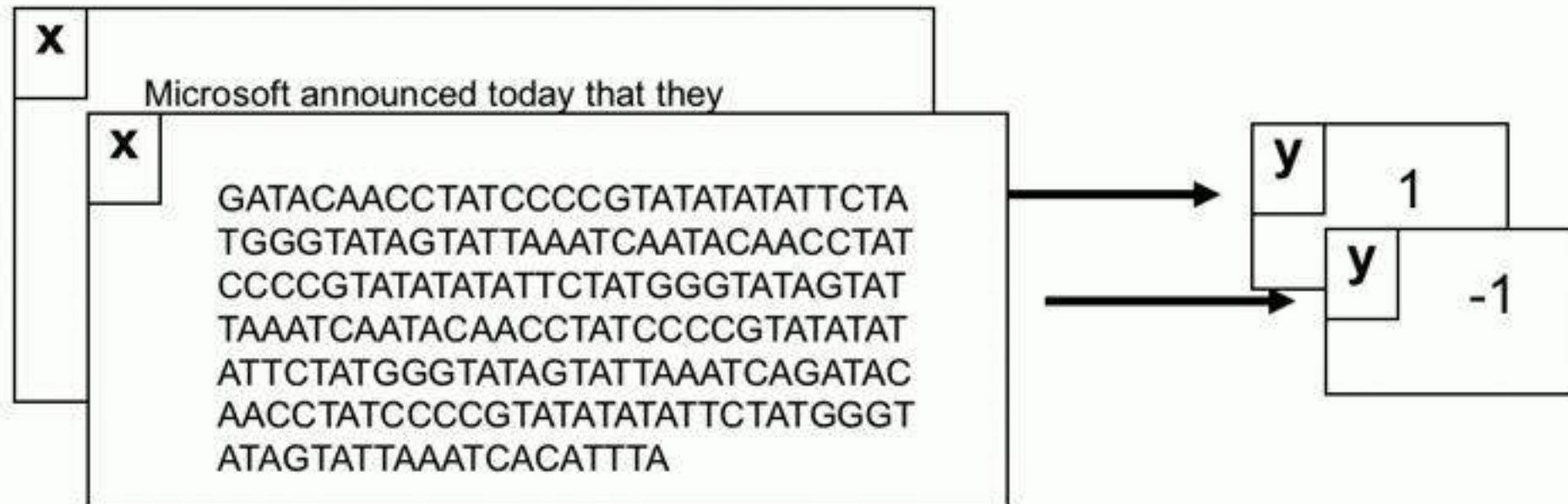


# Warm Up: Supervised Learning

- Find function from input space  $X$  to output space  $Y$

$$h : X \longrightarrow Y$$

such that the prediction error is low.



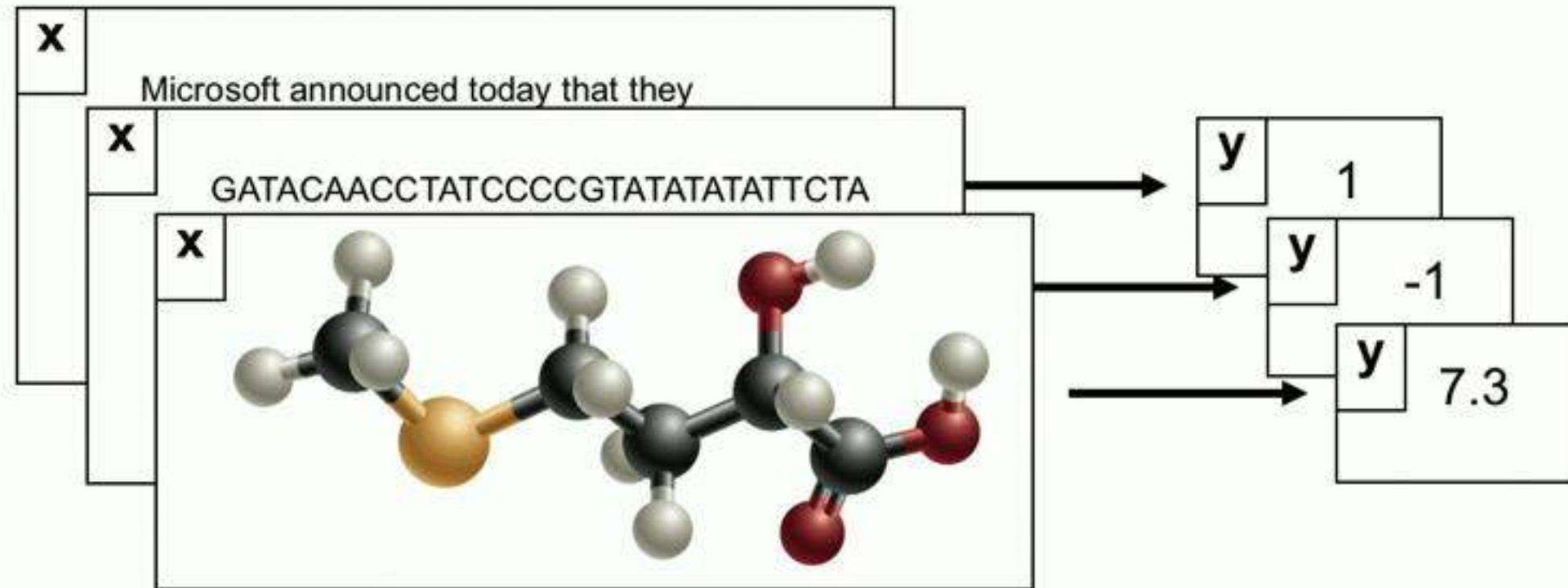


# Warm Up: Supervised Learning

- Find function from input space  $X$  to output space  $Y$

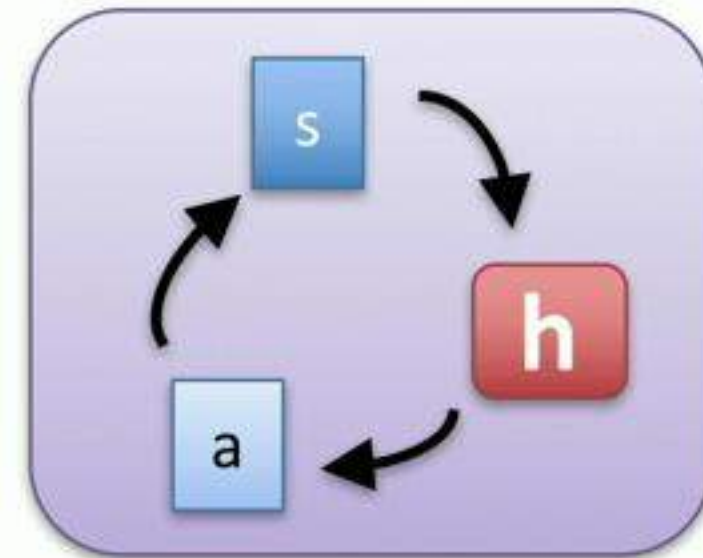
$$h : X \longrightarrow Y$$

such that the prediction error is low.



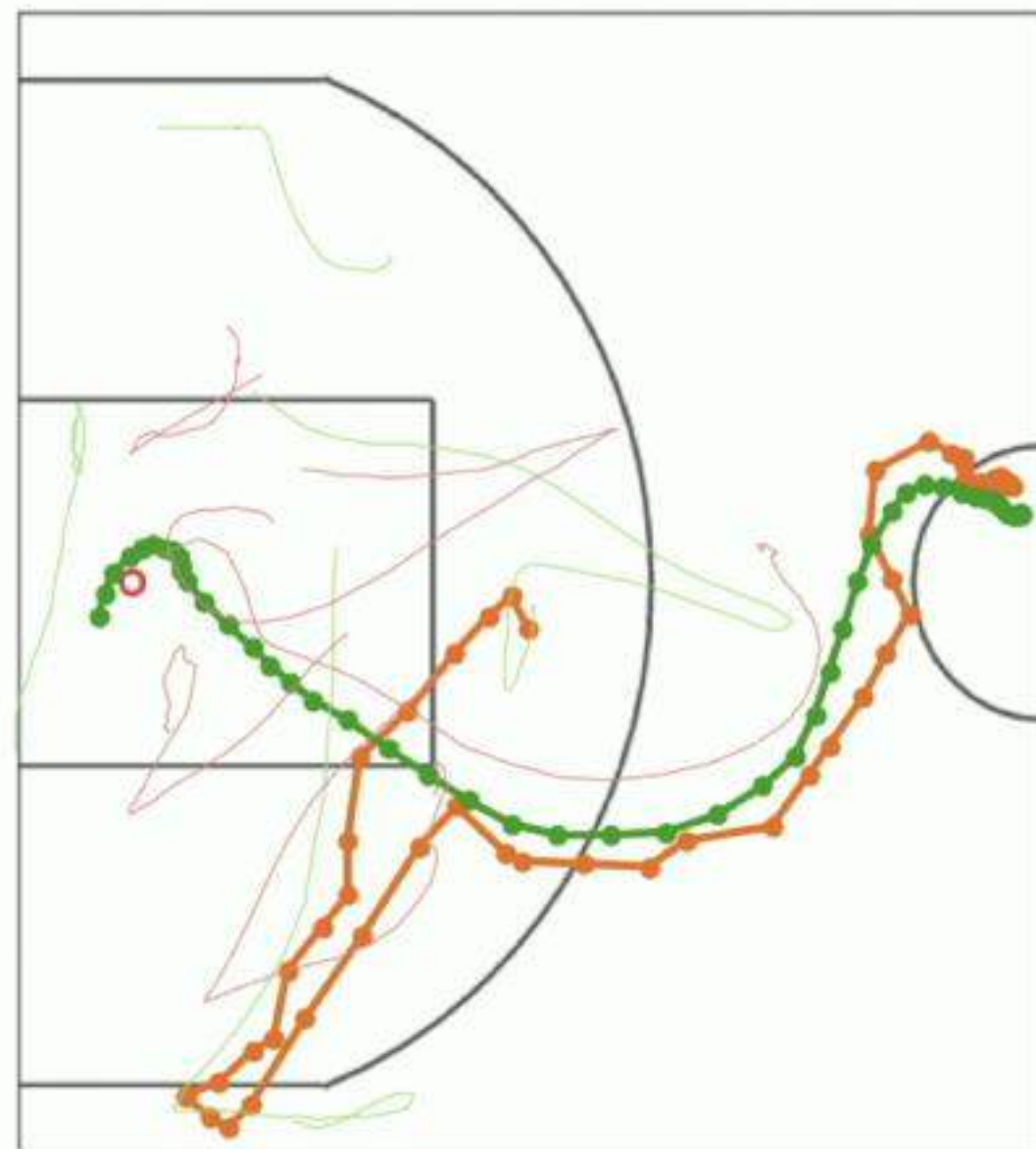
# Imitation Learning

- Input:
  - Sequence of contexts/states:
- Predict:
  - Sequence of actions
- Learn Using:
  - Sequences of demonstrated actions



# Example: Basketball Player Trajectories

- $s$  = location of players & ball
- $a$  = next location of player
- Training set:  $D = \{(\vec{s}, \vec{a})\}$ 
  - $\vec{s}$  = sequence of  $s$
  - $\vec{a}$  = sequence of  $a$
- **Goal:** learn  $h(s) \rightarrow a$





# What to Imitate?

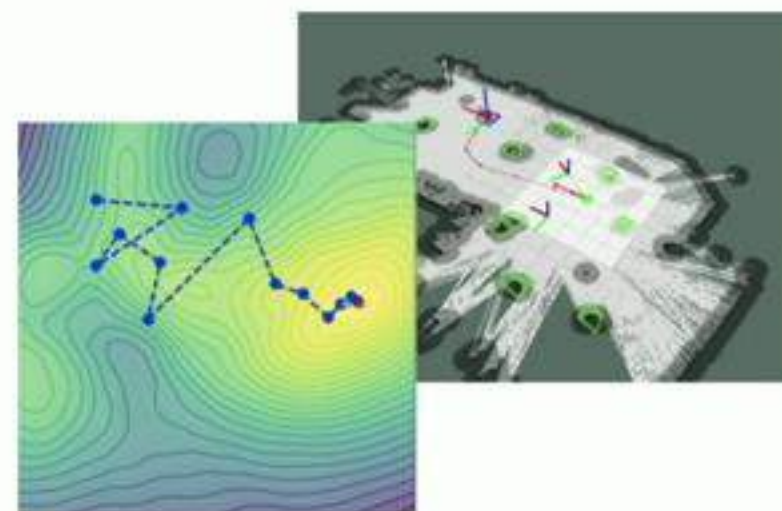
## Human Demonstrations



## Animal Demonstrations



## Computational Oracle



**Pre-collected  
Demonstrations**



**Oracle  
Querying  
& Online**

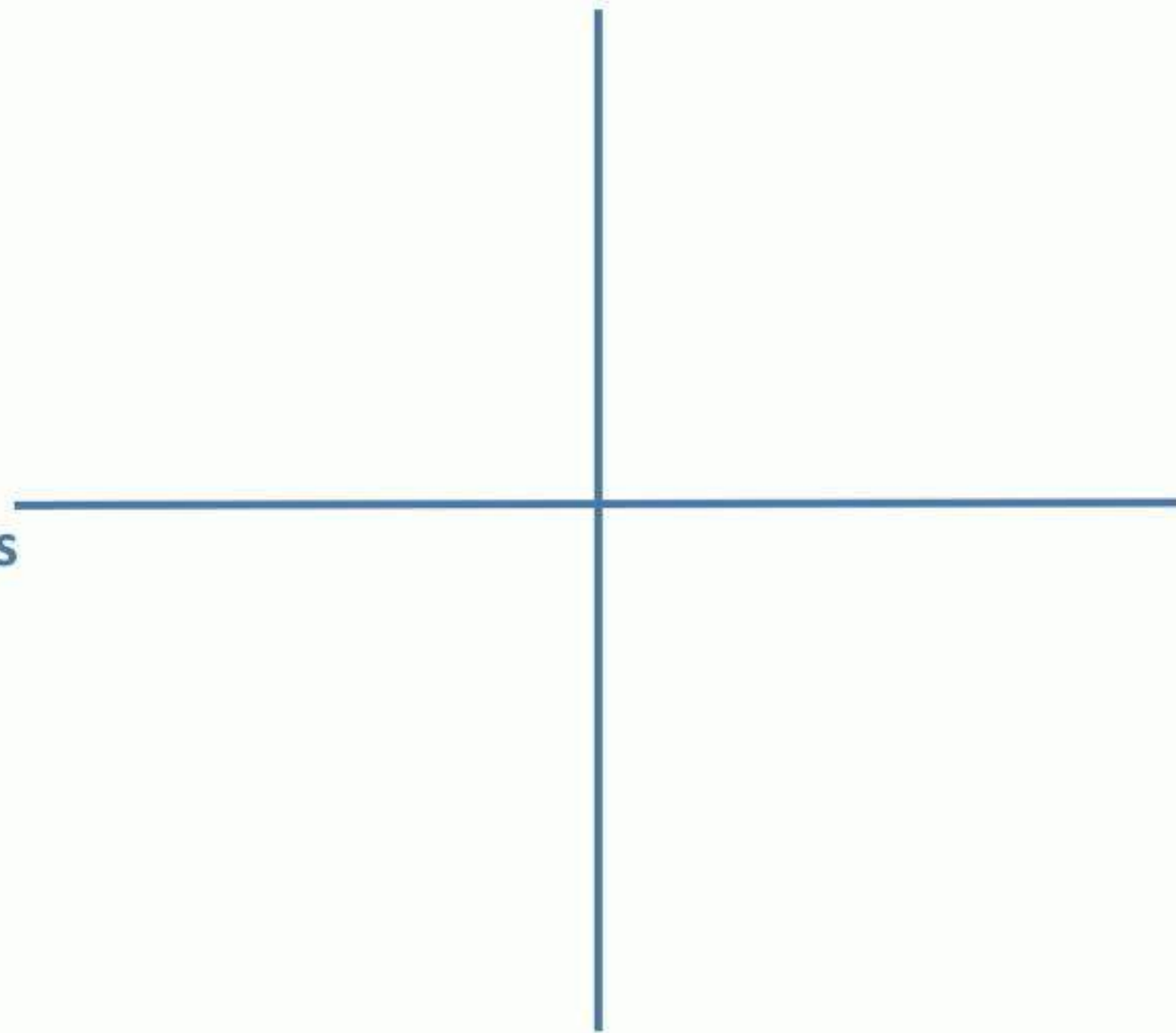


**Policy Learning**

**Pre-collected  
Demonstrations**

**Oracle  
Querying  
& Online**

**Value Function Learning**  
(Inverse Reinforcement Learning)



## Policy Learning

**Reduction to PAC**  
[Syed & Schapire 2007]

**Dagger**  
[Ross et al., 2011]

**Autonomous Navigation**  
[Pomerleau 1991]

**SEARN**  
[Daume et al., 2009]

**Pre-collected  
Demonstrations**

**Oracle  
Querying  
& Online**

**GAIL**  
[Ho & Ermon 2016]

**DARKO**  
[Rhinehart & Kitani, 2016]

**MaxEnt IRL**  
[Ziebart et al., 2008]

**Bellman Gradient Iteration**  
[Li & Burdick, 2017]

**Apprenticeship Learning**  
[Abbeel & Ng, 2004]

**Value Function Learning**  
(Inverse Reinforcement Learning)



## Policy Learning

**Reduction to PAC**  
[Syed & Schapire 2007]

**Dagger**  
[Ross et al., 2011]

**Autonomous Navigation**  
[Pomerleau 1991]

**SEARN**  
[Daume et al., 2009]

### Previous (Deep Imitation) Work:

- Minimal assumptions
- Inefficient in complex & structured settings

**MaxEnt IRL**  
[Ziebart et al., 2008]

[Rhinehart & Kitani, 2016]

**Apprenticeship Learning**  
[Abbeel & Ng, 2004]

**Bellman Gradient Iteration**  
[Li & Burdick, 2017]

## Value Function Learning (Inverse Reinforcement Learning)

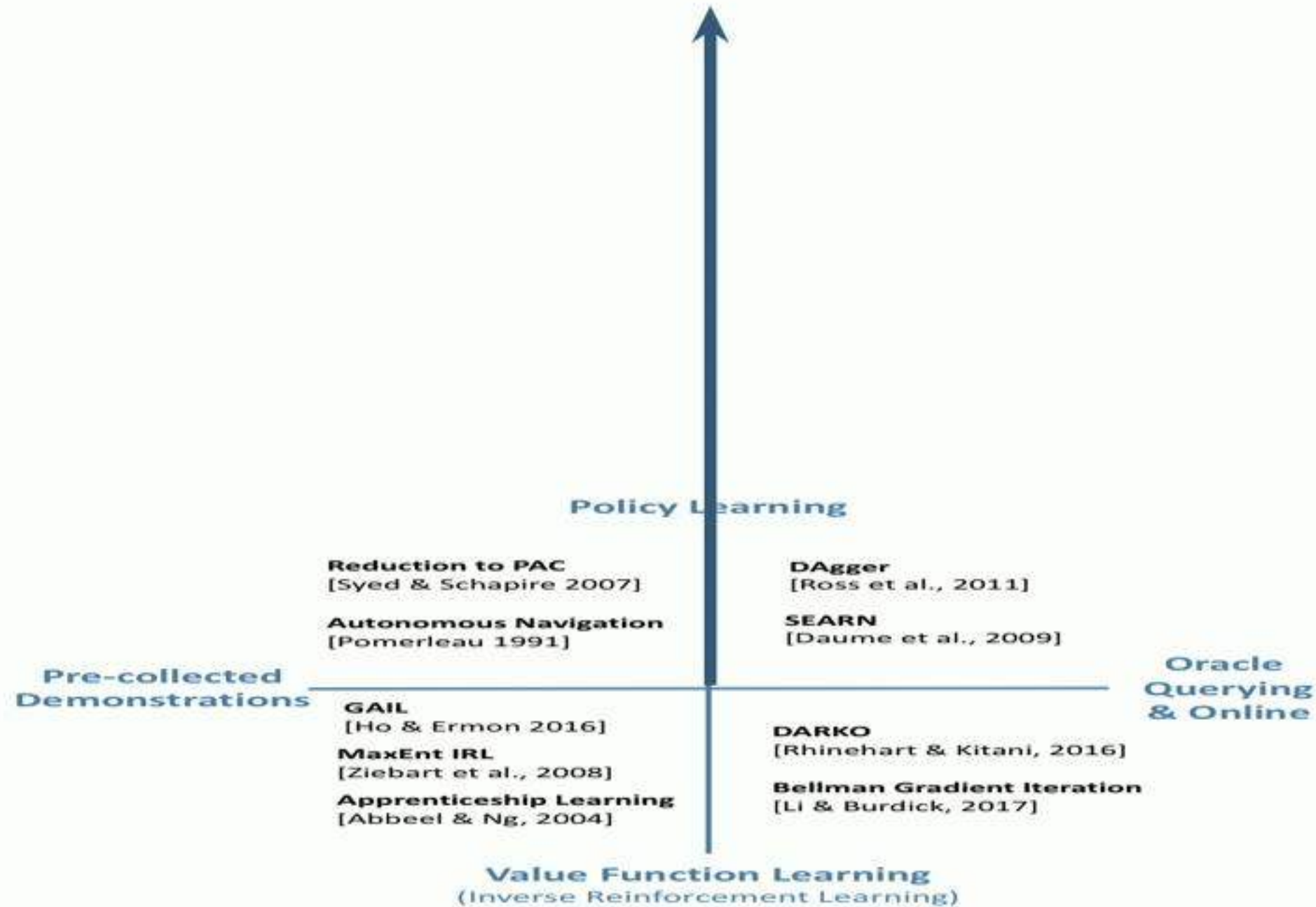
Pre-colle  
Demonstra

acle  
erying  
online

# Structured Imitation Learning



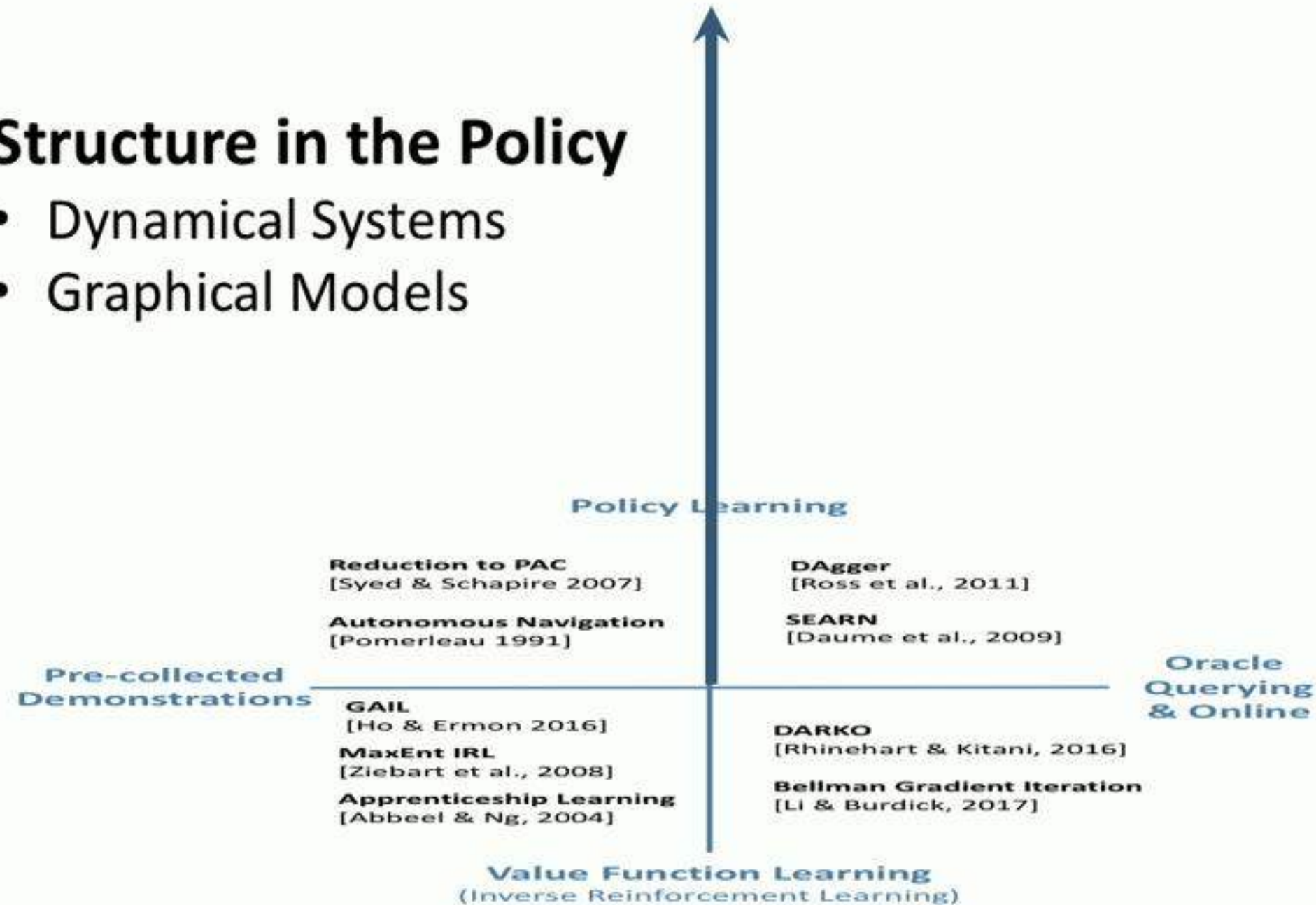
# Structured Imitation Learning



# Structured Imitation Learning

## Structure in the Policy

- Dynamical Systems
- Graphical Models



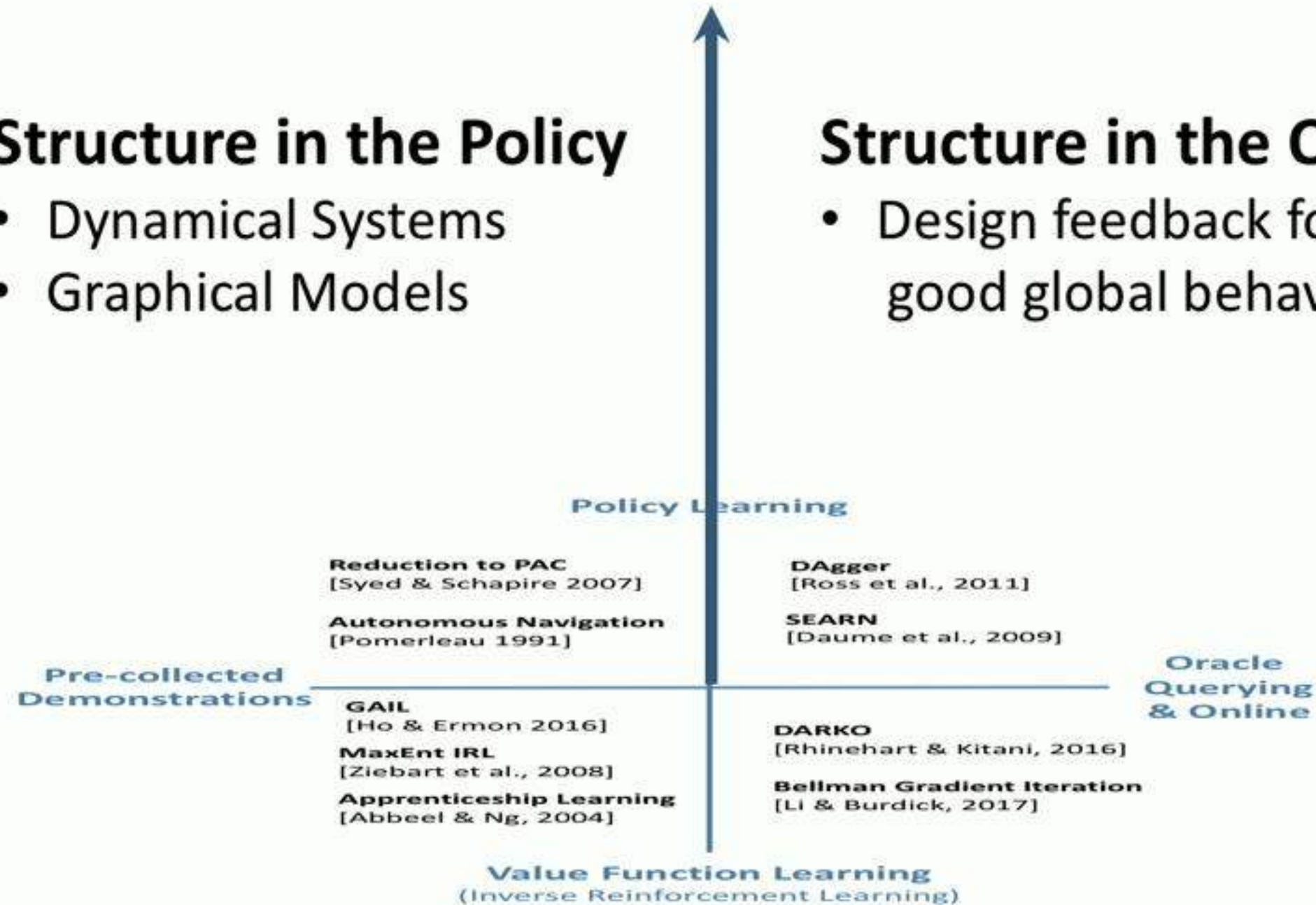
# Structured Imitation Learning

## Structure in the Policy

- Dynamical Systems
- Graphical Models

## Structure in the Oracle

- Design feedback for good global behavior





# Structured Imitation Learning

## Structure in the Policy

- Dynamical Systems
- Graphical Models

## Structure in the Oracle

- Design feedback for good global behavior

## Benefits:

- Better inductive bias
- Reductions to conventional learning
- Composable theoretical guarantees

Value Function Learning  
(Inverse Reinforcement Learning)



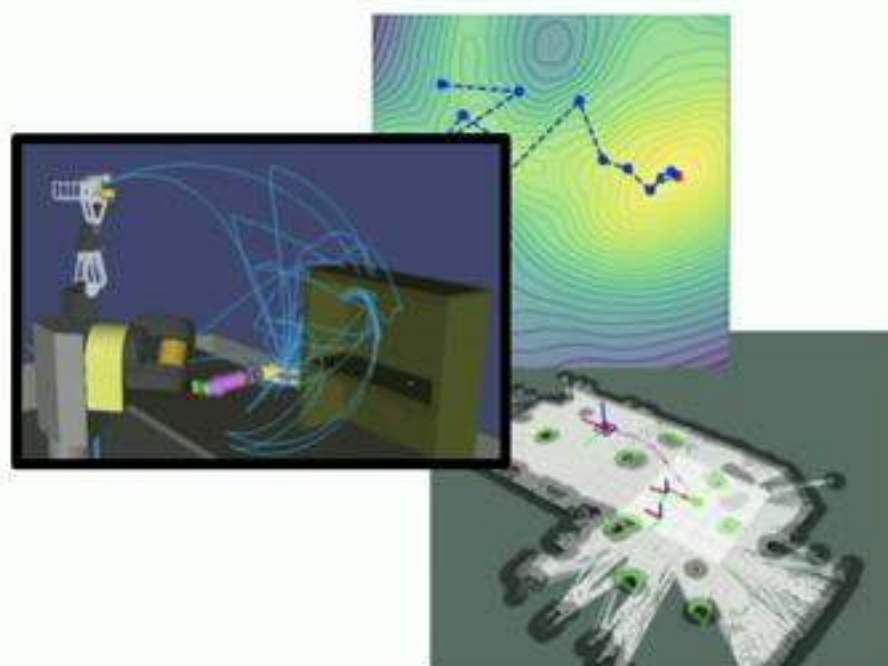
**Speech Animation**



**Coordinated Learning**



**Hierarchical Behaviors  
(Generative)**



**Learning to Optimize**



**Smooth Imitation Learning**





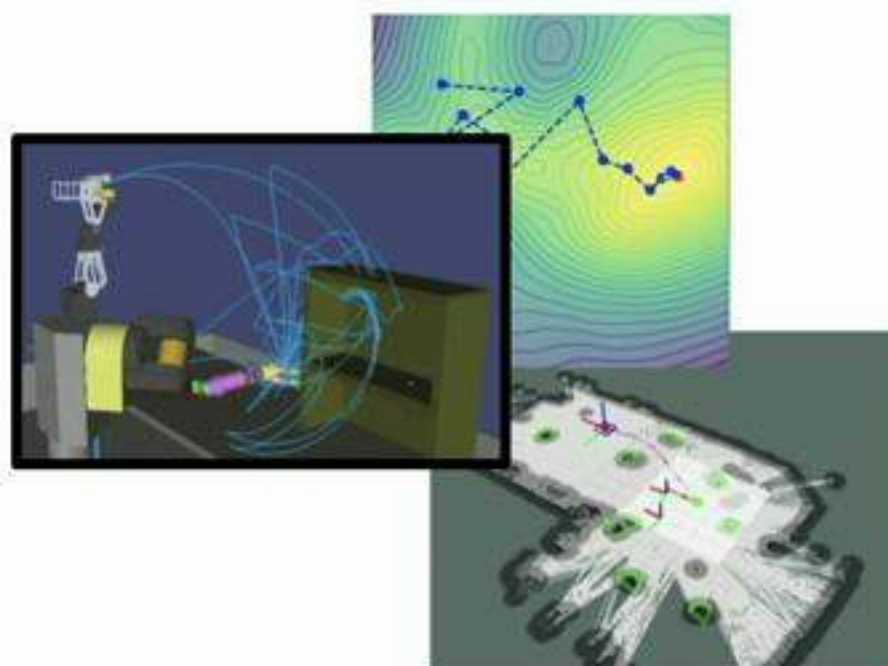
**Speech Animation**



**Coordinated Learning**



**Hierarchical Behaviors  
(Generative)**



**Learning to Optimize**



**Smooth Imitation Learning**

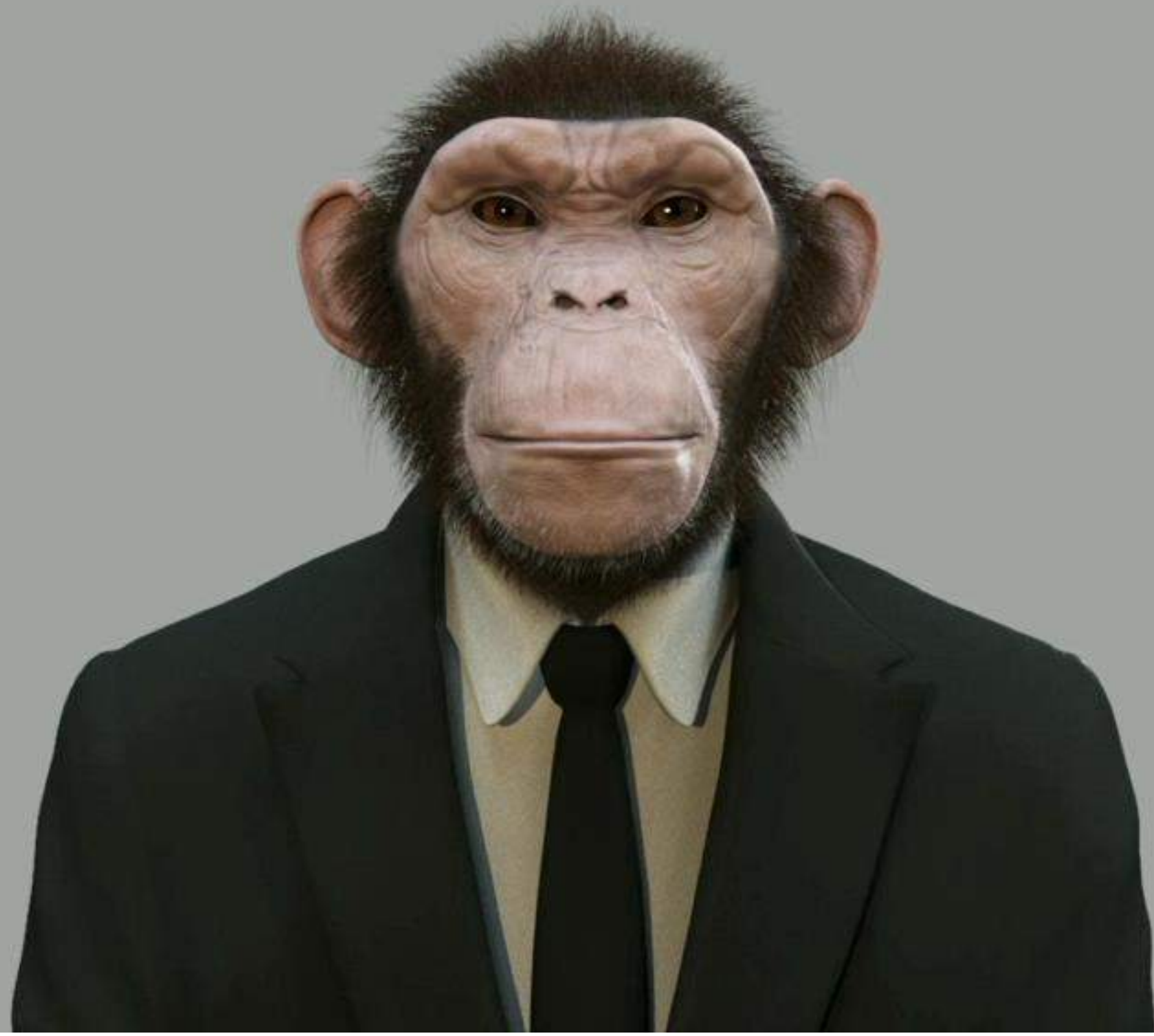




- **Animation artists spend  $\geq 50\%$  time on face**

- Mostly eyes & mouth
- Very tedious

**We'll focus on mouth & speech.**





Sarah  
Taylor

# Prediction Task



Taehwan  
Kim

Input sequence

$$X = \langle x_1, x_2, \dots, x_{|x|} \rangle$$

Output sequence

$$Y = \langle y_1, y_2, \dots, y_{|y|} \rangle, y_t \in \mathbb{R}^D$$

**Goal:** learn predictor

$$h : X \rightarrow Y$$





Sarah Taylor



Taehwan Kim

# Prediction Task

Input sequence

$$X = \langle x_1, x_2, \dots, x_{|x|} \rangle$$

Output sequence

$$Y = \langle y_1, y_2, \dots, y_{|y|} \rangle, y_t \in R^D$$

**Goal:** learn predictor

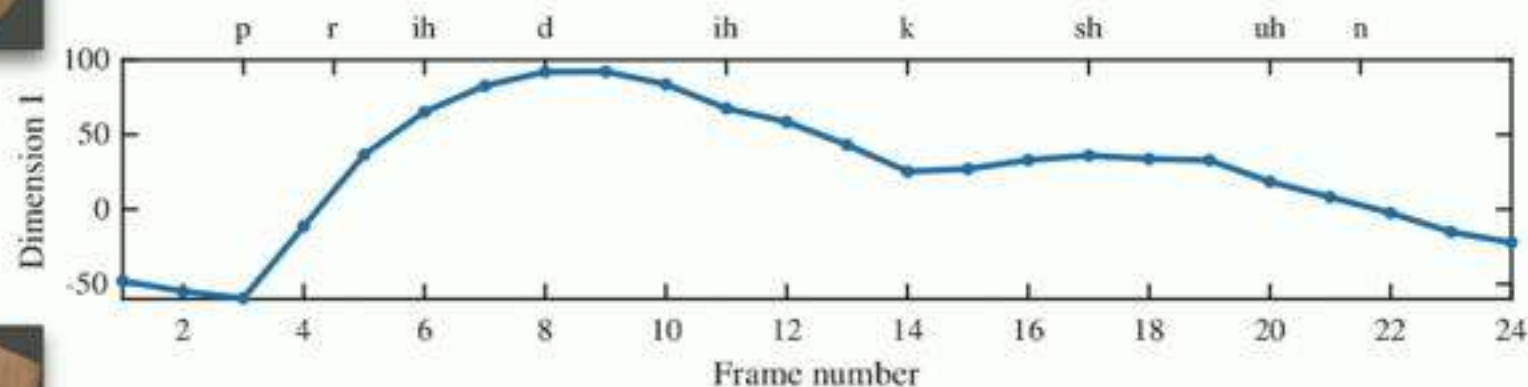
$$h : X \rightarrow Y$$

$X$	Frame	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	Token	-	p	p	r	ih	ih	d	d	ih	ih	ih	ih	k	k	sh	sh	sh	sh	uh	uh	n	-

Phoneme sequence



$Y$



Sequence of face configurations

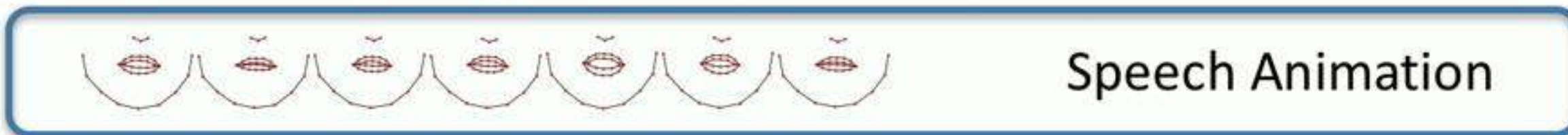


Input Audio



*s s s s s ih ih ih g g r r ae ae ae ae fff*

Speech Recognition



Speech Animation



Retargeting

E.g., [Sumner & Popovic 2004]

(chimp rig courtesy of Hao Li)

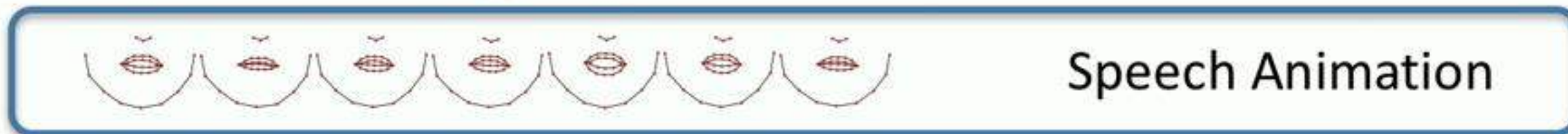


Input Audio



*s s s s s ih ih ih g g r r ae ae ae ae fff*

Speech Recognition



Speech Animation



Retargeting

E.g., [Sumner & Popovic 2004]

(chimp rig courtesy of Hao Li)



Editing



German

© Disney



Sarah Taylor



Taehwan Kim

**A Decision Tree Framework for Spatiotemporal Sequence Prediction**

Taehwan Kim, Yisong Yue, Sarah Taylor, Iain Matthews. KDD 2015

**A Deep Learning Approach for Generalized Speech Animation**

Sarah Taylor, Taehwan Kim, Yisong Yue, et al. SIGGRAPH 2017

Polish

© Disney



Sarah Taylor



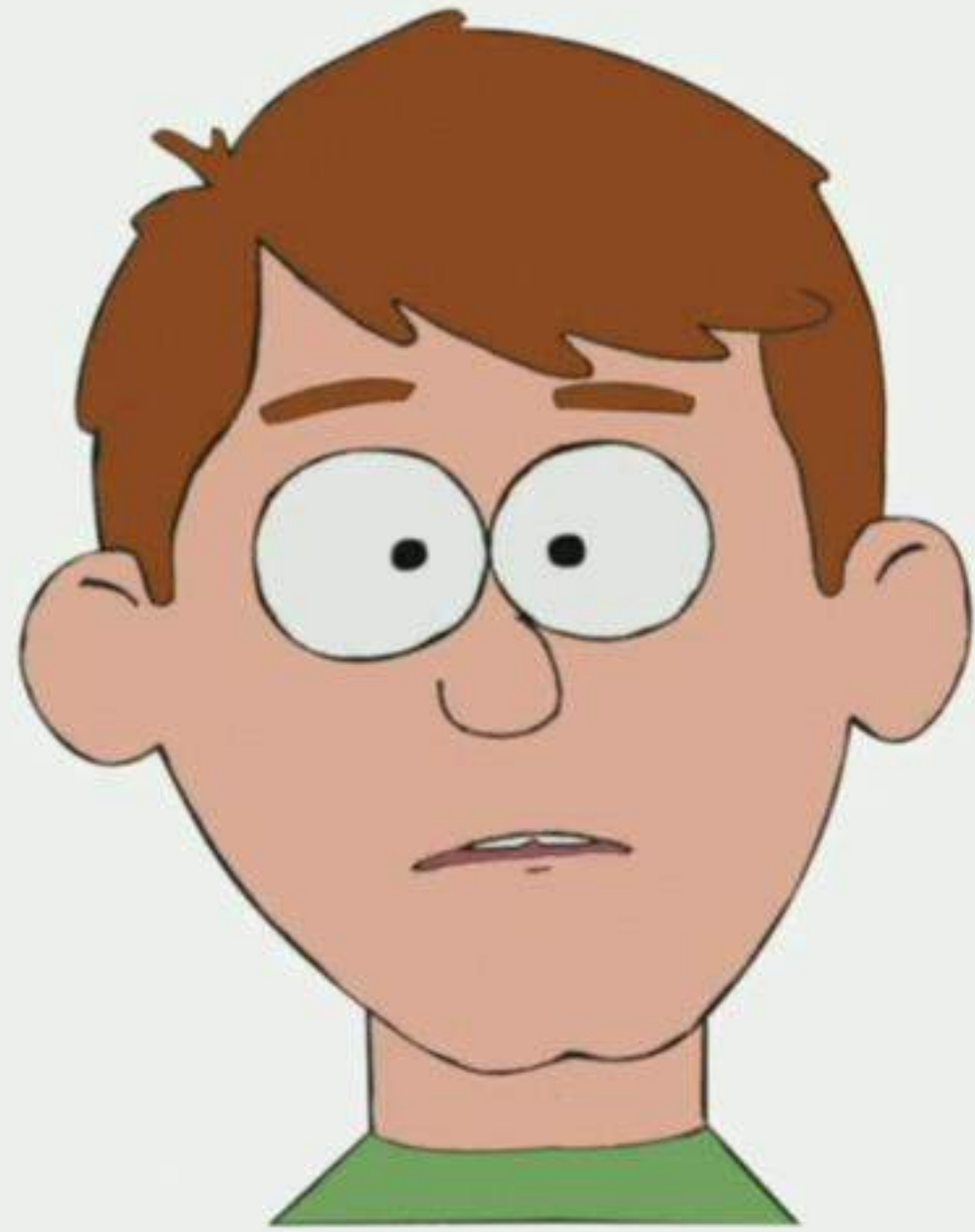
Taehwan Kim

**A Decision Tree Framework for Spatiotemporal Sequence Prediction**

Taehwan Kim, Yisong Yue, Sarah Taylor, Iain Matthews. KDD 2015

**A Deep Learning Approach for Generalized Speech Animation**

Sarah Taylor, Taehwan Kim, Yisong Yue, et al. SIGGRAPH 2017



Sinhalese

© Disney



Sarah Taylor



Taehwan Kim

**A Decision Tree Framework for Spatiotemporal Sequence Prediction**

Taehwan Kim, Yisong Yue, Sarah Taylor, Iain Matthews. KDD 2015

**A Deep Learning Approach for Generalized Speech Animation**

Sarah Taylor, Taehwan Kim, Yisong Yue, et al. SIGGRAPH 2017





Our Prediction



Chinese



Sarah Taylor



Taehwan Kim

**A Decision Tree Framework for Spatiotemporal Sequence Prediction**

Taehwan Kim, Yisong Yue, Sarah Taylor, Iain Matthews. KDD 2015

**A Deep Learning Approach for Generalized Speech Animation**

Sarah Taylor, Taehwan Kim, Yisong Yue, et al. SIGGRAPH 2017



Sarah Taylor



Taehwan Kim

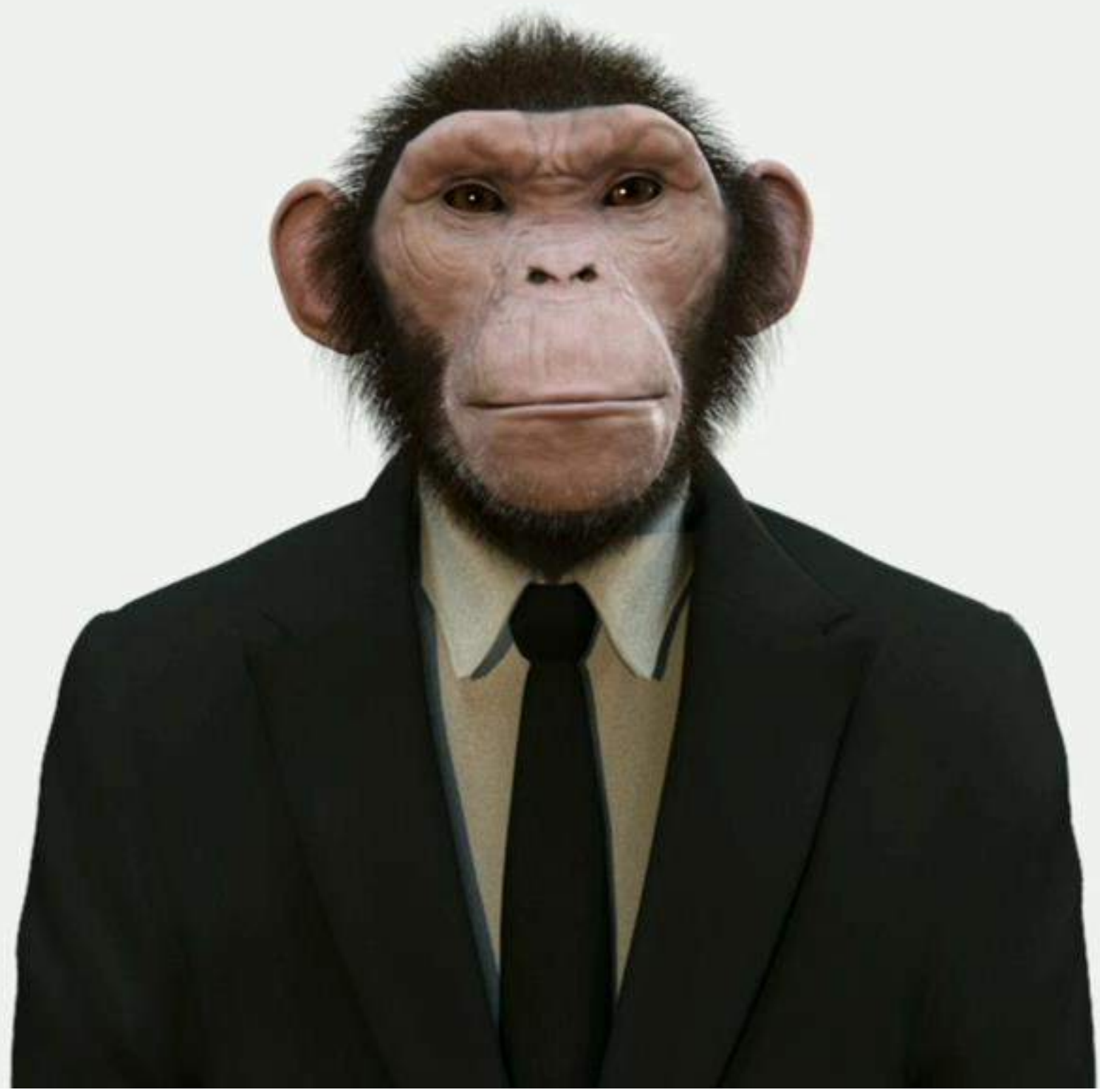
**A Decision Tree Framework for Spatiotemporal Sequence Prediction**

Taehwan Kim, Yisong Yue, Sarah Taylor, Iain Matthews. KDD 2015

**A Deep Learning Approach for Generalized Speech Animation**

Sarah Taylor, Taehwan Kim, Yisong Yue, et al. SIGGRAPH 2017









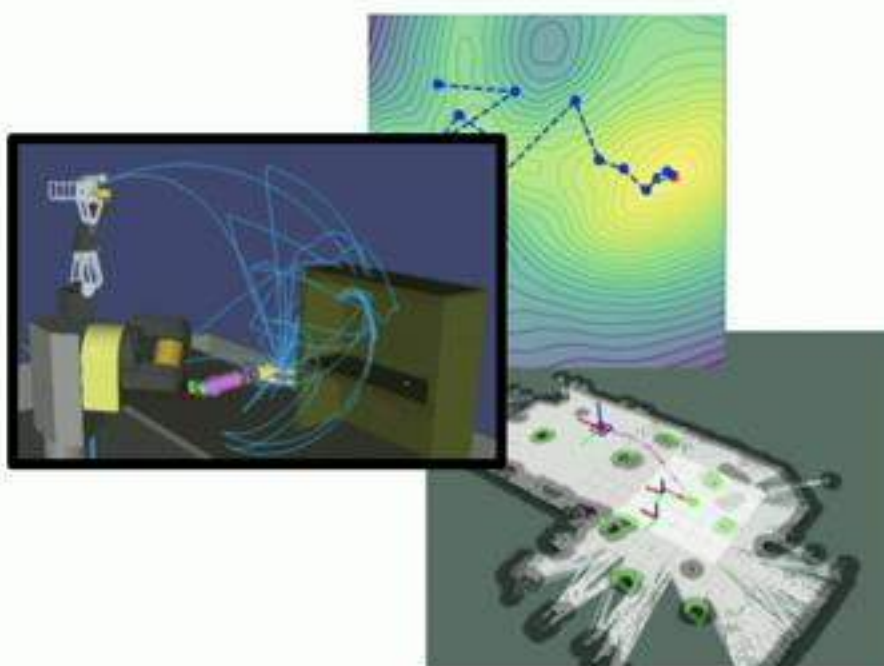
**Speech Animation**



**Coordinated Learning**



**Hierarchical Behaviors  
(Generative)**



**Learning to Optimize**



**Smooth Imitation Learning**



Sarah Taylor



Taehwan Kim

**A Decision Tree Framework for Spatiotemporal Sequence Prediction**

Taehwan Kim, Yisong Yue, Sarah Taylor, Iain Matthews. KDD 2015

**A Deep Learning Approach for Generalized Speech Animation**

Sarah Taylor, Taehwan Kim, Yisong Yue, et al. SIGGRAPH 2017



# Our Approach



English Premier League  
2012-2013

Match date: 04/05/2013

**Data-Driven Ghosting using Deep Imitation Learning**

Hoang Le, Peter Carr, Yisong Yue, Patrick Lucey. SSAC 2017





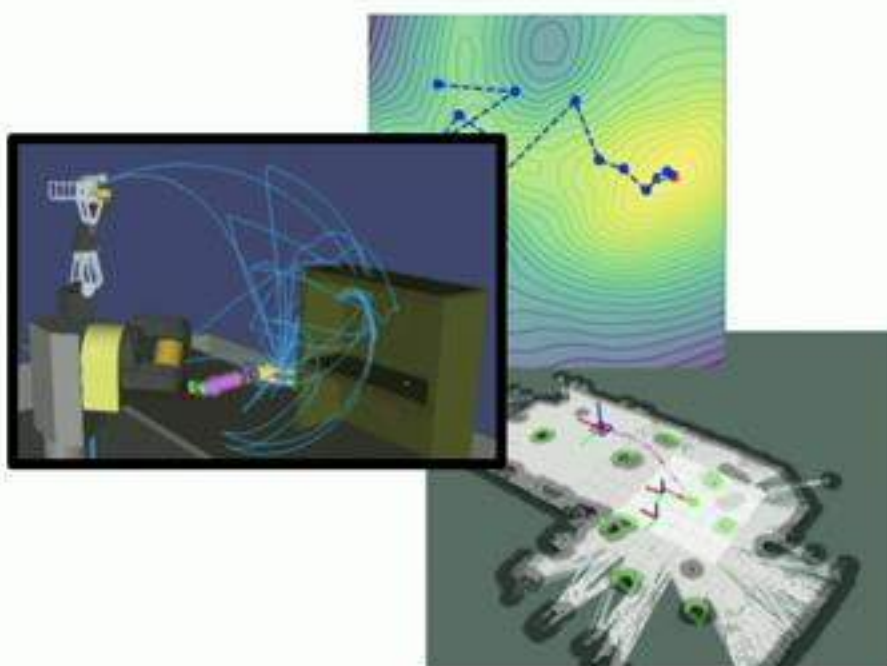
**Speech Animation**



**Coordinated Learning**



**Hierarchical Behaviors  
(Generative)**



**Learning to Optimize**



**Smooth Imitation Learning**

# Our Approach



English Premier League  
2012-2013

Match date: 04/05/2013

**Data-Driven Ghosting using Deep Imitation Learning**

Hoang Le, Peter Carr, Yisong Yue, Patrick Lucey. SSAC 2017

# Naïve Baseline



English Premier League  
2012-2013

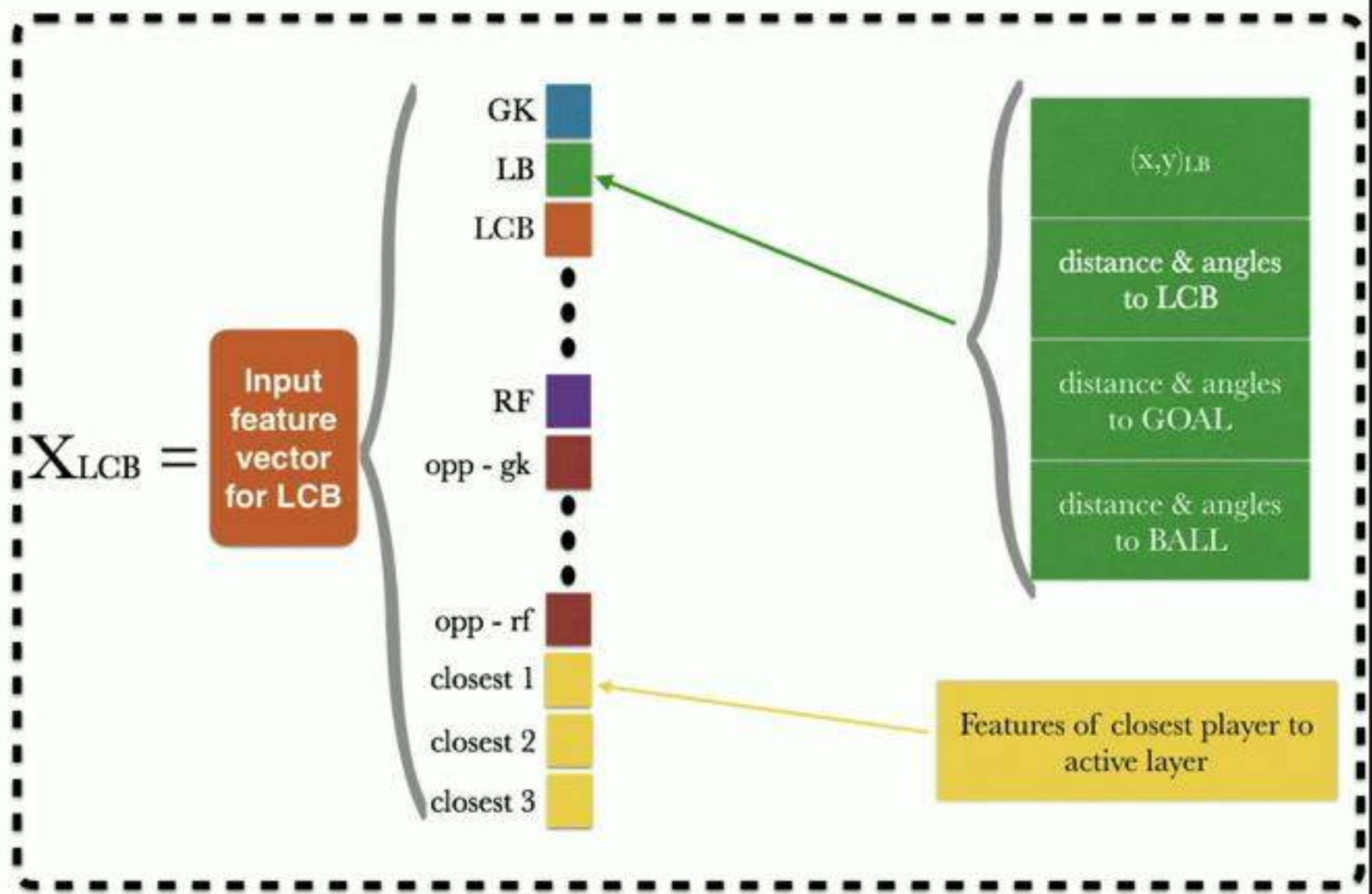
Match date: 04/05/2013



# State Representation

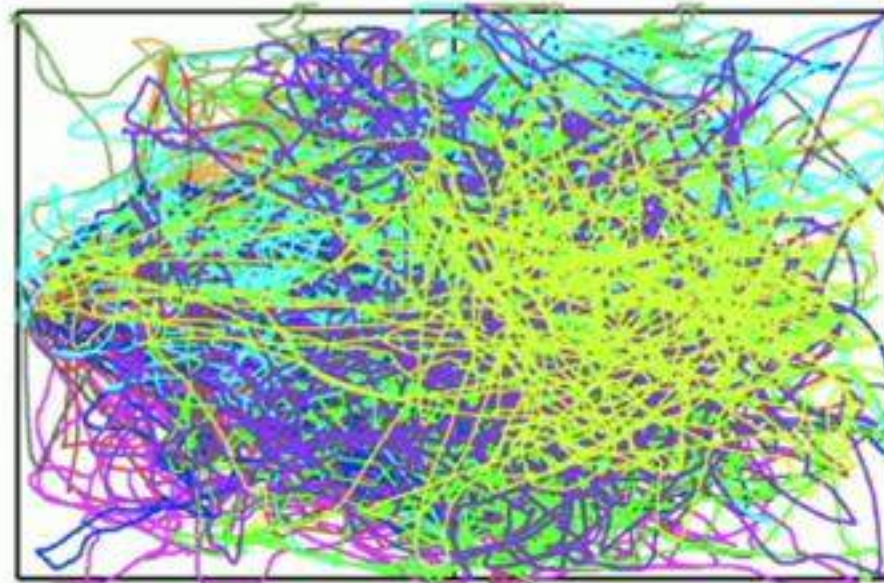


Geometric features computed



## But Who Plays Which Role?

- All we get are trajectories!
  - Don't know which belongs to which role.

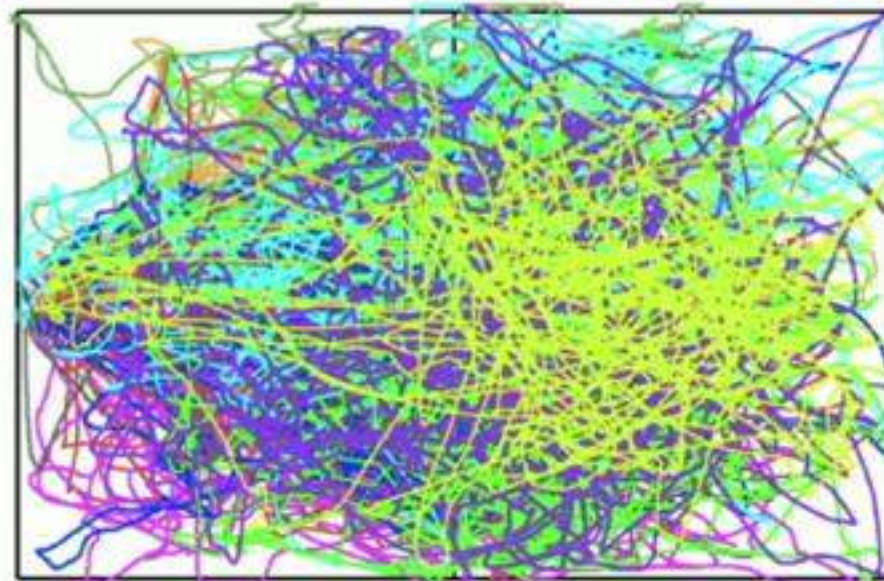


- Need to solve a permutation problem



## But Who Plays Which Role?

- All we get are trajectories!
  - Don't know which belongs to which role.



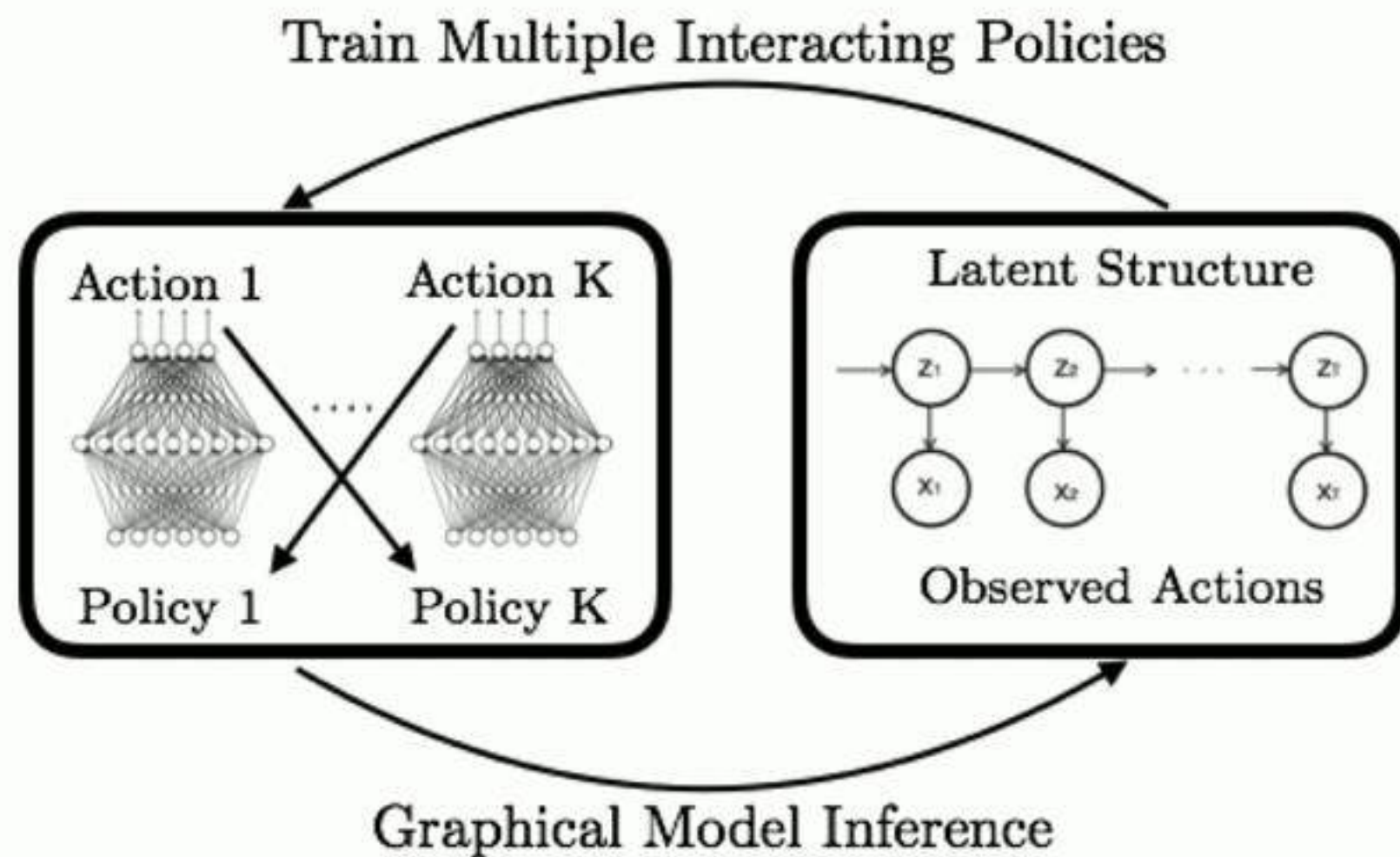
- Need to solve a permutation problem
  - **Naïve baseline ignores this!**





Hoang  
Le

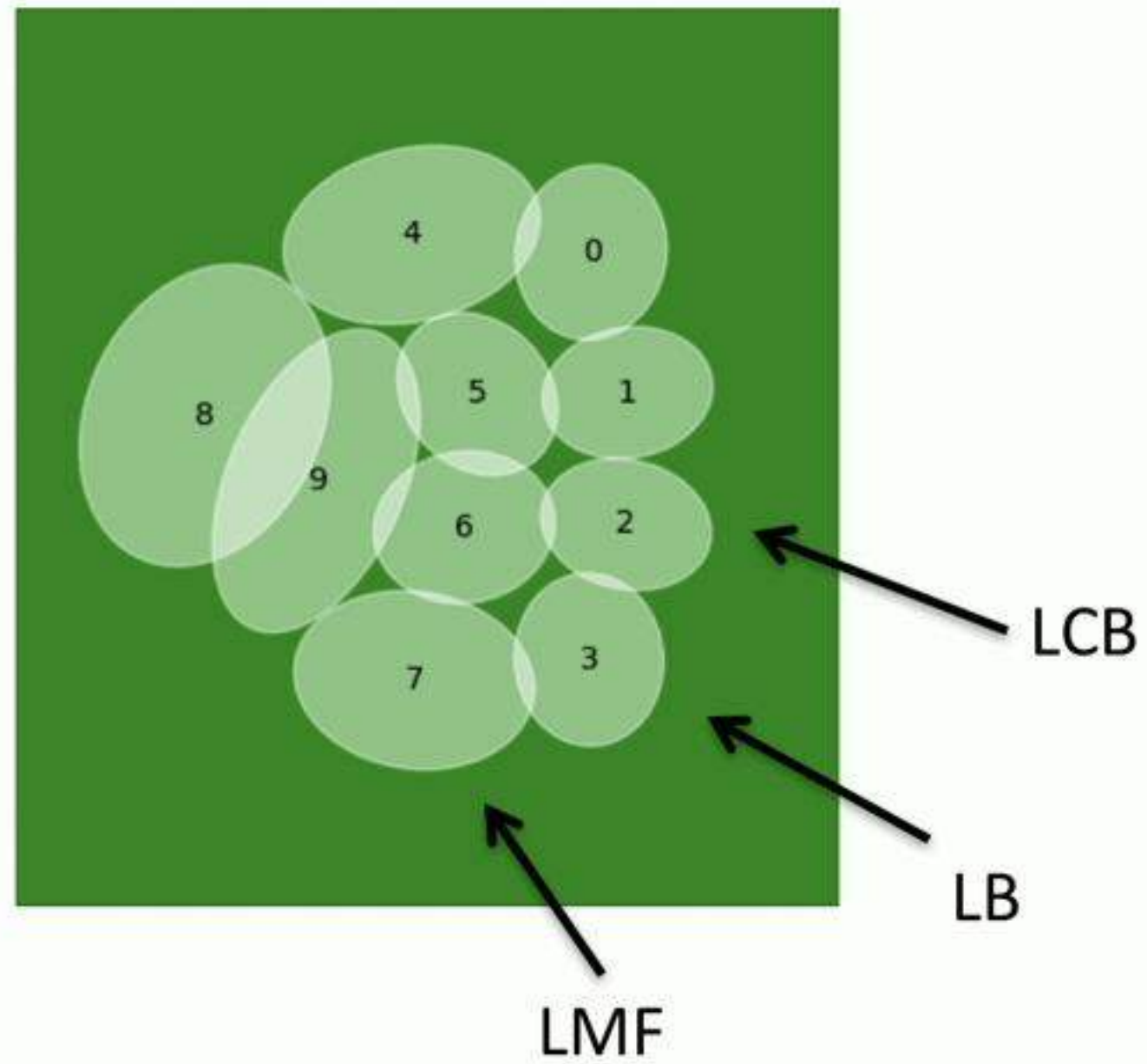
# Coordination Model



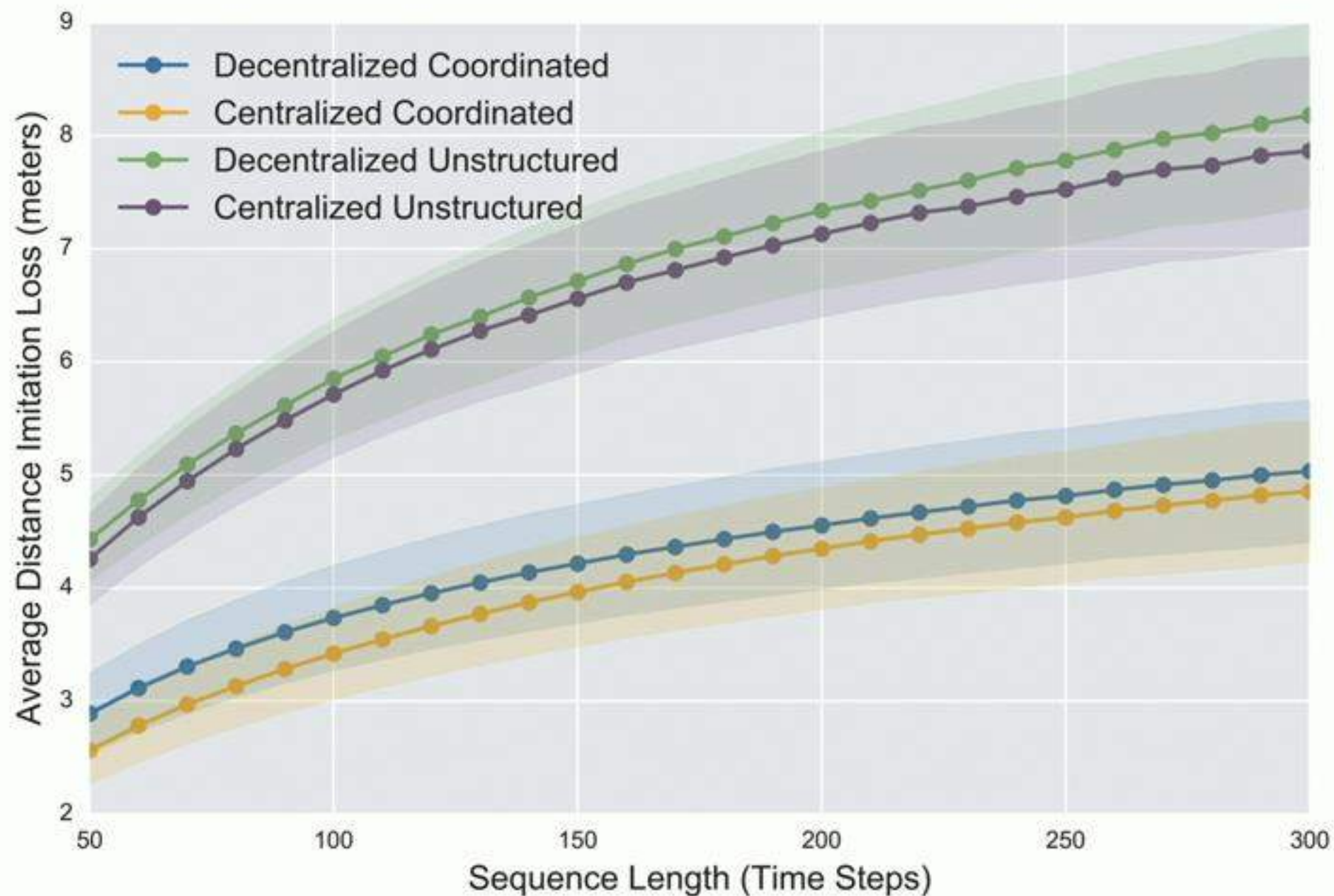
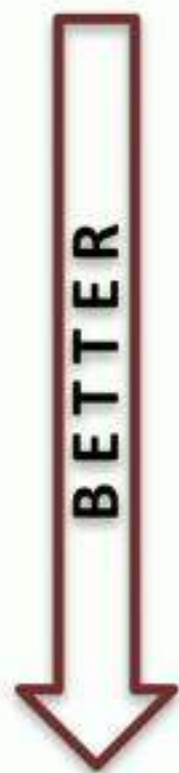
**Coordinated Multi-Agent Imitation Learning**

Hoang Le, Yisong Yue, Peter Carr, Patrick Lucey. ICML 2017

# Learned Roles



# Imitation Error on Test Examples



## Coordinated Multi-Agent Imitation Learning

Hoang Le, Yisong Yue, Peter Carr, Patrick Lucey. ICML 2017



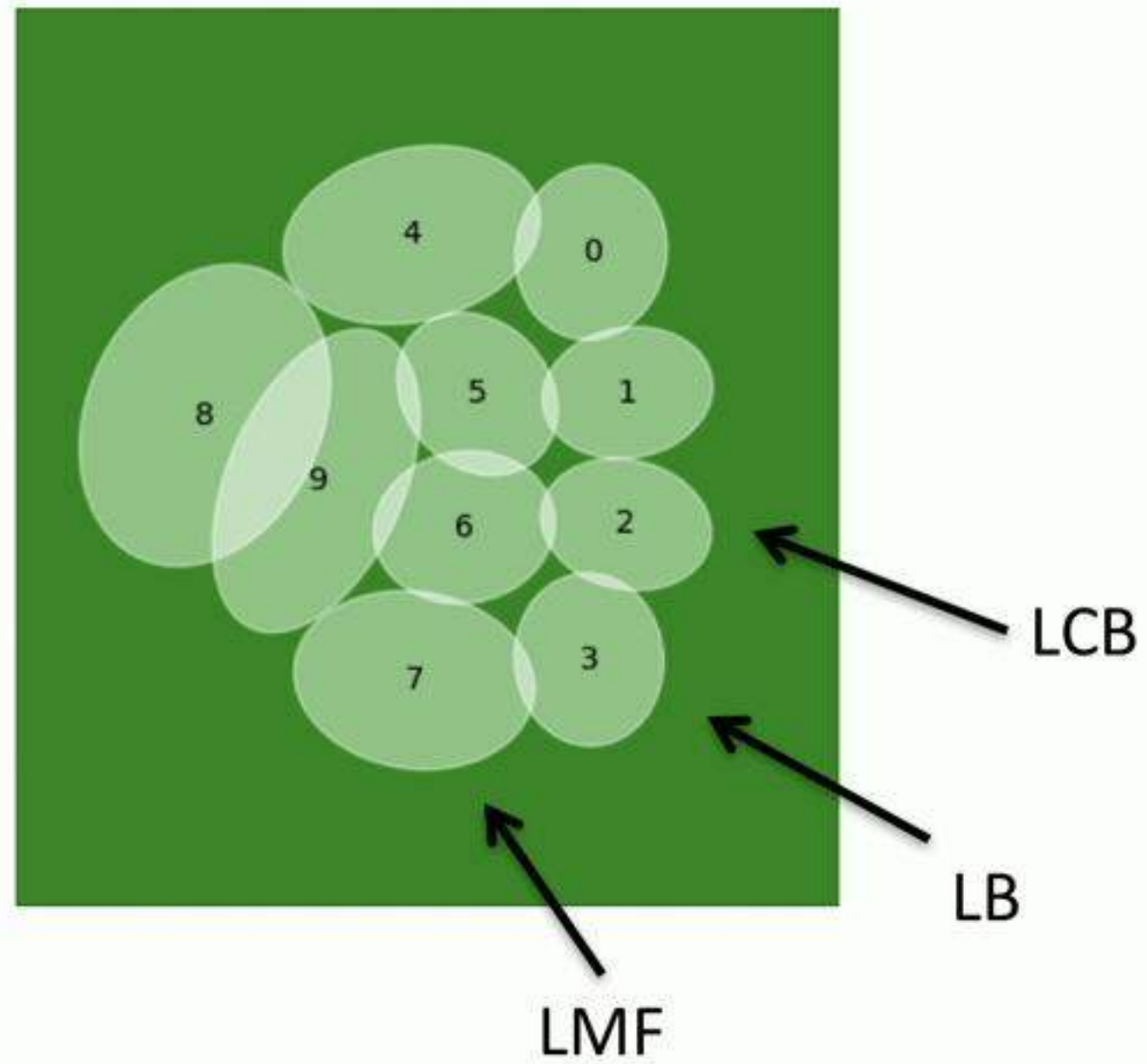
# Naïve Baseline



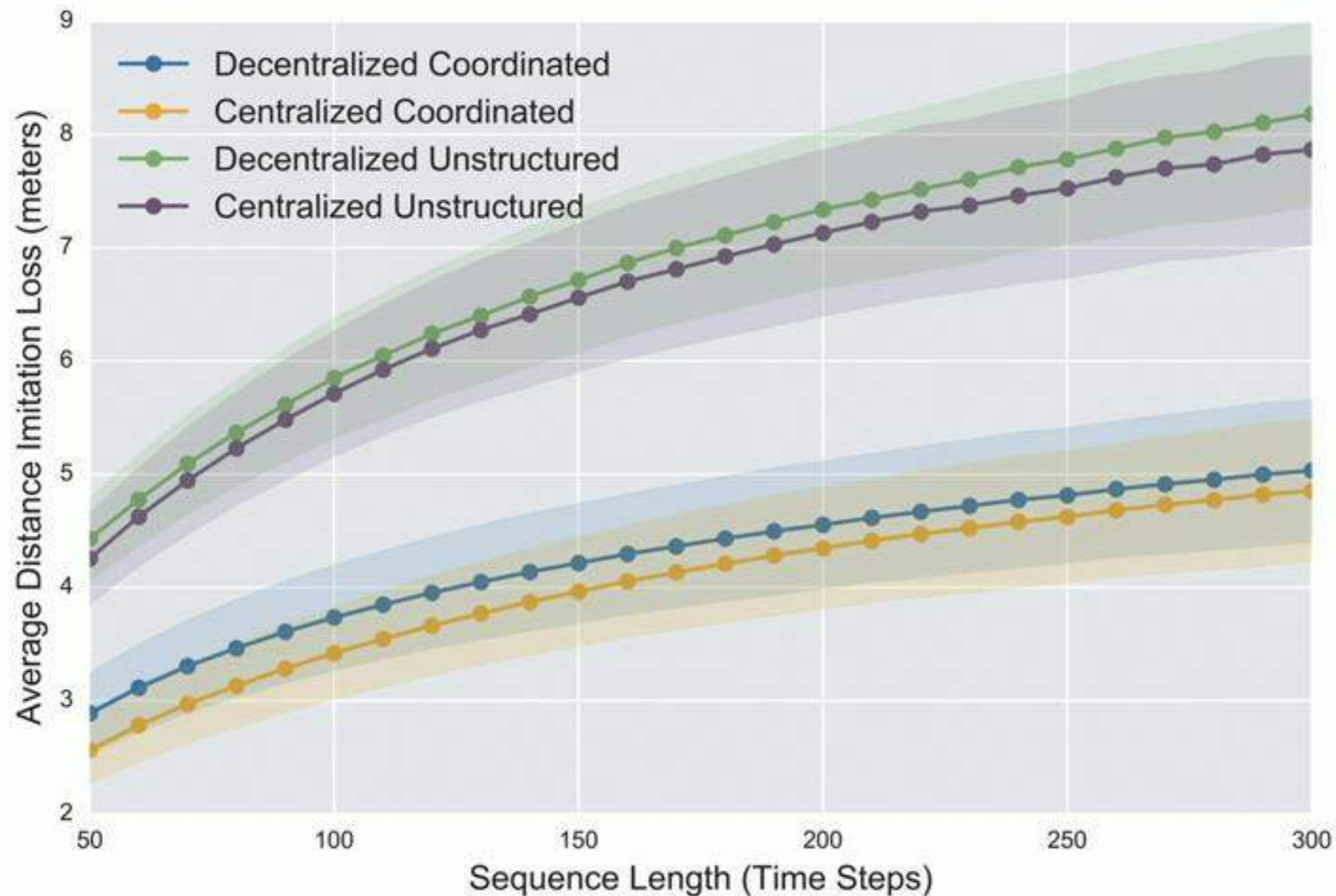
English Premier League  
2012-2013

Match date: 04/05/2013

# Learned Roles



# Imitation Error on Test Examples



## Coordinated Multi-Agent Imitation Learning

Hoang Le, Yisong Yue, Peter Carr, Patrick Lucey. ICML 2017

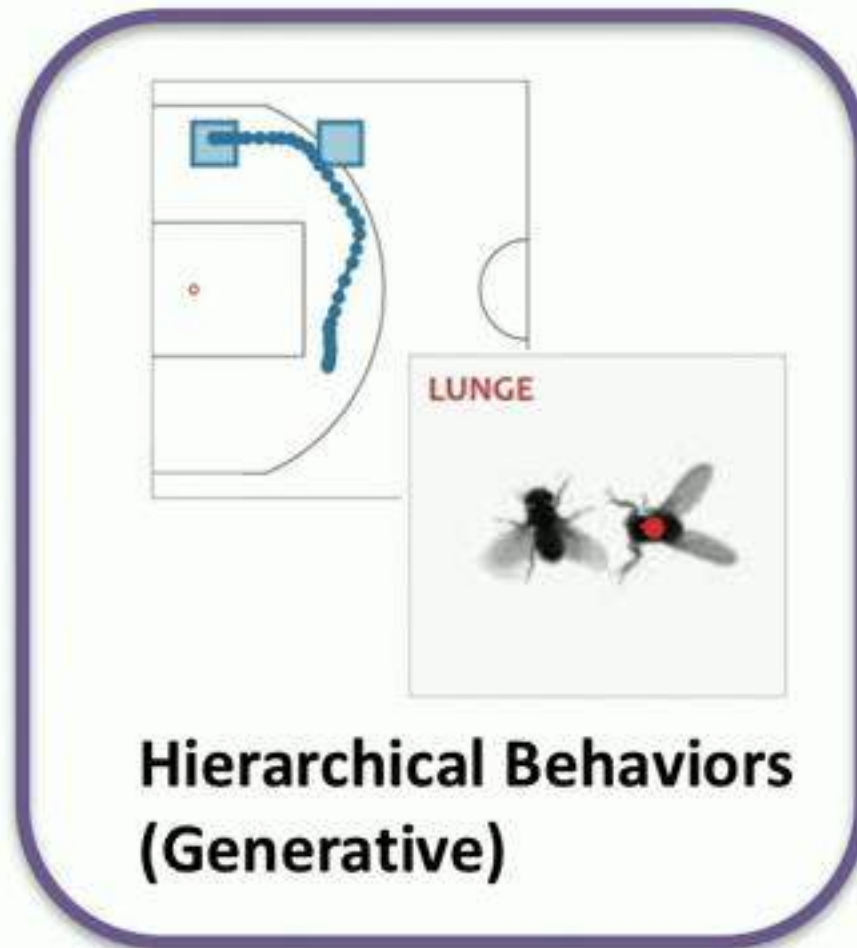




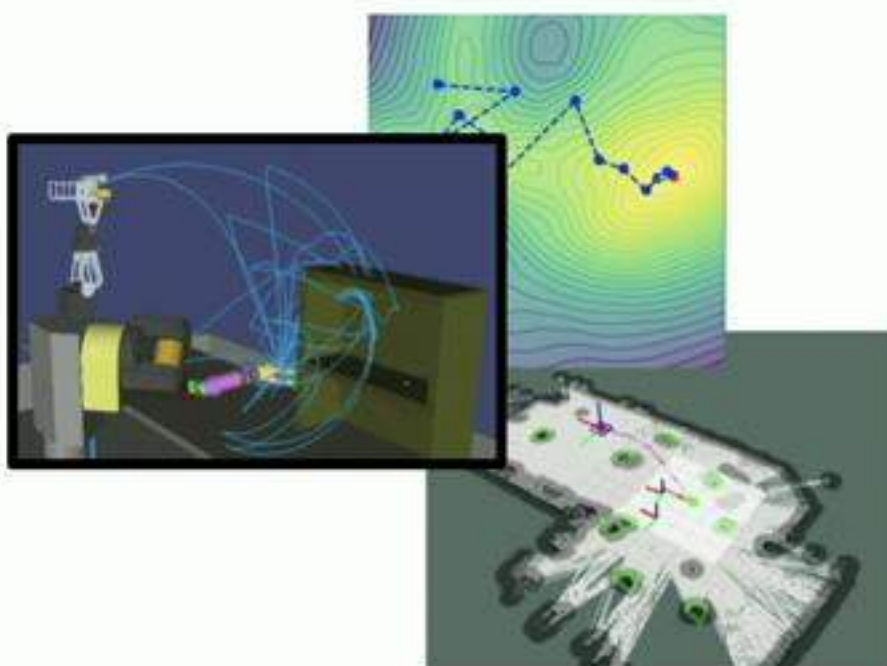
**Speech Animation**



**Coordinated Learning**



**Hierarchical Behaviors  
(Generative)**



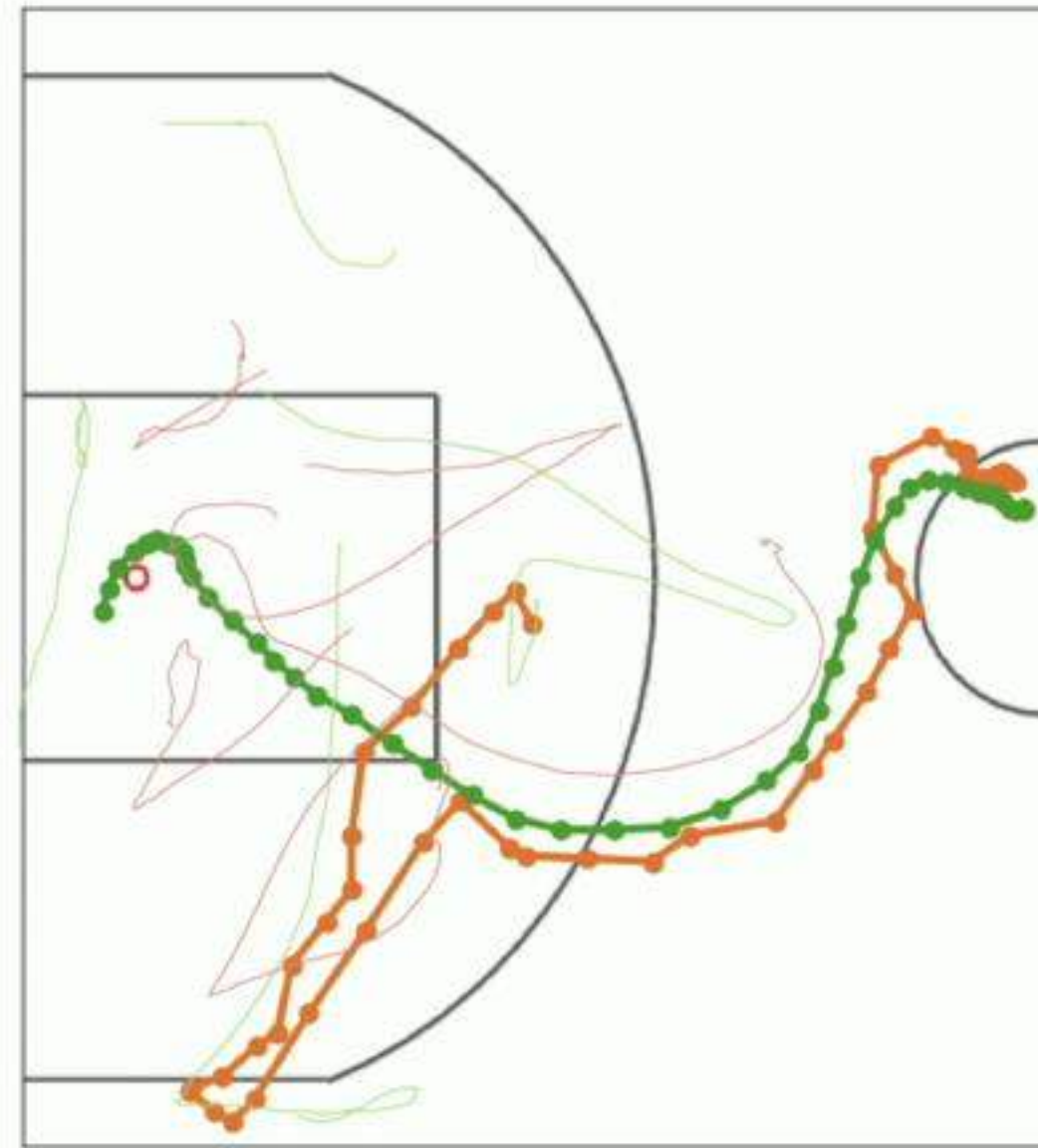
**Learning to Optimize**



**Smooth Imitation Learning**

# Strategy vs Tactics

- Long-term Goal:
  - Curl around basket
- Tactics
  - Drive left w/ ball
  - Pass ball
  - Cut towards basket



Stephan  
Zheng



Eric  
Zhan



# Generative + Hierarchical Imitation Learning

- **Generative Imitation Learning**
  - No single “correct” action
- **Hierarchical**
  - Make predictions at multiple resolutions

**Generating Long-term Trajectories using Deep Hierarchical Networks**

Stephan Zheng, Yisong Yue, Patrick Lucey. NIPS 2016

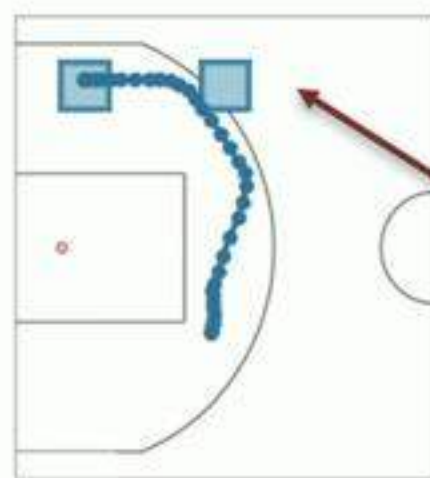
**Generative Multi-Agent Behavioral Cloning**

Eric Zhan, Stephan Zheng, Yisong Yue, Long Sha, Patrick Lucey. arXiv

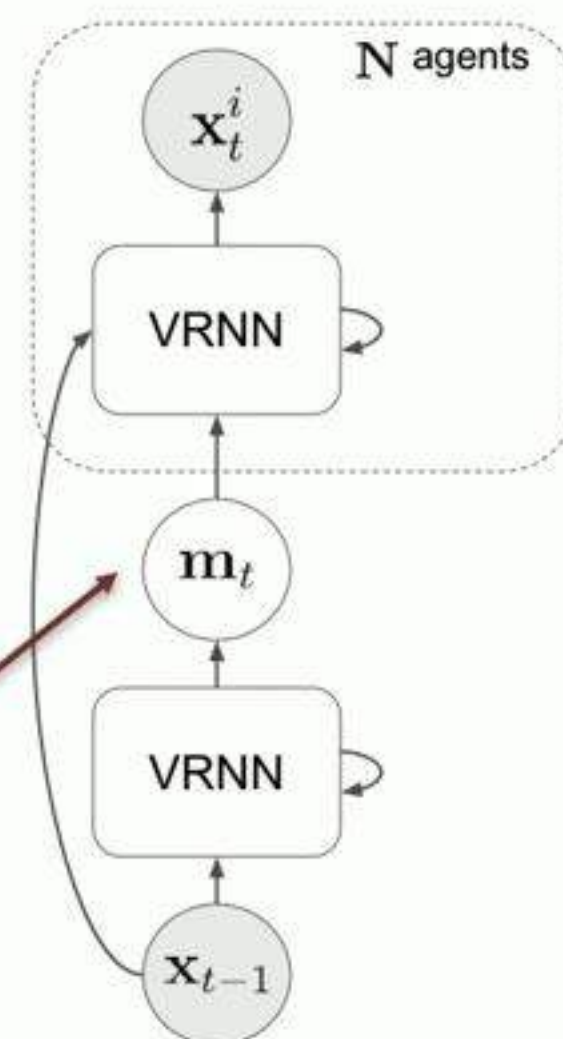


# Generative + Hierarchical Imitation Learning

- **Generative Imitation Learning**
  - No single “correct” action
- **Hierarchical**
  - Make predictions at multiple resolutions



Macro-goals

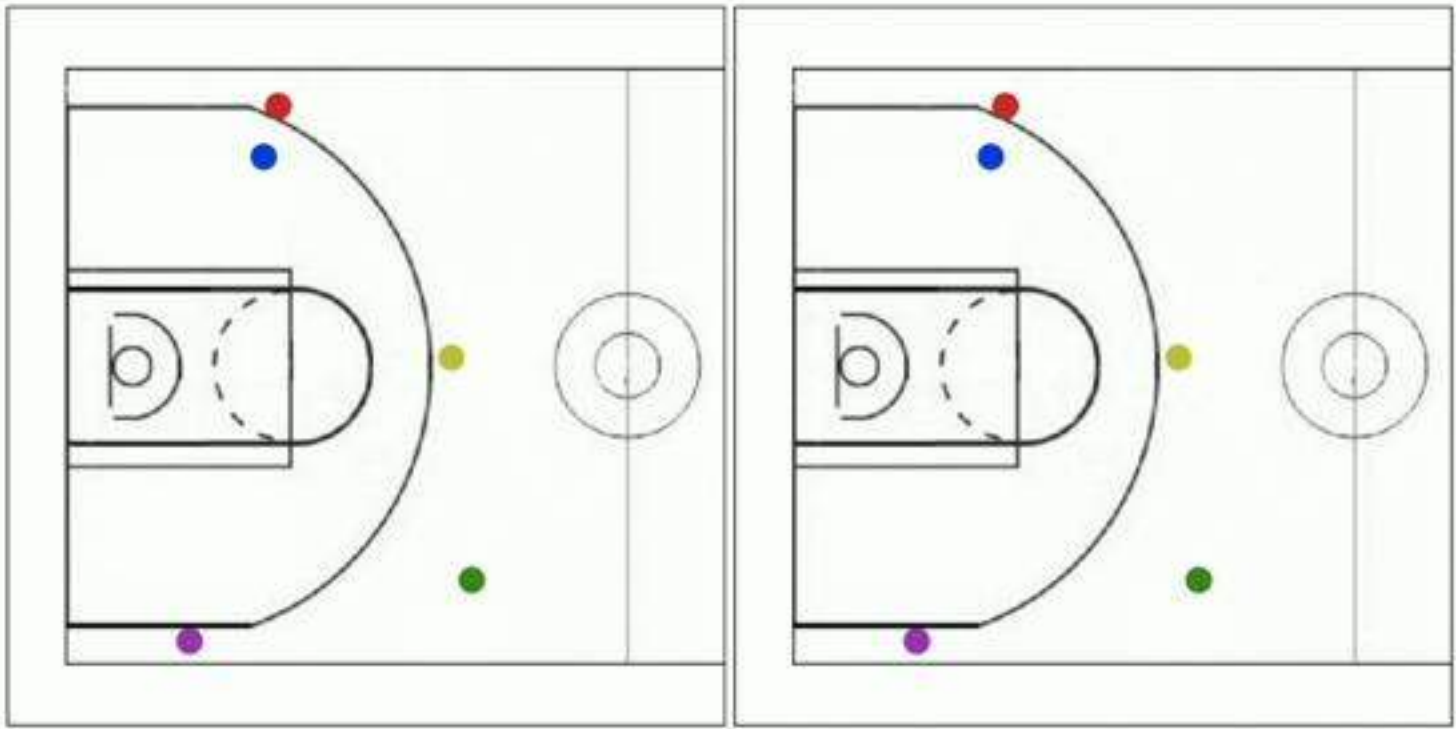


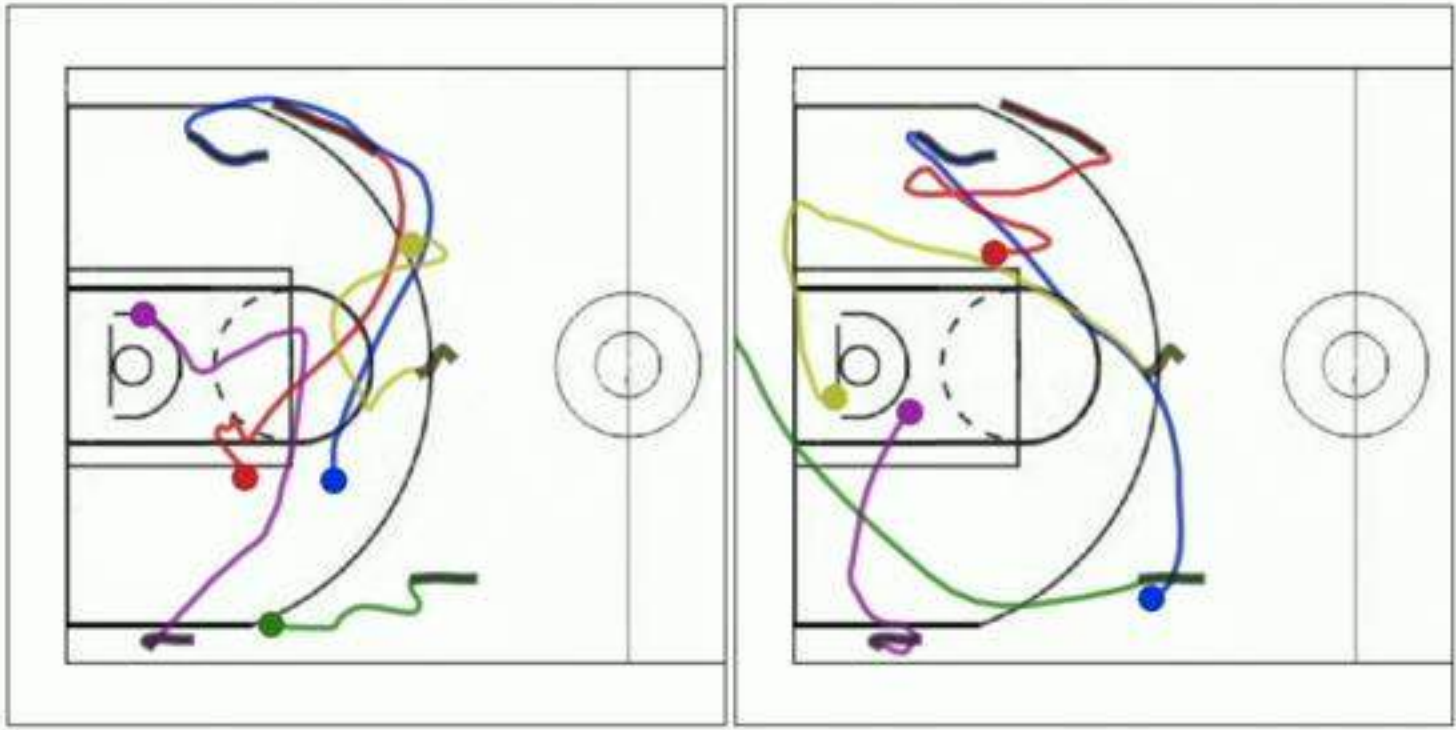
**Generating Long-term Trajectories using Deep Hierarchical Networks**

Stephan Zheng, Yisong Yue, Patrick Lucey. NIPS 2016

**Generative Multi-Agent Behavioral Cloning**

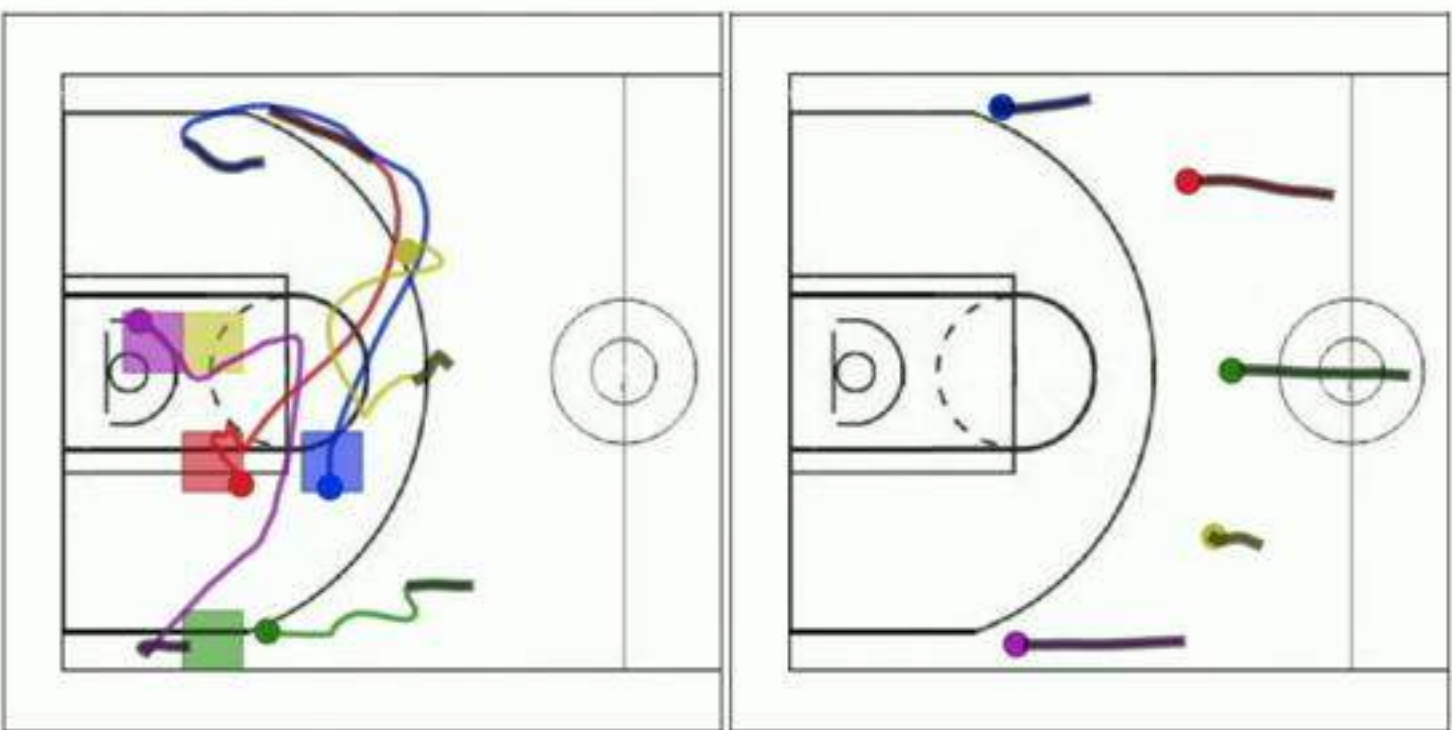
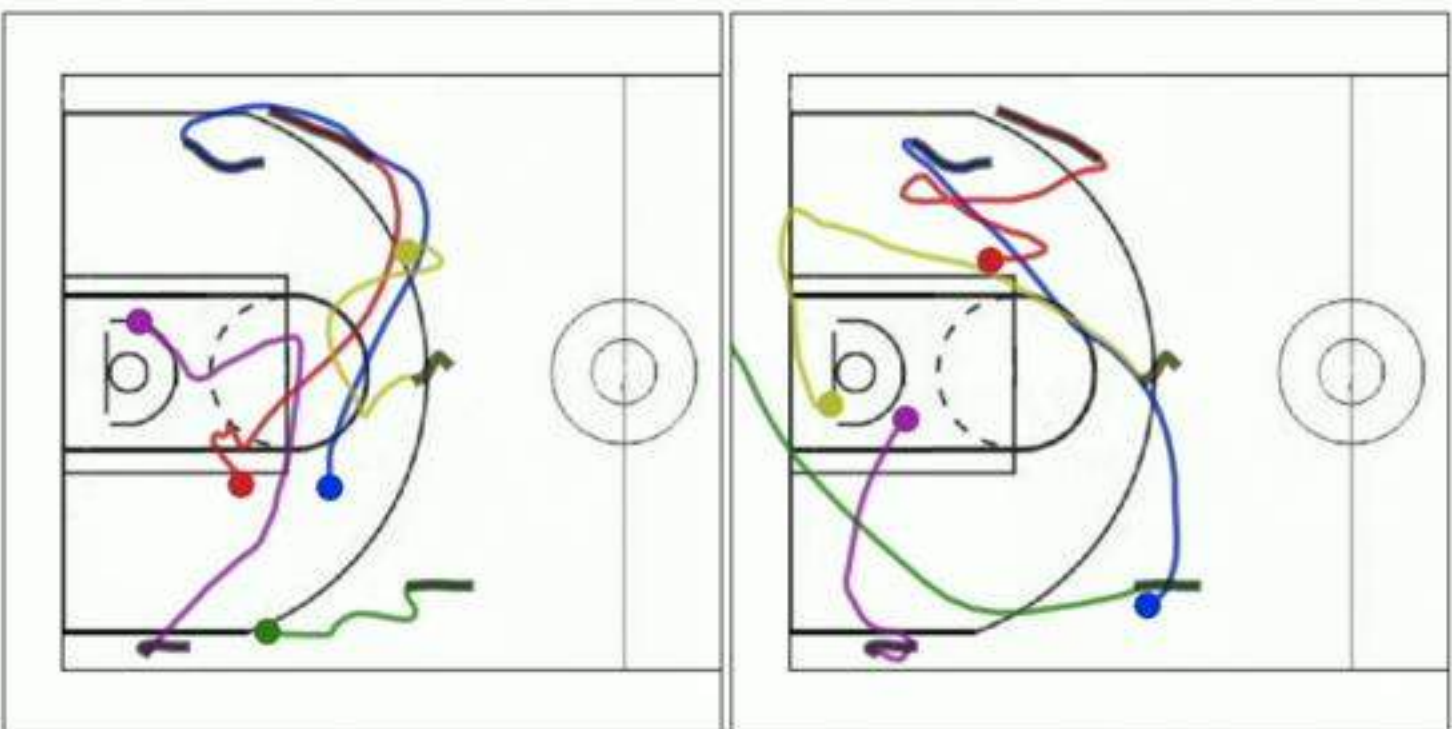
Eric Zhan, Stephan Zheng, Yisong Yue, Long Sha, Patrick Lucey. arXiv

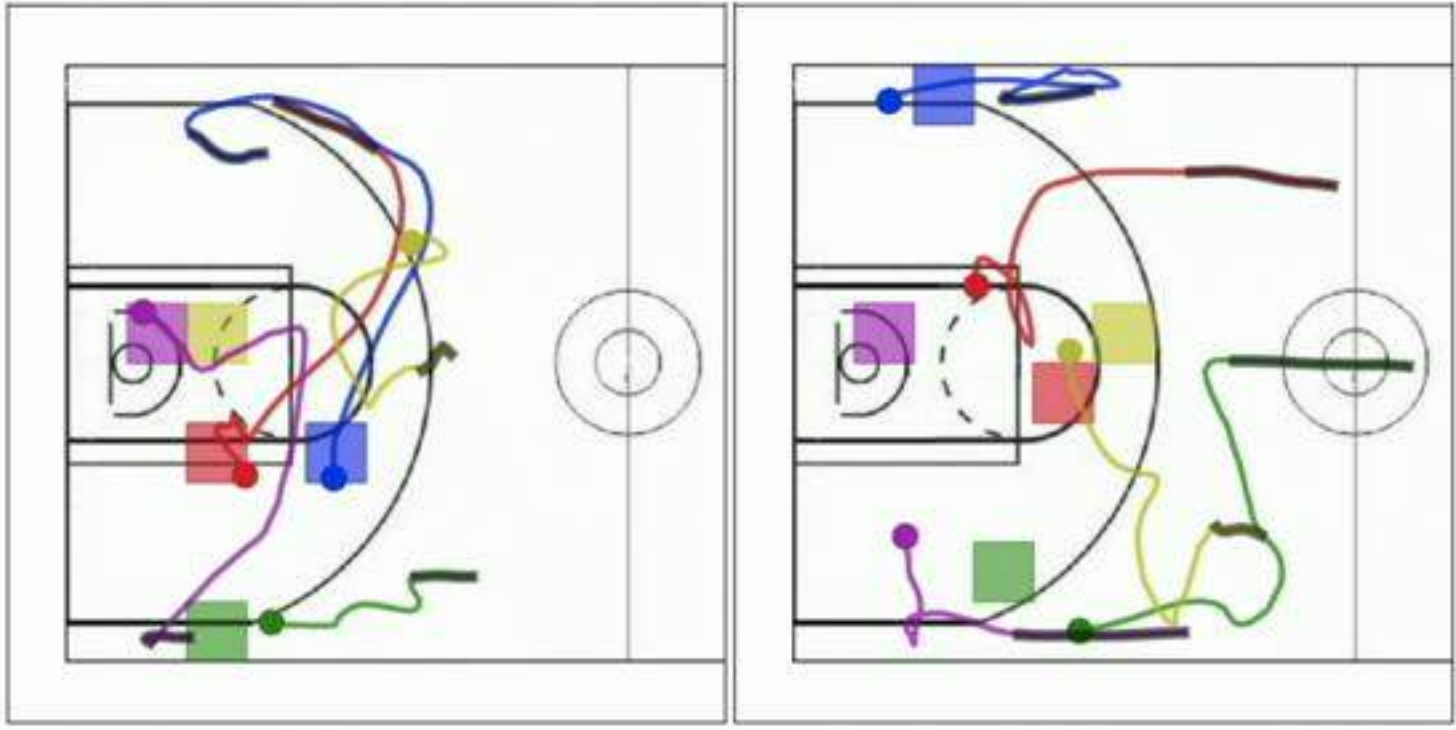
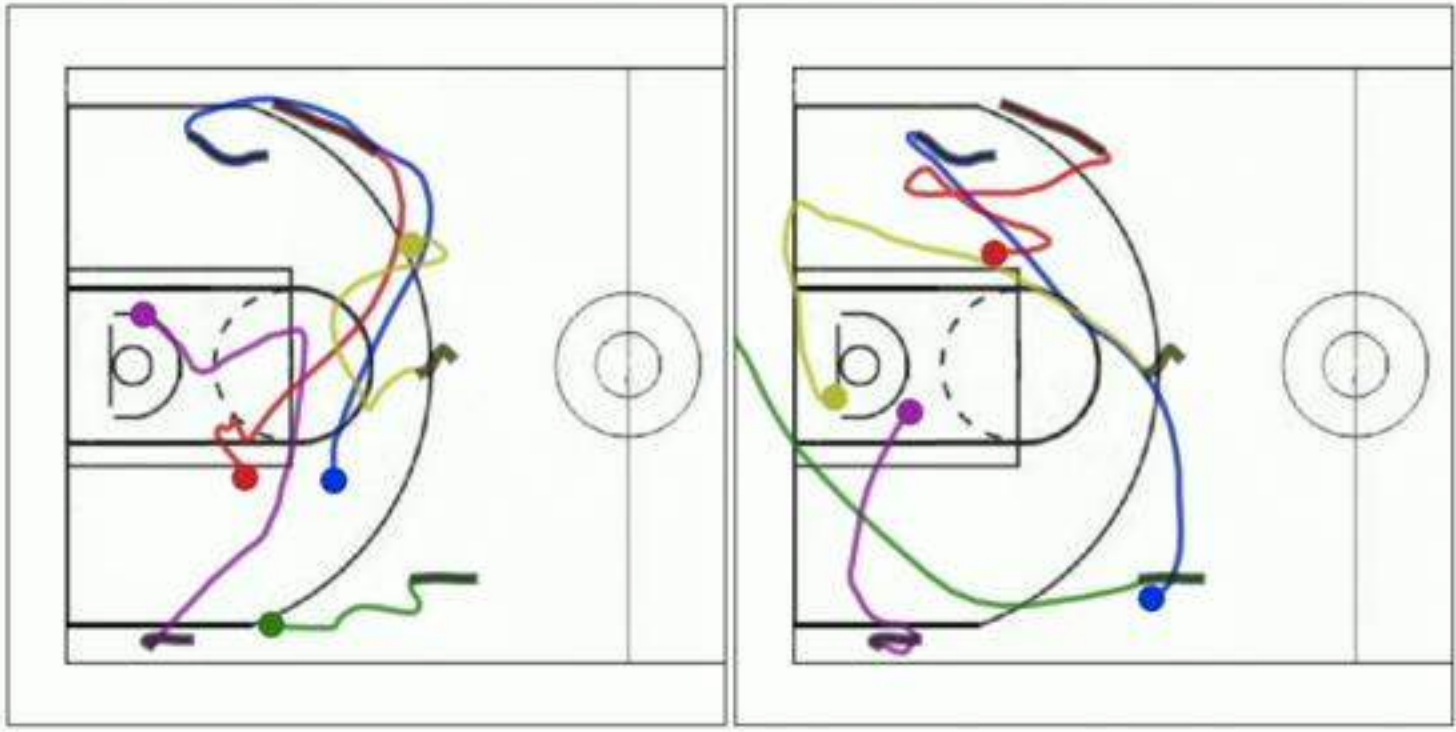




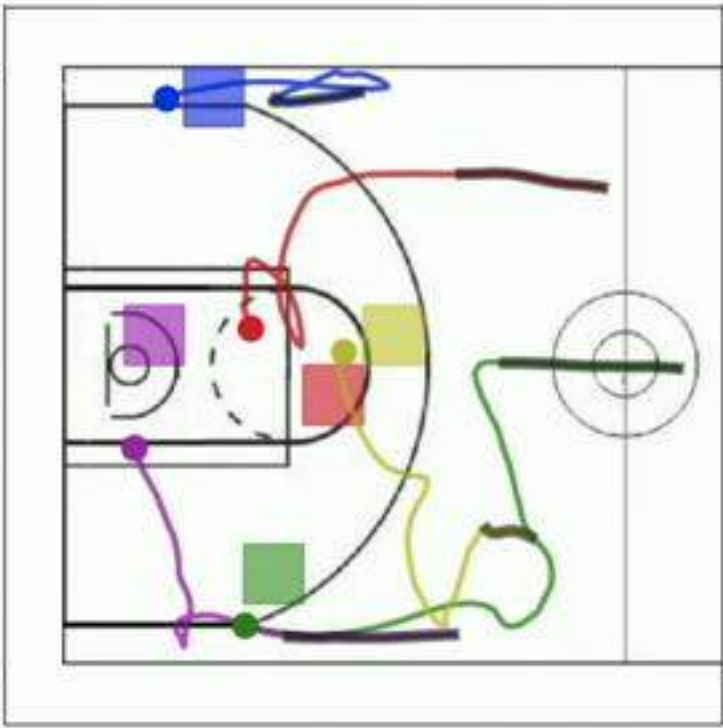
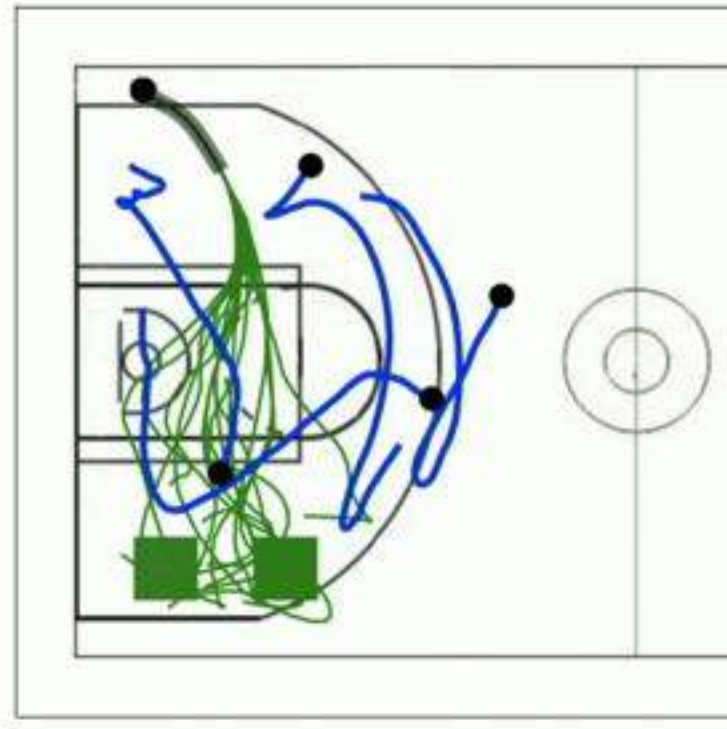


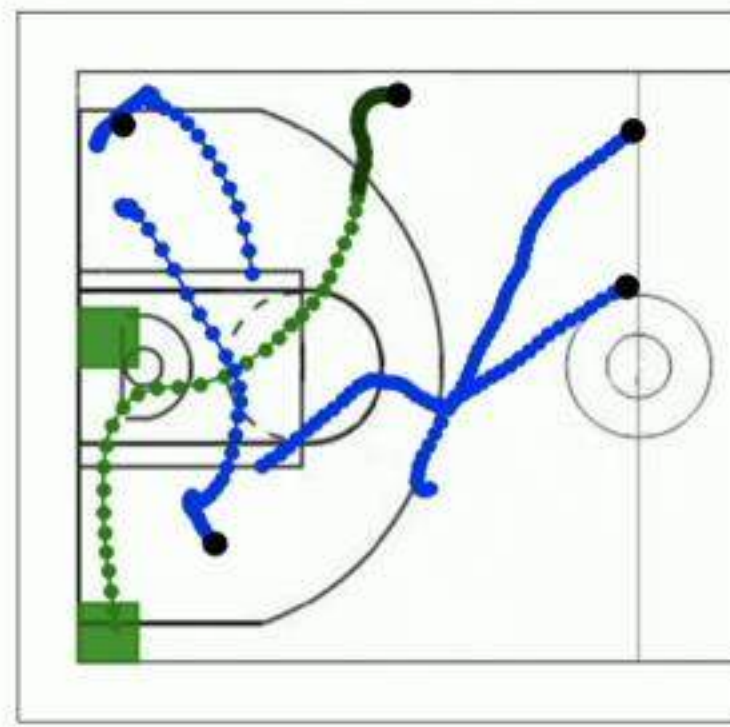
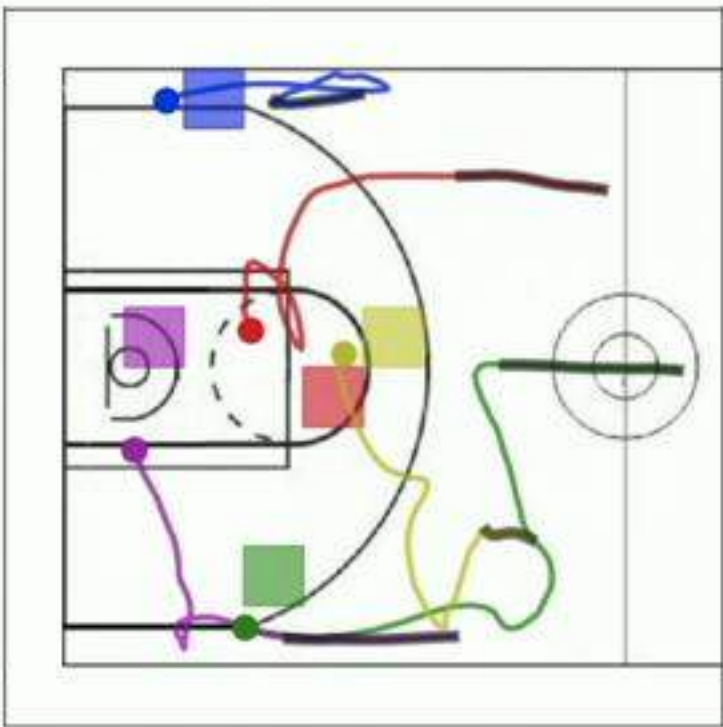
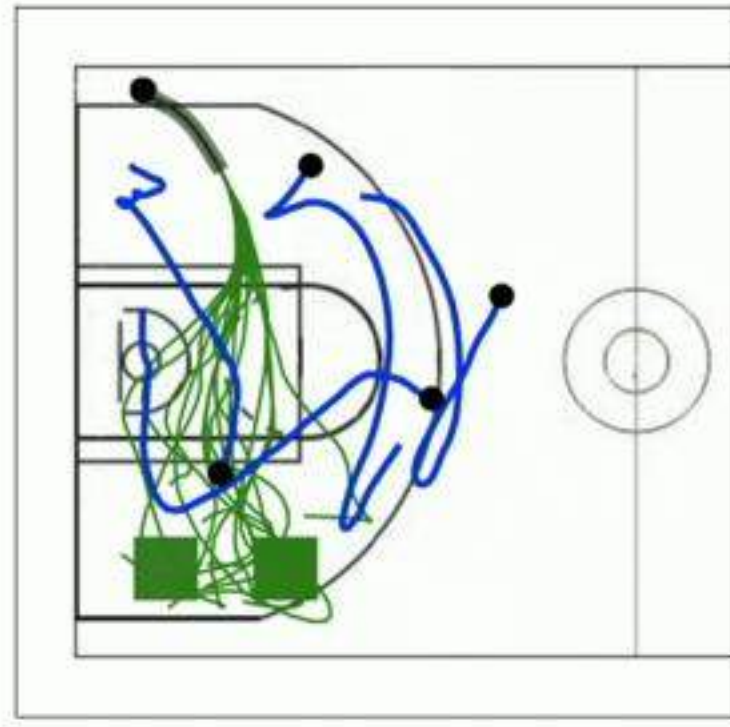








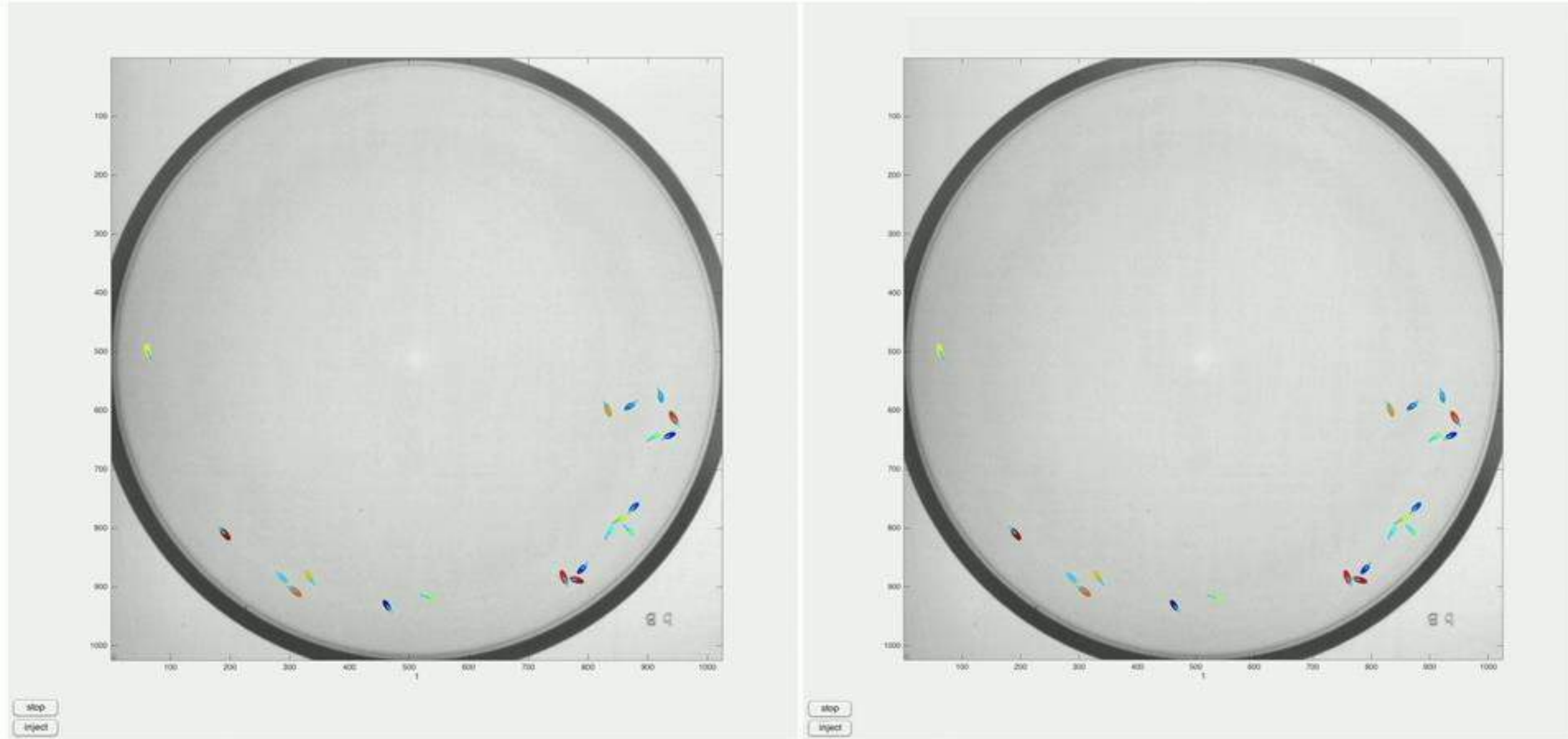






Eyrún  
Eyolfsson

# Drosophila Behavior





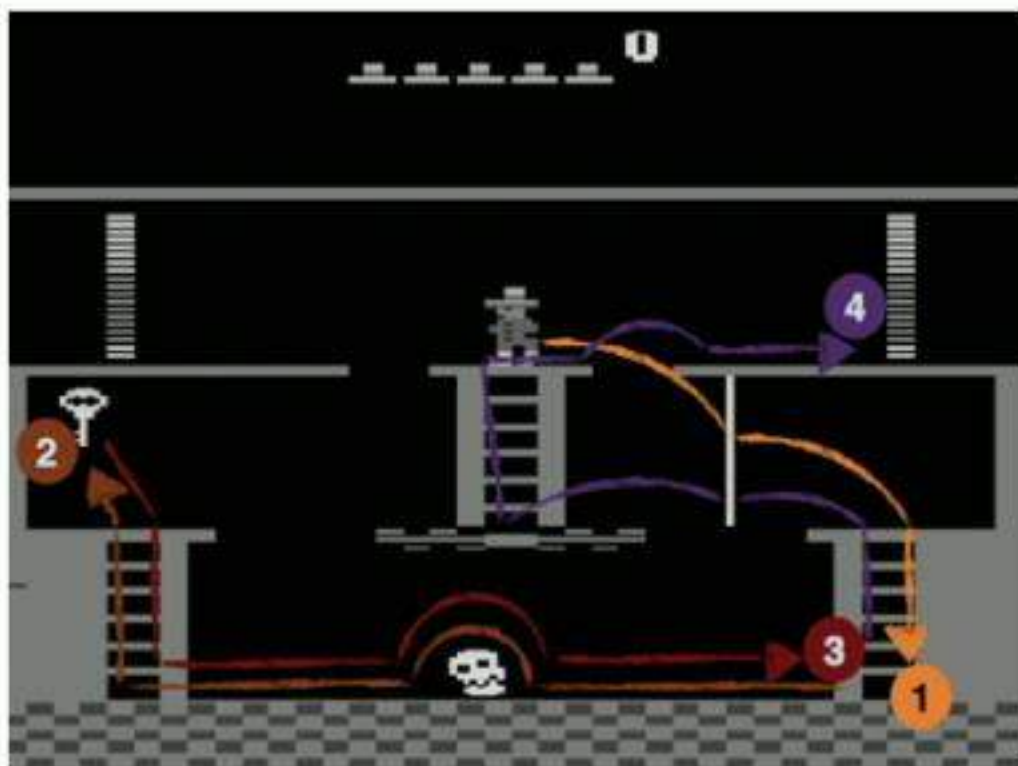
# Activity Labels



**Learning recurrent representations for hierarchical behavior modeling**  
Eyrún Eyolfsson, Kristin Branson, Yisong Yue, Pietro Perona, ICLR 2017

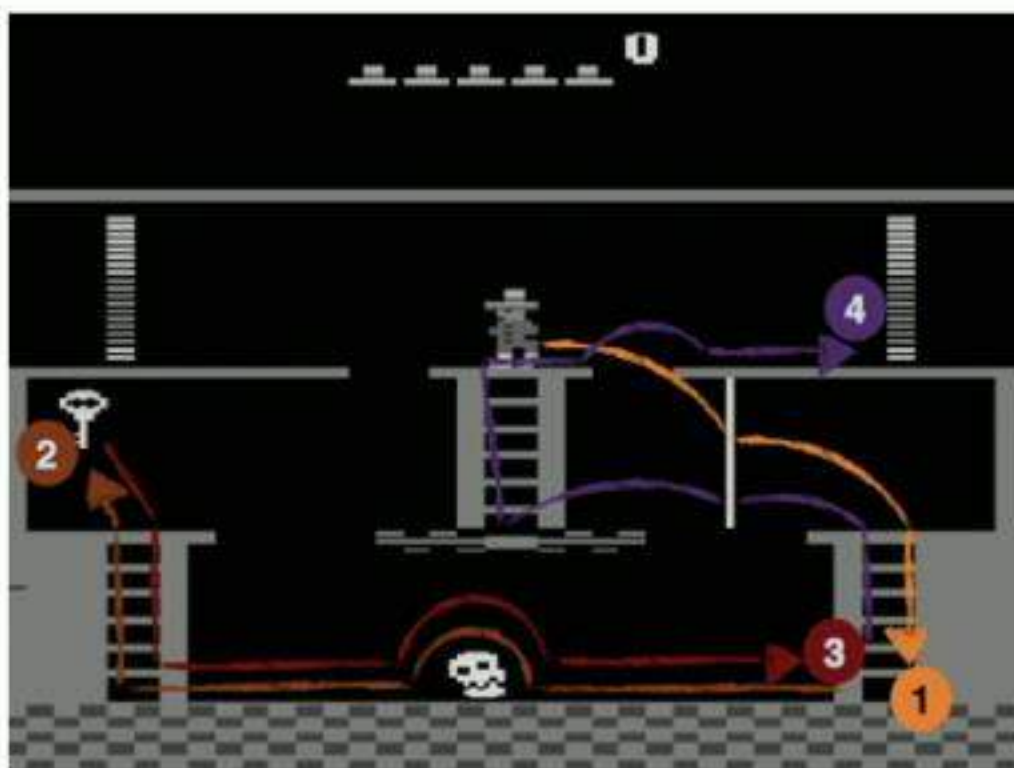
## Aside: Hierarchically Composing IL & RL

- IL for meta-controller (plan sub-goals)
- RL/IL for low-level controllers (individual sub-goals)

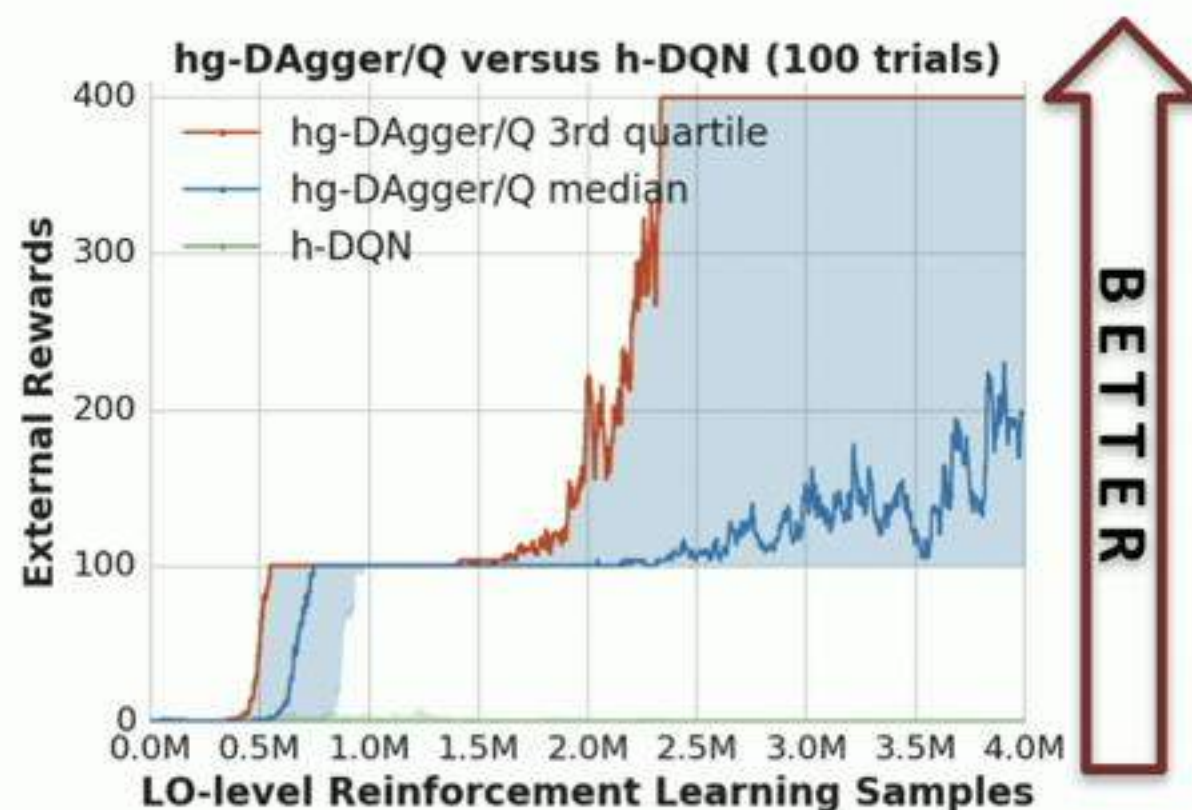


## Aside: Hierarchically Composing IL & RL

- IL for meta-controller (plan sub-goals)
- RL/IL for low-level controllers (individual sub-goals)



- More label efficient than flat IL
- Converge much faster than conventional hierarchical RL



### Hierarchical Imitation and Reinforcement Learning

Hoang Le, Nan Jiang, Alekh Agarwal, Miro Dudik, Yisong Yue, Hal Daume. ICML 2018





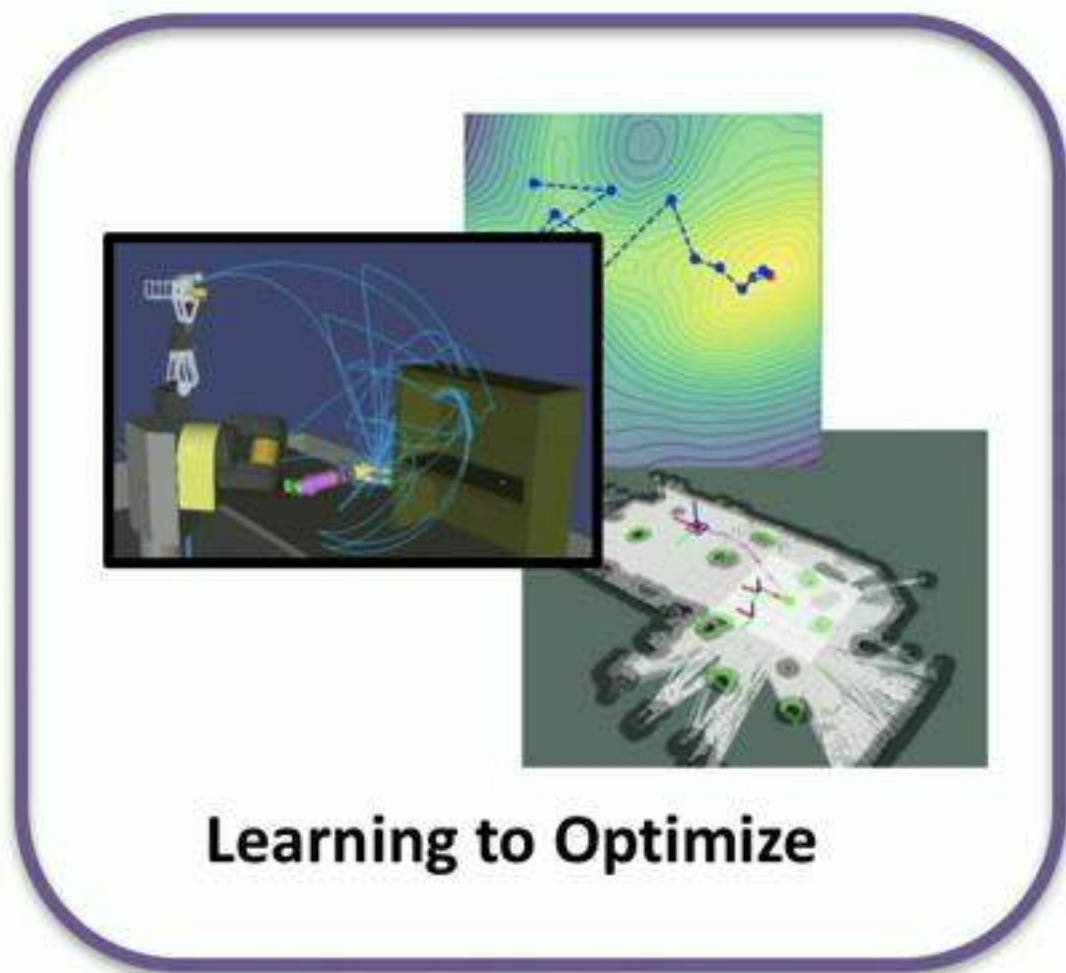
**Speech Animation**



**Coordinated Learning**



**Hierarchical Behaviors  
(Generative)**



**Learning to Optimize**



**Smooth Imitation Learning**

# Optimization as Sequential Decision Making

- Many solvers are sequential:
  - Greedy
  - Search heuristics
  - Gradient Descent
- Can view as solver as “agent”
  - State = intermediate solution
  - Find a state with high reward (solution)

# Optimization as Sequential Decision Making

## Contextual Submodular Maximization

- Training set:  $(x, F_x)$
- Greedily maximize  $F_x$  using only  $x$
- **Learning Policies for Contextual Submodular Prediction [ICML 2013]**



Stephane Ross



# Optimization as Sequential Decision Making

## Contextual Submodular Maximization

- Training set:  $(x, F_x)$
- Greedily maximize  $F_x$  using only  $x$
- **Learning Policies for Contextual Submodular Prediction [ICML 2013]**



Stephane Ross

## Learning to Search

- Training set:  $(x=\text{MILP}, y=\text{solution/search-trace})$
- Find  $y$  (or better solution)
- **Learning to Search via Retrospective Imitation [arXiv]**



Jialin Song

# Optimization as Sequential Decision Making

## Contextual Submodular Maximization

- Training set:  $(x, F_x)$
- Greedily maximize  $F_x$  using only  $x$
- **Learning Policies for Contextual Submodular Prediction [ICML 2013]**



Stephane Ross

## Learning to Search

- Training set:  $(x=\text{MILP}, y=\text{solution/search-trace})$
- Find  $y$  (or better solution)
- **Learning to Search via Retrospective Imitation [arXiv]**



Jialin Song

## Learning to Infer

- Training set:  $(x=\text{data/model}, L=\text{likelihood})$
- Iteratively optimize  $L$  (generalizes VAEs)
- **Iterative Amortized Inference [ICML 2018]**
- **A General Method for Amortizing Variational Filtering [NIPS 2018]**



Joe Marino





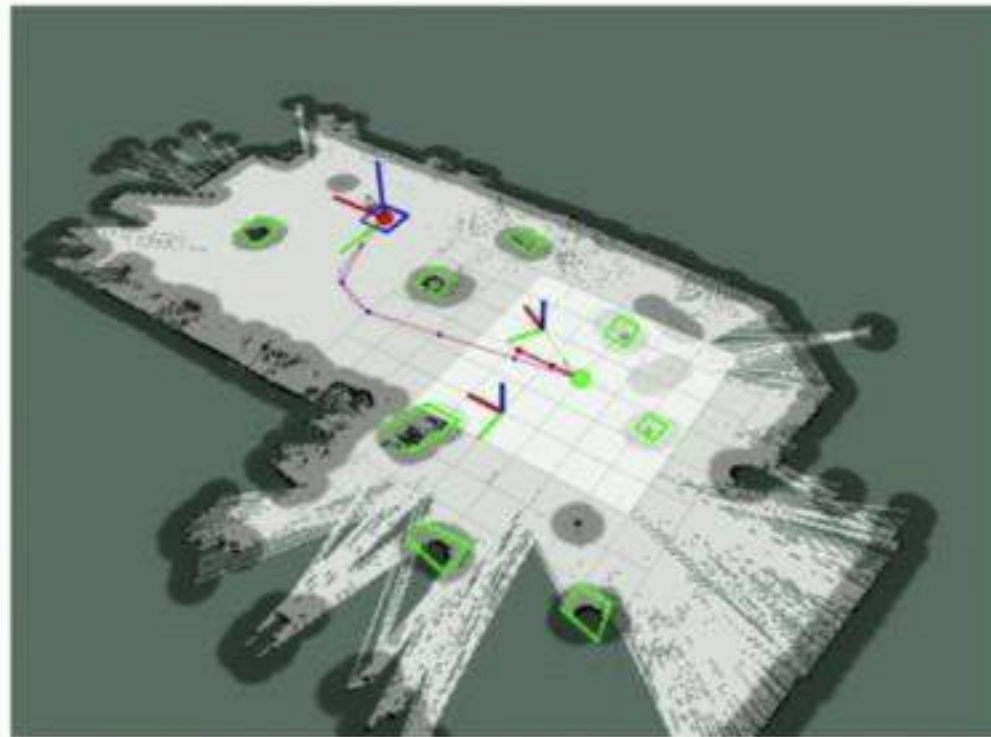
Ravi  
Lanka

# Ongoing Research

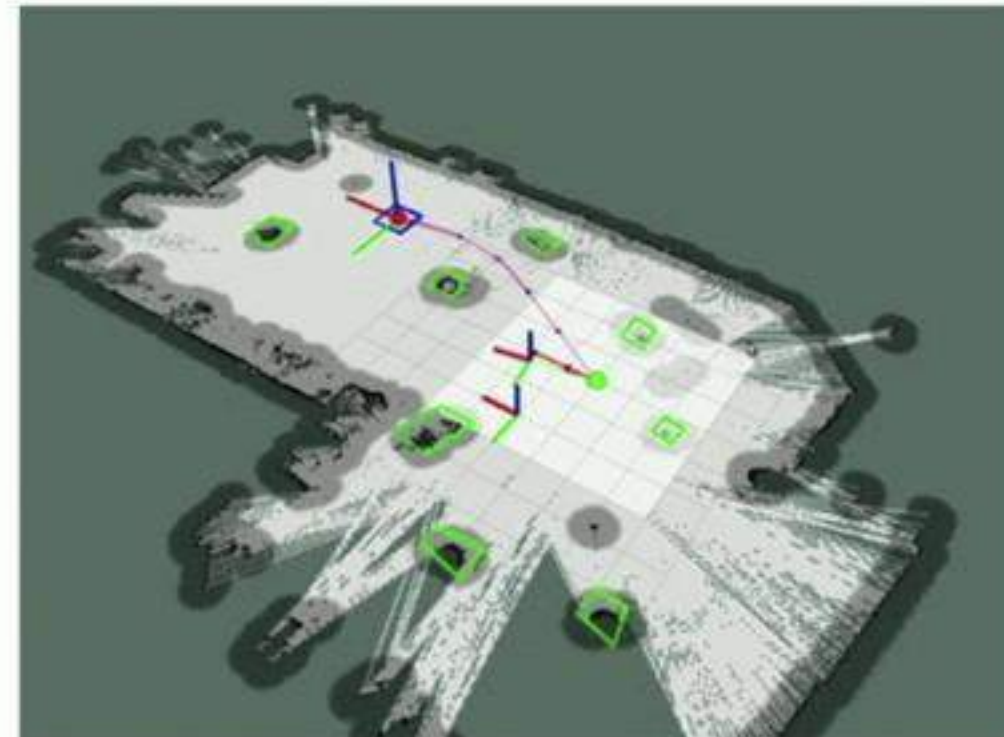
## Risk-Aware Planning



Jialin  
Song



Low Risk

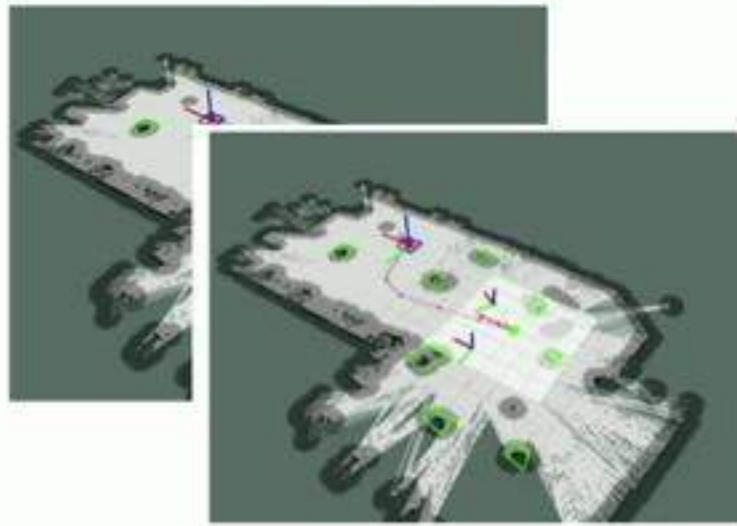


High Risk

- Compiled as mixed integer program
- Challenging optimization problem







Distribution of Planning Problems

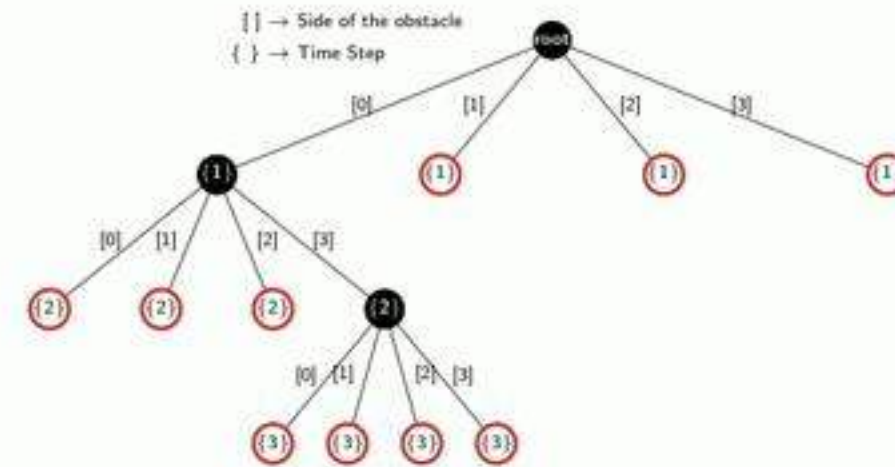


$$\min_{\mathbf{U}} J(\mathbf{U}, \mathbf{X})$$

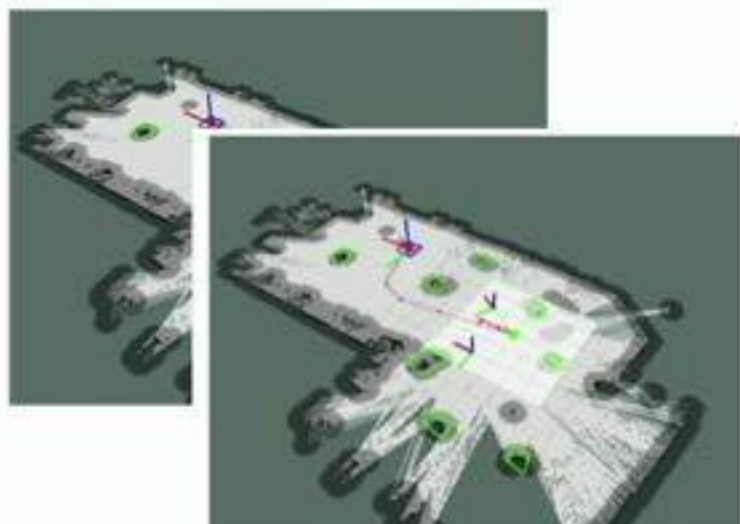
subject to,

(Dynamic Constraint)  $x_{t+1} = Ax_t + Bu_t,$

(Safety Constraints)  $h_t^{iT} x_t \leq g_t^i$



Compiled as Combinatorial Search Problems



Distribution of Planning Problems

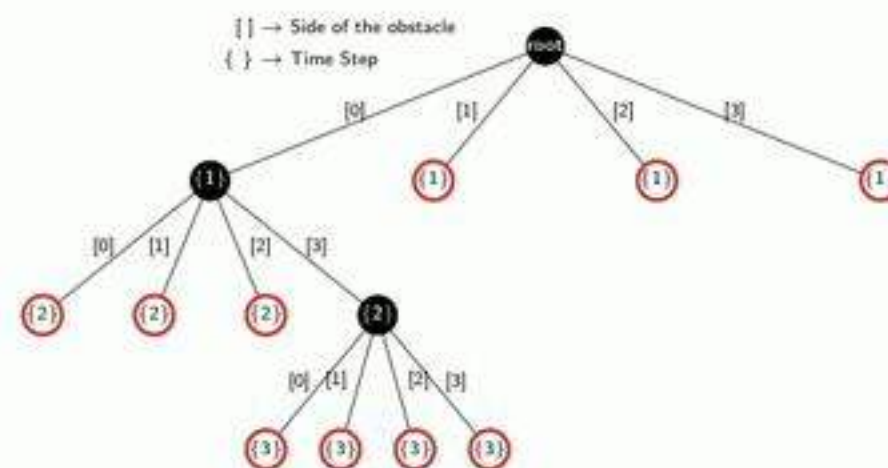


$$\min_{\mathbf{U}} J(\mathbf{U}, \mathbf{X})$$

subject to,

(Dynamic Constraint)  $x_{t+1} = Ax_t + Bu_t,$

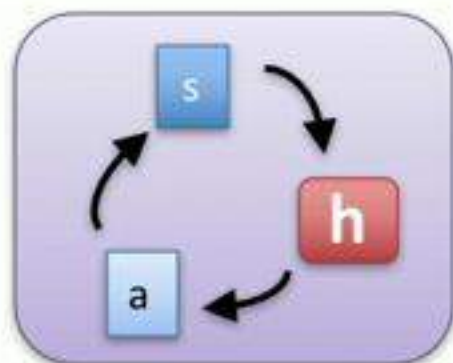
(Safety Constraints)  $h_t^{iT} x_t \leq g_t^i$



Compiled as Combinatorial Search Problems

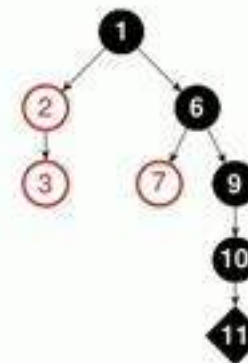


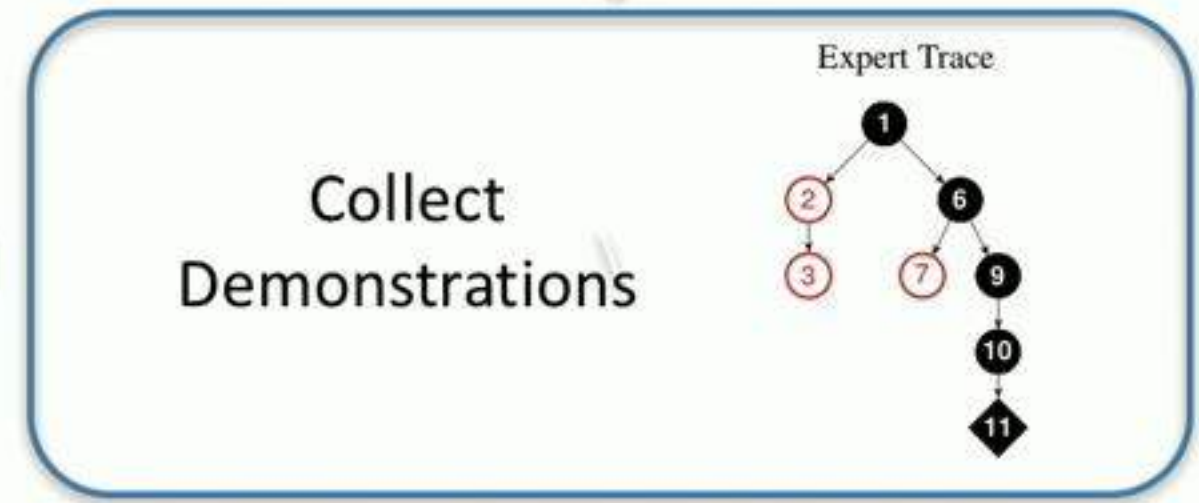
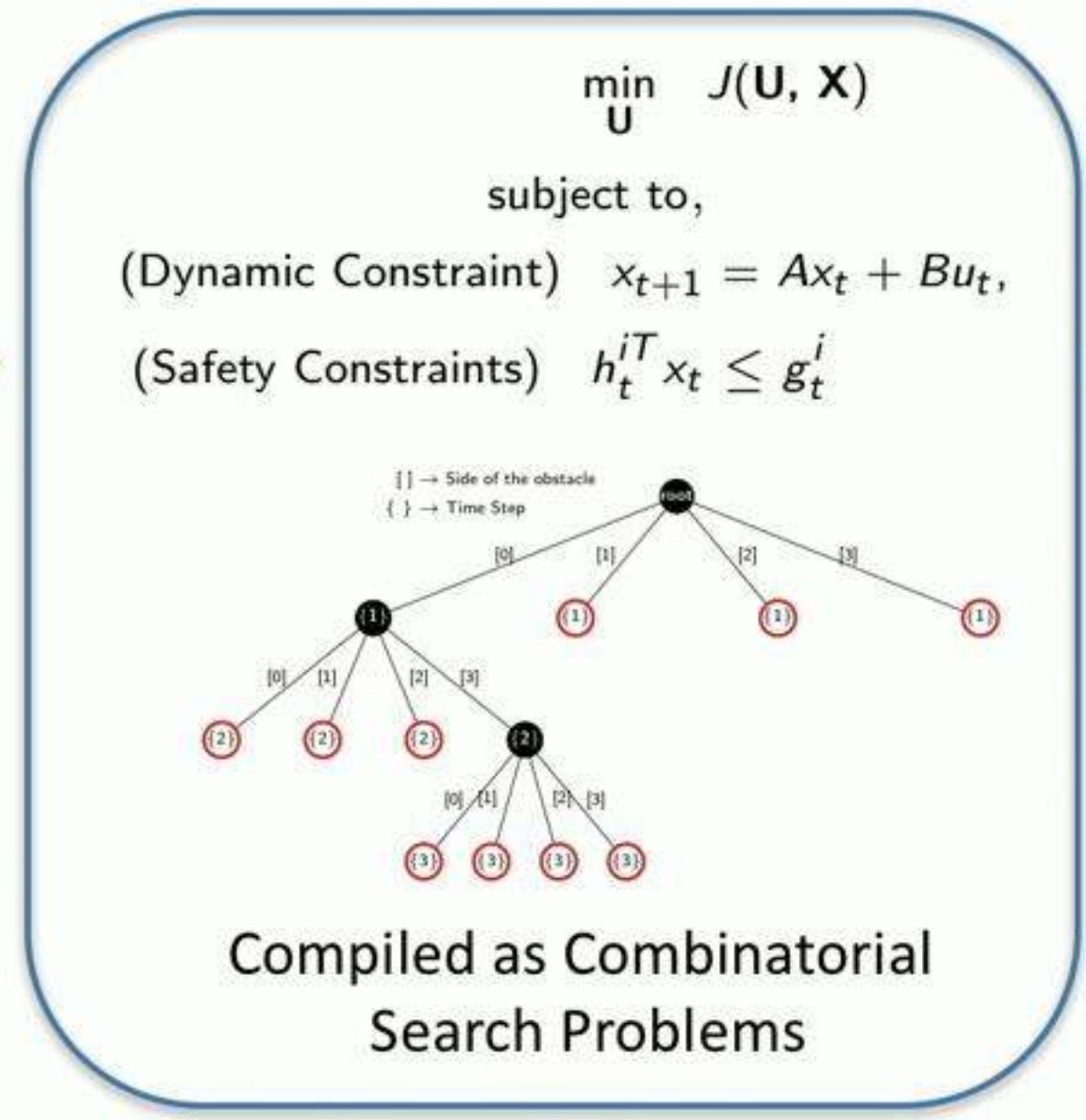
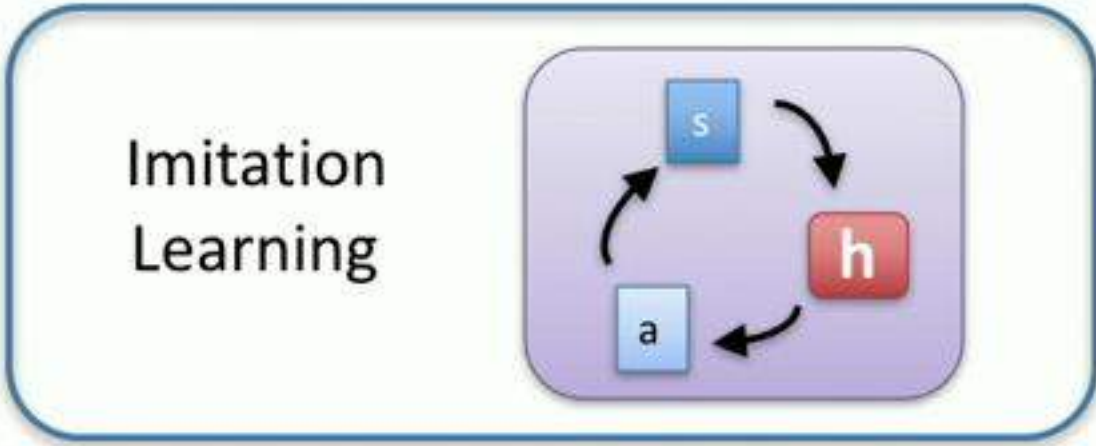
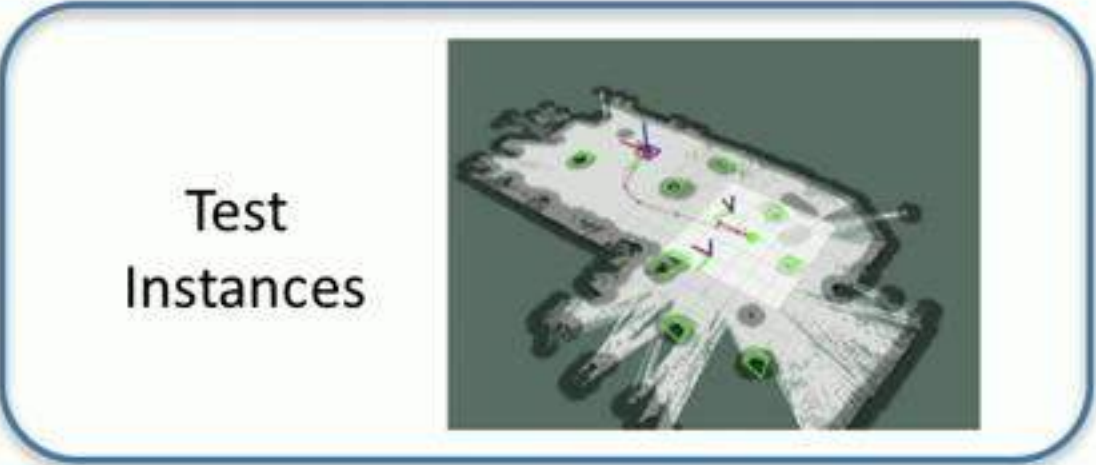
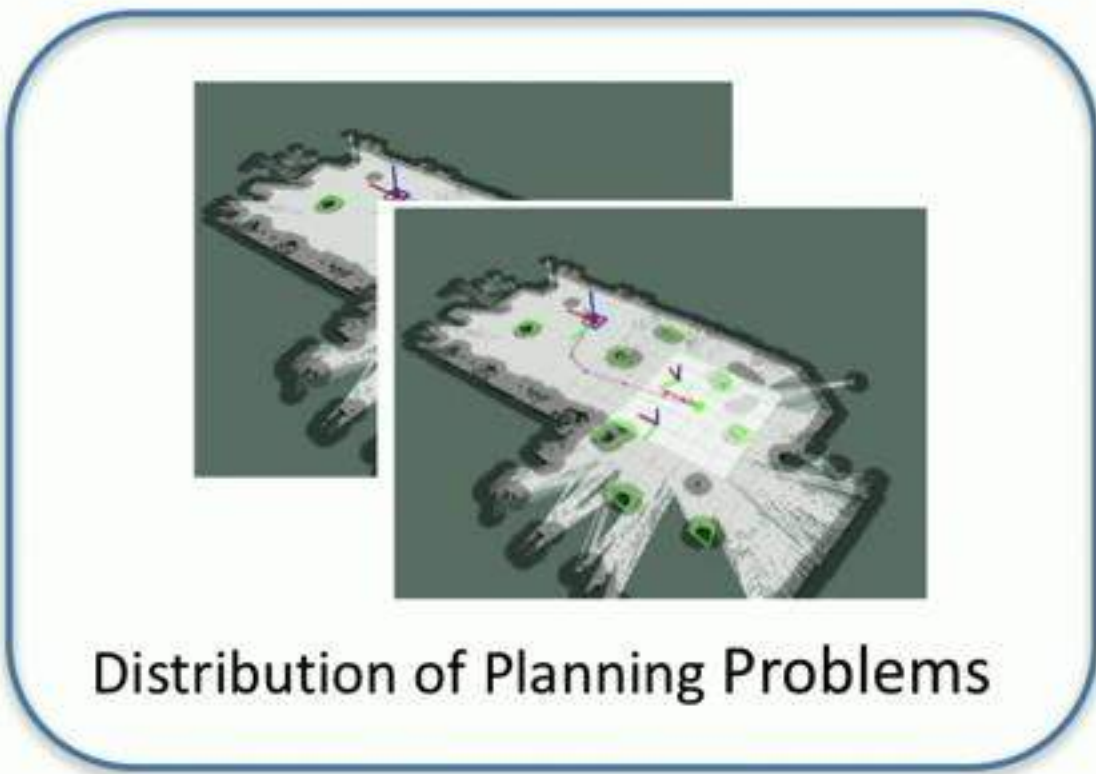
Imitation Learning



Collect Demonstrations

Expert Trace

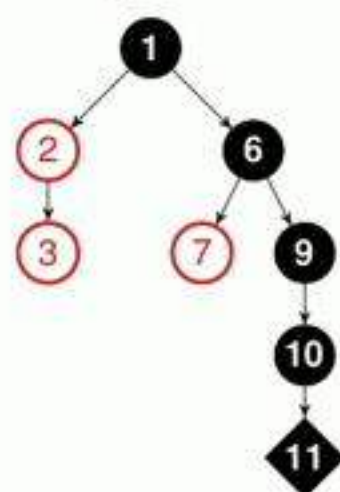






# Learning to Search via Retrospective Imitation

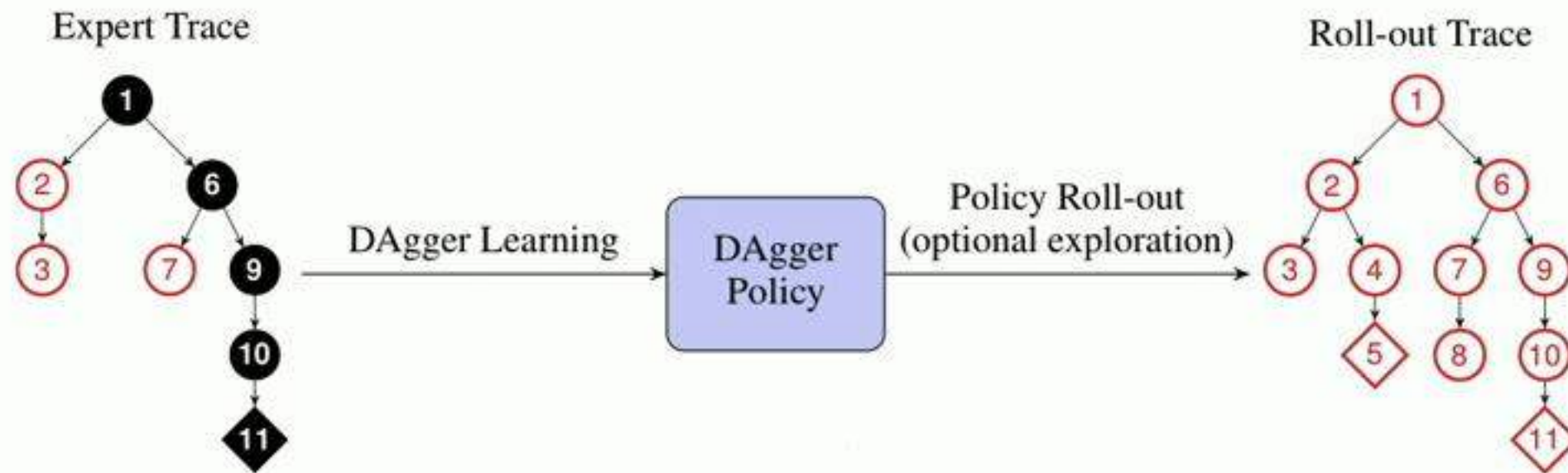
Expert Trace



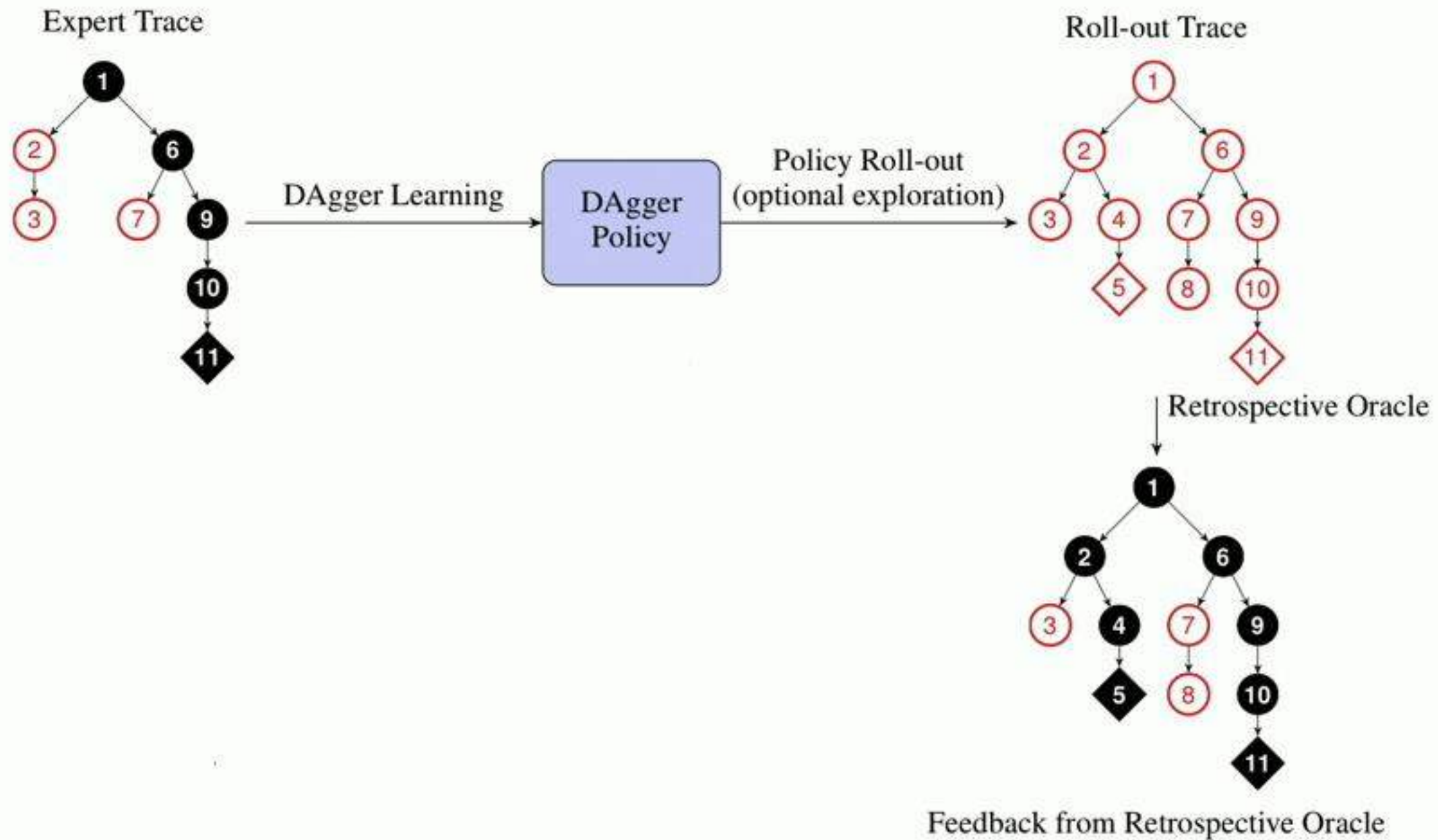
**Learning to Search via Retrospective Imitation**

R. Lanka, J. Song, A. Zhao, Y. Yue, M. Ono. arXiv

# Learning to Search via Retrospective Imitation



# Learning to Search via Retrospective Imitation

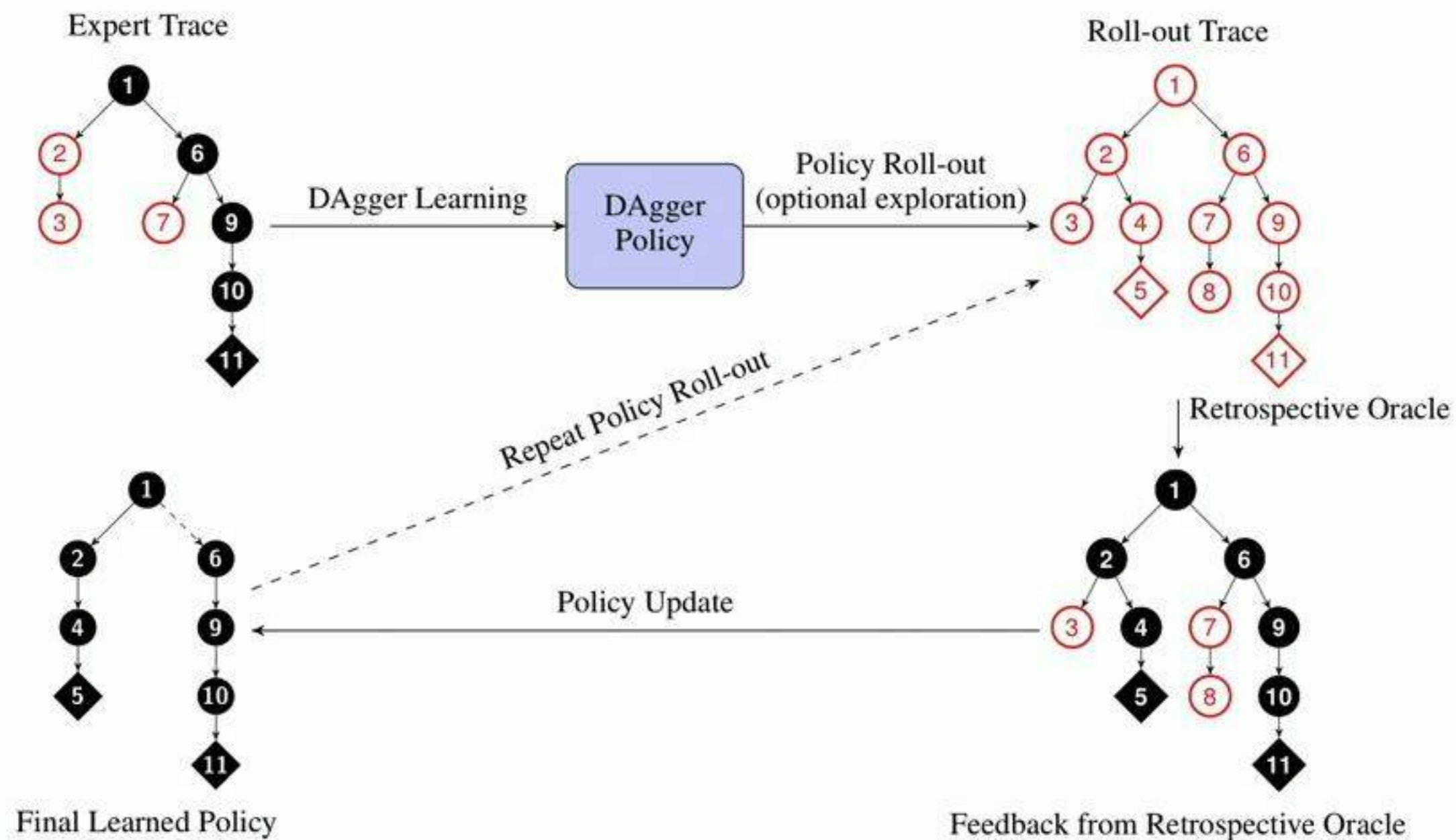


**Learning to Search via Retrospective Imitation**

R. Lanka, J. Song, A. Zhao, Y. Yue, M. Ono. arXiv



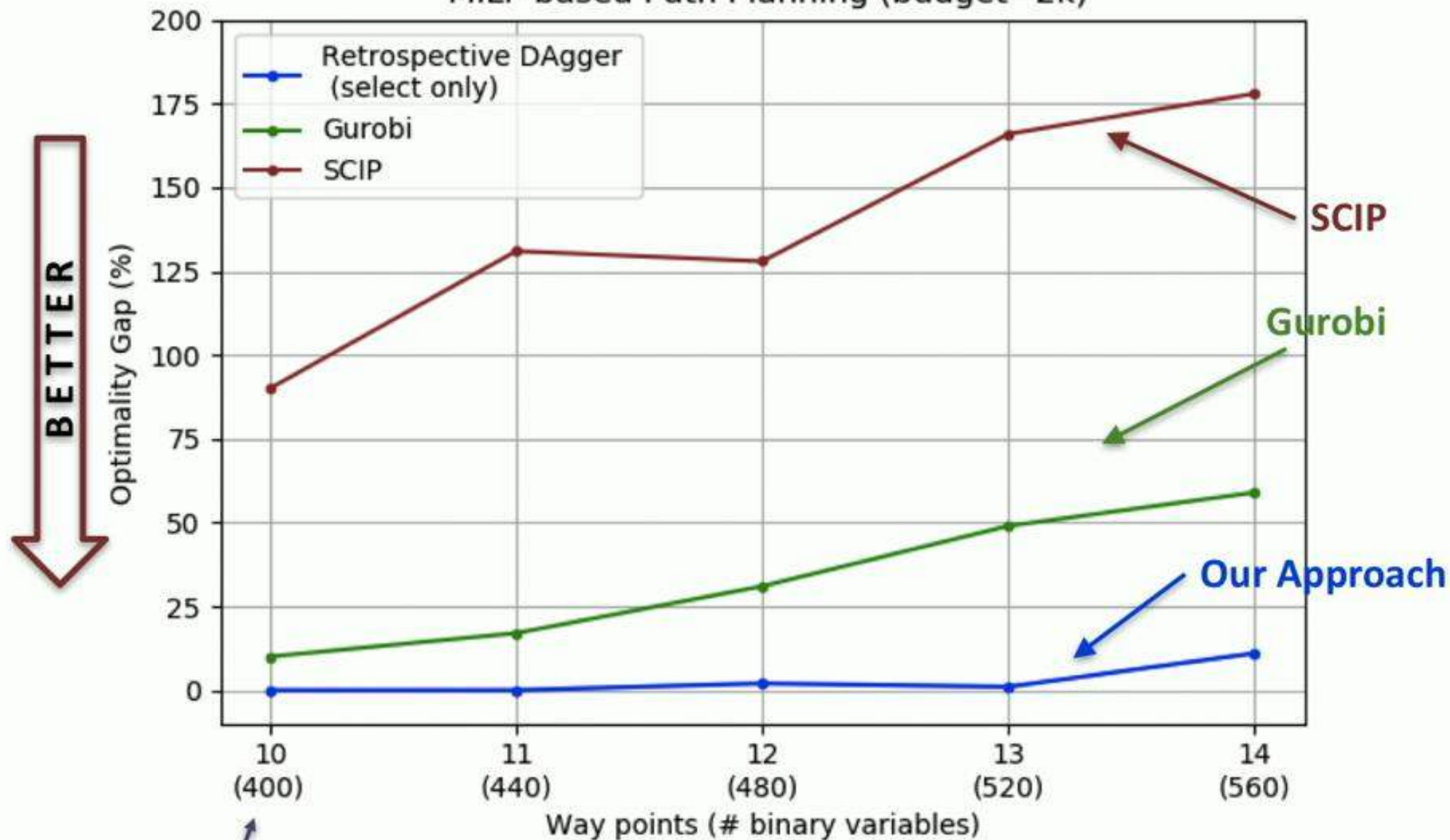
# Learning to Search via Retrospective Imitation



Learning to Search via Retrospective Imitation

R. Lanka, J. Song, A. Zhao, Y. Yue, M. Ono. arXiv

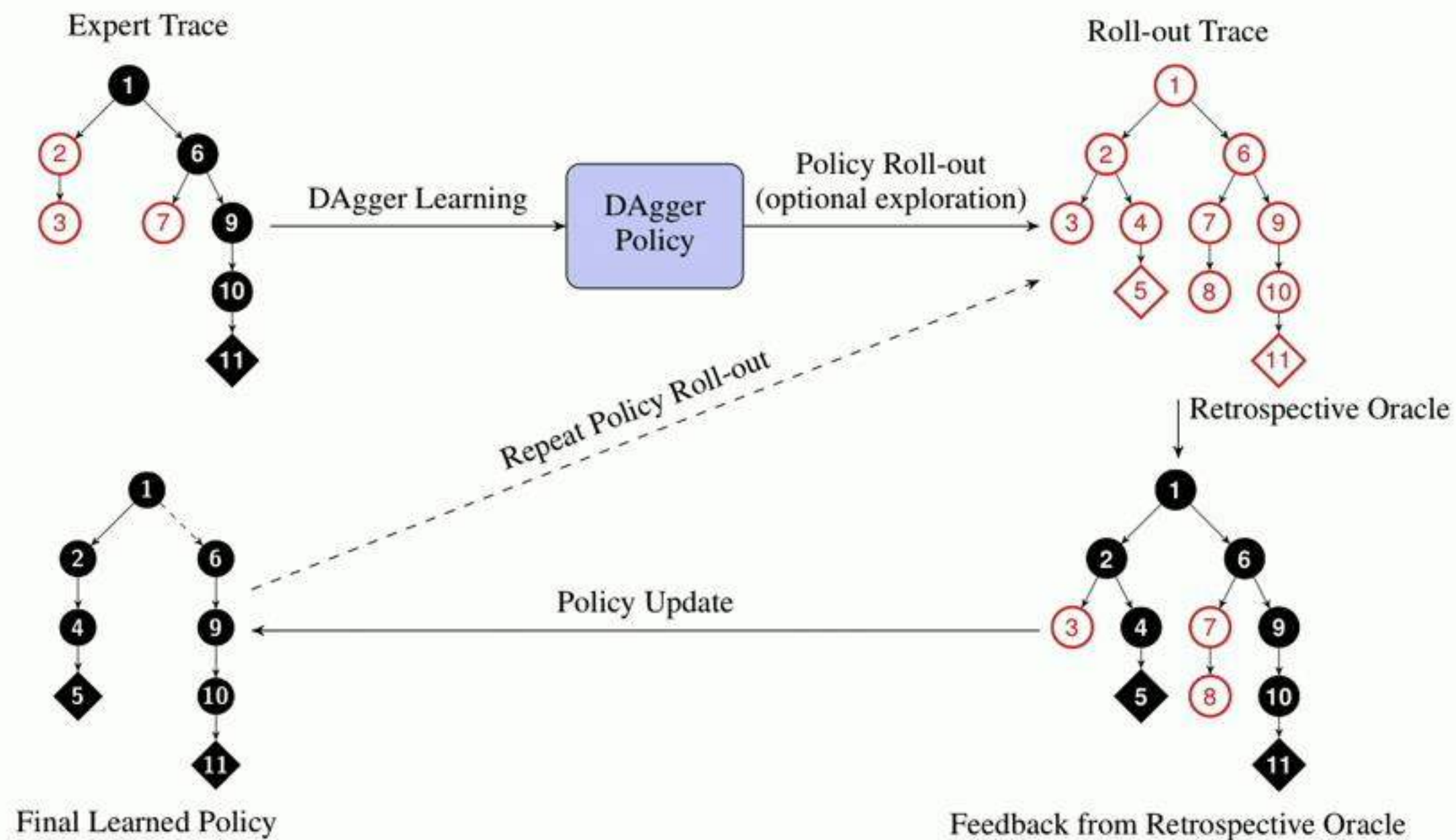
### Retrospective DAgger vs Heuristics for MILP based Path Planning (budget=2k)



Initial demonstrations only at smallest size!

**Learning to Search via Retrospective Imitation**  
R. Lanka, J. Song, A. Zhao, Y. Yue, M. Ono. arXiv

# Learning to Search via Retrospective Imitation

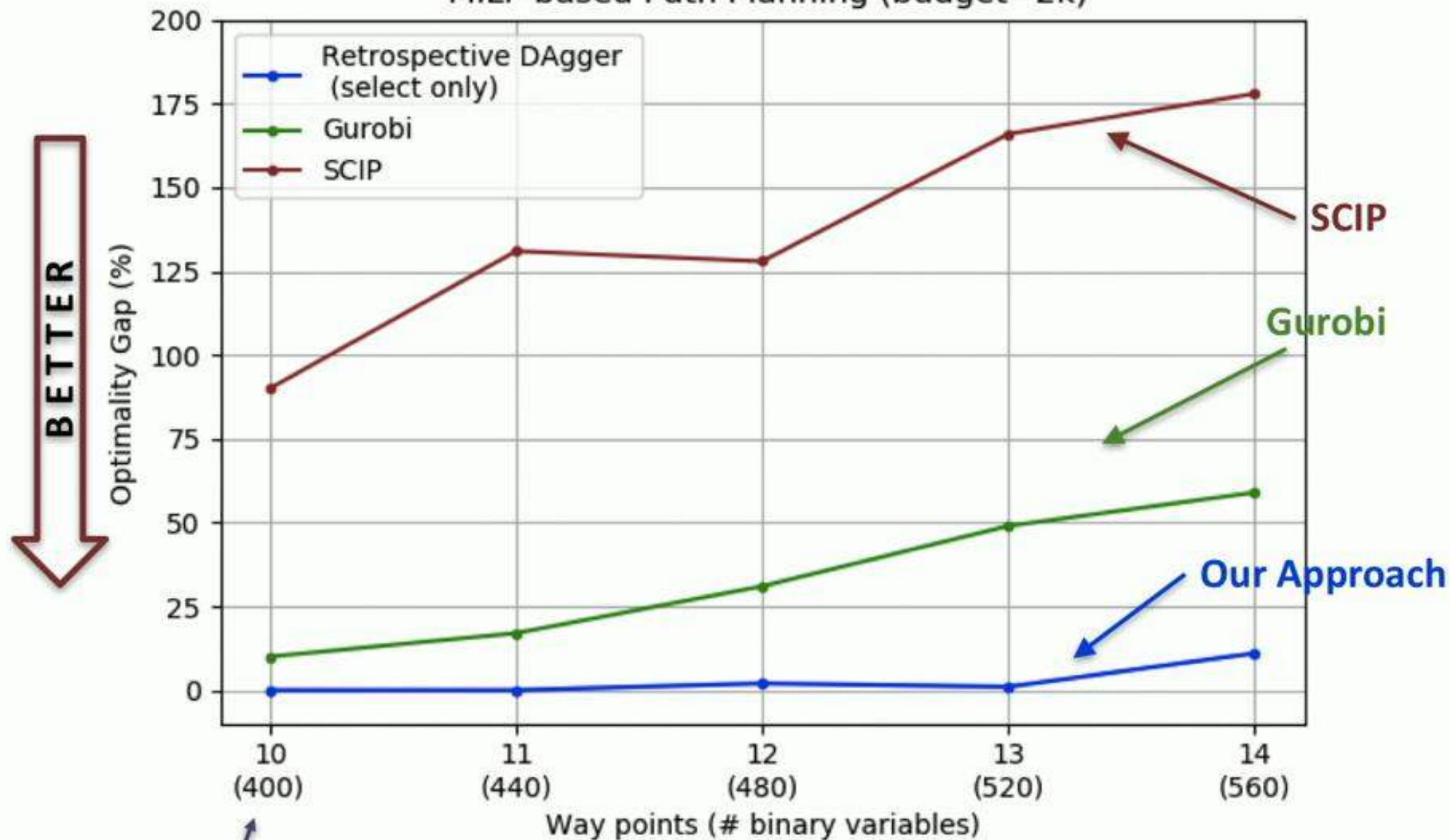


Learning to Search via Retrospective Imitation

R. Lanka, J. Song, A. Zhao, Y. Yue, M. Ono. arXiv

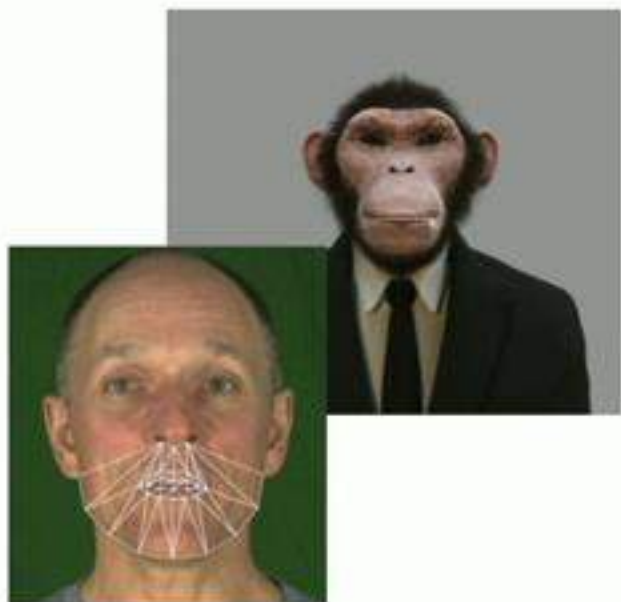


### Retrospective DAgger vs Heuristics for MILP based Path Planning (budget=2k)



Initial demonstrations only at smallest size!

**Learning to Search via Retrospective Imitation**  
R. Lanka, J. Song, A. Zhao, Y. Yue, M. Ono. arXiv



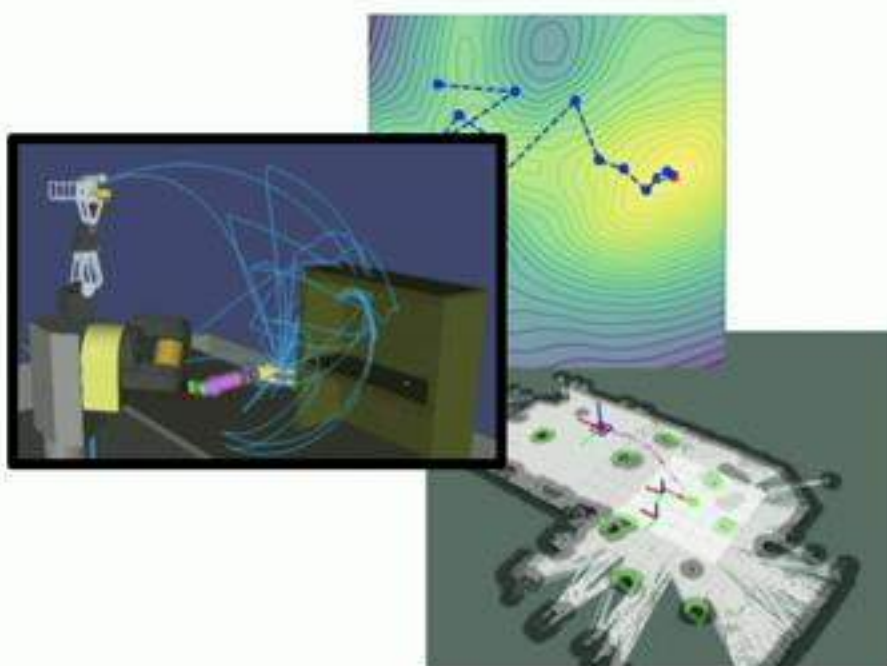
**Speech Animation**



**Coordinated Learning**



**Hierarchical Behaviors  
(Generative)**



**Learning to Optimize**



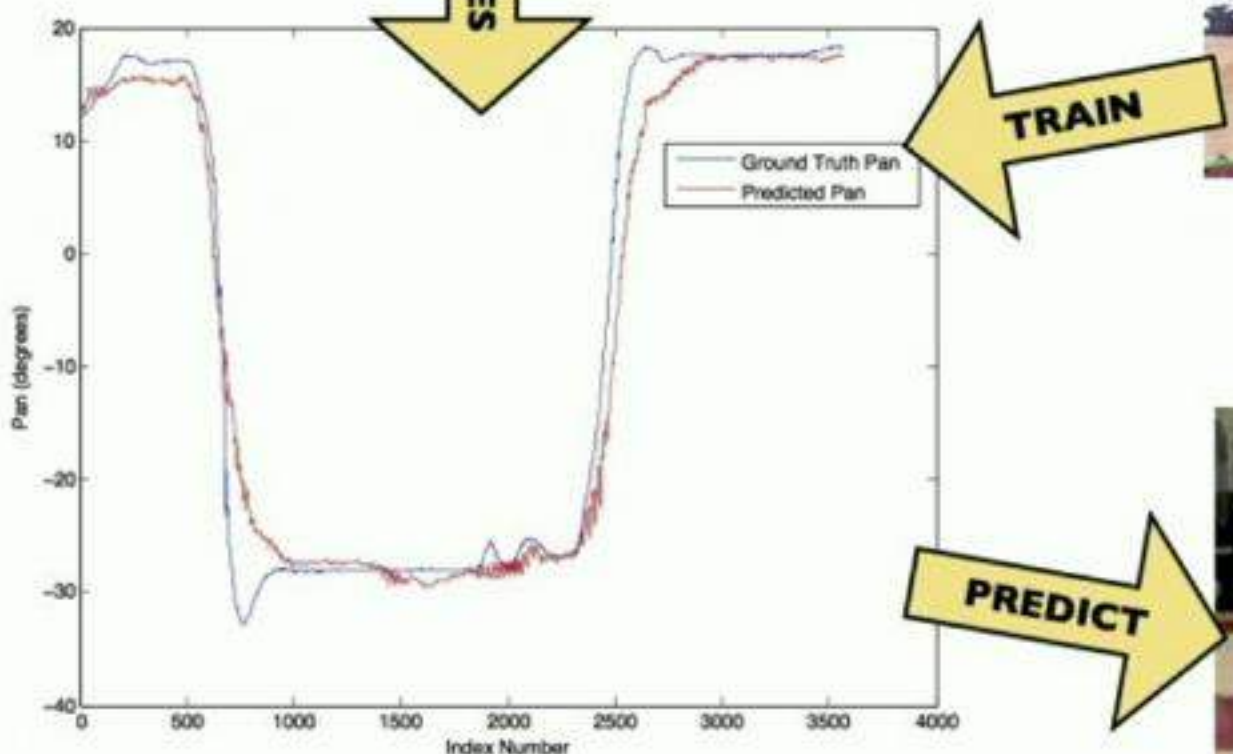
**Smooth Imitation Learning**



# Realtime Player Detection and Tracking



FEATURES



Learned Regressor

## Human Operated Camera

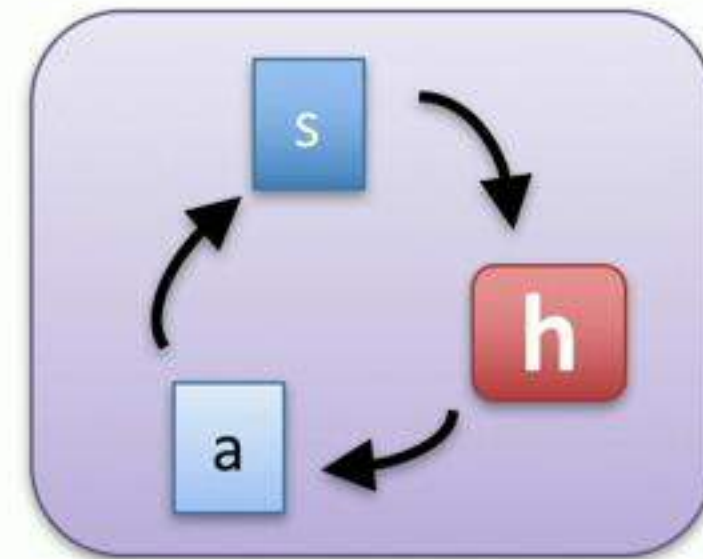


Autonomous Robotic Camera



# Problem Formulation

- Input: stream of  $x_t$ 
  - E.g., noisy player detections
- State  $s_t = (x_{t:t-K}, a_{t-1:t-K})$ 
  - Recent detections and actions
- Goal: learn  $h(s_t) \rightarrow a_t$ 
  - Imitate expert

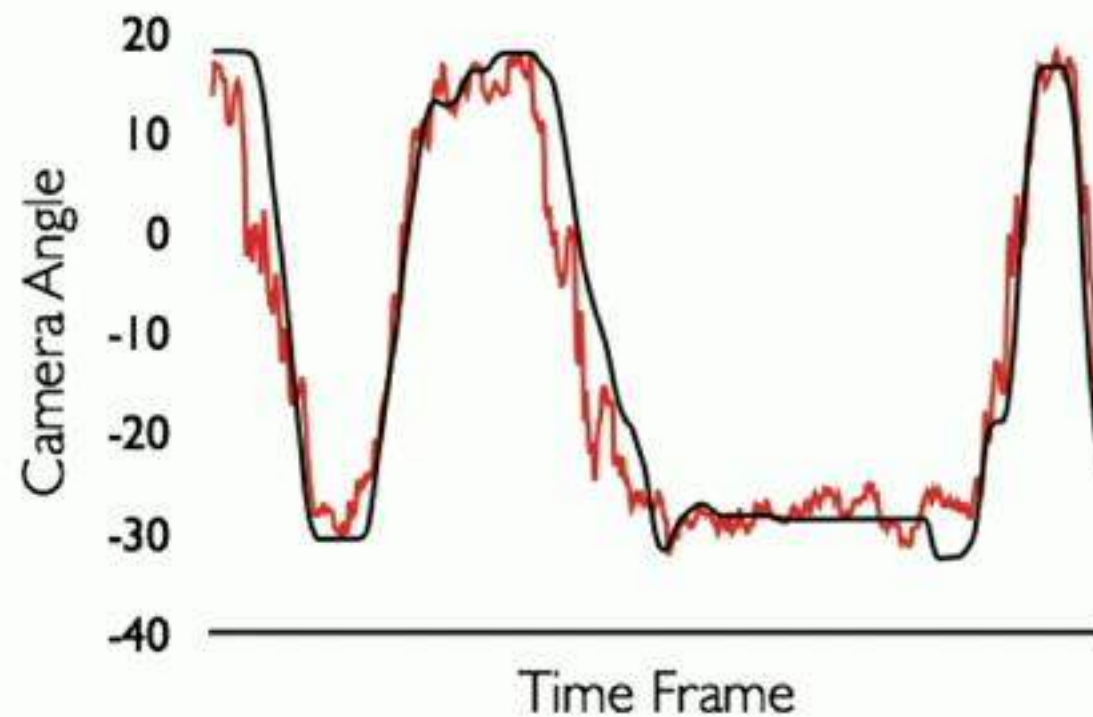


# Naïve Approach

- Supervised learning of demonstration data
  - Train predictor per frame
  - Predict per frame

# Naïve Approach

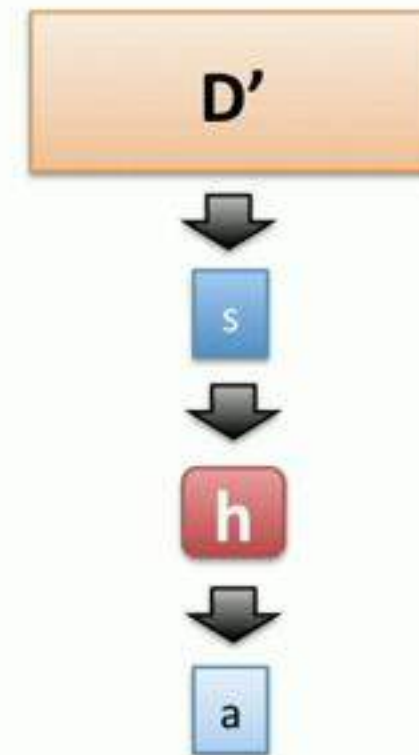
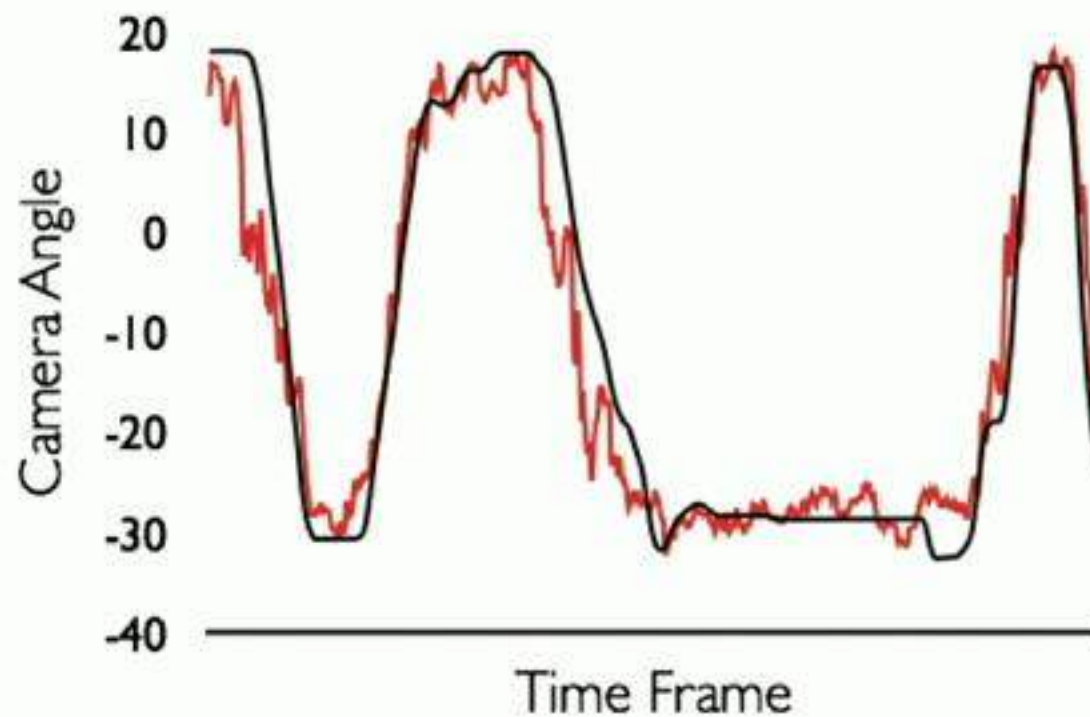
- Supervised learning of demonstration data
  - Train predictor per frame
  - Predict per frame





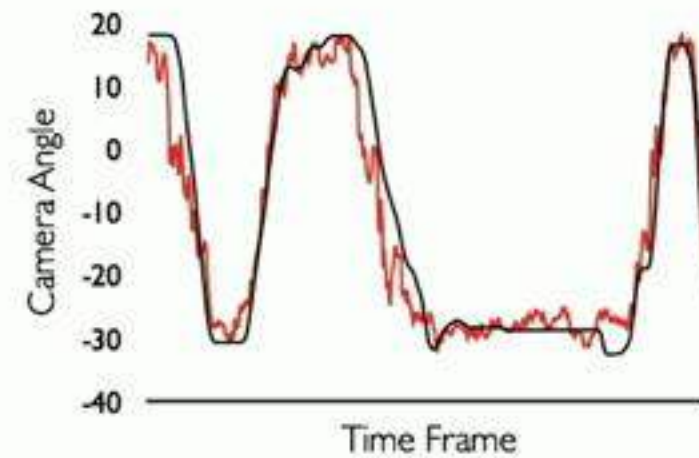
# Naïve Approach

- Supervised learning of demonstration data
  - Train predictor per frame
  - Predict per frame



# What is the Problem?

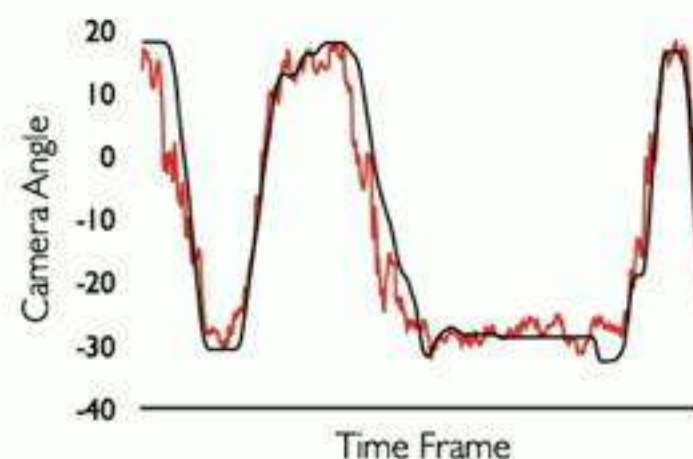
- Basically takes “infinite” training data to train smooth model.
  - Via input/output examples



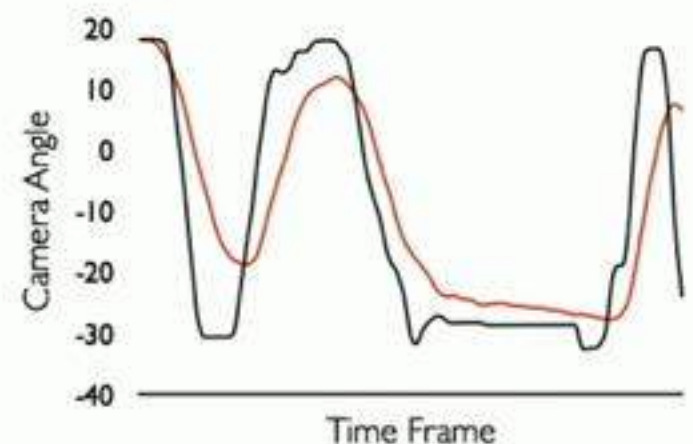
# What is the Problem?

- Basically takes “infinite” training data to train smooth model.

– Via input/output examples



- In practice, people do post-hoc smoothing





# Cannot Rely 100% on Learning!

- People have models of smoothness!
  - Kalman Filters
  - Linear Autoregressors
  - Etc...
- Pure ML approach throws them away!
  - "black box"

# Hybrid Model-Based + Black-Box

- Model-based approaches
  - Strong assumptions, well specified
  - Lacks flexibility
  - E.g., Kalman Filter, Linear Autoregressor
- Black-box approaches
  - Assumption free, underspecified
  - Requires a lot of training data
  - E.g., random forest, deep neural network
- **Best of both worlds?**

**Conventional  
Models**

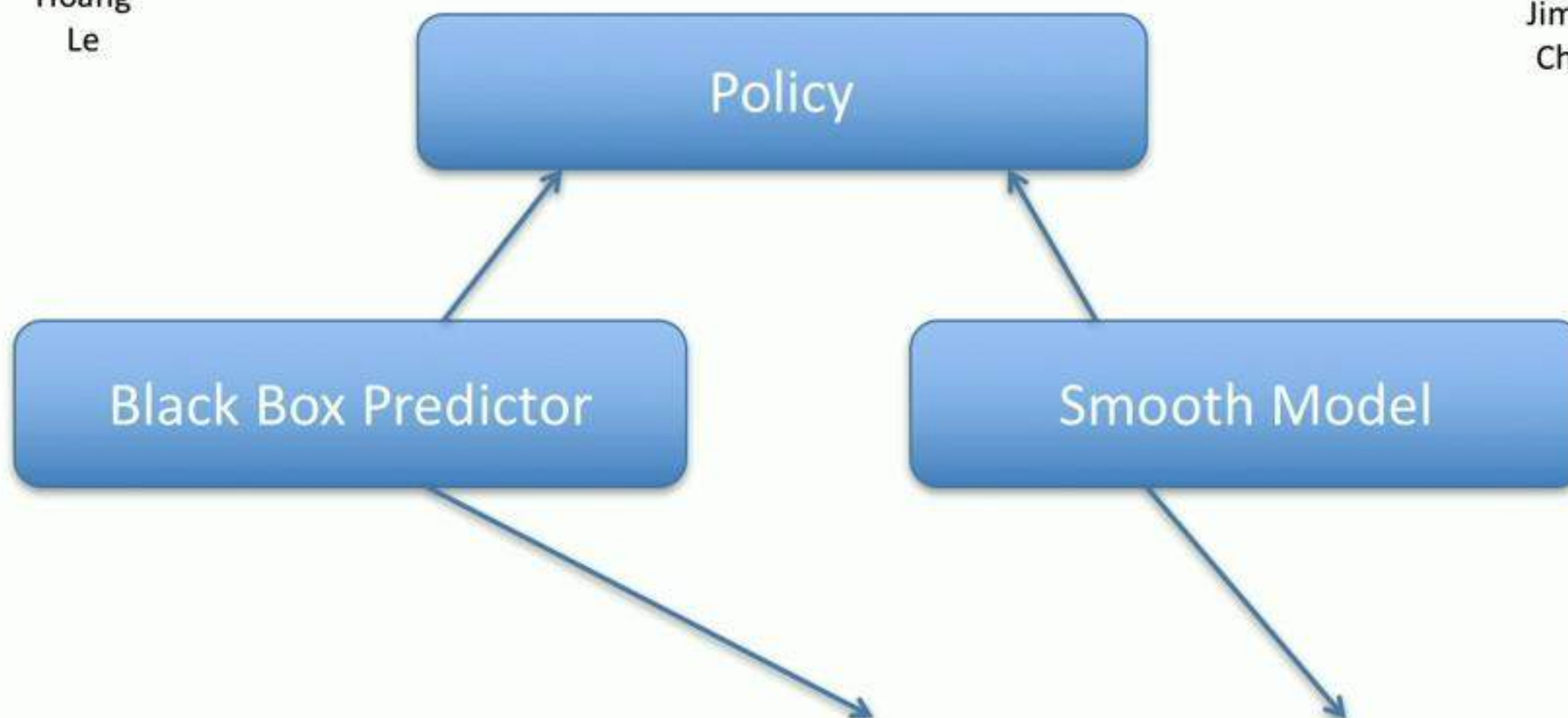


Hoang  
Le



Jimmy  
Chen

# New Policy Class

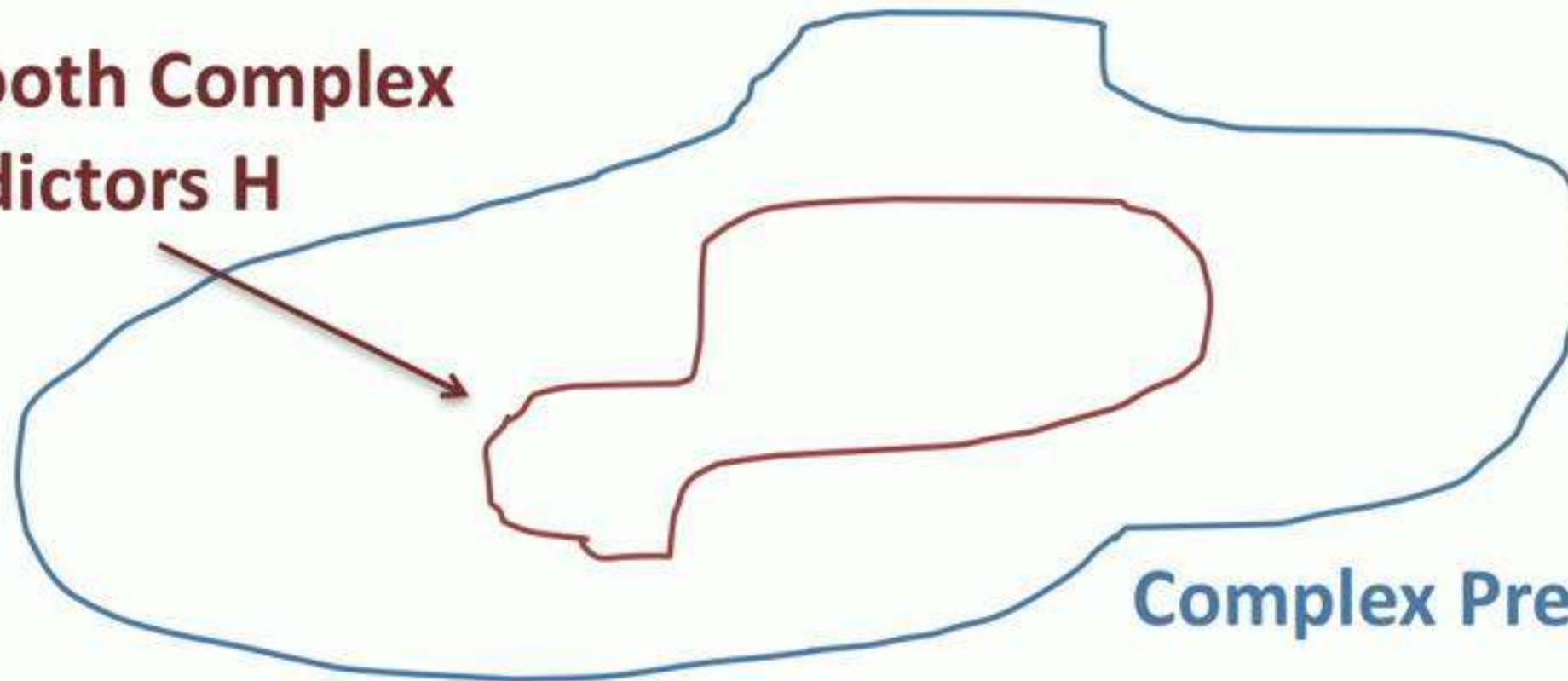


$$h(s_t \equiv (x_{t:t-K}, a_{t-1:t-K})) = \operatorname{argmin}_{a'} (f(s_t) - a')^2 + \lambda (g(a_{t-1:t-K}) - a')^2$$
$$= \frac{f(s_t) + \lambda g(a_{t-1:t-K})}{1 + \lambda}$$



# Functional Regularization

**Smooth Complex Predictors H**



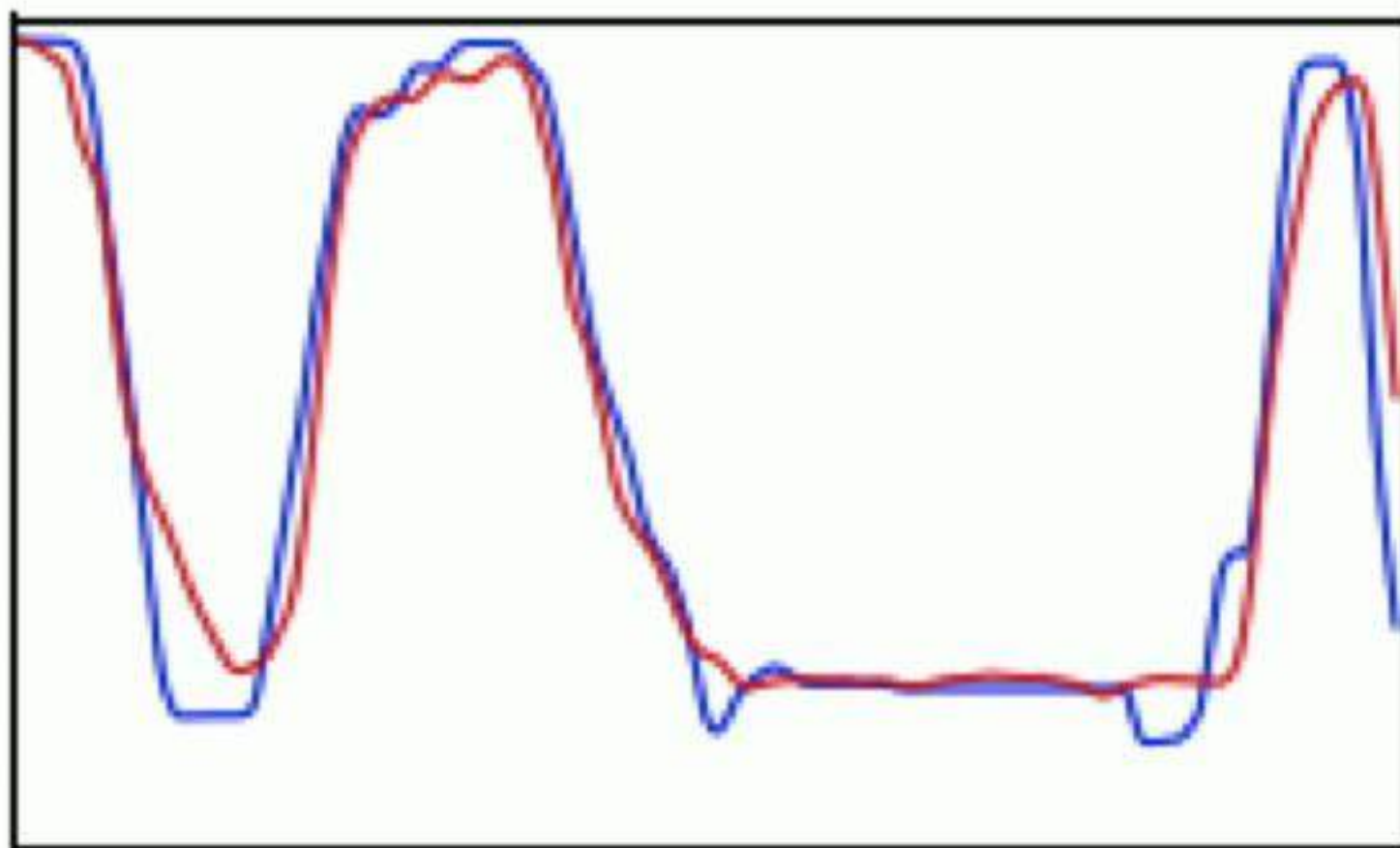
**Complex Predictors F**

$$\begin{aligned} h(s_t \equiv (x_{t:t-K}, a_{t-1:t-K})) &= \operatorname{argmin}_{a'} (f(s_t) - a')^2 + \lambda (g(a_{t-1:t-K}) - a')^2 \\ &= \frac{f(s_t) + \lambda g(a_{t-1:t-K})}{1 + \lambda} \end{aligned}$$

**Smooth Imitation Learning for Online Sequence Prediction**

Hoang Le, Andrew Kang, Yisong Yue, Peter Carr. ICML 2016

# Our Result



$$h(s_t \equiv (x_{t:t-K}, a_{t-1:t-K})) = \frac{f(s_t) + \lambda g(a_{t-1:t-K})}{1 + \lambda}$$

**Smooth Imitation Learning for Online Sequence Prediction**

Hoang Le, Andrew Kang, Yisong Yue, Peter Carr. ICML 2016



# Qualitative Comparison



Baseline



Our Approach

**Learning Online Smooth Predictors for Real-time Camera Planning using Recurrent Decision Trees**  
Jianhui Chen, Hoang Le, Peter Carr, Yisong Yue, Jim Little. CVPR 2016



# Qualitative Comparison



Baseline



Our Approach

**Learning Online Smooth Predictors for Real-time Camera Planning using Recurrent Decision Trees**  
Jianhui Chen, Hoang Le, Peter Carr, Yisong Yue, Jim Little. CVPR 2016



# Qualitative Comparison



TECH & MEDIA  
**Disney using human operators to train automatic cameras for broadcasts**



SHARE

**BEN RAINS**  
Thursday June 23rd, 2016

Read about the latest sports tech news, innovations, ideas and products that impact players, fans and the sports industry overall at [SportTechie.com](#).

The Walt Disney Company recently announced they would be enhancing their basketball and soccer television coverage by improving their automated camera technology. Computer engineers are helping automated cameras learn from human operators to help create a smoother and cleaner broadcast.



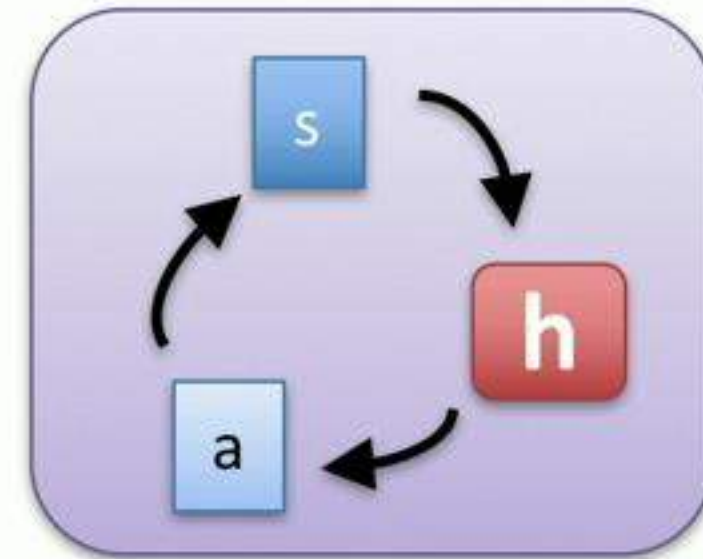
approach

Recurrent Decision Trees

Learning Online  
Jianhui Chen, Hoang

# Definition: Rollout

- Execute  $h$  sequentially





# Definition: Learning Reduction

- Original Learning Problem:
  - Sequential Decision Making

$$D = \{(\vec{s}, \vec{a})\}$$

- Converted Learning Problem:
  - Classification / Regression

$$D' = \{(s', a')\}$$

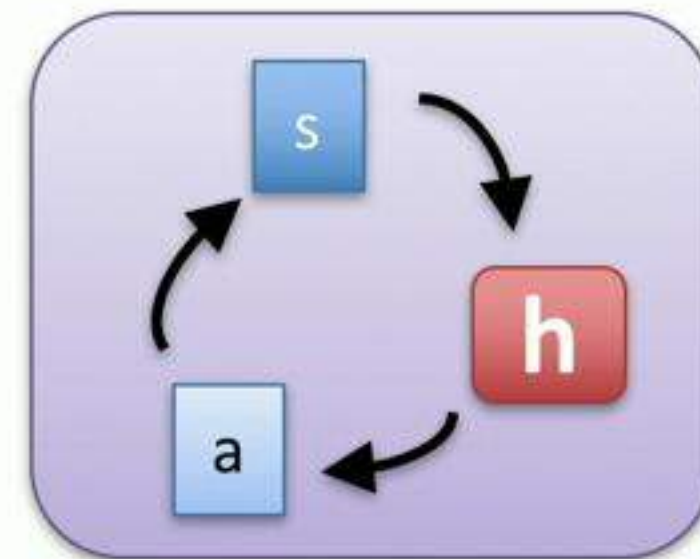
- Train  $h$  on  $D'$  (easy to do)

- **Theoretical Goals:**

- Guarantees on  $D'$  lift to  $D$

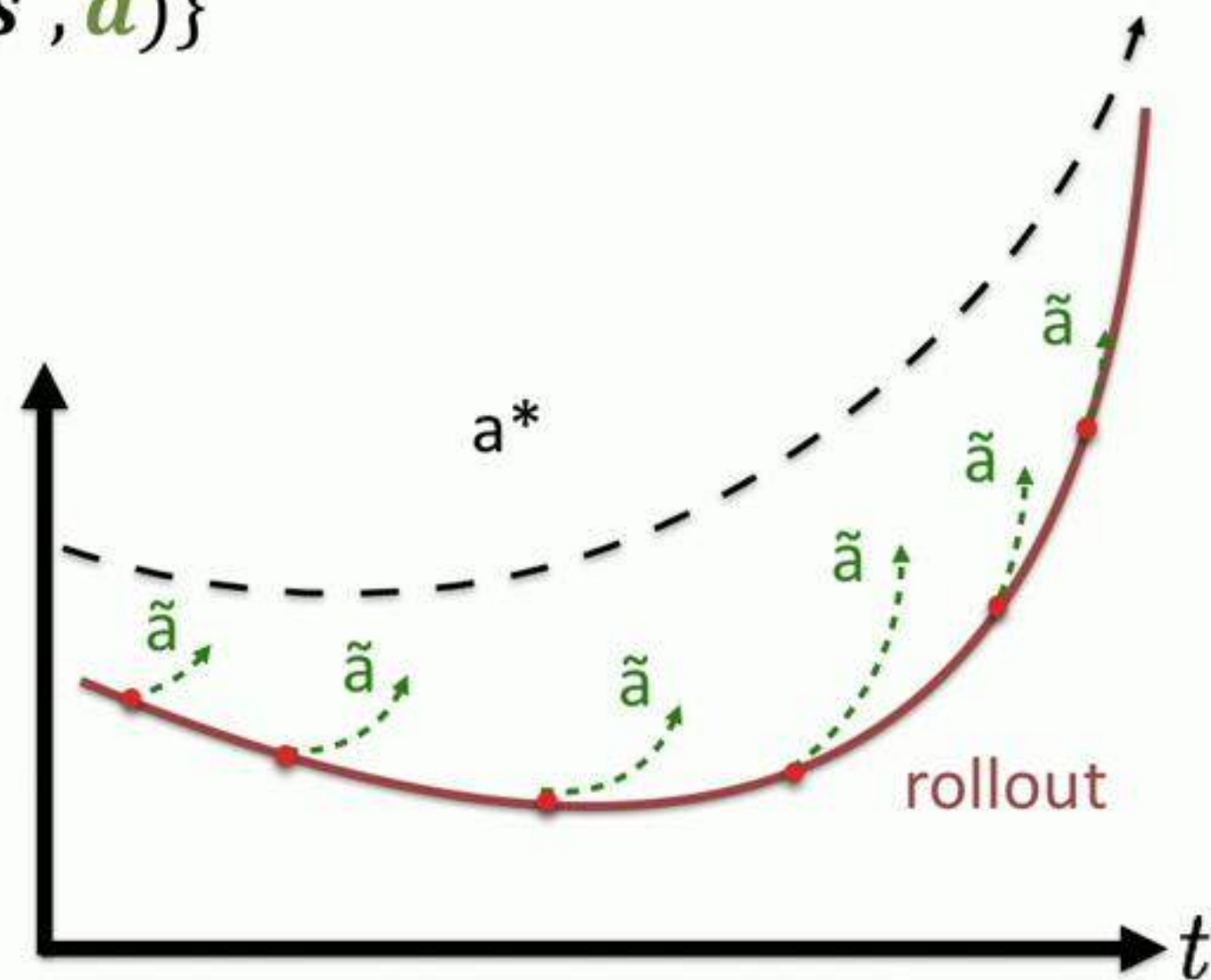
- **Practical Goals:**

- Works well in practice =)



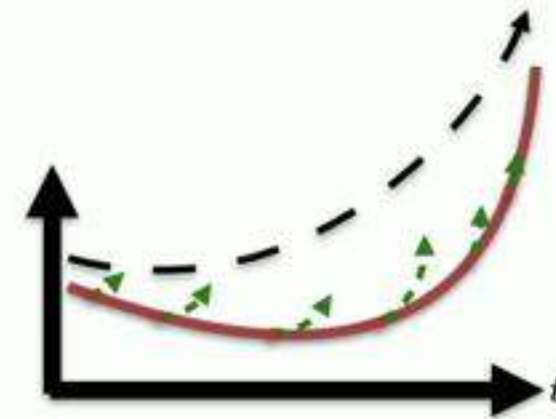
# SIMILE: Supervised Training Signal

$$D' = \{(s', \tilde{a})\}$$



# SIMILE: Theoretical Guarantees

- Always Smooth
- Guaranteed Improvement

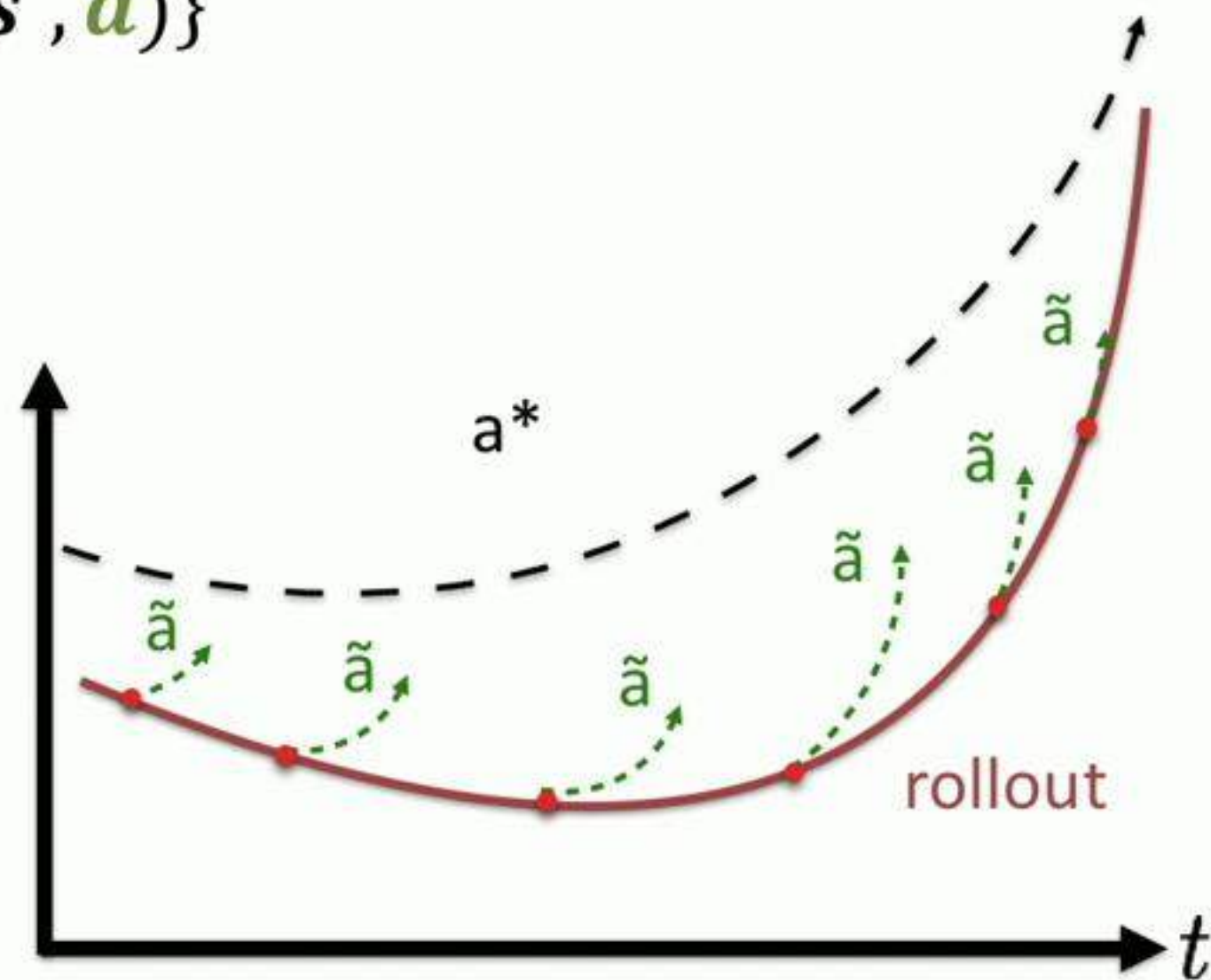


$$h(s_t \equiv (x_{t:t-K}, a_{t-1:t-K})) = \frac{f(s_t) + \lambda g(a_{t-1:t-K})}{1 + \lambda}$$



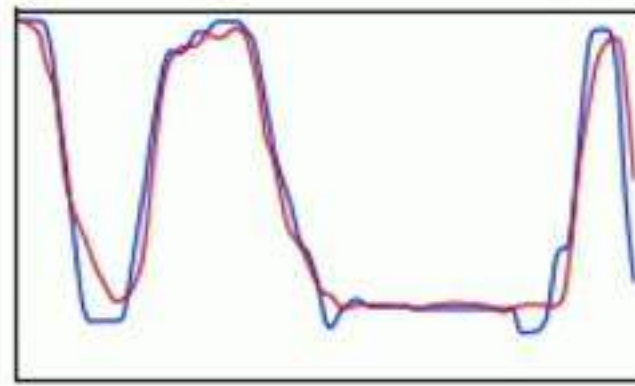
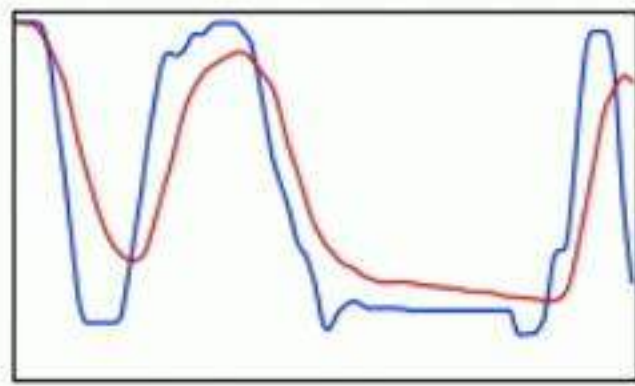
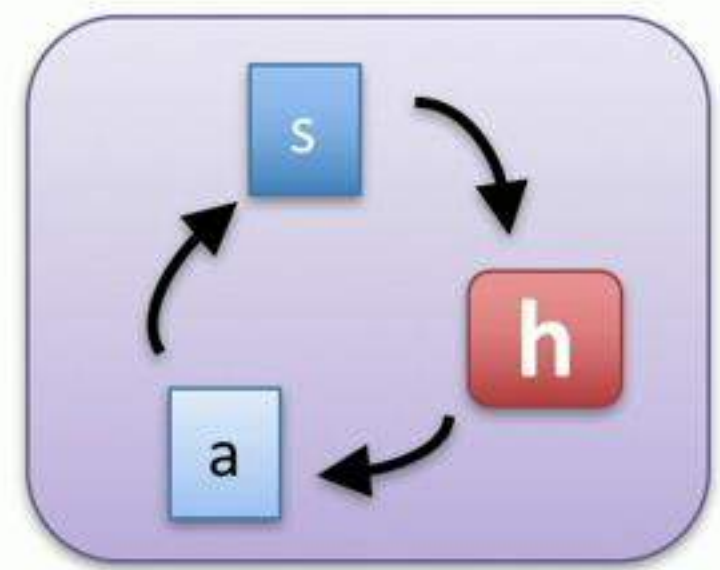
# SIMILE: Supervised Training Signal

$$D' = \{(s', \tilde{a})\}$$



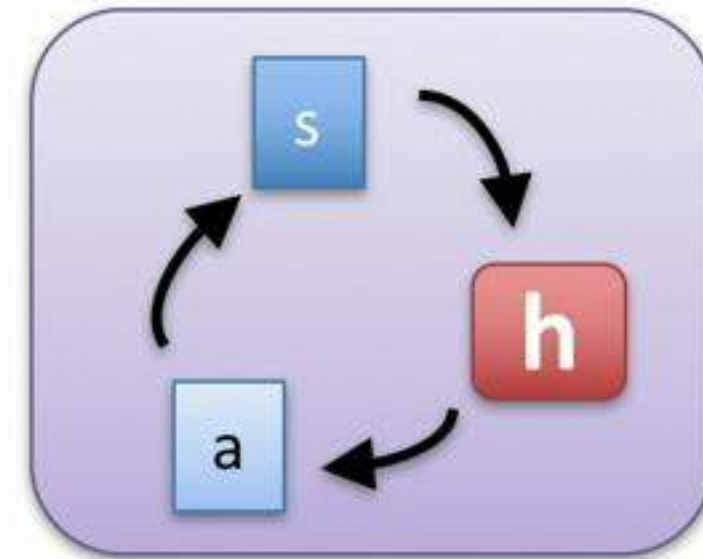
# Definition: Rollout

- Execute h sequentially
- Collect relevant statistics
  - RL: reward distribution
  - This talk: state distribution



# Definition: Rollout

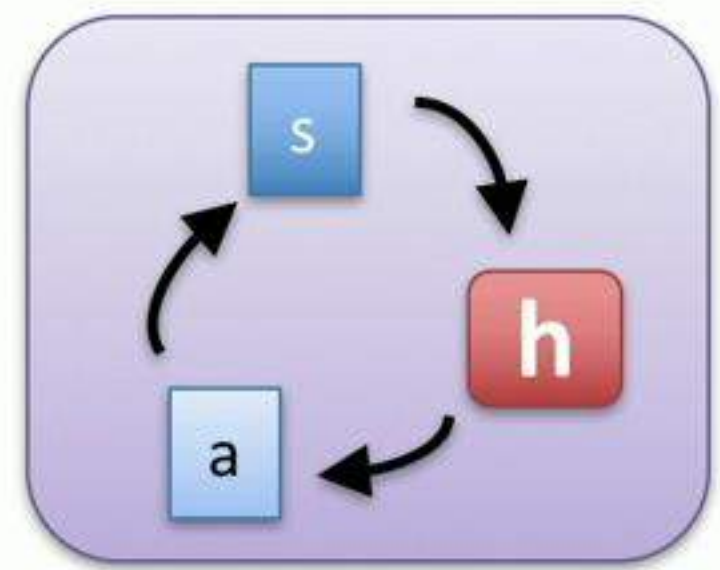
- Execute h sequentially





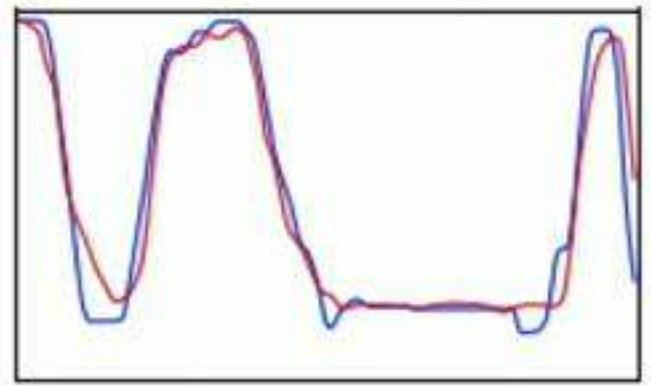
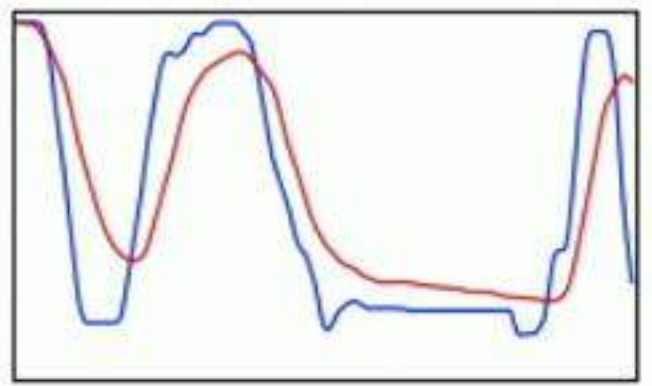
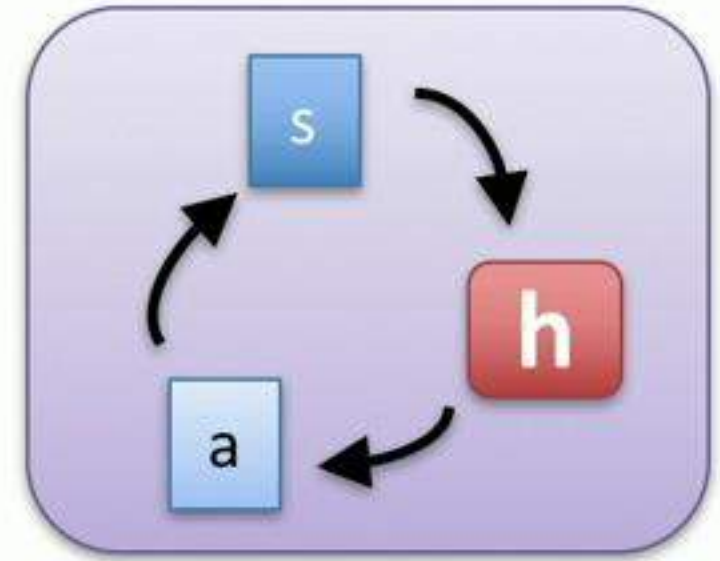
# Definition: Rollout

- Execute h sequentially
- Collect relevant statistics
  - RL: reward distribution
  - This talk: state distribution



# Definition: Rollout

- Execute h sequentially
- Collect relevant statistics
  - RL: reward distribution
  - This talk: state distribution



# Definition: Learning Reduction

- Original Learning Problem:
  - Sequential Decision Making

$$D = \{(\vec{s}, \vec{a})\}$$

- Converted Learning Problem:
  - Classification / Regression

$$D' = \{(s', a')\}$$

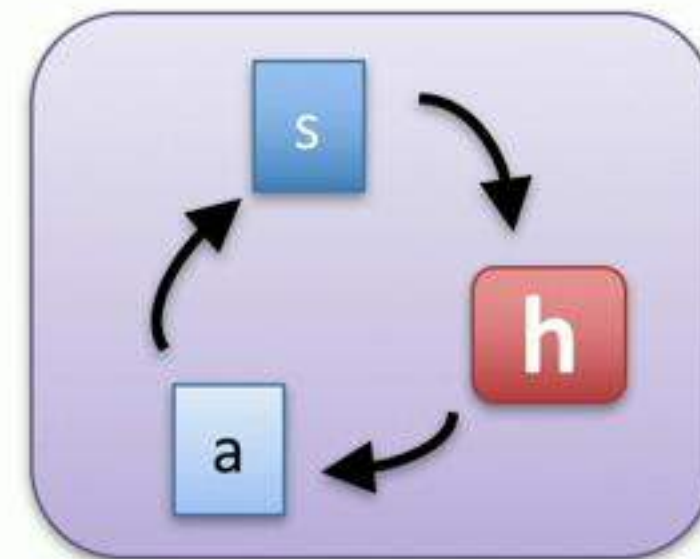
- Train  $h$  on  $D'$  (easy to do)

- **Theoretical Goals:**

- Guarantees on  $D'$  lift to  $D$

- **Practical Goals:**

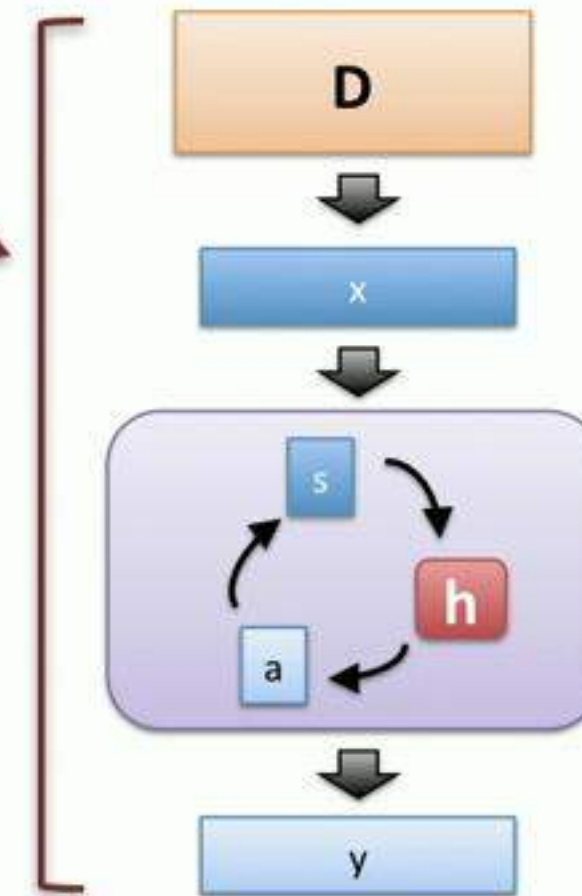
- Works well in practice =)





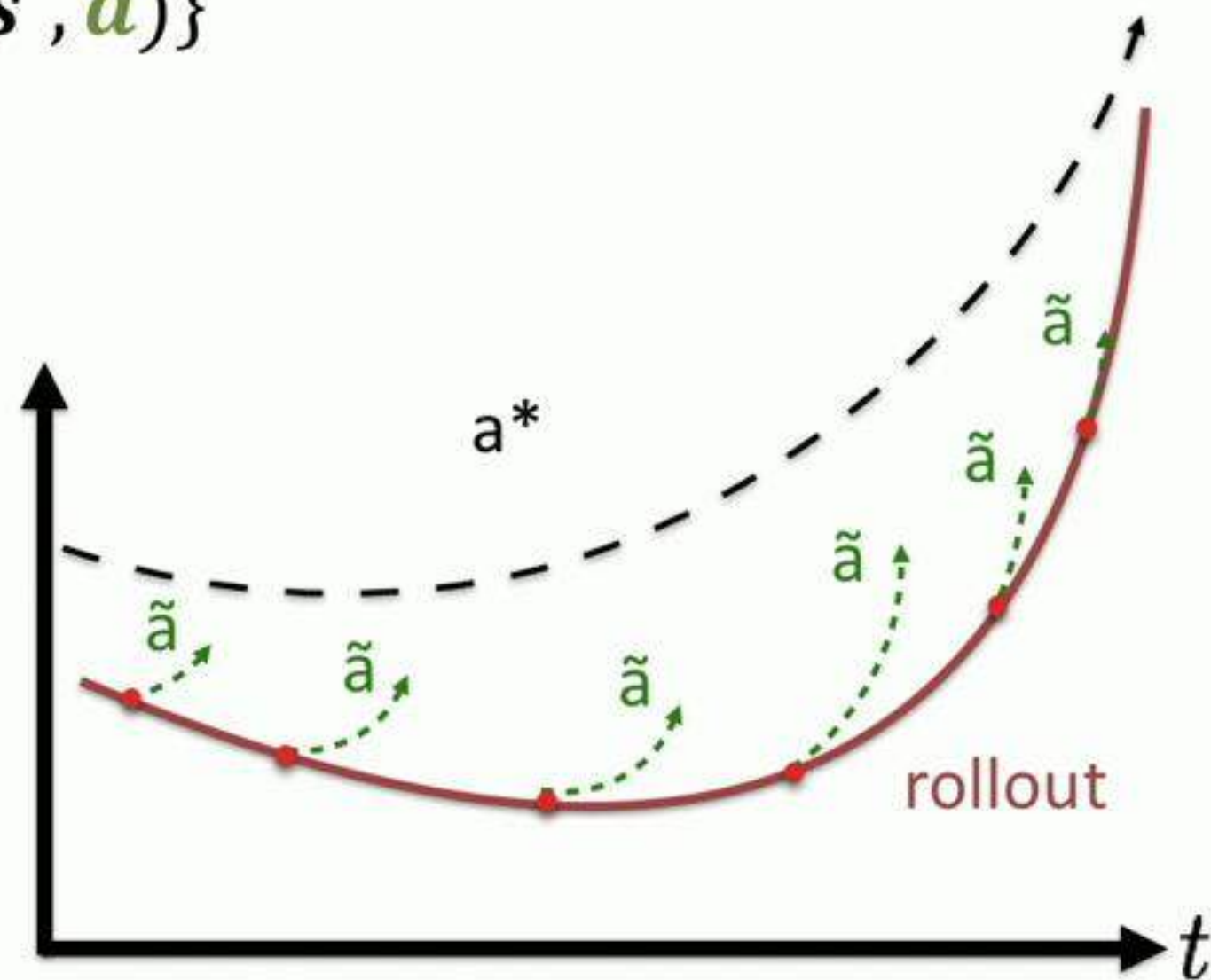
# SIMILE Learning Algorithm

- Initial Predictor:  $h_0$  ← Memorize Demonstrations
- For  $m = 1, \dots$ 
  - Rollout  $h_{m-1}$  on stream of  $x$
  - Collect training data  $D'_m$ 
    - Smooth feedback
  - Train new policy  $h_m$ 
    - $h'_m \leftarrow$  regression on  $D'_m$
    - Interpolate to obtain  $h_m$ 
      - $h_m \leftarrow \beta_m h'_m + (1 - \beta_m) h_{m-1}$



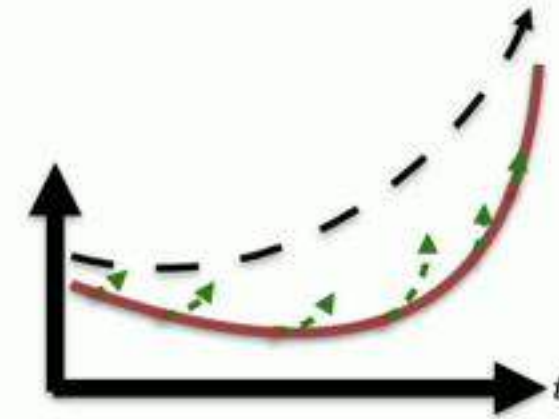
# SIMILE: Supervised Training Signal

$$D' = \{(s', \tilde{a})\}$$



# SIMILE: Theoretical Guarantees

- Always Smooth
- Guaranteed Improvement
  - Converge to optimal smooth model
- Adaptive learning rate  $\beta_m$ 
  - Converge exponentially faster than SEARN
  - Exploit (Lipschitz) smoothness property of policy class



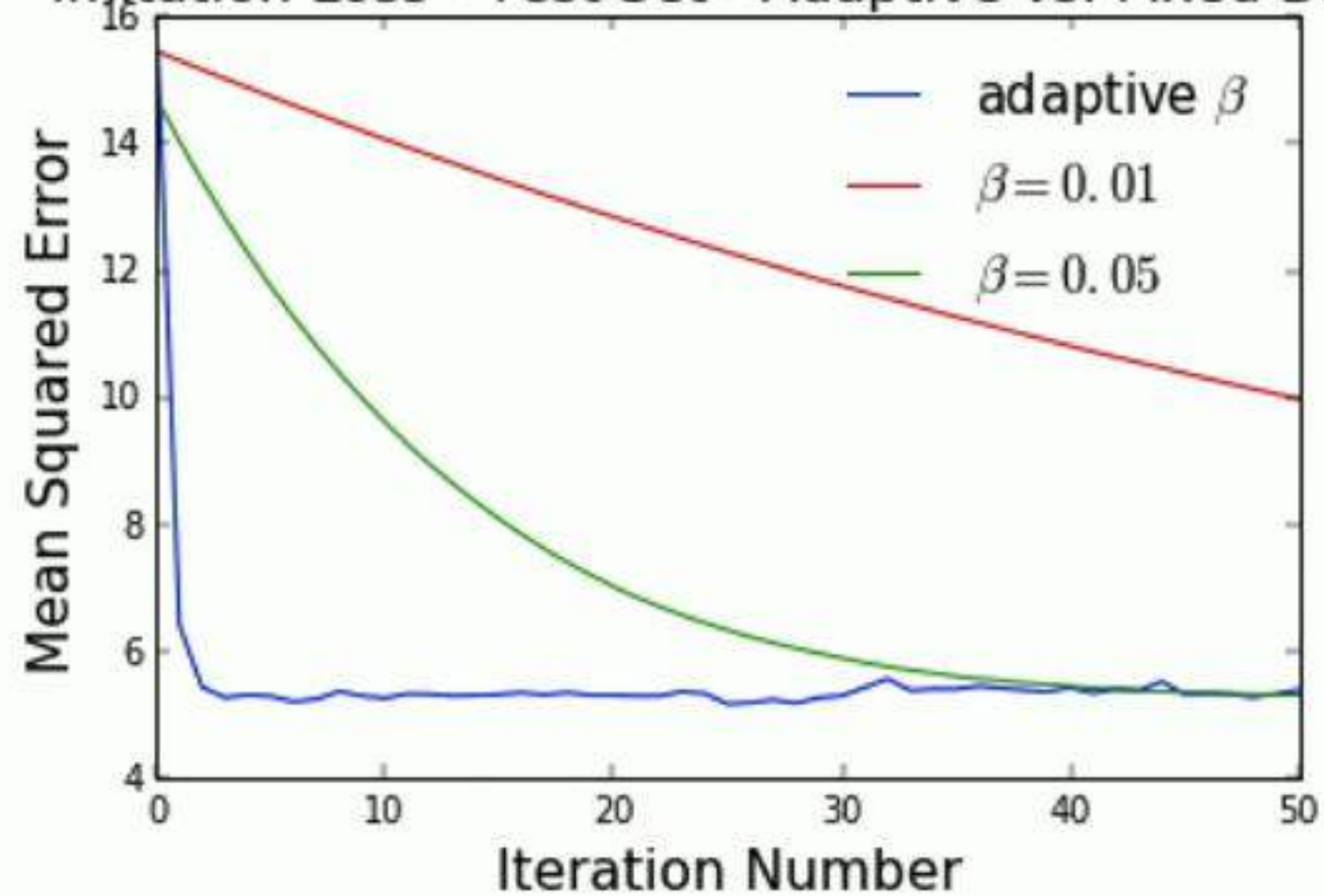
$$h(s_t \equiv (x_{t:t-K}, a_{t-1:t-K})) = \frac{f(s_t) + \lambda g(a_{t-1:t-K})}{1 + \lambda}$$



# Adaptive Learning Rate



Imitation Loss - Test Set - Adaptive vs. Fixed Beta




# Lessons Learned

- **Intuition:** Let model do most of work
  - Black box (deep neural net) adds flexibility
  - “Regularization” improves learning
    - Exponentially faster convergence compared to SEARN

# Lessons Learned

- **Intuition:** Let model do most of work
  - Black box (deep neural net) adds flexibility
  - “Regularization” improves learning
    - Exponentially faster convergence compared to SEARN
- Applicable to other approaches?



Exploit Lipschitz  
from smooth  
temporal dynamics



# Lessons Learned

- **Intuition:** Let model do most of work
  - Black box (deep neural net) adds flexibility
  - “Regularization” improves learning
    - Exponentially faster convergence compared to SEARN
- Applicable to other approaches?
  - Deep learning + robust control?

Exploit Lipschitz  
from smooth  
temporal dynamics



Aaron  
Ames



Soon-Jo  
Chung



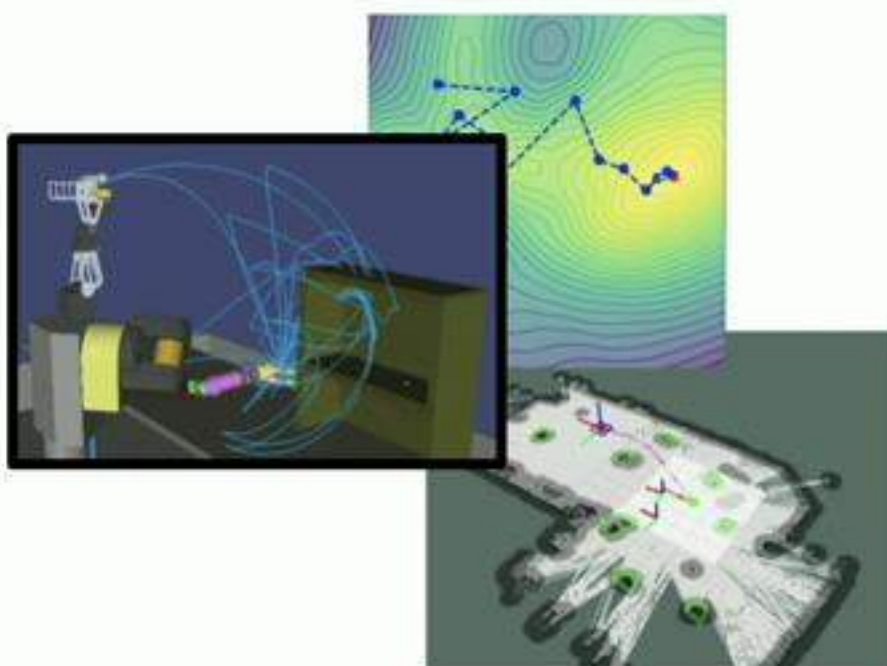
**Speech Animation**



**Coordinated Learning**



**Hierarchical Behaviors  
(Generative)**



**Learning to Optimize**



**Smooth Imitation Learning**

# New Frontiers in Imitation learning

- **Incorporating Structure**
  - Smoothness of output space
  - Latent structure of input space
  - New feedback oracles



# New Frontiers in Imitation learning

- **Incorporating Structure**
  - Smoothness of output space
  - Latent structure of input space
  - New feedback oracles
- **New Algorithmic Frameworks**
  - Black Box + Dynamics Model
  - Black Box + Latent Graphical Model
  - Retrospective Imitation

# New Frontiers in Imitation learning

- **Incorporating Structure**
  - Smoothness of output space
  - Latent structure of input space
  - New feedback oracles
- **New Algorithmic Frameworks**
  - Black Box + Dynamics Model
  - Black Box + Latent Graphical Model
  - Retrospective Imitation
- **Cool Applications!**



# New Frontiers in Imitation learning

- **Incorporating Structure**

- Smoothness of output space
- Latent structure of input space
- New feedback oracles



- **New Algorithmic Frameworks**

- Black Box + Dynamics Model
- Black Box + Latent Graphical Model
- Retrospective Imitation



- **Cool Applications!**







Eyrun  
Eyolfsson



Eric  
Zhan



Stephan  
Zheng



Hoang  
Le



Taehwan  
Kim



Sarah  
Taylor



Stephane  
Ross



Jialin  
Song



Joe  
Marino



Andrew  
Kang



Debadeepta  
Dey



Robin  
Zhou



Albert  
Zhao



Jimmy  
Chen



Milan  
Cvitkovic



Ravi  
Lanka



Kristin  
Branson



Peter  
Carr



Patrick  
Lucey



Iain  
Matthews



Jim  
Little



Pietro  
Perona



Drew  
Bagnell



Miro  
Dudik



Hal  
Daume



Alekh  
Agarwal



Nan  
Jiang



Masahiro  
Ono



Stephan  
Mandt



**Smooth Imitation Learning for Online Sequence Prediction**, Hoang Le et al., ICML 2016

**Learning Smooth Online Predictors for Real-Time Camera Planning using Recurrent Decision Trees**, Jianhui Chen et al., CVPR 2016

**A Decision Tree Framework for Spatiotemporal Sequence Prediction**, Taehwan Kim et al., KDD 2015

**A Deep Learning Approach for Generalized Speech Animation**, Sarah Taylor et al., SIGGRAPH 2017

**Generating Long-term Trajectories using Deep Hierarchical Networks**, Stephan Zheng et al., NIPS 2016

**Generative Multi-Agent Behavioral Cloning**, Eric Zhan et al. arXiv

**Learning recurrent representations for hierarchical behavior modeling**, Eyrun Eyolfsson et al., ICLR 2017

**Data-Driven Ghosting using Deep Imitation Learning**, Hoang Le et al., SSAC 2017 (*Best Paper Runner Up*)

**Coordinated Multi-agent Imitation Learning**, Hoang Le et al., ICML 2017

**Learning Policies for Contextual Submodular Prediction**, Stephane Ross et al., ICML 2013

**Learning to Search via Retrospective Imitation**, Jialin Song, Ravi Lanka, et al., arXiv

**Iterative Amortized Inference**, Joseph Marino et al., ICML 2018

**A General Method for Amortizing Variational Filtering**, Joseph Marino et al., NIPS 2018

**Hierarchical Imitation and Reinforcement Learning**, Hoang Le et al., ICML 2018