Efficient Robot Skill Learning: Grounded Simulation Learning and Imitation Learning from Observation

Peter Stone

Learning Agents Research Group (LARG)
Department of Computer Science
The University of Texas at Austin

(Also, Cogitai Inc.)

Research Question

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?

Research Question

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?

Research Areas

- Autonomous agents
- Multiagent systems
- **Robotics**

Research Question

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?

Research Areas

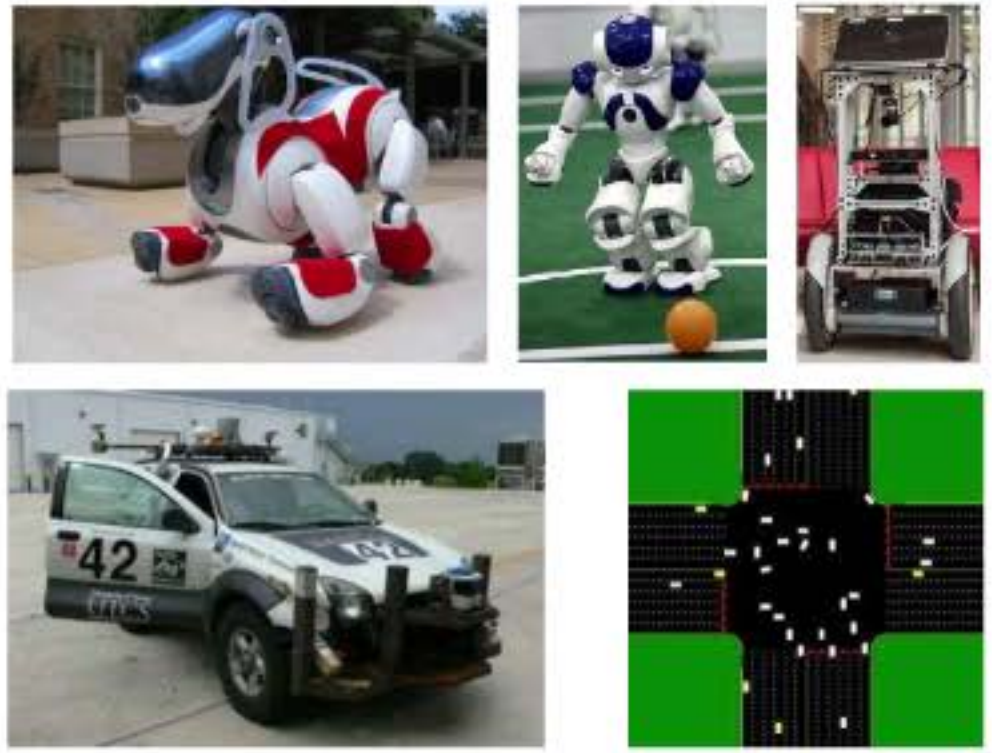
- Autonomous agents
- Multiagent systems
- **Robotics**
- Machine learning
  - **Reinforcement learning**

Research Question

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?

Research Areas

- Autonomous agents
- Multiagent systems
- **Robotics**
- Machine learning
  - **Reinforcement learning**



Learning to interpret natural-language commands through human-robot dialog

Jesse Thomason, Shiqi Zhang, Raymond Mooney, and Peter Stone

Department of Computer Science
The University of Texas at Austin, Austin, TX 78712 USA

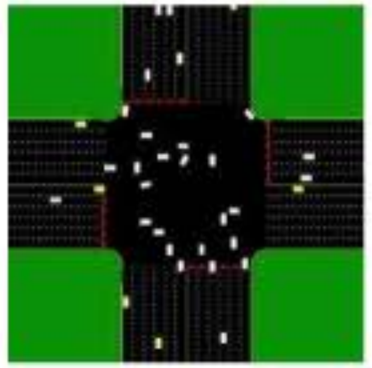
Research Question

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?

Research Areas

- Autonomous agents
- Multiagent systems
- **Robotics**
- Machine learning
  - **Reinforcement learning**



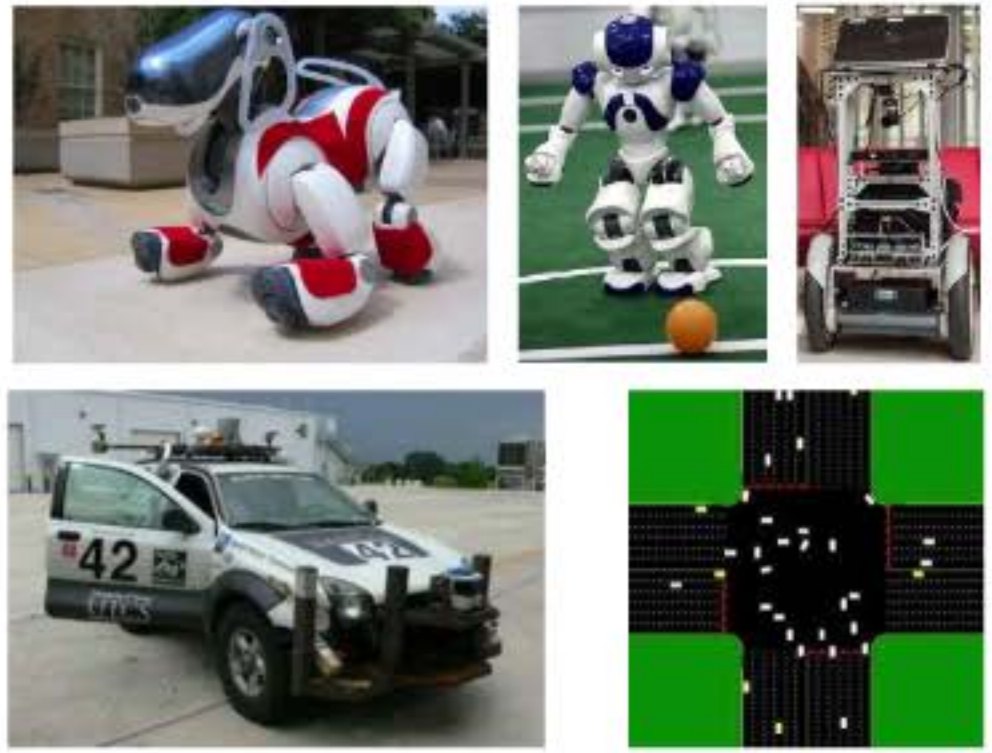



Research Question

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?

Research Areas

- Autonomous agents
- Multiagent systems
- **Robotics**
- Machine learning
  - **Reinforcement learning**
  - **Cogitai**



Cogitai

Self-learning Actionable AI

More than 60 Years Combined AI R&D

Leadership Team

MARK RING
CEO & Cofounder
"Continual Learning"

PETER STONE
President & COO
Cofounder

PETER WURMAN
VP Engineering

DENNIS CRESPO
VP Marketing &
Business Dev

"Brain Trust" Technical Advisory Board —The people who created Reinforcement Learning

SUTTON
U of Alberta

LITTMAN
Brown
University

ISBELL
Georgia Tech

ZHANG
U of Hamberg

VAN ROY
Stanford

SATINDER SINGH
Co-founder

BARTO
U. of Mass.

PRECUP
McGill

BOWLING
U of Alberta

PARKES
Harvard

DAYAN
Gatsby, UCL

Full Time Team

15 PHDs

20 Total

Continua™ SaaS Platform improves any process, software bot, system

<u>First Markets:</u>

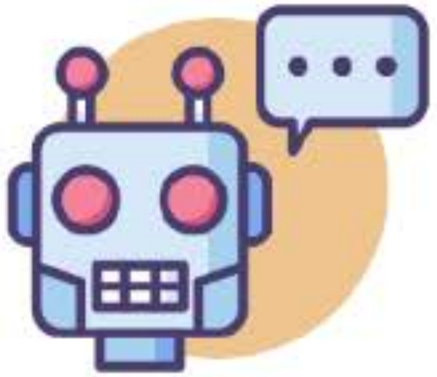
Automotive Engine Control

Robotics Control

Semiconductor Control

Use Cases are Endless

Easy to Replicate Across Industries



Decision Making
Customer service bots

Web marketing

Fitness coaches

Video game agents

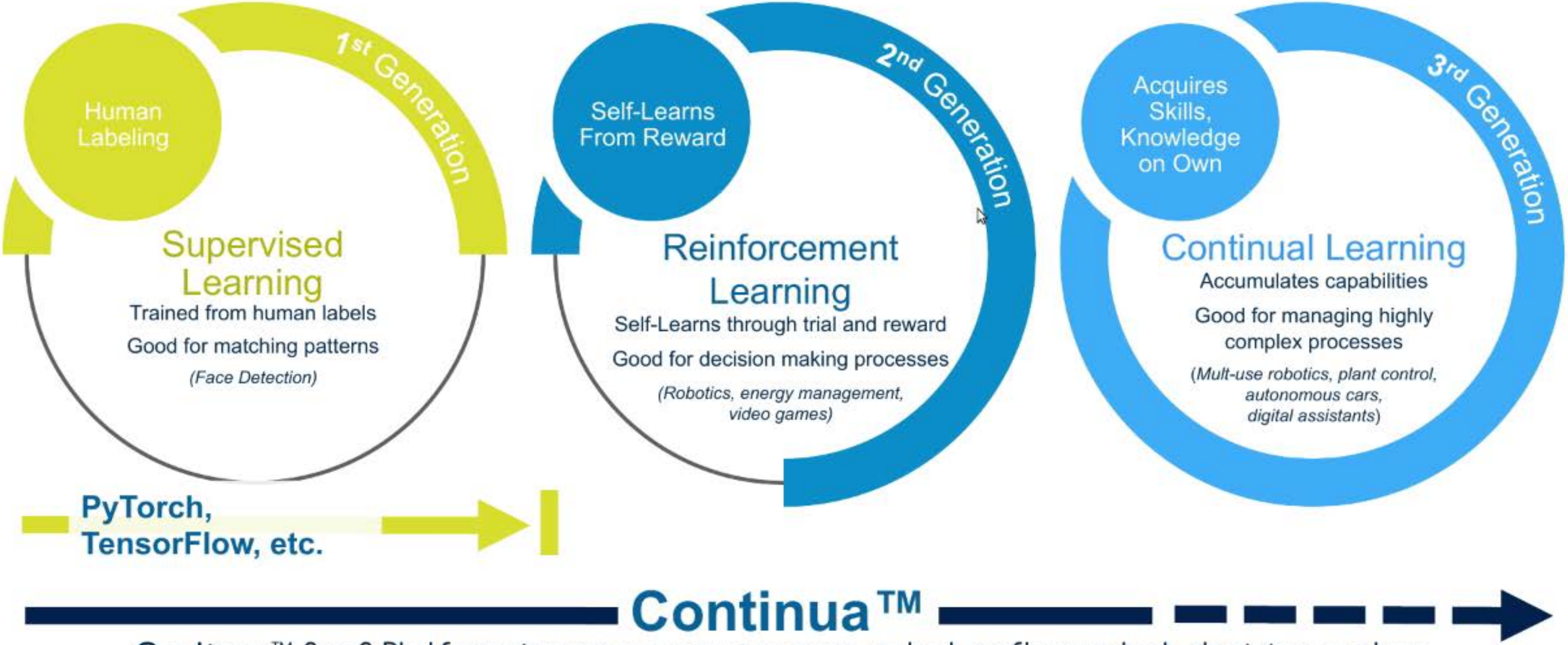
Manufacturing Processes

Robotic process automation

Building management

Self-learning vehicle

CogitAI's Aggressive Roadmap to Continual Learning



Continua™ SaaS Platform improves any process, robot, software bot, decision system

Cogitai

Cogitai

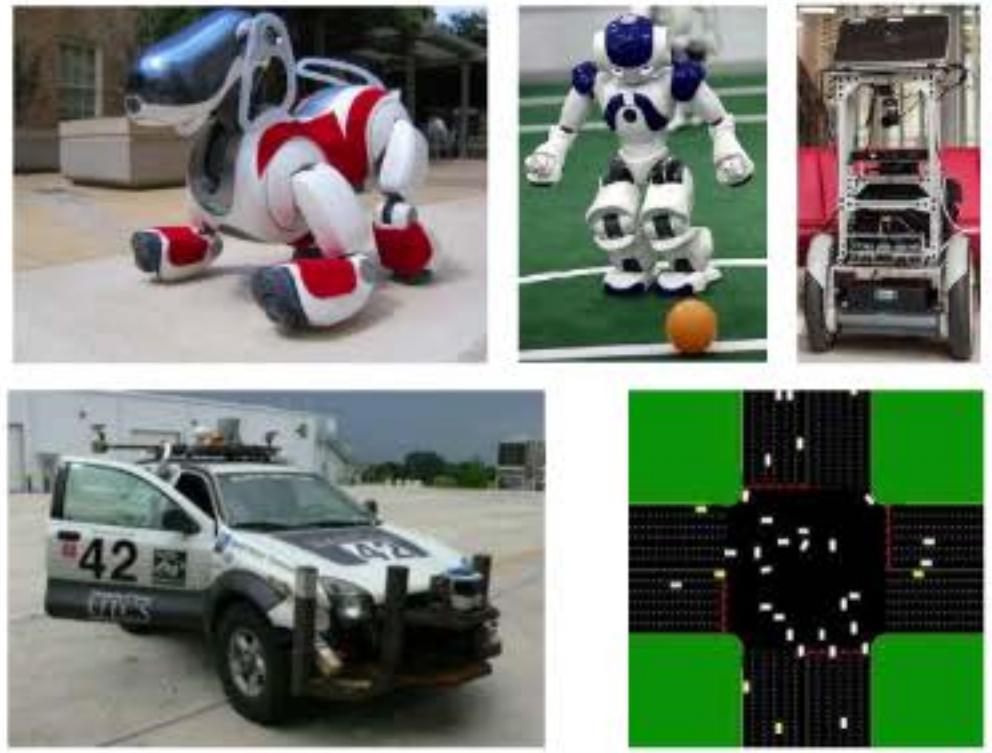
Self-learning Actionable AI

Research Question

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?

Research Areas

- Autonomous agents
- Multiagent systems
- **Robotics**
- Machine learning
  - **Reinforcement learning**
  - **Cogitai**



Efficient Robot Skill Learning

- **Motivation:**

Efficient Robot Skill Learning

- **Motivation:** RoboCup

Efficient Robot Skill Learning

- **Motivation:** RoboCup
- **Sim2Real:**

Efficient Robot Skill Learning

- **Motivation:** RoboCup
- **Sim2Real:** Grounded Simulation Learning

Efficient Robot Skill Learning

- **Motivation:** RoboCup
- **Sim2Real:** Grounded Simulation Learning
- **Imitation Learning from Observation:**

Efficient Robot Skill Learning

- **Motivation:** RoboCup
- **Sim2Real:** Grounded Simulation Learning
- **Imitation Learning from Observation:** BCO and GAIfO

RoboCup Soccer

RoboCup Soccer

- Grand challenge: beat World Cup champions by 2050

RoboCup Soccer

- Grand challenge: beat World Cup champions by 2050
- Still in relatively early stages

RoboCup Soccer

- Grand challenge: beat World Cup champions by 2050
- Still in relatively early stages
- Many virtues as a challenge problem:
  - Incremental challenges, closed loop at each stage
  - Robot design to multi-robot systems
  - Relatively easy entry
  - Inspiring to many

RoboCup Soccer

- Grand challenge: beat World Cup champions by 2050
- Still in relatively early stages
- Many virtues as a challenge problem:
  - Incremental challenges, closed loop at each stage
  - Robot design to multi-robot systems
  - Relatively easy entry
  - Inspiring to many
- Visible progress  

Small-sized League

Middle-sized League

Legged Robot League

Simulation League

Humanoid League

RoboCup 1997–1998

RoboCup Soccer

- Grand challenge: beat World Cup champions by 2050
- Still in relatively early stages
- Many virtues as a challenge problem:
  - Incremental challenges, closed loop at each stage
  - Robot design to multi-robot systems
  - Relatively easy entry
  - Inspiring to many
- Visible progress  


Small-sized League


Middle-sized League


Legged Robot League


Simulation League


Humanoid League

RoboCup Soccer

- Grand challenge: beat World Cup champions by 2050
- Still in relatively early stages
- Many virtues as a challenge problem:
  - Incremental challenges, closed loop at each stage
  - Robot design to multi-robot systems
  - Relatively easy entry
  - Inspiring to many
- Visible progress  

UT Austin Villa 3D Simulation Team RoboCup 2017 Highlights

World Champions

Record: 23-0

Goals For: 171, Goals Against: 0

AUSTIN VILLA
ROBOT SOCCER TEAM
THE UNIVERSITY OF TEXAS AT AUSTIN

RoboCup Soccer

- Grand challenge: beat World Cup champions by 2050
- Still in relatively early stages
- Many virtues as a challenge problem:
  - Incremental challenges, closed loop at each stage
  - Robot design to multi-robot systems
  - Relatively easy entry
  - Inspiring to many
- Visible progress  

RoboCup@Home

RoboCup@Home

Open-world Reasoning for Service Robots

Yuqian Jiang*, Nick Walker*, Justin Hart, Peter Stone

RoboCup@Home

Efficient Robot Skill Learning

- Motivation: RoboCup
- **Sim2Real:** Grounded Simulation Learning
- Imitation Learning from Observation: BCO and GAIfO

Reinforcement Learning for Physical Robots

Patrick
MacAlpine

Josiah
Hanna

Reinforcement Learning for Physical Robots

Patrick
MacAlpine

Josiah
Hanna

Learning on physical robots:

- Not data-efficient
- Requires supervision
- Manual resets
- Robots break
- Wear and tear make learning non-stationary

Reinforcement Learning in Simulation

Learning in simulation:

- Thousands of trials in parallel
- No supervision needed
- Automatic resets
- Robots don't break

Reinforcement Learning in Simulation

Learning in simulation:

- Thousands of trials in parallel
- No supervision needed
- Automatic resets
- Robots don’t break

But, policies **learned in simulation** often **fail in the real world**.


UTAustinVilla
0:0
<Right>
0.0
Playmode: BeforeKickOff

Reinforcement Learning in Simulation

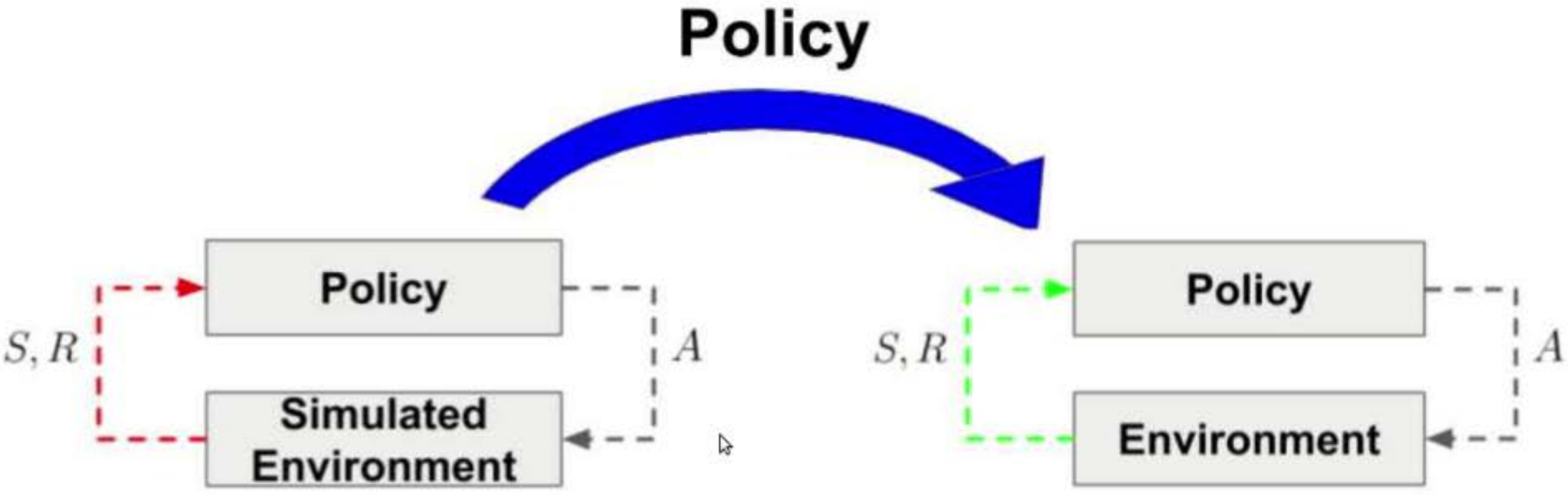
Learning in simulation:

- Thousands of trials in parallel
- No supervision needed
- Automatic resets
- Robots don't break

But, policies **learned in simulation** often **fail in the real world**.

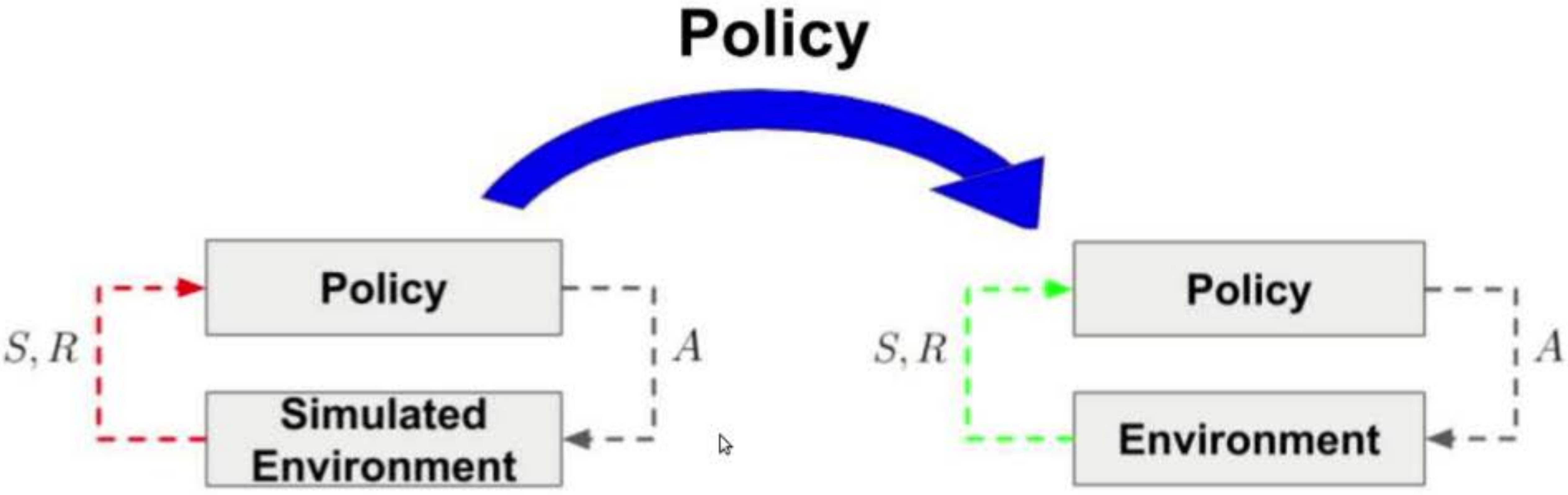
Sim2Real

Sim2Real



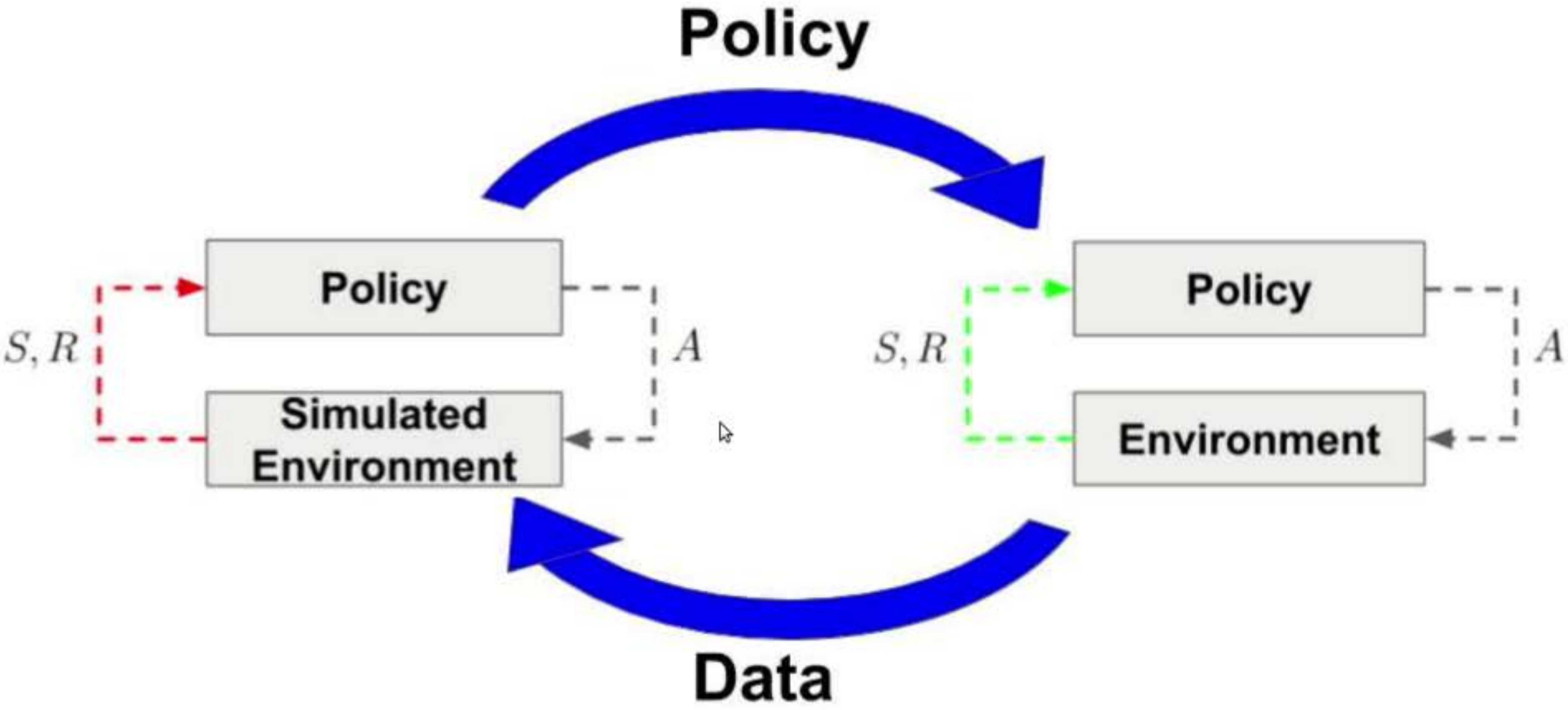
(Cutler and How, "Efficient Reinforcement Learnng for Robots using Informative Simulated Priors");
(Cully et al., "Robots that can adapt like animals");
(Rusu et al., "Sim-to-real robot learning from pixels with progressive nets")

Sim2Real



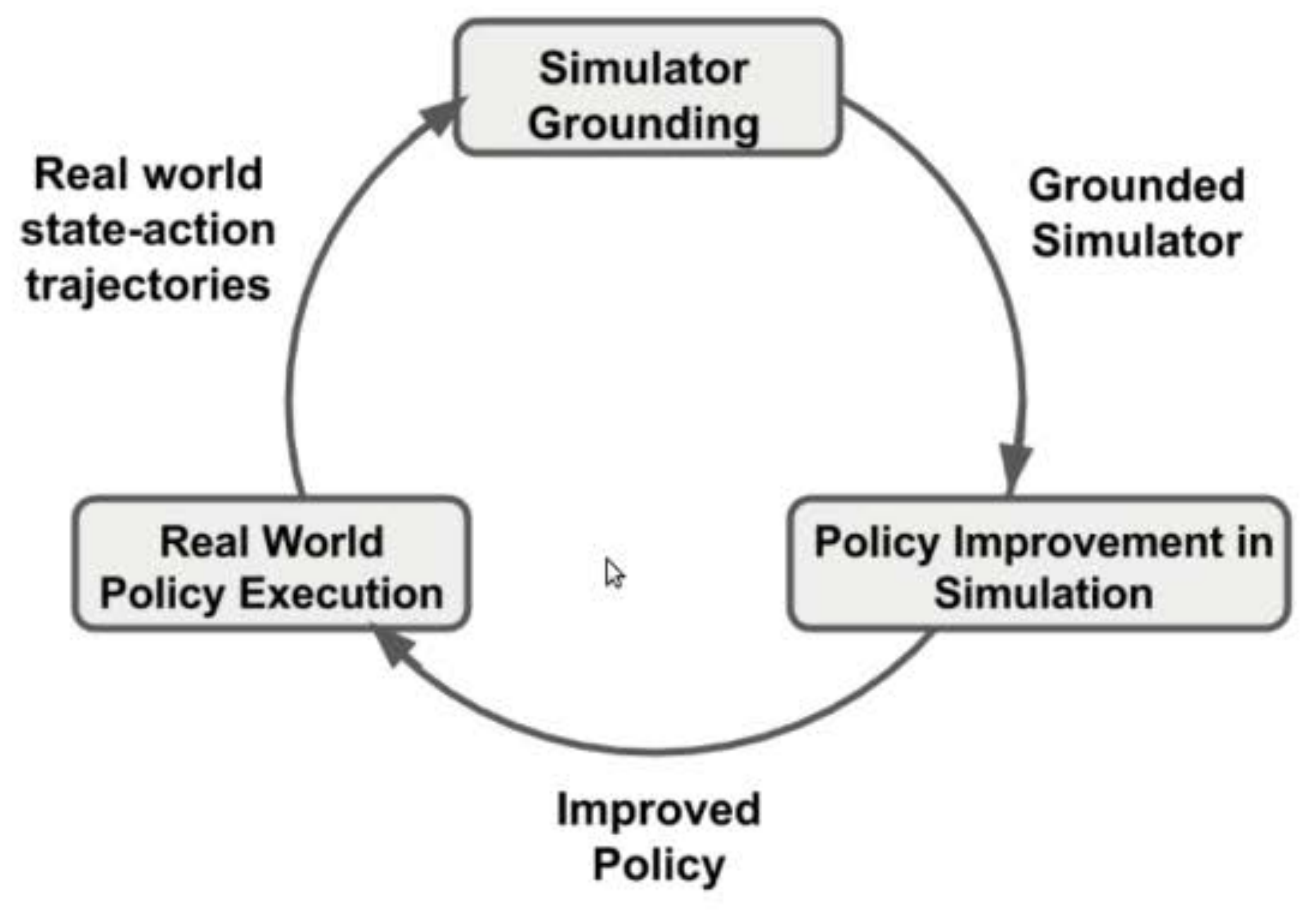
(Jakobi, Husbands, and Harvey, “Noise and the reality gap: The use of simulation in evolutionary robotics”); (Peng et al., “Sim-to-Real Transfer of Robotic Control with Dynamics Randomization”); (Tobin et al., “Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World”)

Sim2Real



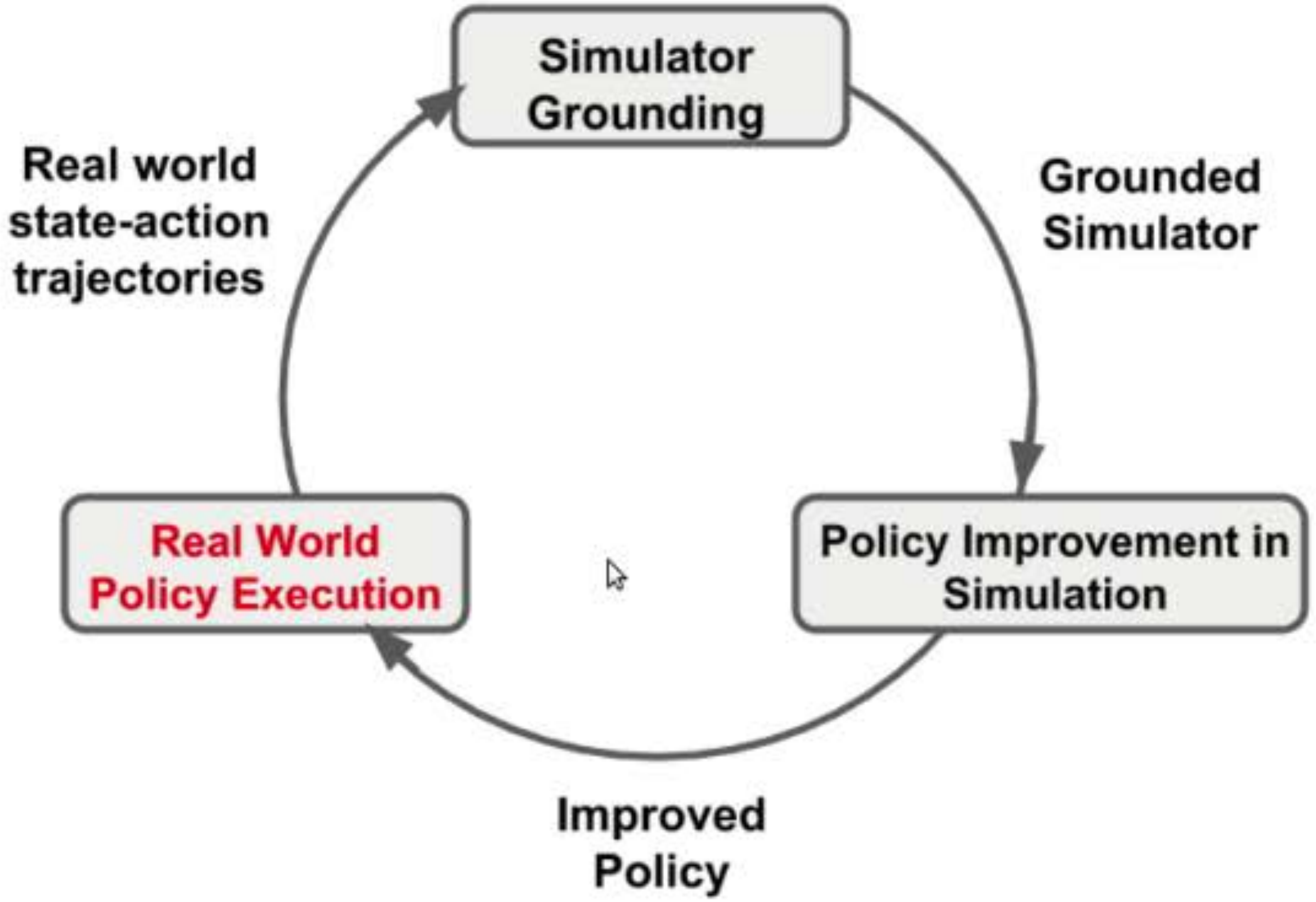
(Abbeel, Quigley, and Ng, "Using Inaccurate Models in Reinforcement Learning");
(Ross and Bagnell, "Agnostic System Identification for Model-Based Reinforcement Learning")

Grounded Simulation Learning



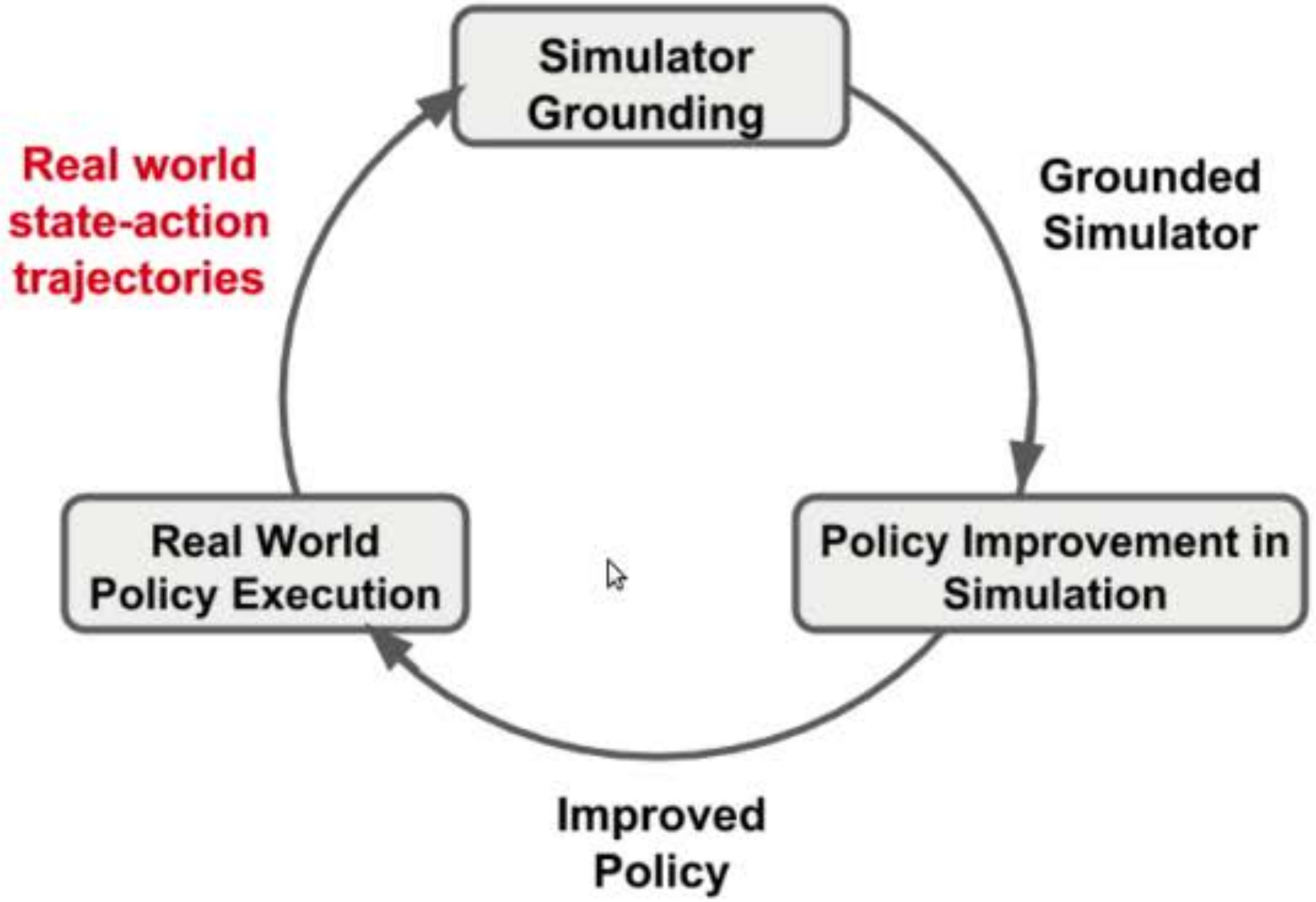
Farchy, Barrett, MacAlpine, and Stone, AAMAS 2013

Grounded Simulation Learning



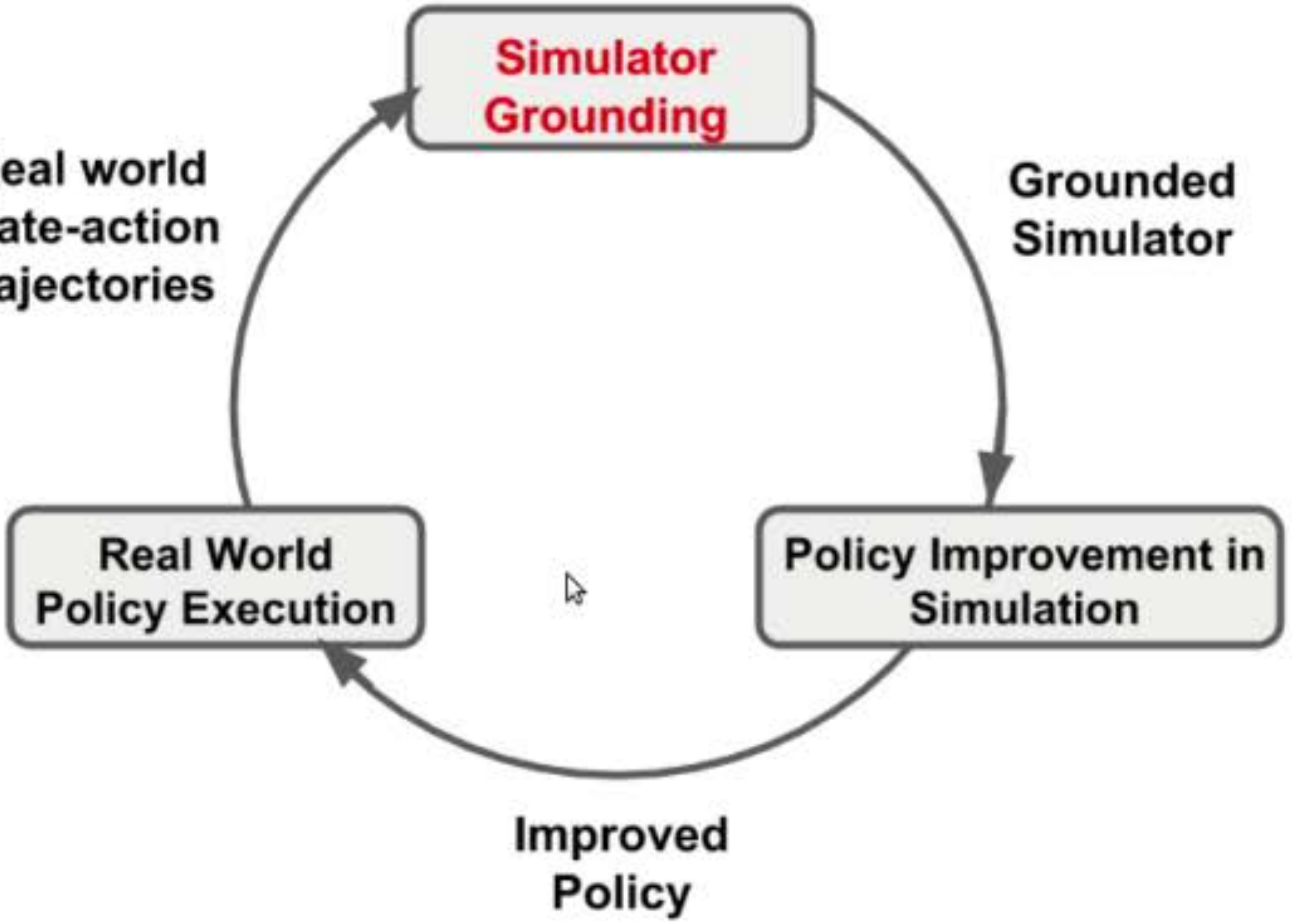
Farchy, Barrett, MacAlpine, and Stone, AAMAS 2013

Grounded Simulation Learning



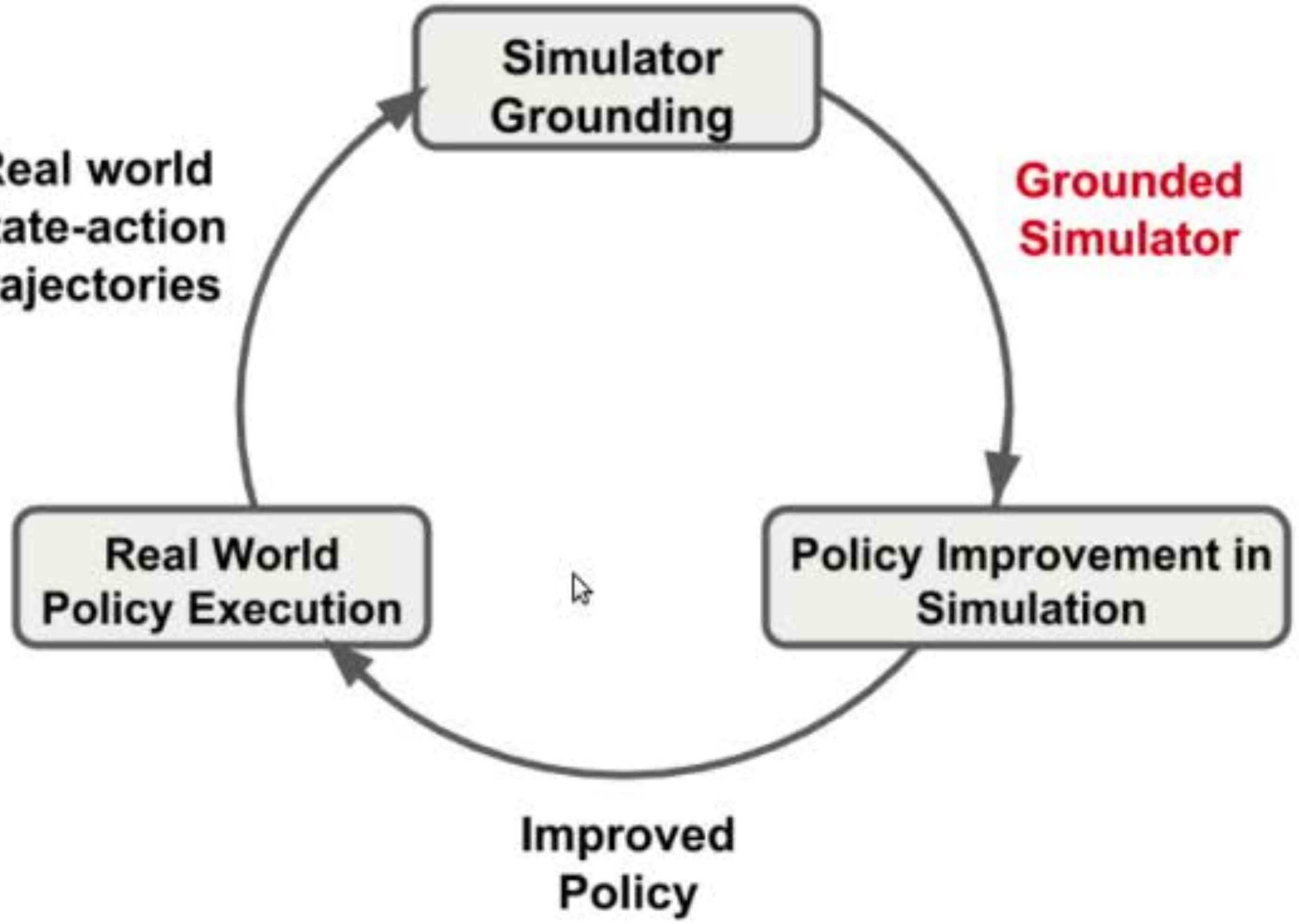
Farchy, Barrett, MacAlpine, and Stone, AAMAS 2013

Grounded Simulation Learning



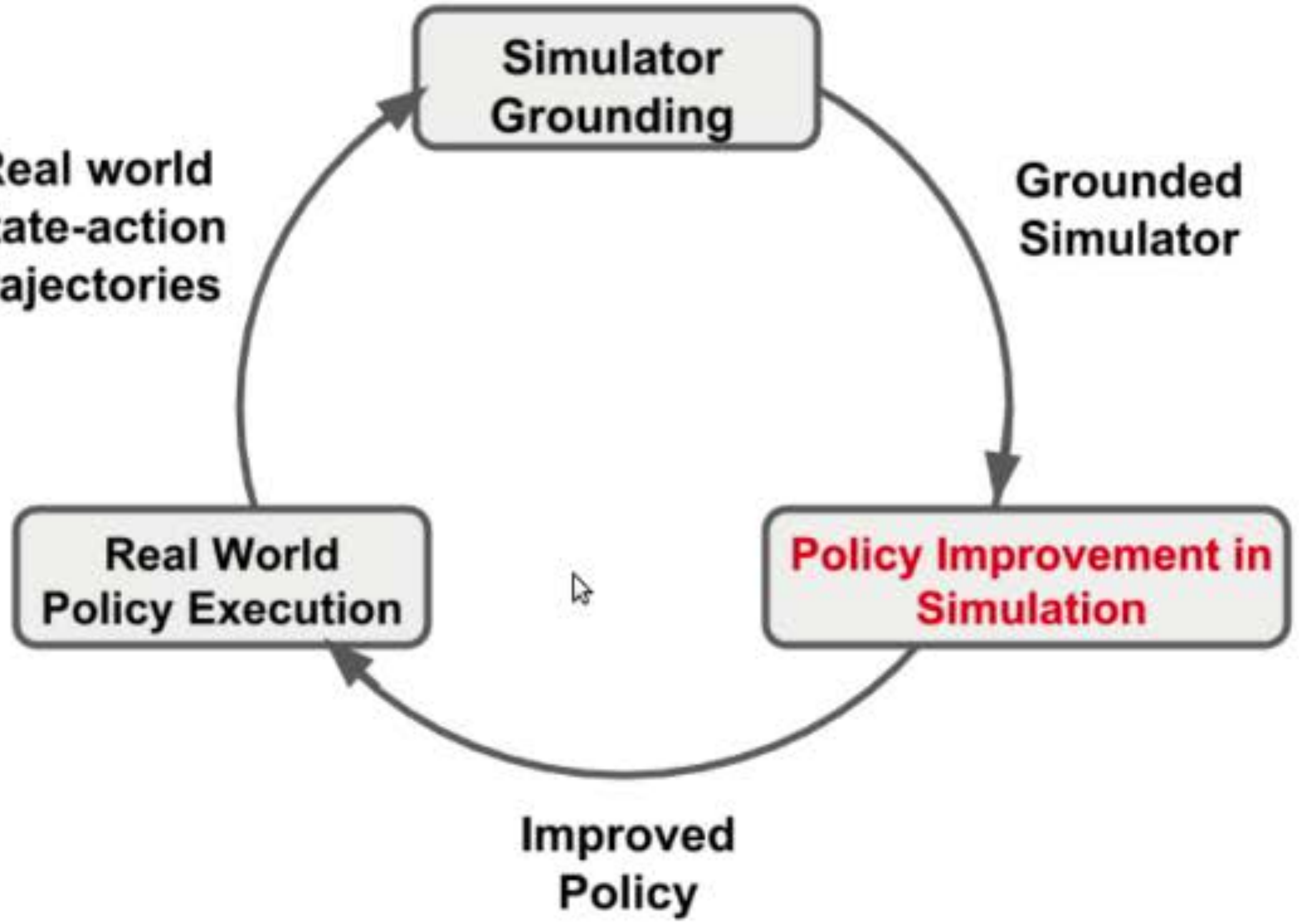
Farchy, Barrett, MacAlpine, and Stone, AAMAS 2013

Grounded Simulation Learning



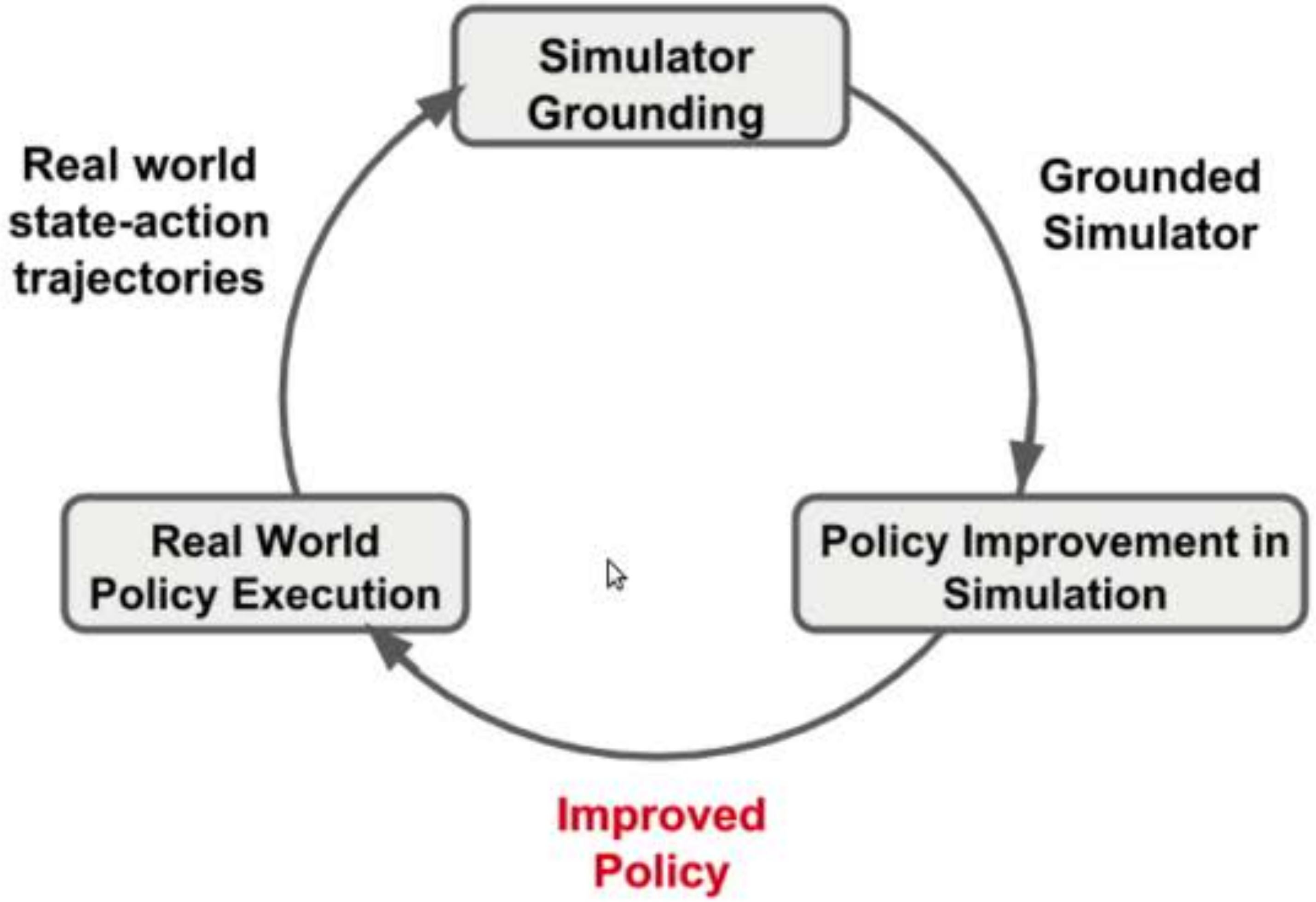
Farchy, Barrett, MacAlpine, and Stone, AAMAS 2013

Grounded Simulation Learning



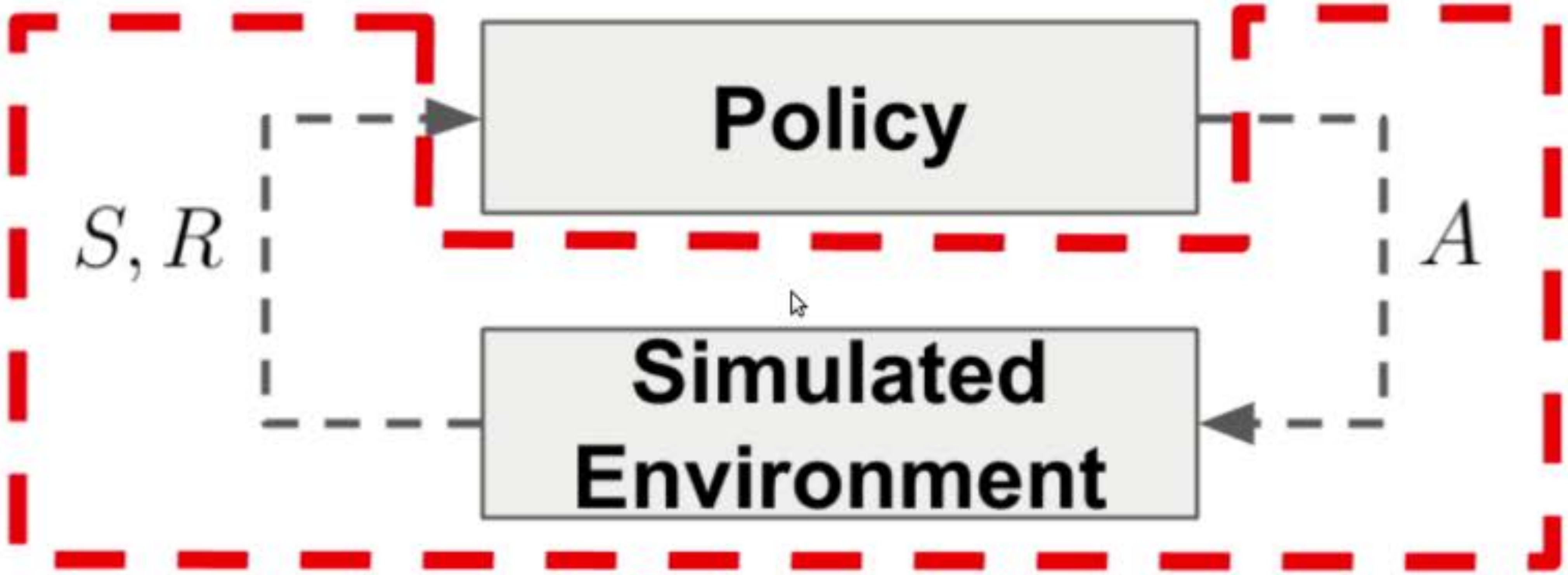
Farchy, Barrett, MacAlpine, and Stone, AAMAS 2013

Grounded Simulation Learning

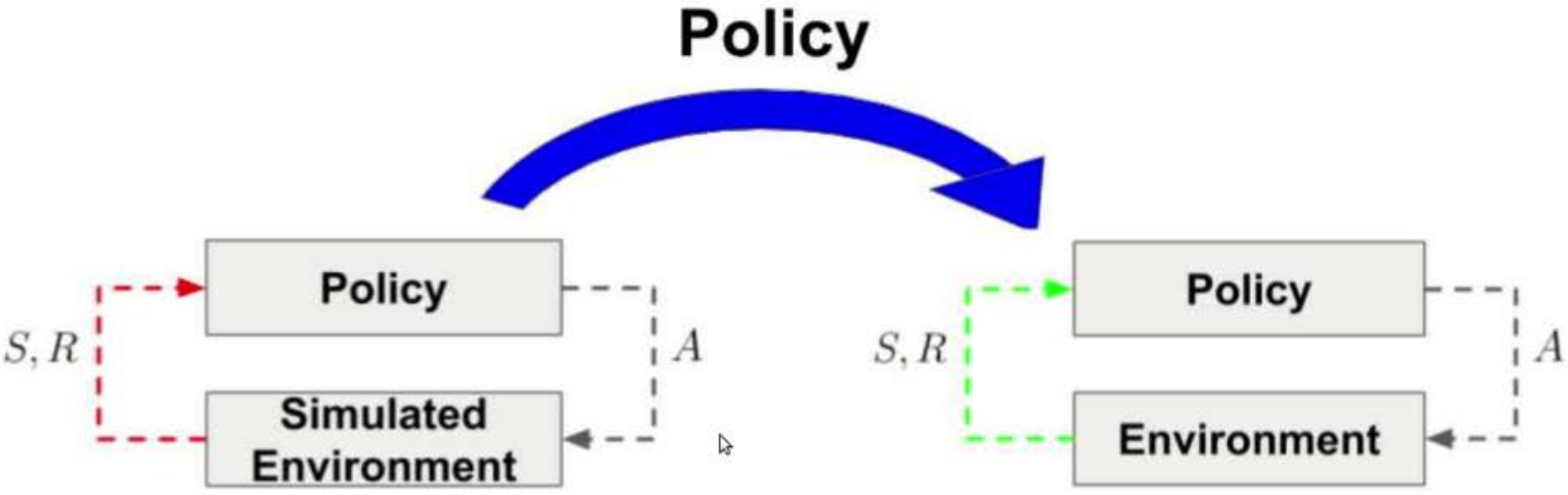


Farchy, Barrett, MacAlpine, and Stone, AAMAS 2013

Simulator Grounding

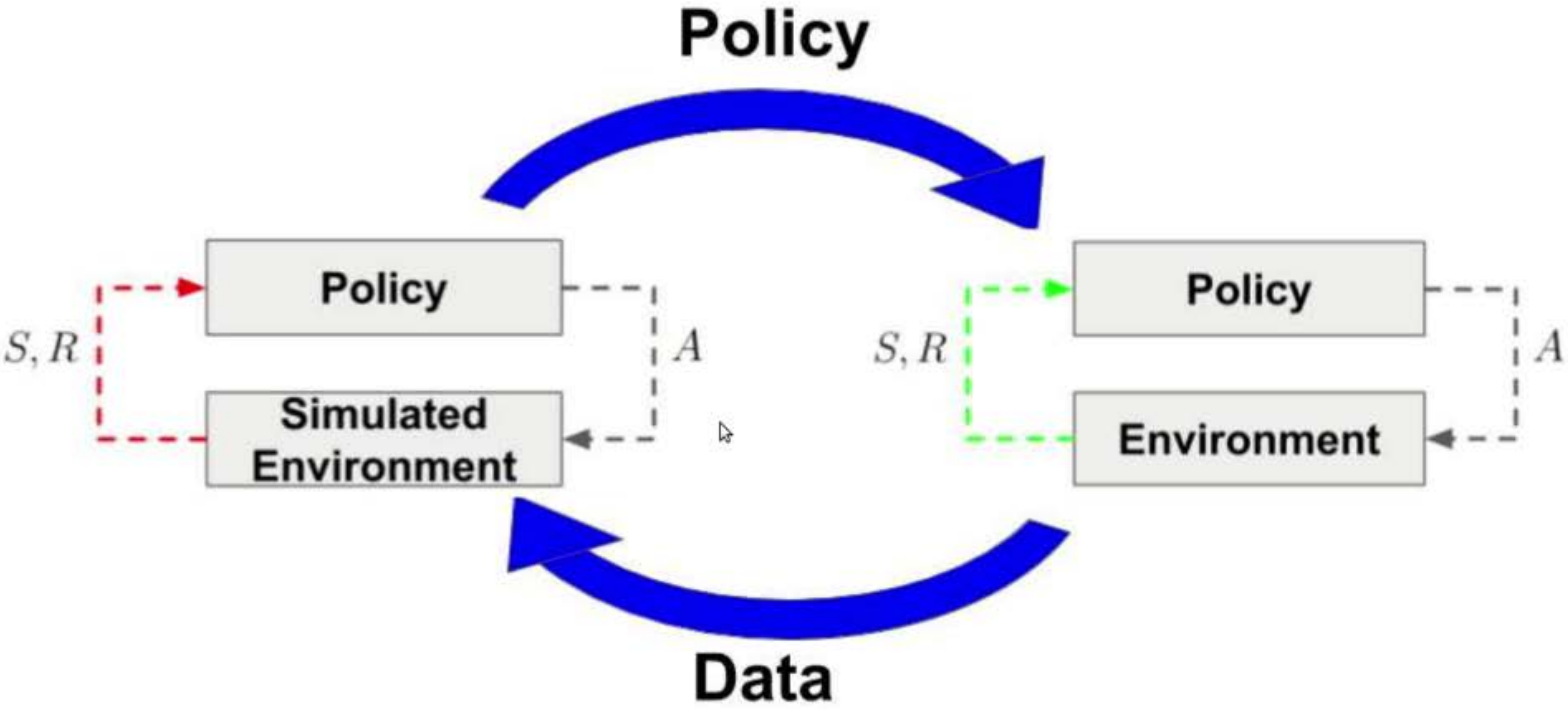


Sim2Real



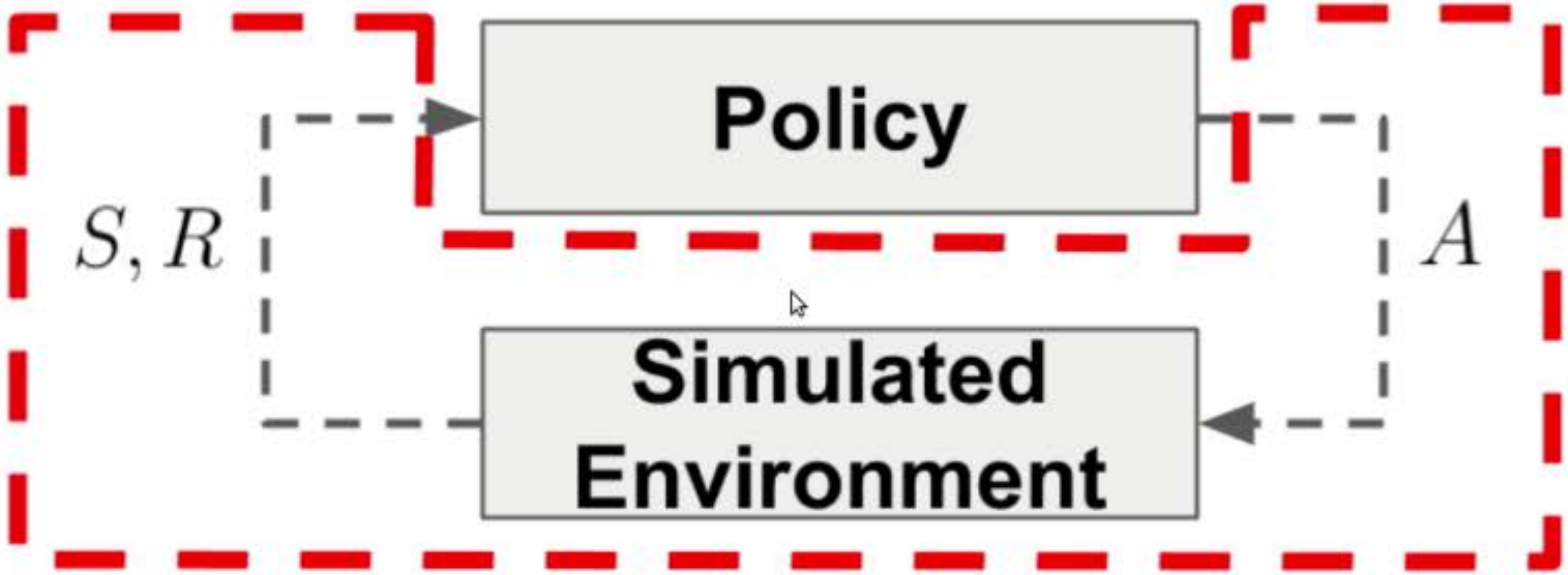
(Jakobi, Husbands, and Harvey, “Noise and the reality gap: The use of simulation in evolutionary robotics”);
(Peng et al., “Sim-to-Real Transfer of Robotic Control with Dynamics Randomization”);
(Tobin et al., “Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World”)

Sim2Real

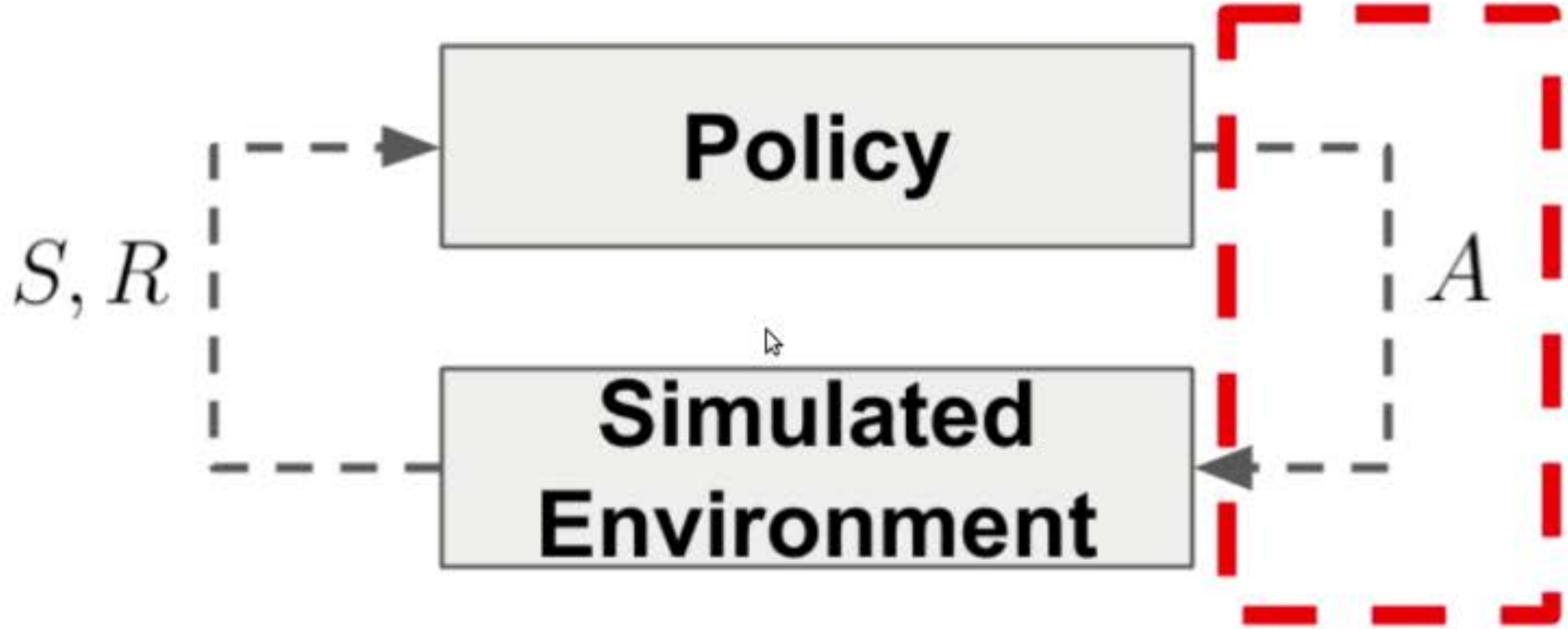


(Abbeel, Quigley, and Ng, "Using Inaccurate Models in Reinforcement Learning");
(Ross and Bagnell, "Agnostic System Identification for Model-Based Reinforcement Learning")

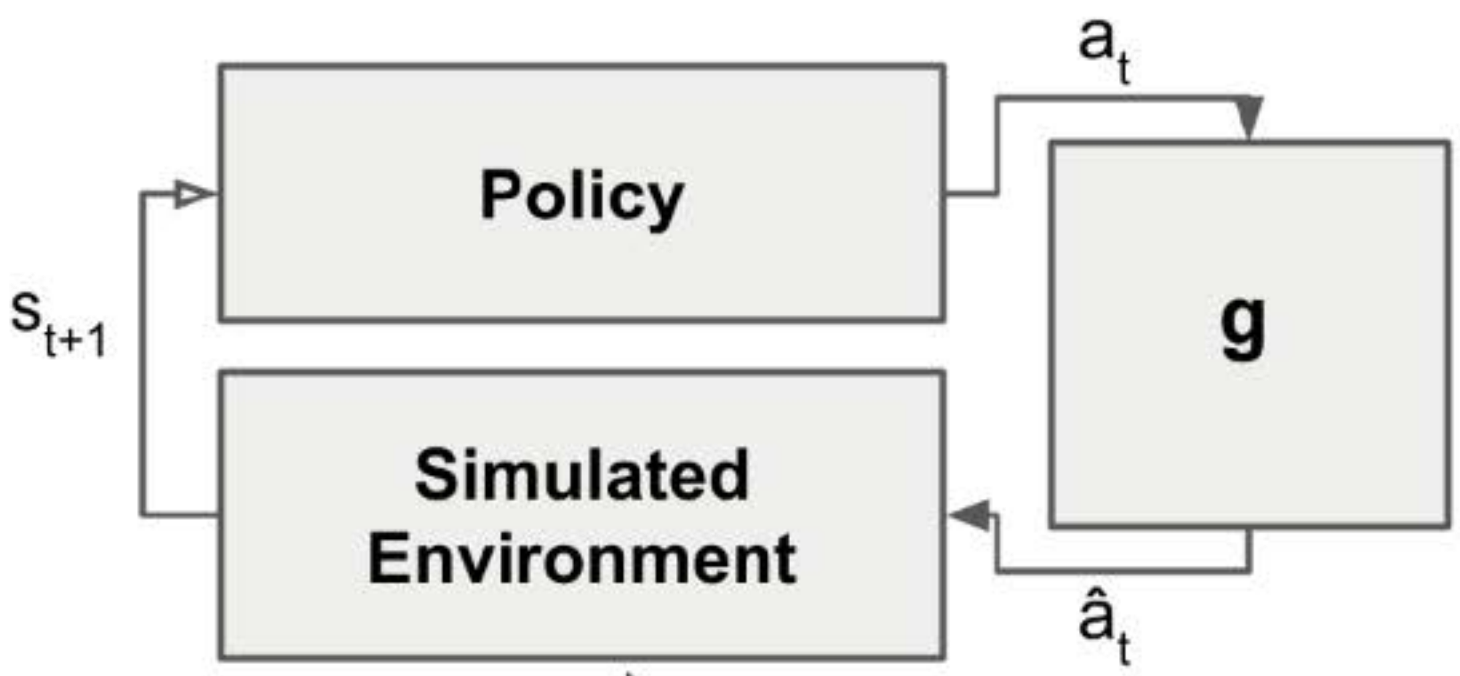
Simulator Grounding



Simulator Grounding



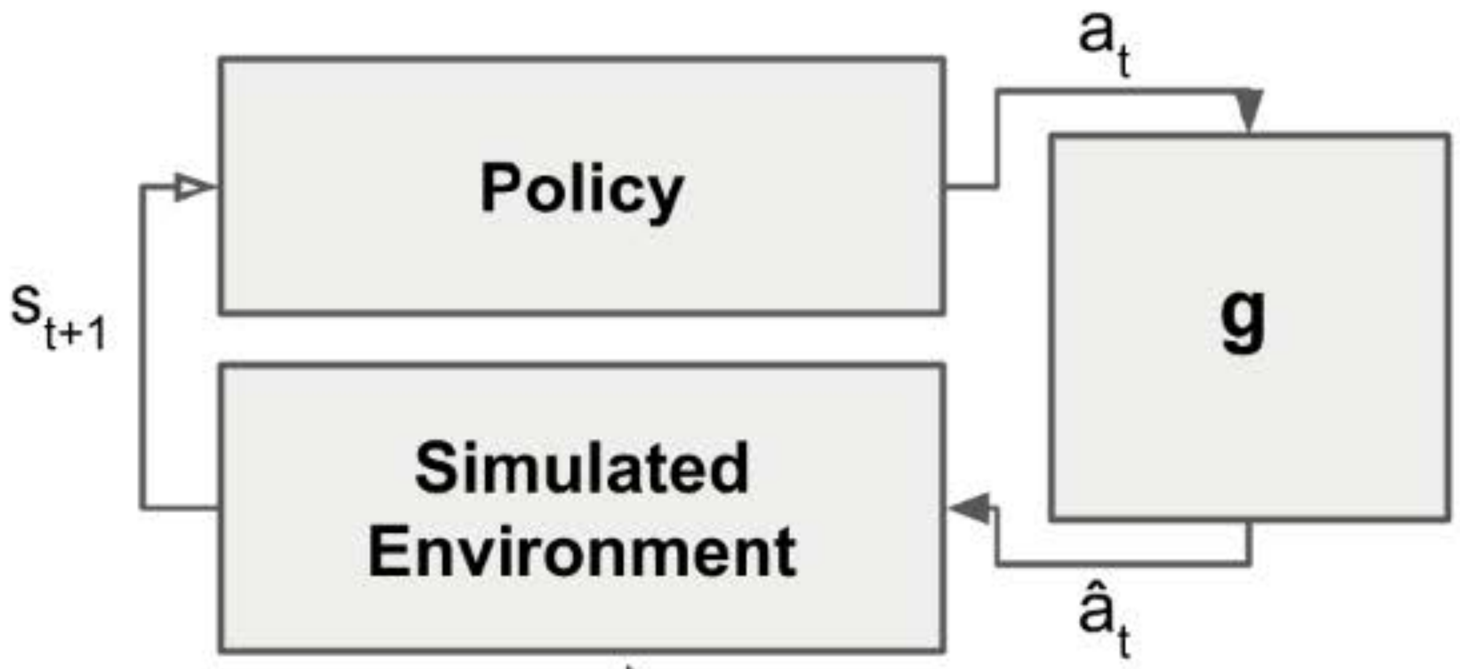
Grounded Action Transformation



Replace robot's action $\mathbf{a}_t$ with an action that produces a more "realistic" transition.

Hanna and Stone, AAAI 2017

Grounded Action Transformation

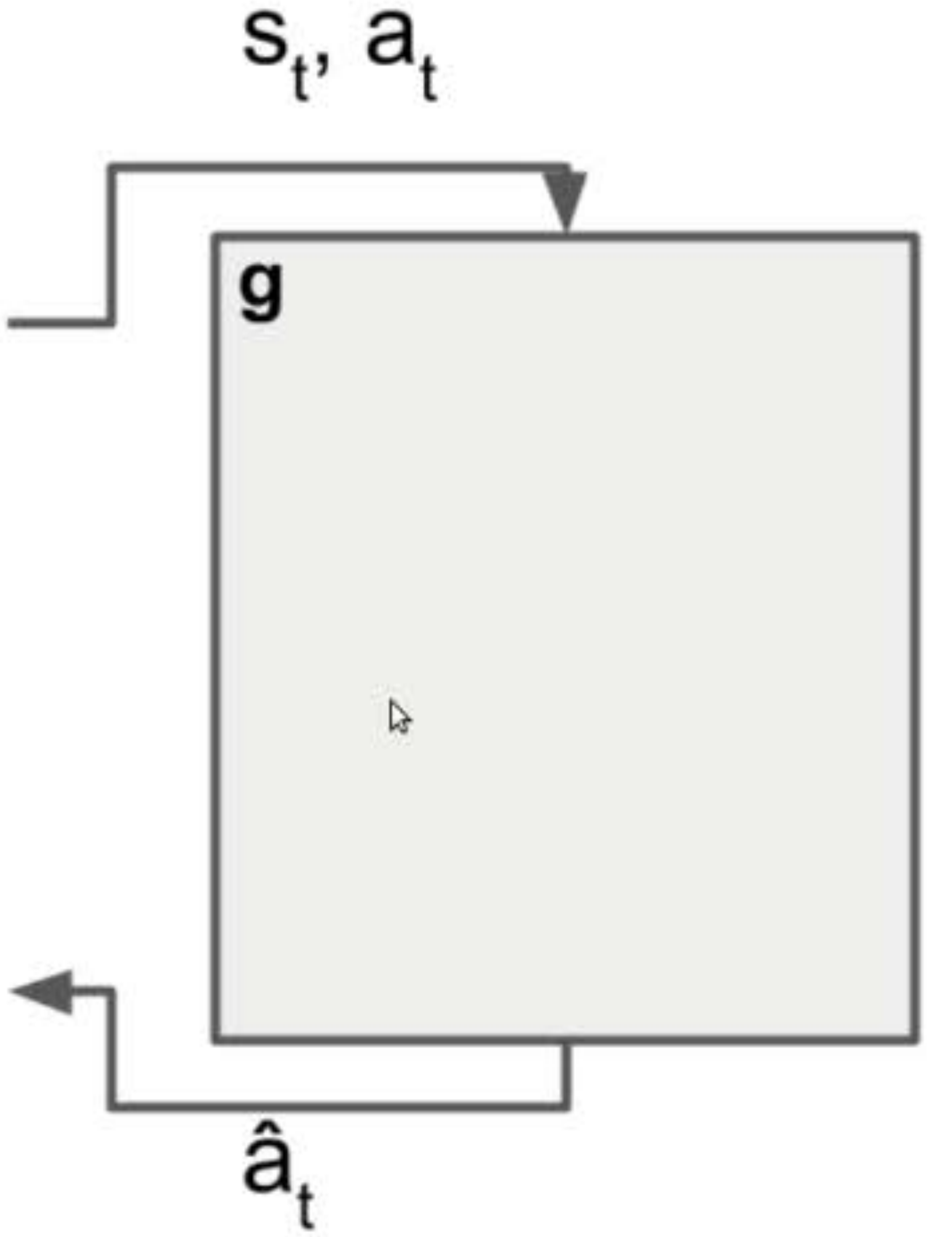


Replace robot's action $\mathbf{a}_t$ with an action that produces a more "realistic" transition.

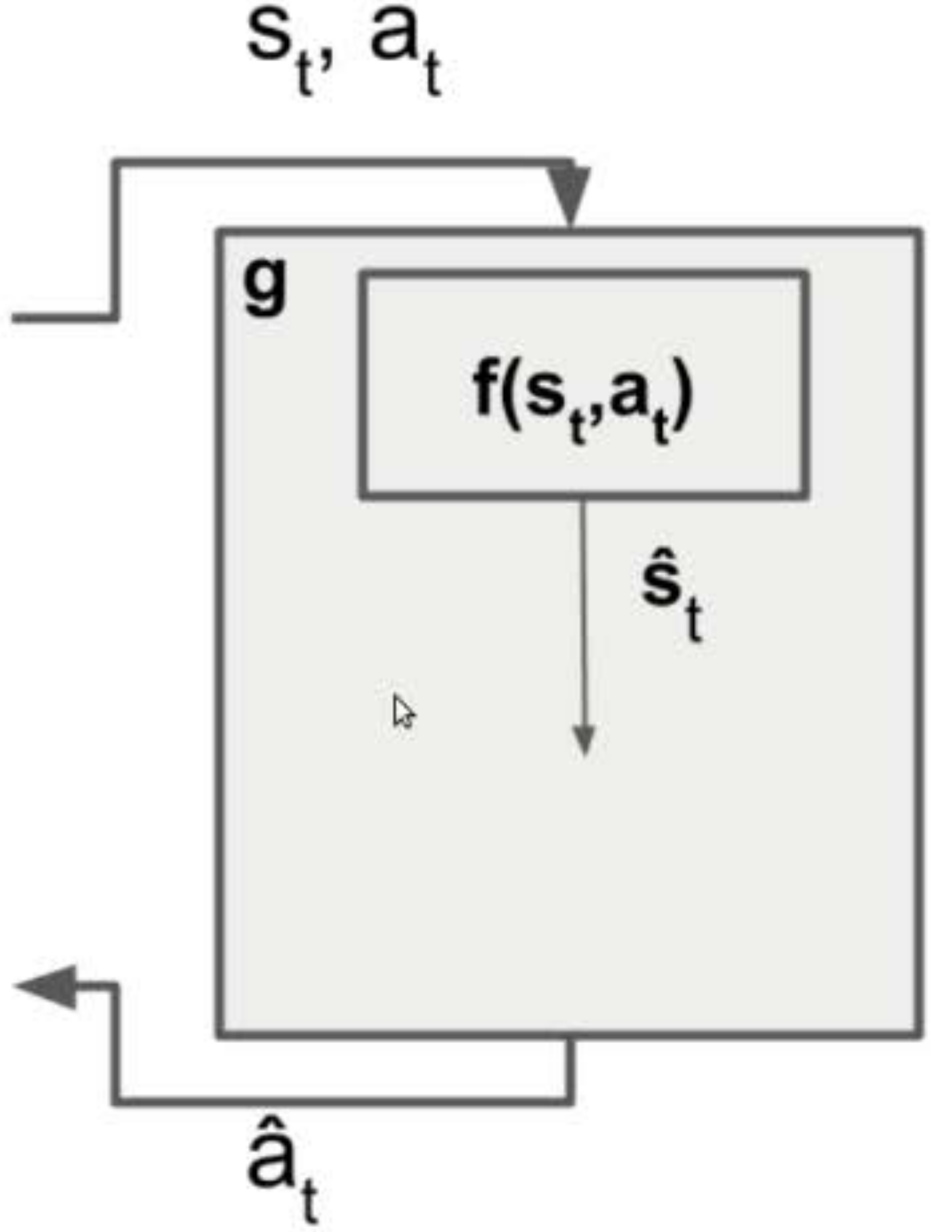
Learn this action as a function $g(\mathbf{s}_t, \mathbf{a}_t)$.

Hanna and Stone, AAAI 2017

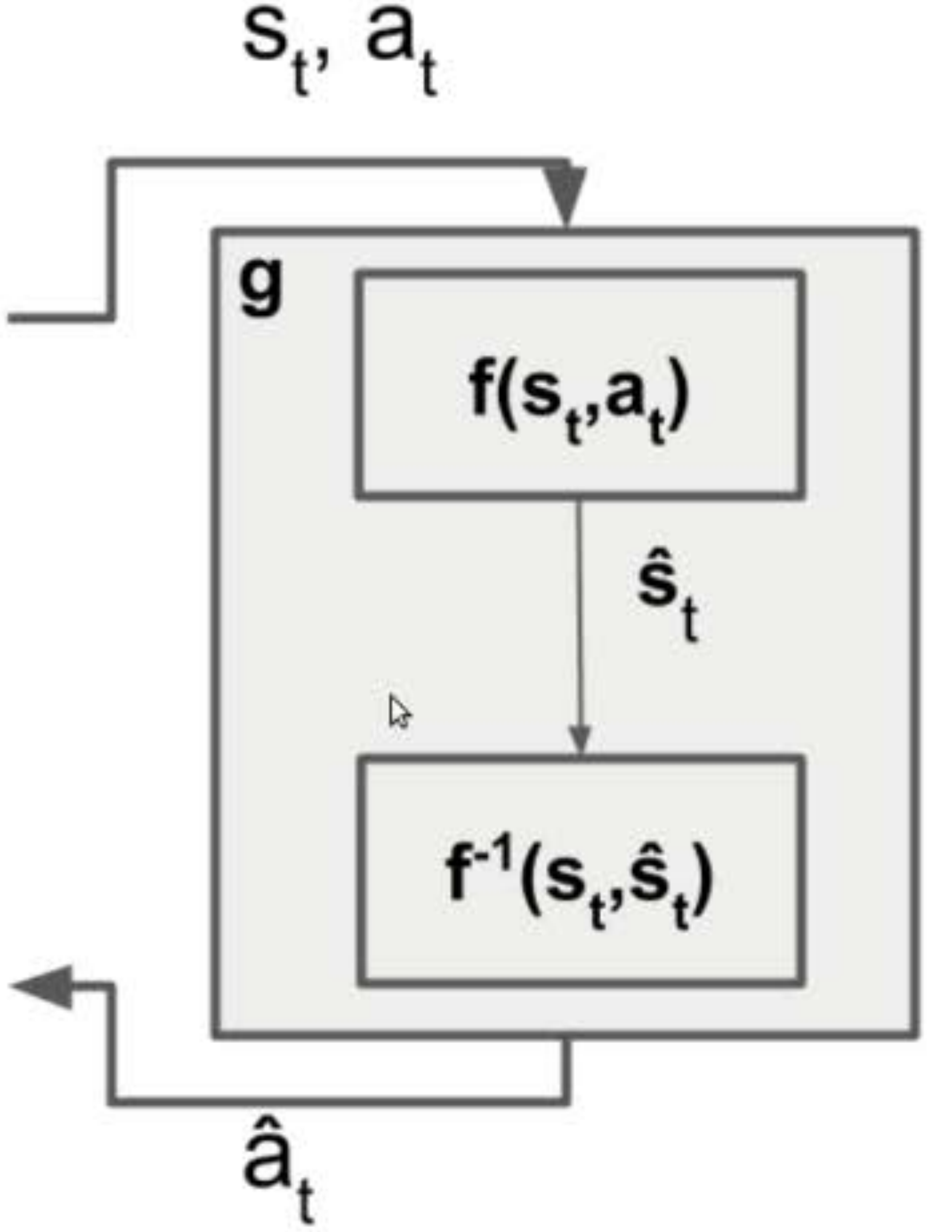
Grounded Action Transformation



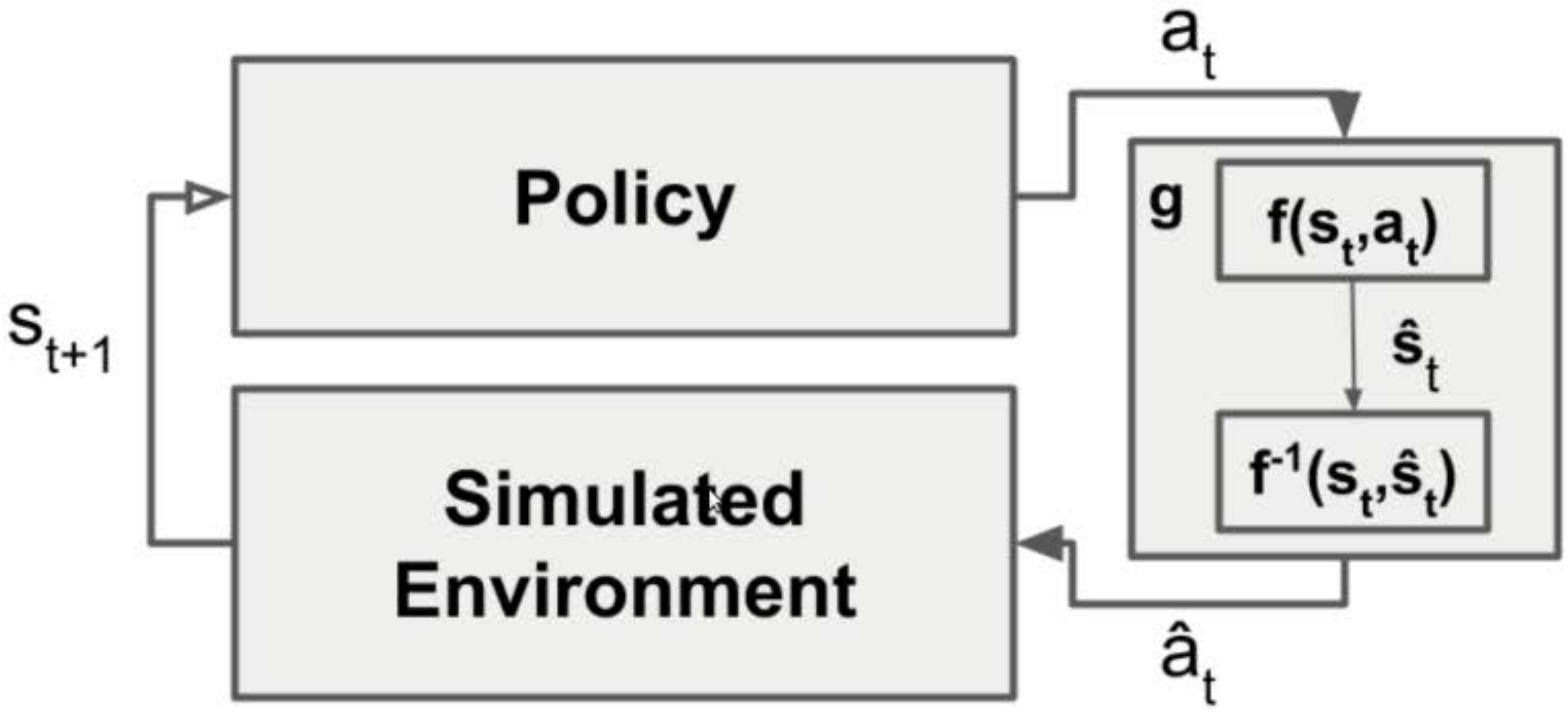
Grounded Action Transformation



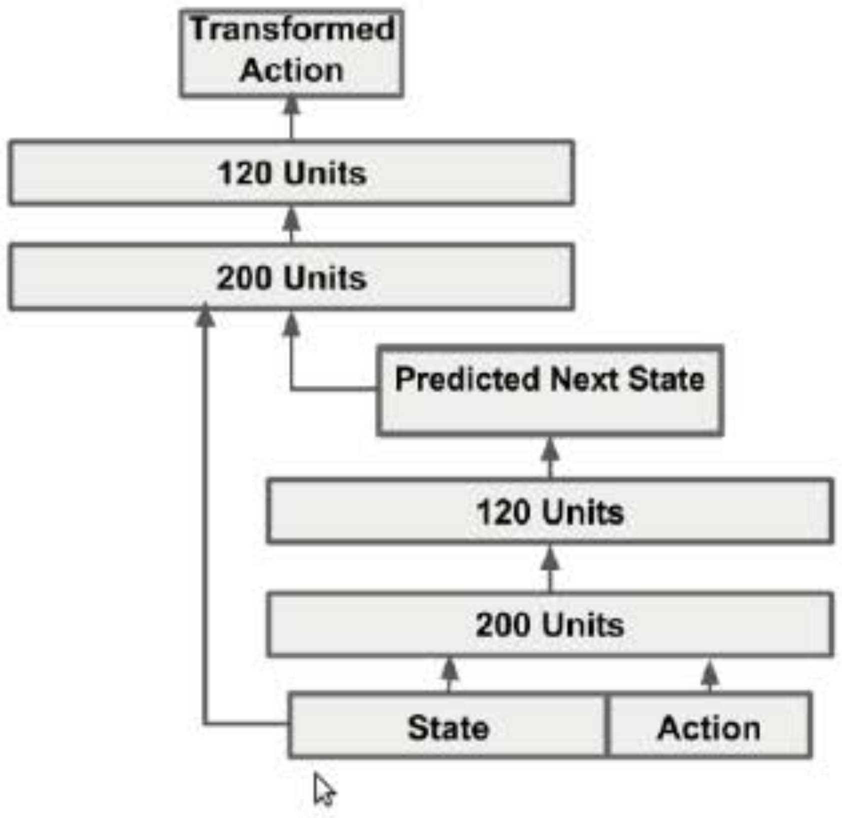
Grounded Action Transformation



Grounded Action Transformation

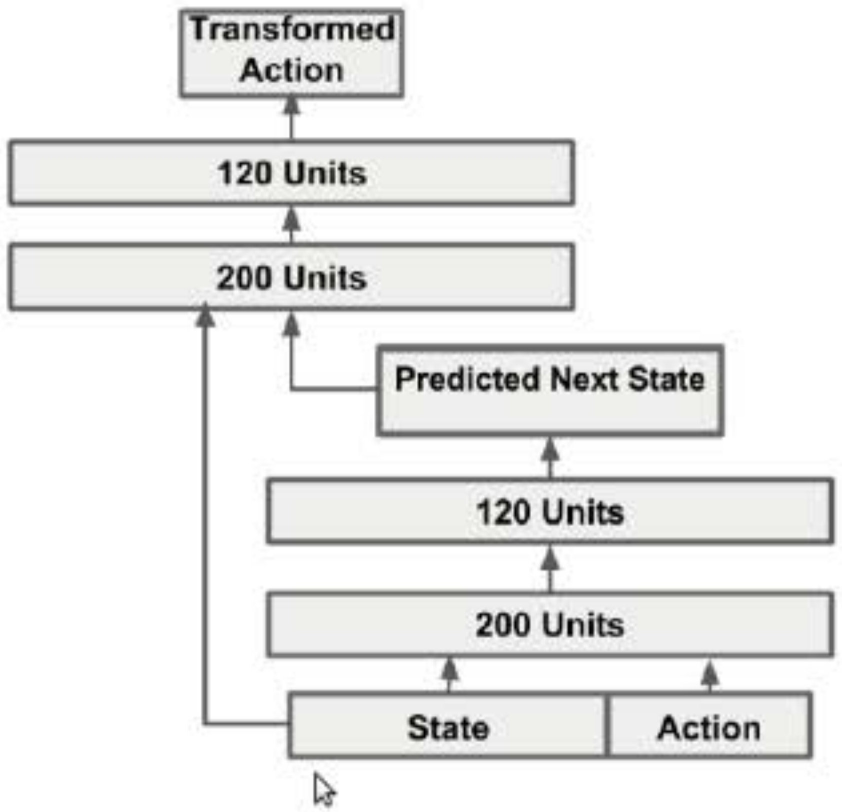


Supervised Implementation



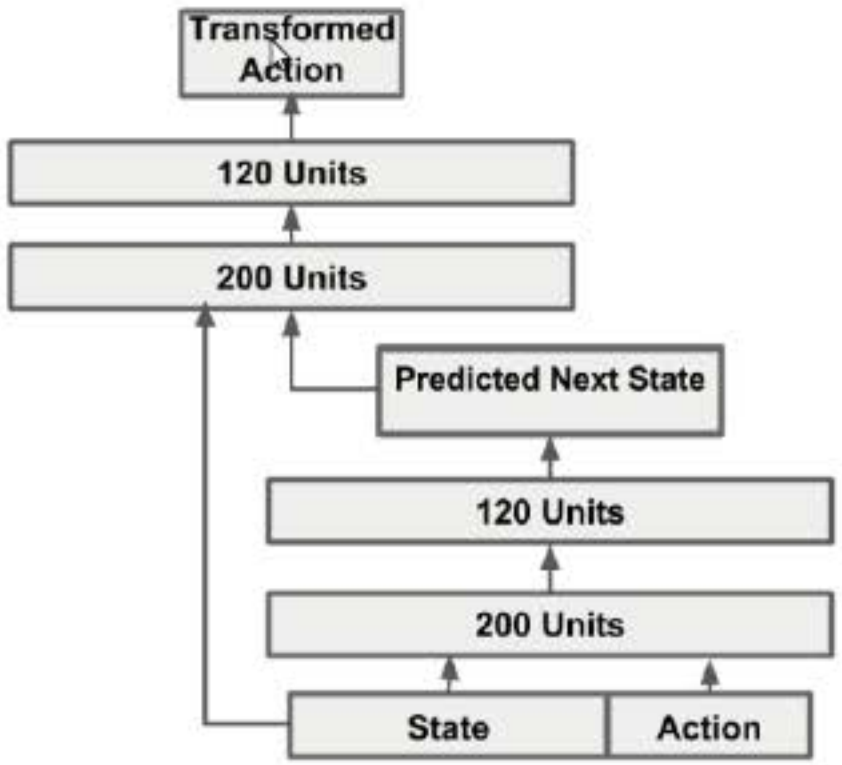
- Forward model:
  - trained with 15 real world trajectories of 2000 time-steps

Supervised Implementation



- Forward model:
  - trained with 15 real world trajectories of 2000 time-steps
- Inverse model:
  - trained with 50 simulated trajectories of 1000 time-steps

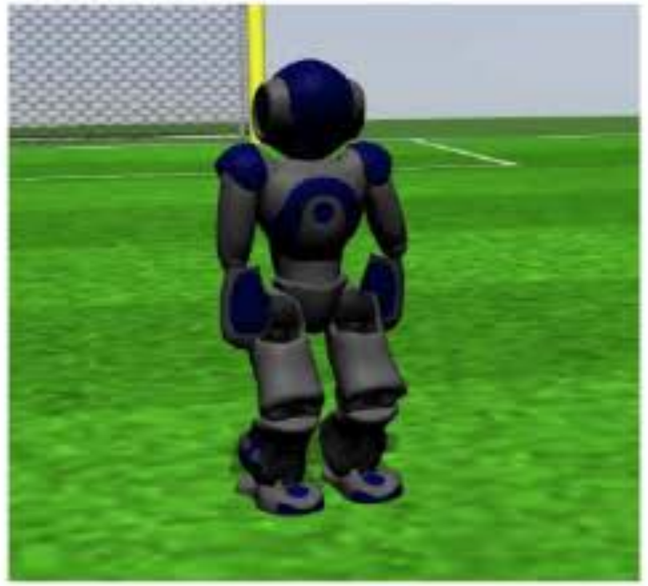
Supervised Implementation



- Forward model:
  - trained with 15 real world trajectories of 2000 time-steps
- Inverse model:
  - trained with 50 simulated trajectories of 1000 time-steps
- Initial policy in Initial vs. grounded simulator

Empirical Results

(a) Softbank NAO



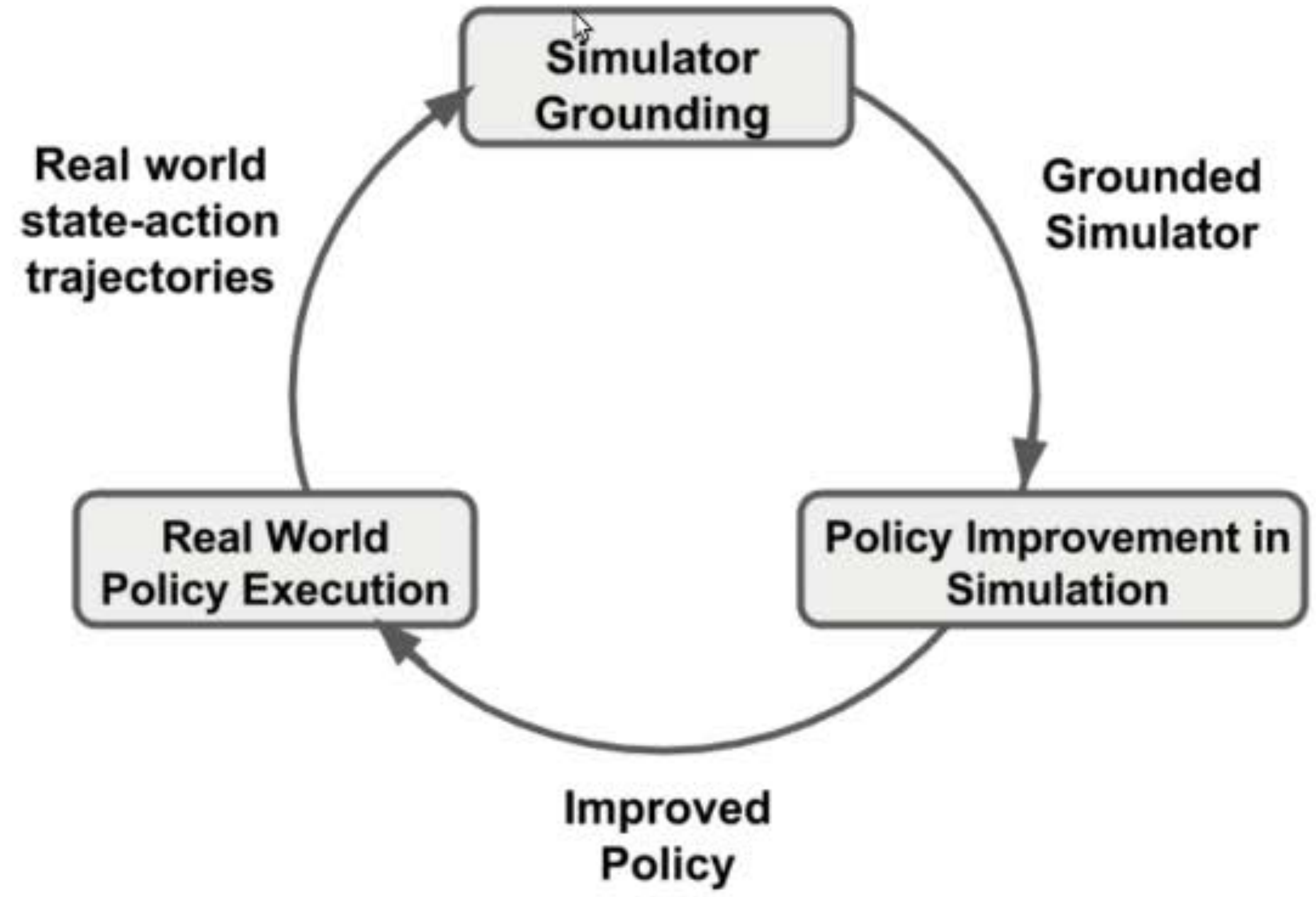
(b) Gazebo NAO

(c) SimSpark NAO

Applied GAT to learning fast bipedal walks for the Nao robot.

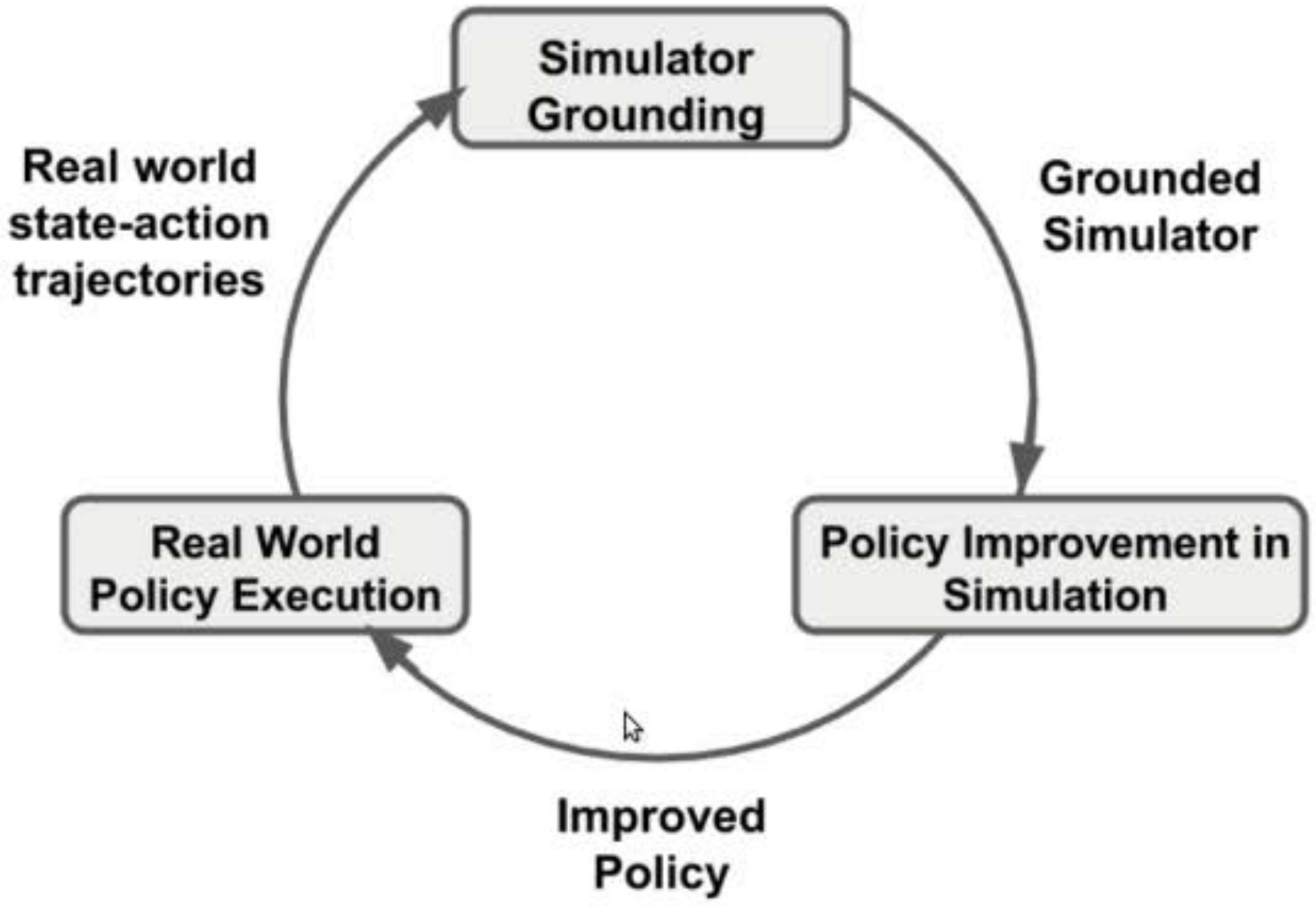
- Initial policy: University of New South Wales Walk Engine.
- Policy Search Algorithm: CMA-ES stochastic search method.

Empirical Results



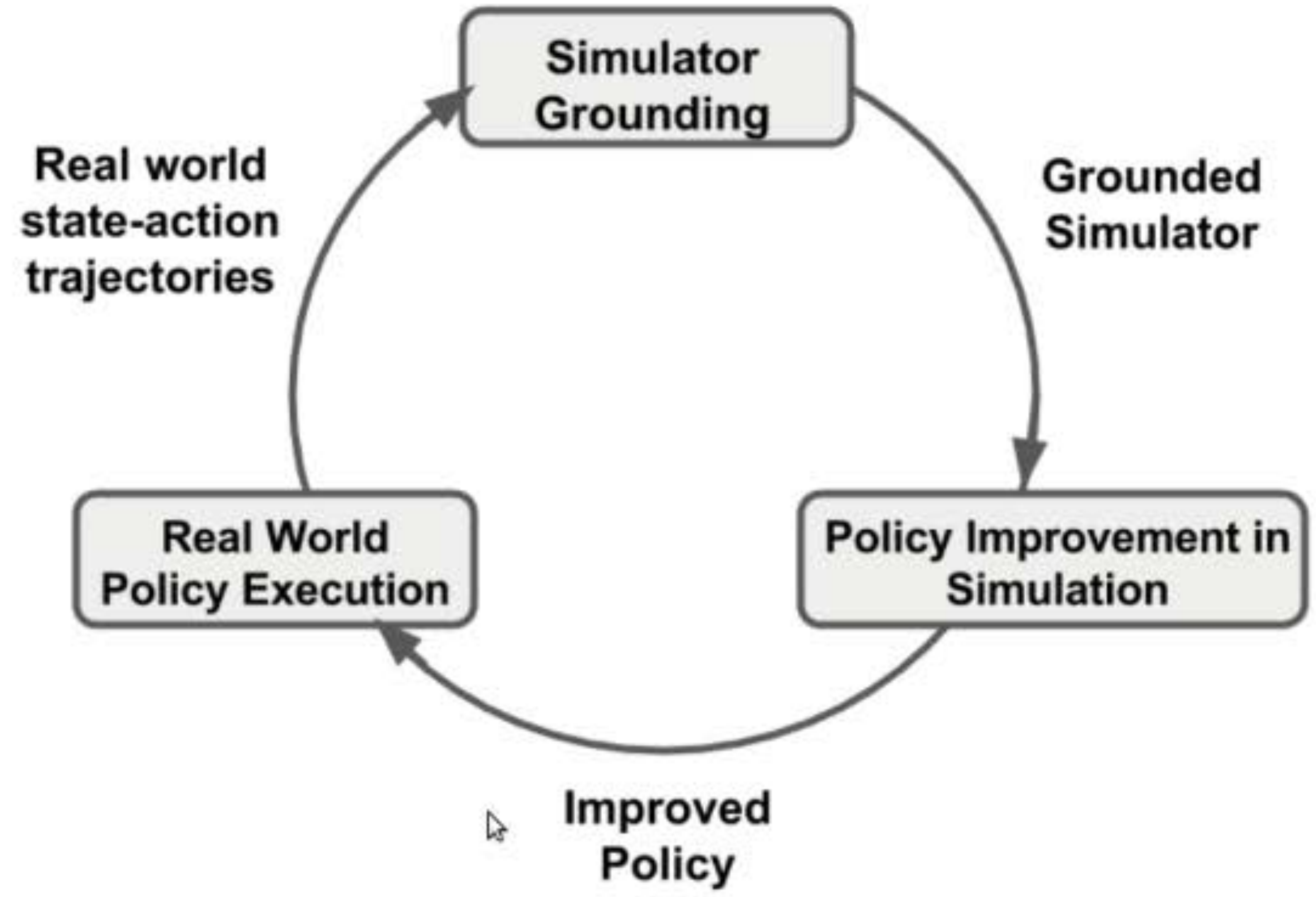
Method	Velocity (cm/s)	% Improve
Initial policy	19.3	0.0

Empirical Results



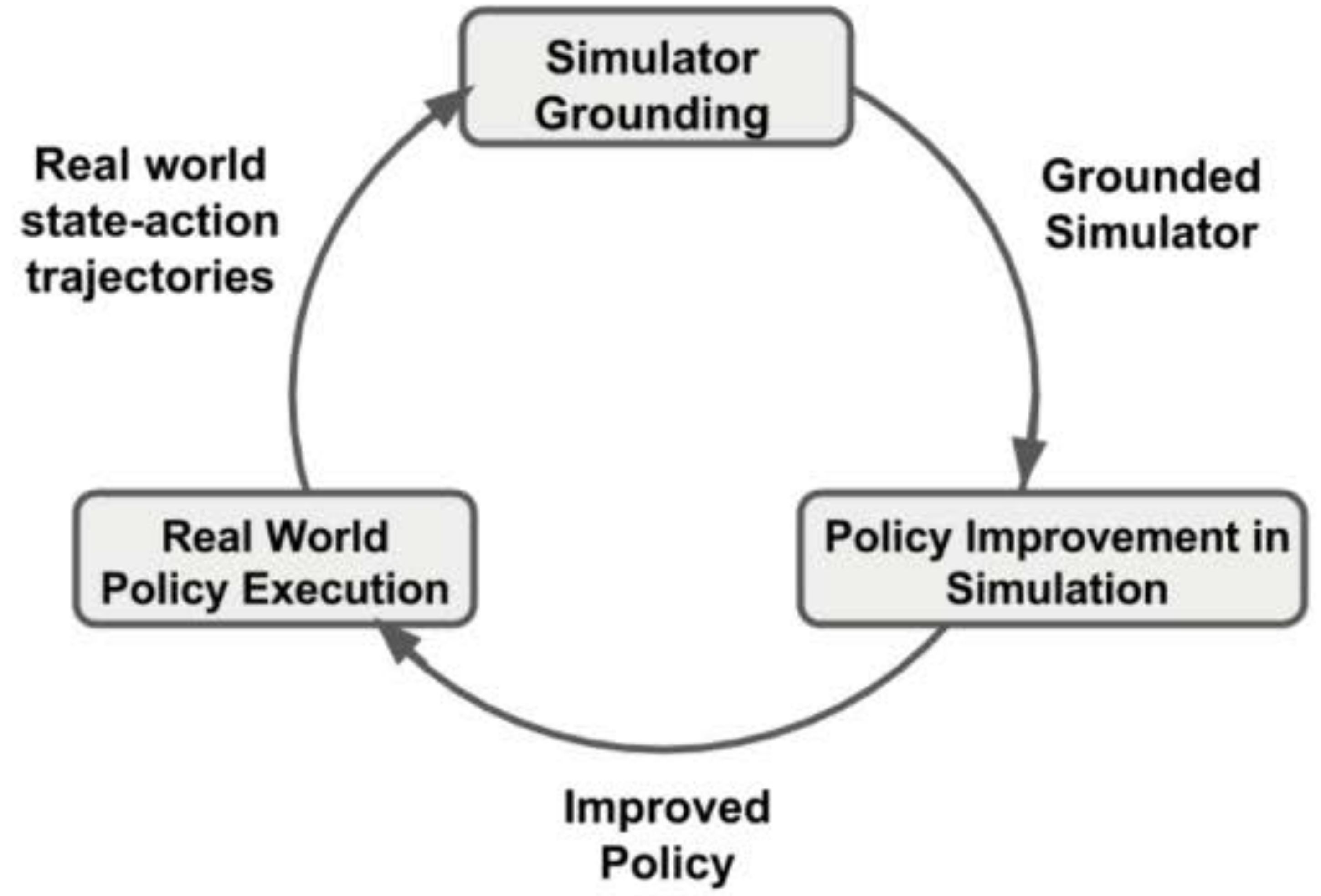
Method	Velocity (cm/s)	% Improve
Initial policy	19.3	0.0

Empirical Results



Method	Velocity (cm/s)	% Improve
Initial policy	19.3	0.0
1st iteration	26.3	34.6

Empirical Results



Method	Velocity (cm/s)	% Improve
Initial policy	19.3	0.0
1st iteration	26.3	34.6
2nd iteration	**28.0**	43.3

GSL Summary

- Introduced **Grounded Simulation Learning** for Sim2Real.

Hanna and Stone, AAAI 2017

GSL Summary

- Introduced **Grounded Simulation Learning** for Sim2Real.
- Improved walk speed of Nao robot by over 40% compared to state-of-the-art walk engine.
- Fastest known stable walk on the Nao

Patrick MacAlpine

Josiah Hanna

Hanna and Stone, AAAI 2017

GSL Summary

- Introduced **Grounded Simulation Learning** for Sim2Real.
- Improved walk speed of Nao robot by over 40% compared to state-of-the-art walk engine.
- Fastest known stable walk on the Nao

Patrick MacAlpine

Josiah Hanna

Ongoing Work:

- Extending to other robotics tasks and platforms
- When does grounding actions work and when does it not?

Hanna and Stone, AAAI 2017

GSL Summary

- Introduced **Grounded Simulation Learning** for Sim2Real.
- Improved walk speed of Nao robot by over 40% compared to state-of-the-art walk engine.
- Fastest known stable walk on the Nao

Patrick MacAlpine

Josiah Hanna

Ongoing Work:

- Extending to other robotics tasks and platforms
- When does grounding actions work and when does it not?
- Connecting to **off-policy evaluation** and **safe learning**

Hanna and Stone, AAAI 2017

Robot Skill Learning: Real World to Sim and Back

- Motivation: RoboCup
- Sim2Real: Grounded Simulation Learning
- **Imitation Learning from Observation:**
  - Model-based approach: BCO

GSL Summary

- Introduced **Grounded Simulation Learning** for Sim2Real.
- Improved walk speed of Nao robot by over 40% compared to state-of-the-art walk engine.
- Fastest known stable walk on the Nao

Patrick MacAlpine

Josiah Hanna

Ongoing Work:

- Extending to other robotics tasks and platforms
- When does grounding actions work and when does it not?
- Connecting to **off-policy evaluation** and **safe learning**

Hanna and Stone, AAAI 2017

Robot Skill Learning: Real World to Sim and Back

- Motivation: RoboCup
- Sim2Real: Grounded Simulation Learning
- **Imitation Learning from Observation:**
  - Model-based approach: BCO

Robot Skill Learning: Real World to Sim and Back

- Motivation: RoboCup
- Sim2Real: Grounded Simulation Learning
- **Imitation Learning from Observation:**
  - Model-based approach: BCO
  - Model-free approach: GAIfO

Robot Skill Learning: Real World to Sim and Back

- Motivation: RoboCup
- Sim2Real: Grounded Simulation Learning
- **Imitation Learning from Observation:**
  - Model-based approach: BCO
  - Model-free approach: GAIfO

Faraz Torabi

Garrett Warnell

Imitation Learning

Goal:

- Learn how to make decisions by trying to imitate another agent.

Imitation Learning

Goal:

- Learn how to make decisions by trying to imitate another agent.

Conventional Imitation Learning:

- Observations of other agent (demonstrations) consist of state-action pairs.[1]

[1] Niekum et al., "Learning and generalization of complex tasks from unstructured demonstrations".

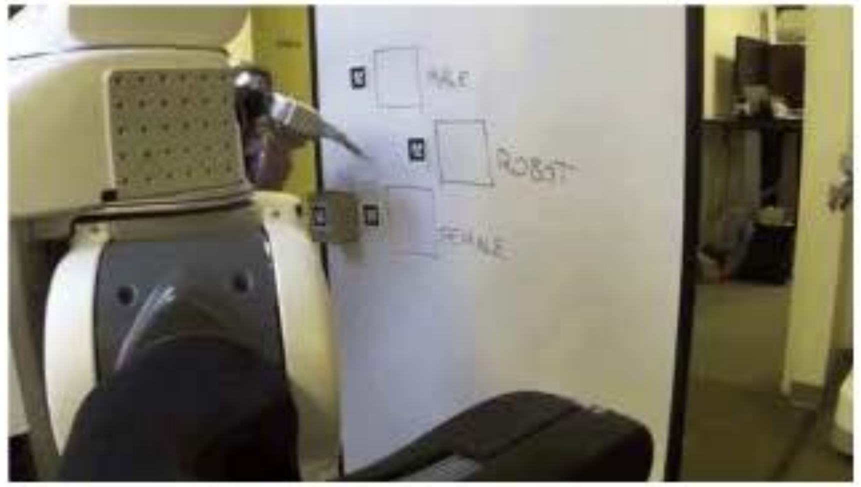
Imitation Learning

Goal:

- Learn how to make decisions by trying to imitate another agent.

Conventional Imitation Learning:

- Observations of other agent (demonstrations) consist of state-action pairs.[1]



Imitation Learning

Goal:

- Learn how to make decisions by trying to imitate another agent.

Conventional Imitation Learning:

- Observations of other agent (demonstrations) consist of state-action pairs.[1]

Challenge:

- Precludes using a large amount of demonstration data where action sequences are not given (e.g. YouTube videos).

[1] Niekum et al., "Learning and generalization of complex tasks from unstructured demonstrations".

Imitation Learning

Algorithms:

Imitation Learning

Algorithms:

- Behavioral Cloning:

Imitation Learning

Algorithms:

- Behavioral Cloning:
  - End to End Learning for Self-Driving Cars.[2]

[2] Zhang and Cho, "Query-Efficient Imitation Learning for End-to-End Simulated Driving."

Imitation Learning

Algorithms:

- Behavioral Cloning:
  - End to End Learning for Self-Driving Cars.[2]
- Inverse Reinforcement Learning:

[2] Zhang and Cho, "Query-Efficient Imitation Learning for End-to-End Simulated Driving."

Imitation Learning

Algorithms:

- Behavioral Cloning:
  - End to End Learning for Self-Driving Cars.[2]
- Inverse Reinforcement Learning:
  - Generative Adversarial Imitation Learning.[3]
  - Guided Cost Learning.[4]

---

[2] Zhang and Cho, "Query-Efficient Imitation Learning for End-to-End Simulated Driving."

[3] Ho and Ermon, "Generative adversarial imitation learning".

[4] Finn, Levine, and Abbeel, "Guided cost learning: Deep inverse optimal control via policy optimization".

Imitation from Observation

Goal:

- Learn how to perform a task given state-only demonstrations.

Imitation from Observation

Goal:

- Learn how to perform a task given state-only demonstrations.

Imitation from Observation

Goal:

- Learn how to perform a task given state-only demonstrations.

Formulation:

- Given:
  - $D_{demo} = (s_0, s_1, ...)$
- Learn:
  - $\pi : \mathcal{S} \to \mathcal{A}$

Imitation from Observation

Previous work:

Imitation from Observation

Previous work:

- Time Contrastive Networks (TCN).[5]
- Imitation from observation: Learning to imitate behaviors from raw video via context translation.[6]
- Learning invariant feature spaces to transfer skills with reinforcement learning.[7]

[5] Sermanet et al., “Time-contrastive networks: Self-supervised learning from multi-view observation”.

[6] Liu et al., “Imitation from observation: Learning to imitate behaviors from raw video via context translation”.

[7] Gupta et al., “Learning invariant feature spaces to transfer skills with reinforcement learning”.

Imitation from Observation

Previous work:

- Time Contrastive Networks (TCN).[5]
- Imitation from observation: Learning to imitate behaviors from raw video via context translation.[6]
- Learning invariant feature spaces to transfer skills with reinforcement learning.[7]

Concentrate on perception; require time-aligned demonstrations.

[5] Sermanet et al., "Time-contrastive networks: Self-supervised learning from multi-view observation".

[6] Liu et al., "Imitation from observation: Learning to imitate behaviors from raw video via context translation".

[7] Gupta et al., "Learning invariant feature spaces to transfer skills with reinforcement learning".

Efficient Robot Skill Learning

- Motivation: RoboCup
- Sim2Real: Grounded Simulation Learning
- Imitation Learning from Observation:
  - **Model-based approach:** BCO
  - Model-free approach: GAIfO

Model-based Approach

- Imitation Learning: $D_{demo} = \{(s_0, a_0), (s_1, a_1), ...\}$

Model-based Approach

- Imitation Learning: $D_{demo} = \{(s_0, a_0), (s_1, a_1), ...\}$
- Imitation from Observation: $D_{demo} = \{(s_0, ?), (s_1, ?), ...\}$

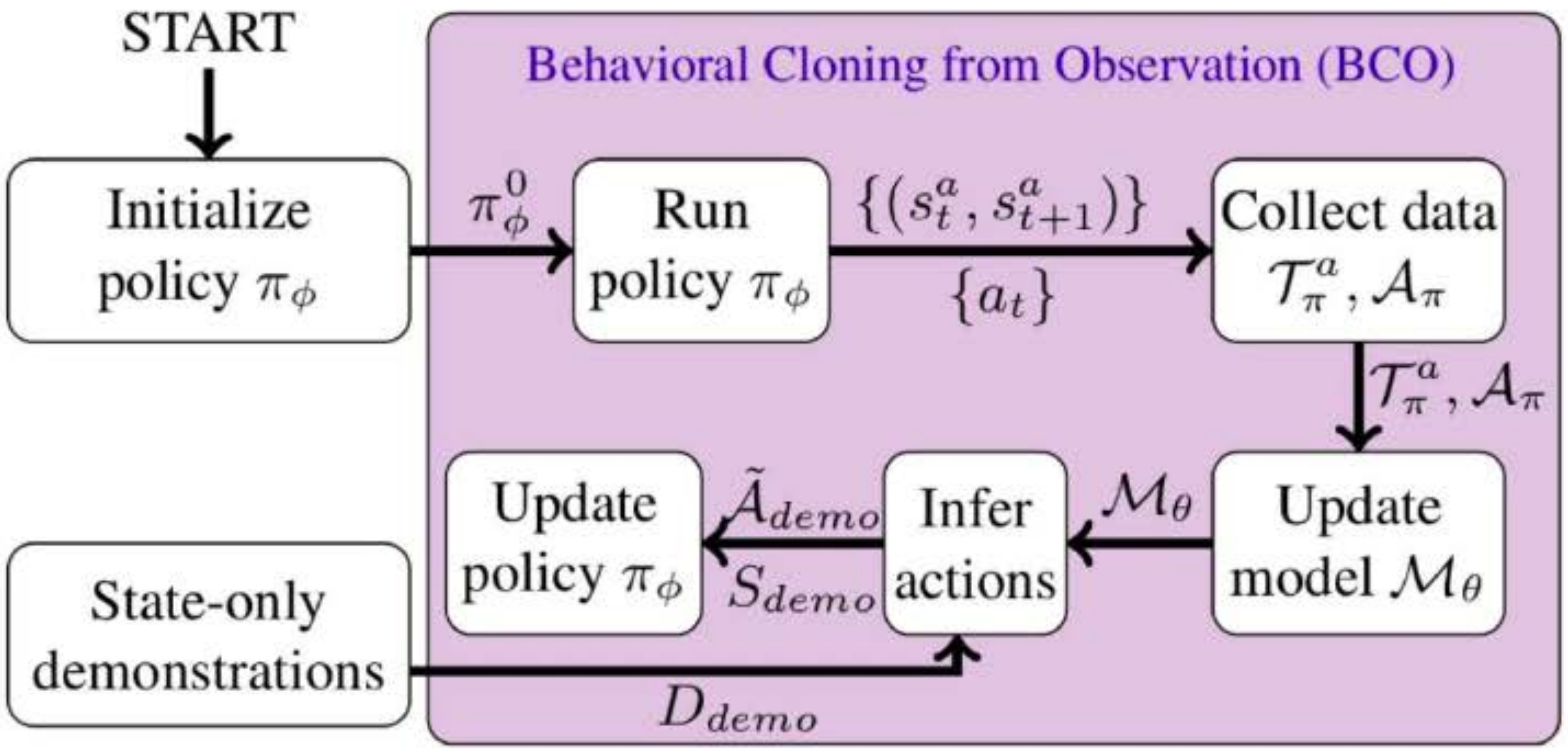
Model-based Approach

- Imitation Learning: $D_{demo} = \{(s_0, a_0), (s_1, a_1), ...\}$
- Imitation from Observation: $D_{demo} = \{(s_0, ?), (s_1, ?), ...\}$

Model-based Approach:

Behavioral Cloning from Observation (BCO)

Algorithm:



Torabi, Warnell, and Stone, IJCAI 2018

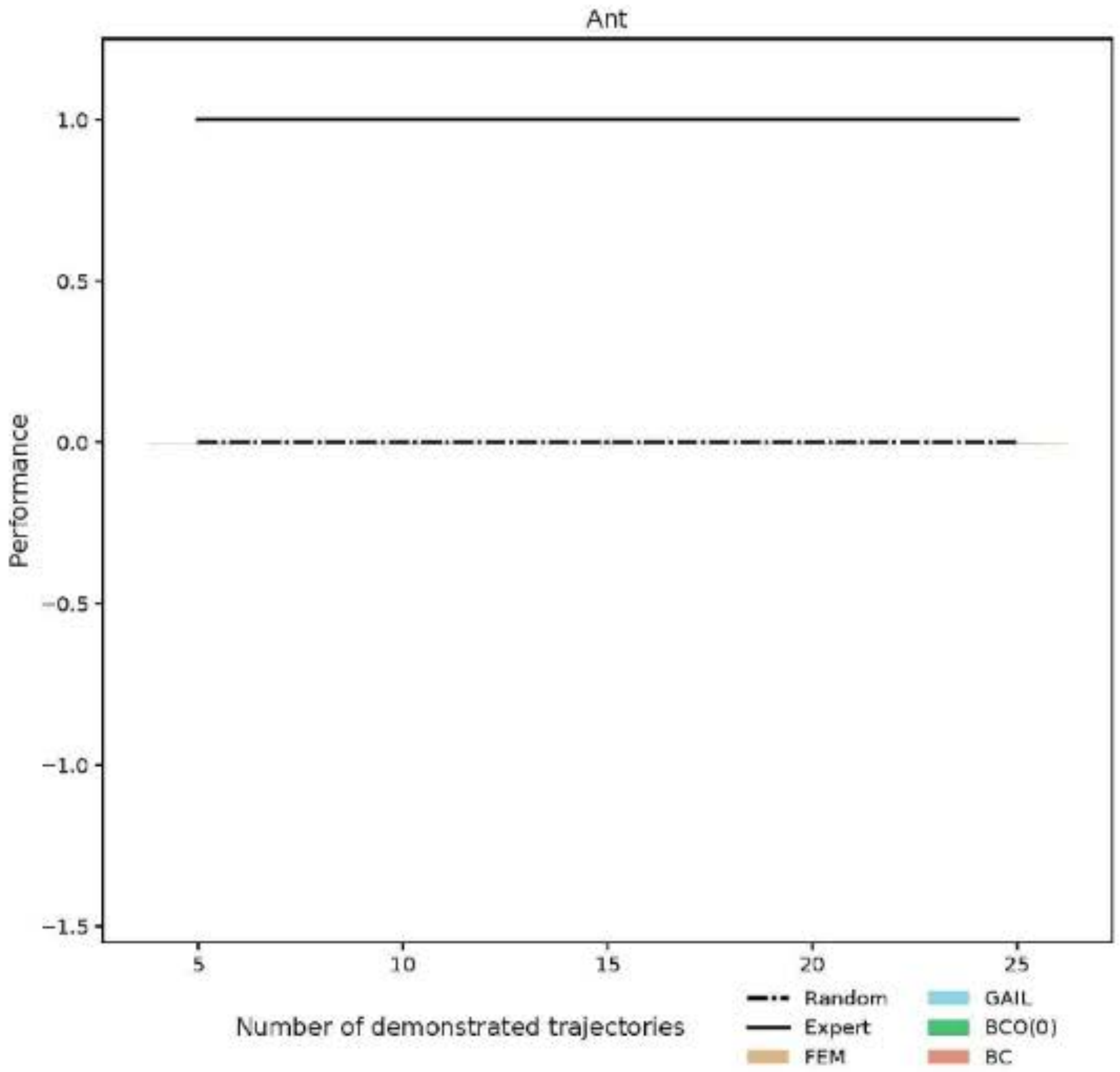
Behavioral Cloning from Observation (BCO)

Experimental Results:

- Domain:
  - Mujoco domain "Ant" with 111 dimensional state space and 8 dimensional action space.

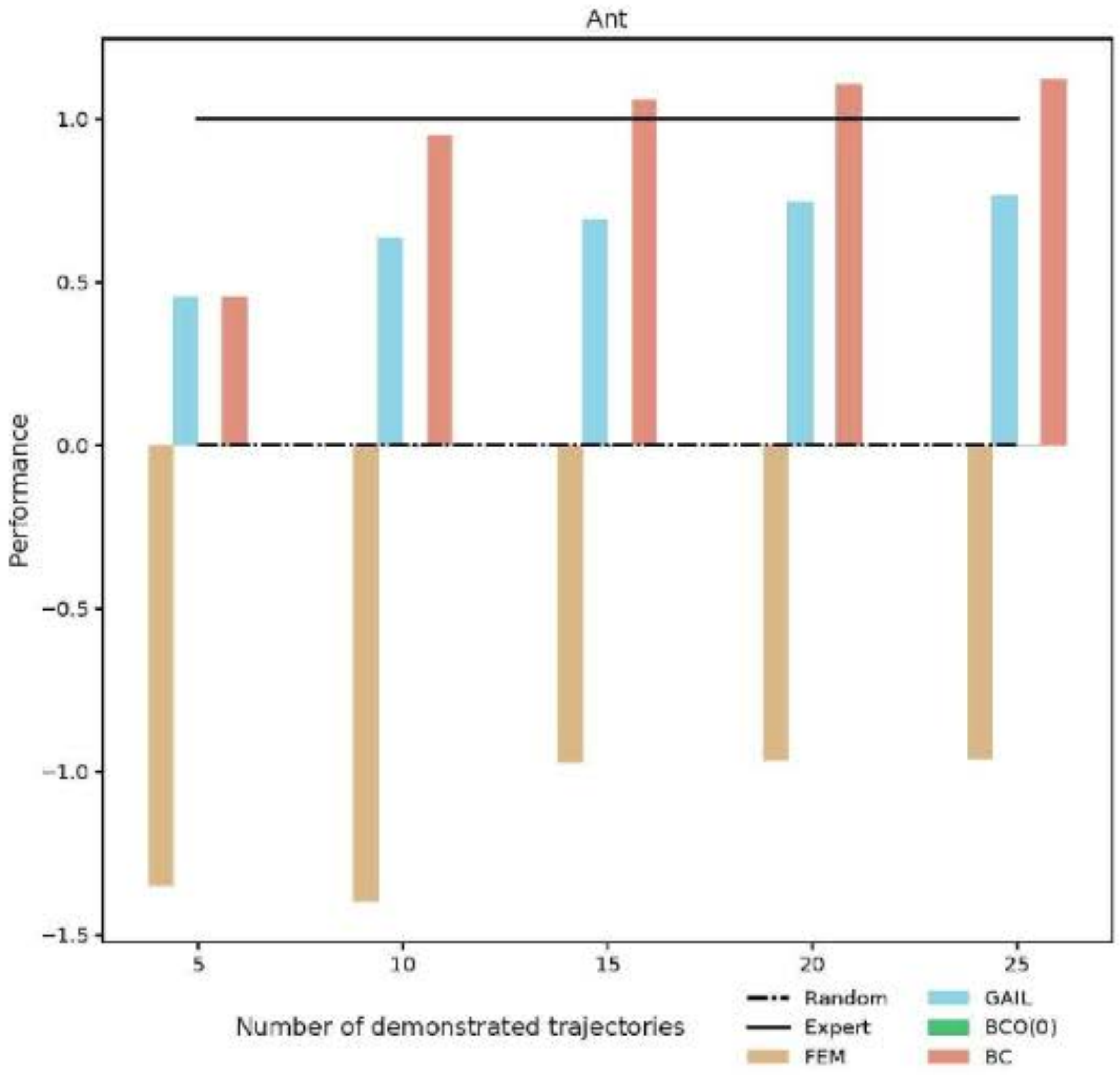
Behavioral Cloning from Observation (BCO)

Experimental Results:



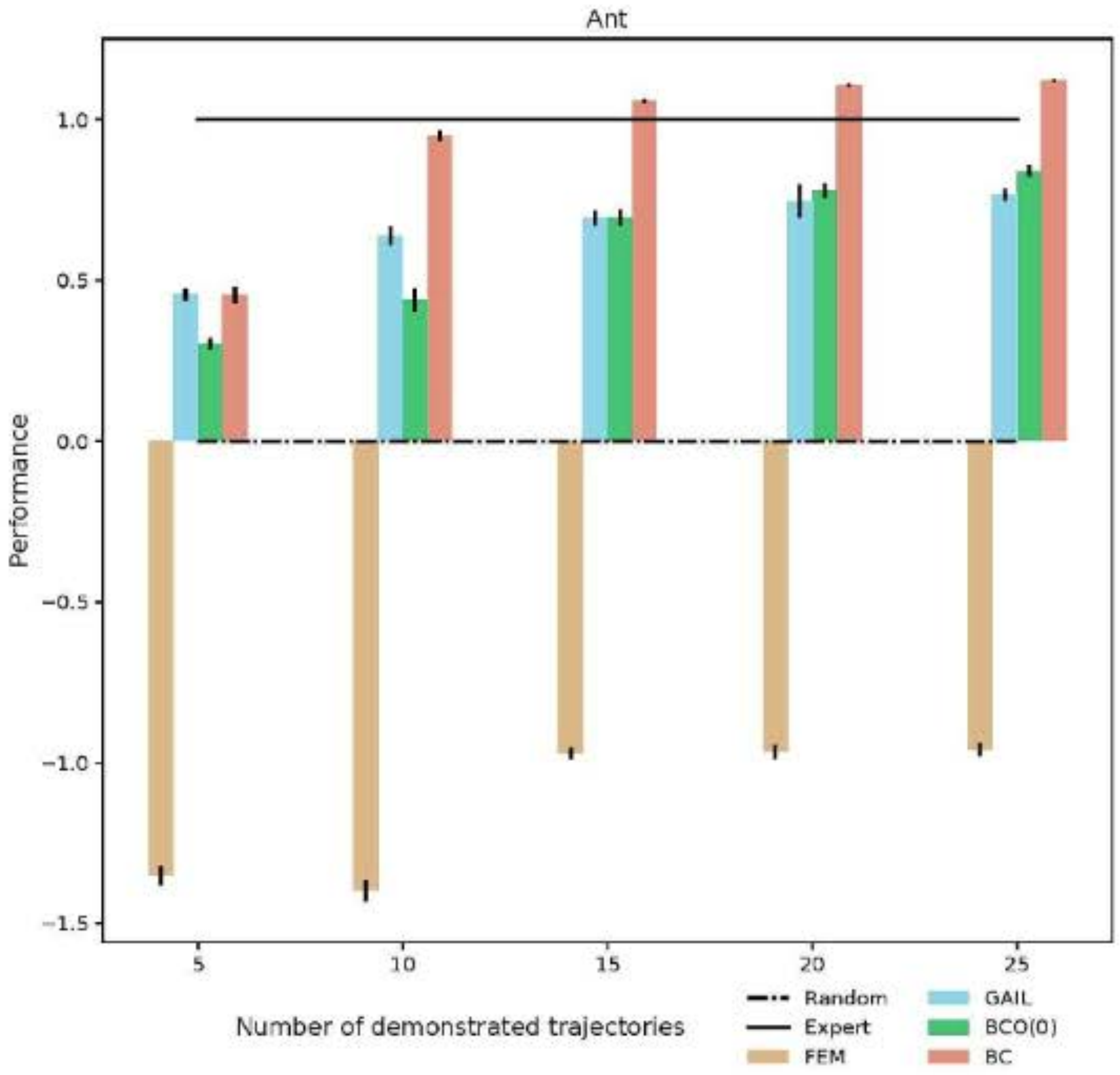
Behavioral Cloning from Observation (BCO)

Experimental Results:



Behavioral Cloning from Observation (BCO)

Experimental Results:



Behavioral Cloning from Observation (BCO(α))

Issue:

- Inverse dynamics model is learned using a random policy.

Behavioral Cloning from Observation (BCO(α))

Issue:

- Inverse dynamics model is learned using a random policy.

Solution: BCO(α)

Behavioral Cloning from Observation (BCO(α))

Issue:

- Inverse dynamics model is learned using a random policy.

Solution: BCO(α)

- Update the model with the learned policy.

Behavioral Cloning from Observation (BCO(α))

Issue:

- Inverse dynamics model is learned using a random policy.

Solution: BCO(α)

- Update the model with the learned policy.
- Parameter α controls tradeoff between performance and environment interactions

Behavioral Cloning from Observation (BCO(α))

Issue:

- Inverse dynamics model is learned using a random policy.

Solution: BCO(α)

- Update the model with the learned policy.
- Parameter α controls tradeoff between performance and environment interactions
  - $\alpha = 0$: no post-demonstration interaction.

Behavioral Cloning from Observation (BCO(α))

Issue:

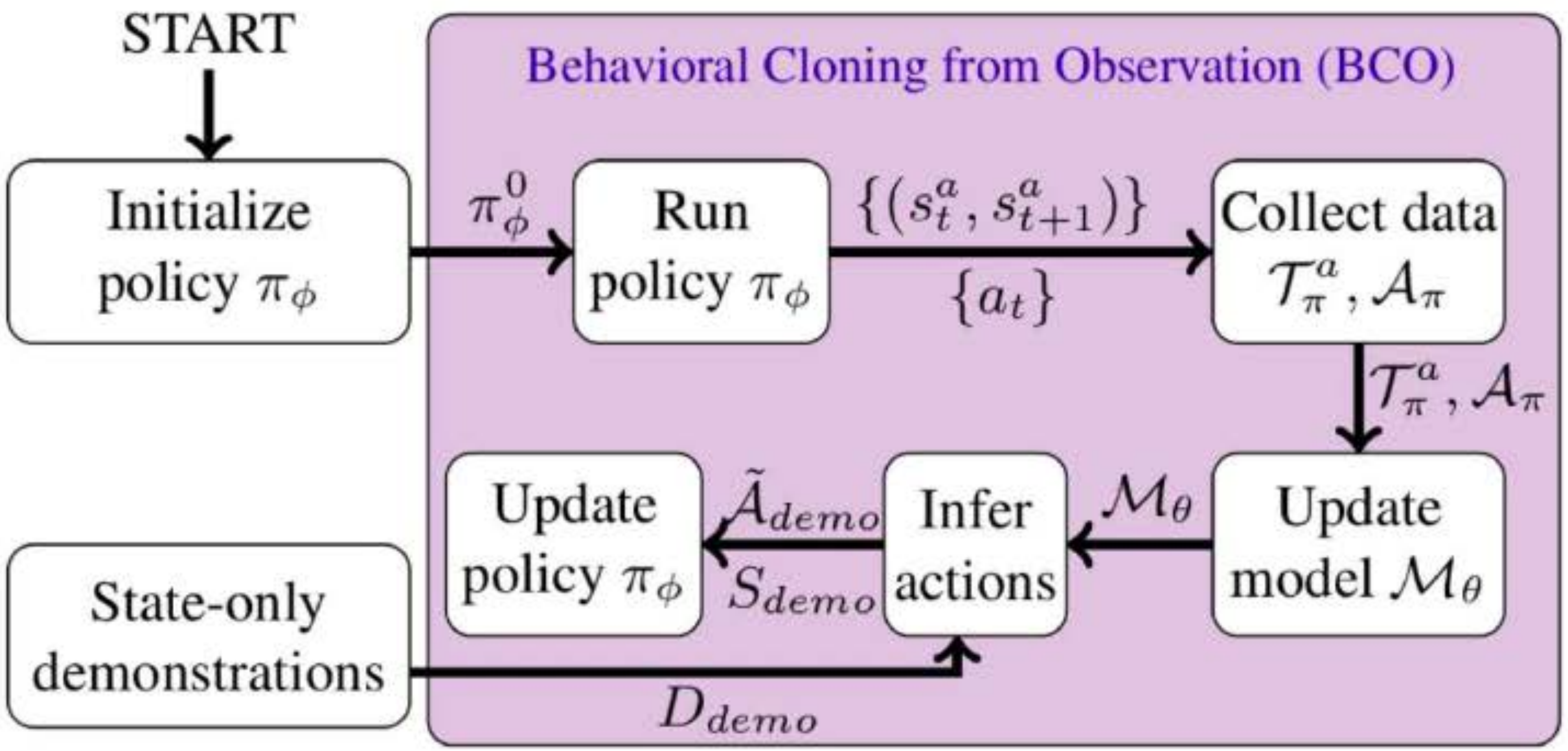
- Inverse dynamics model is learned using a random policy.

Solution: BCO(α)

- Update the model with the learned policy.
- Parameter α controls tradeoff between performance and environment interactions
  - $\alpha = 0$: no post-demonstration interaction.
  - Increasing α: increasing the number of interactions allowed at each iteration.

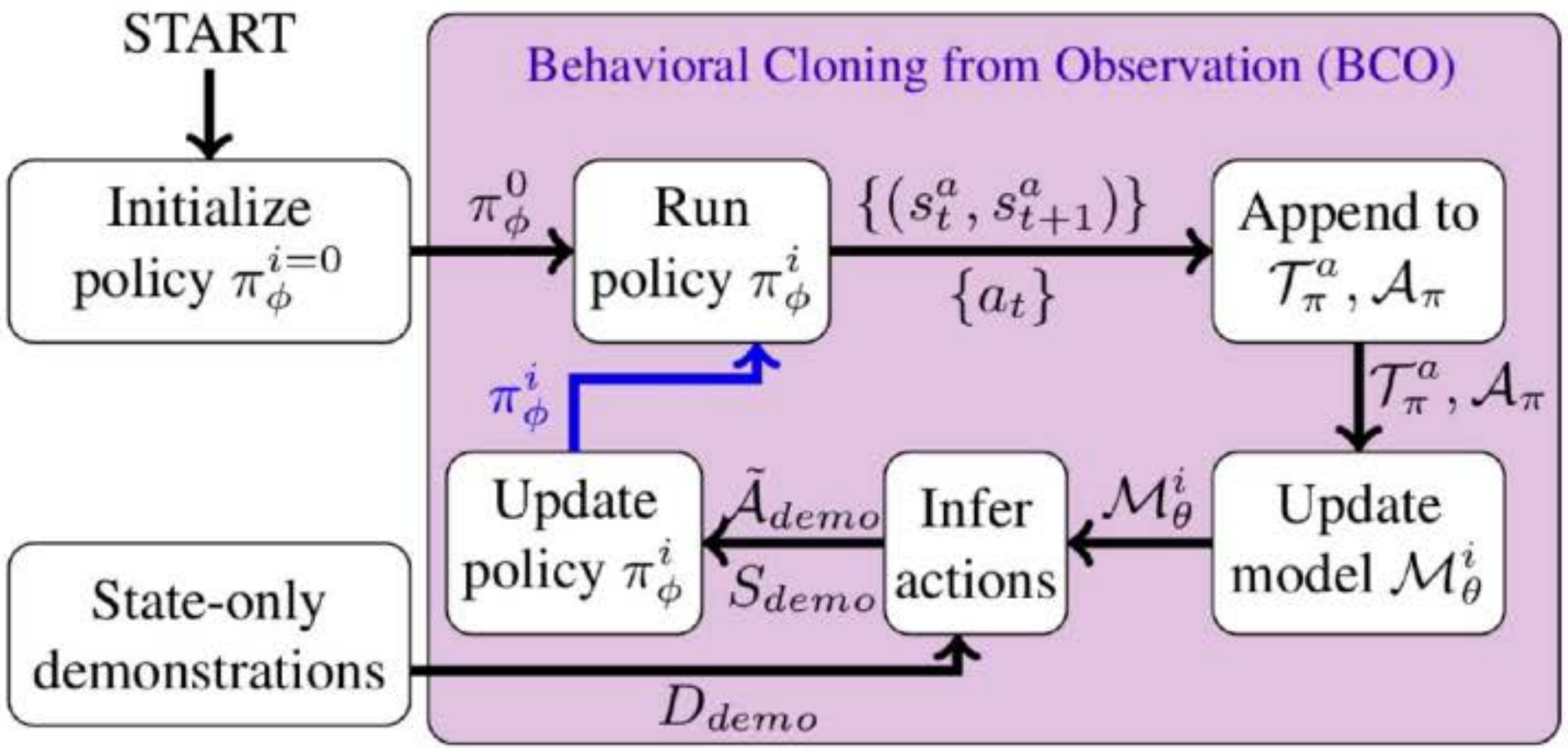
Behavioral Cloning from Observation (BCO(α))

Algorithm:



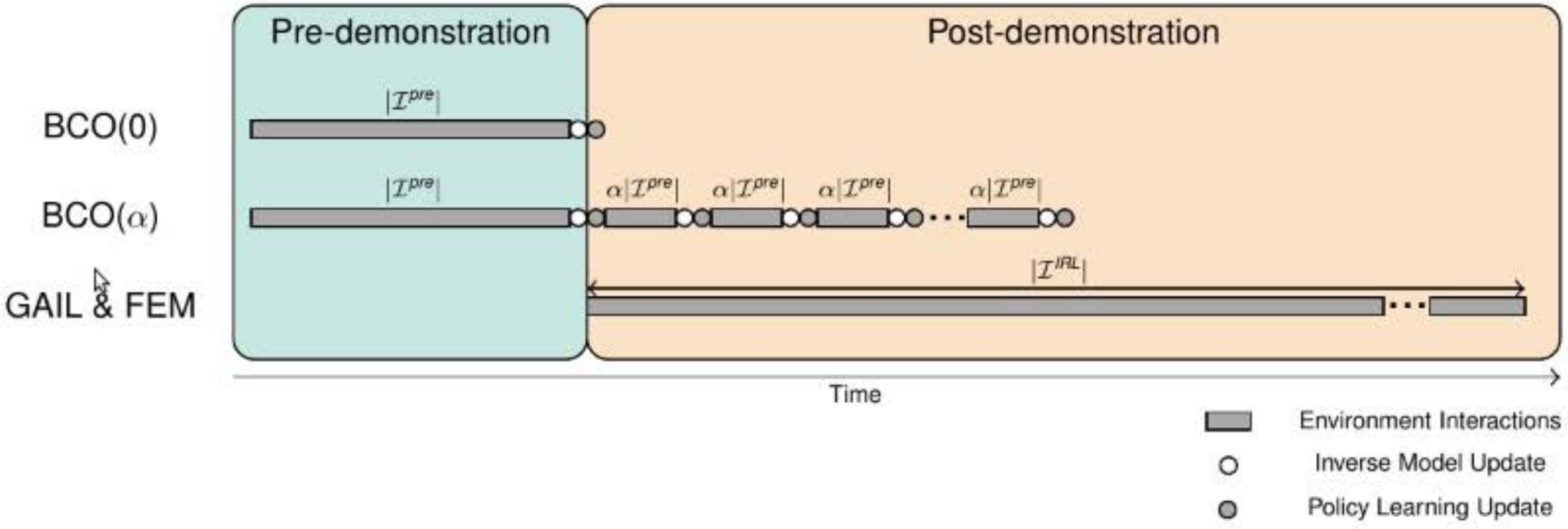
Behavioral Cloning from Observation (BCO(α))

Algorithm:



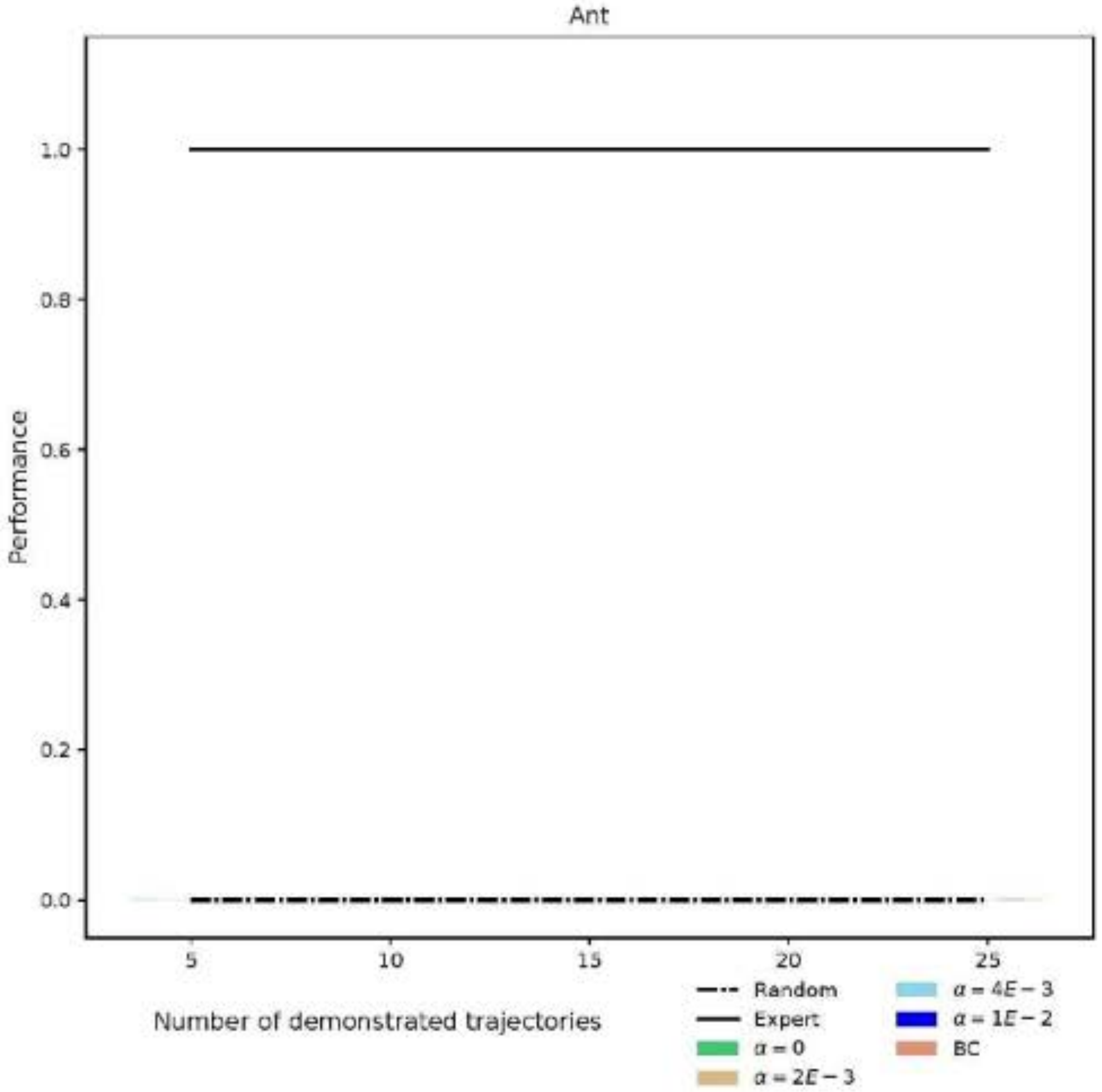
Behavioral Cloning from Observation (BCO(α))

Interaction time:



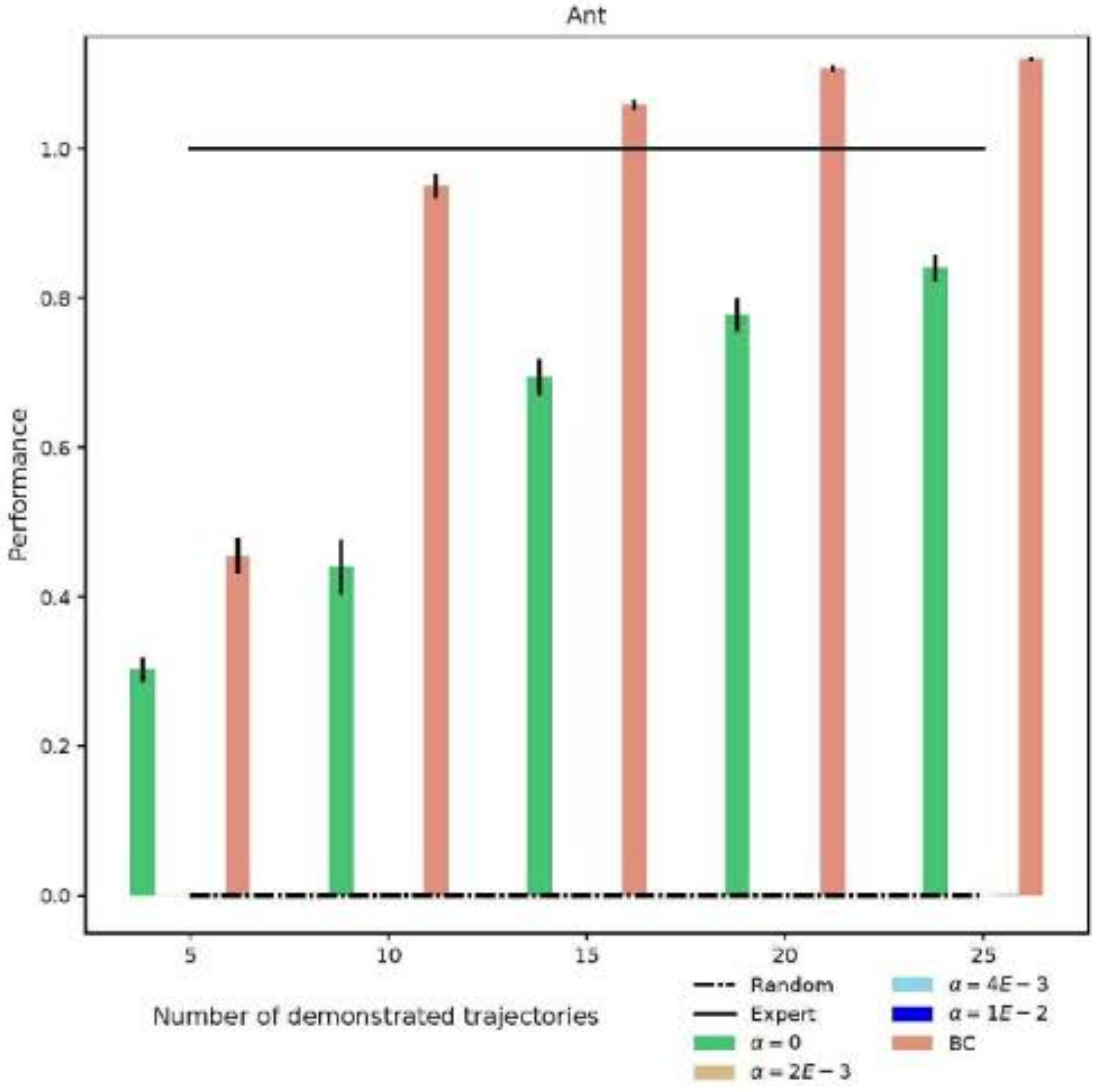
Behavioral Cloning from Observation (BCO(α))

Effect of varying α on BCO(α):



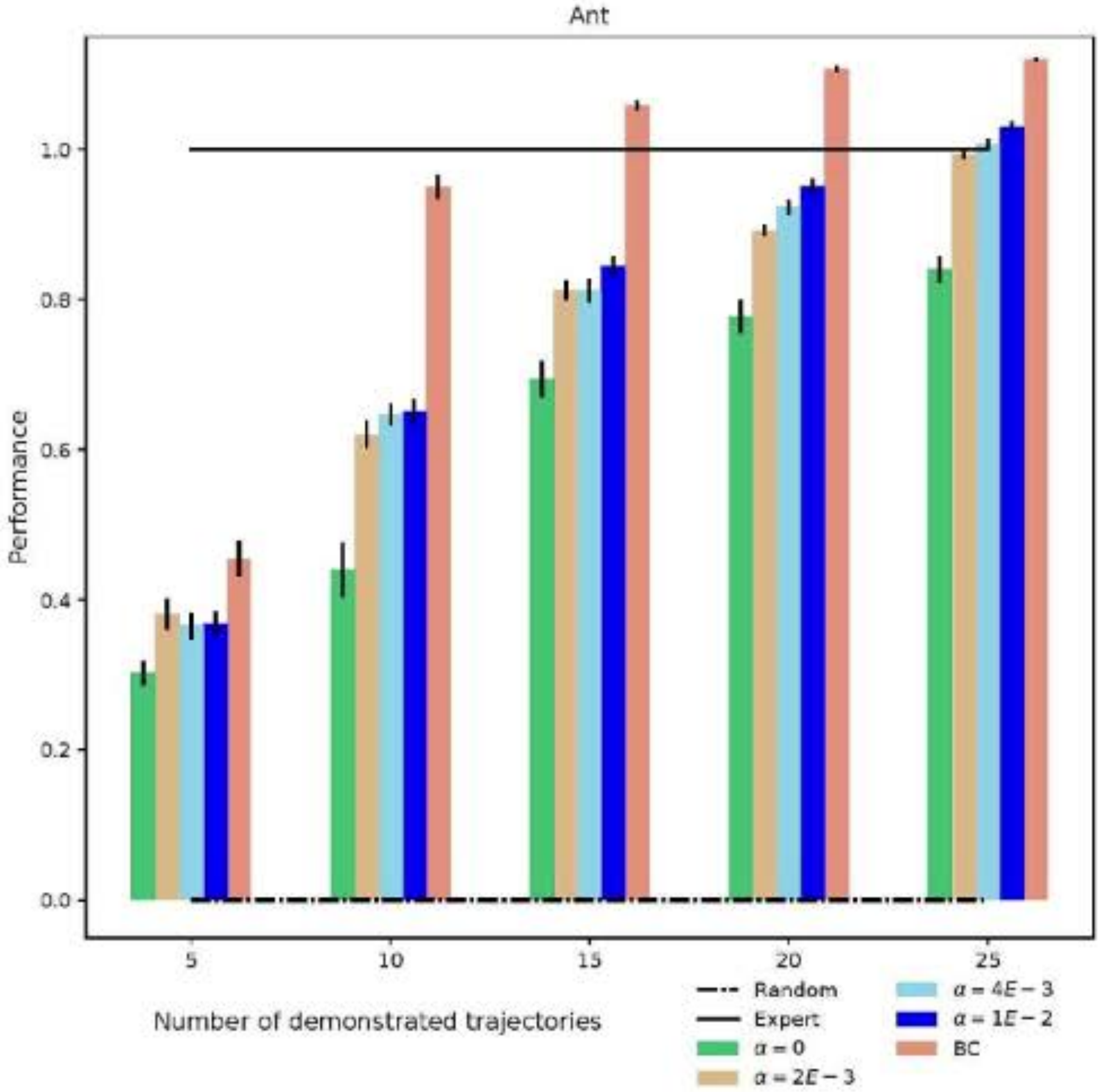
Behavioral Cloning from Observation (BCO(α))

Effect of varying α on BCO(α):



Behavioral Cloning from Observation (BCO(α))

Effect of varying α on BCO(α):

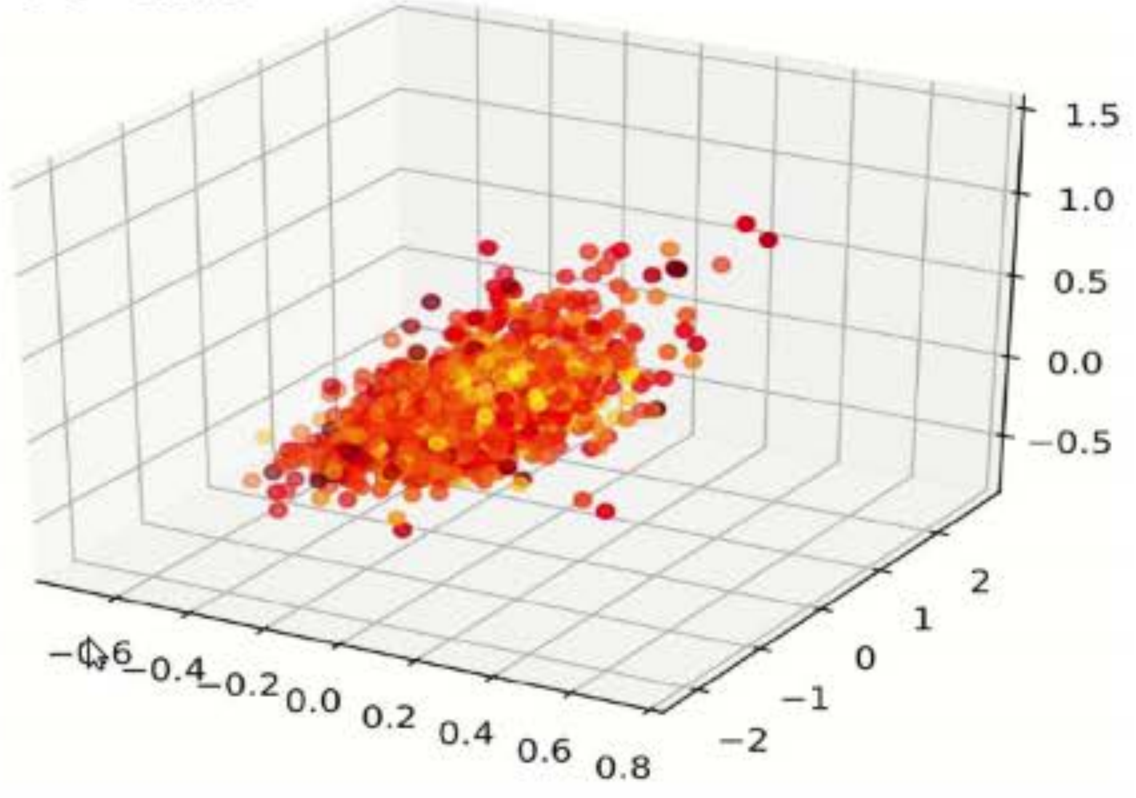


Efficient Robot Skill Learning

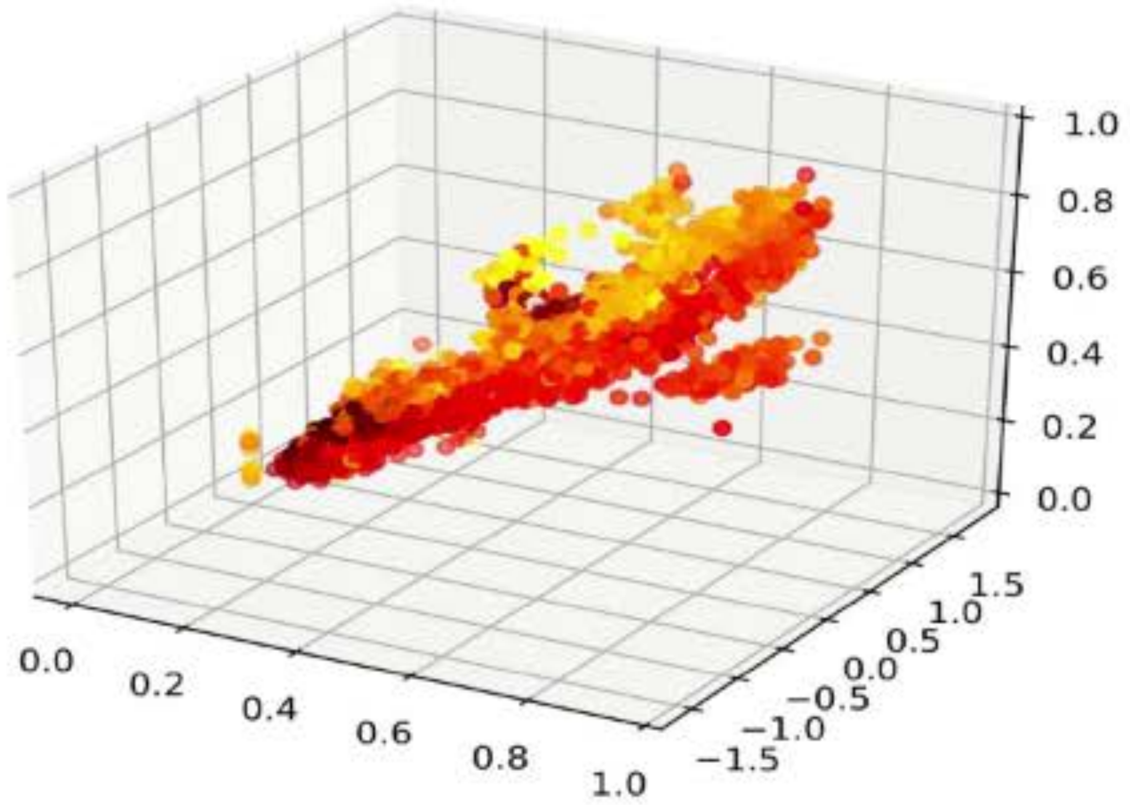
- Motivation: RoboCup
- Sim2Real: Grounded Simulation Learning
- Imitation Learning from Observation:
  - Model-based approach: BCO
  - **Model-free approach**: GAIfO

Gen. Adversarial Imitation from Observation (GAIfO)

Motivation:



(a) Random Policy

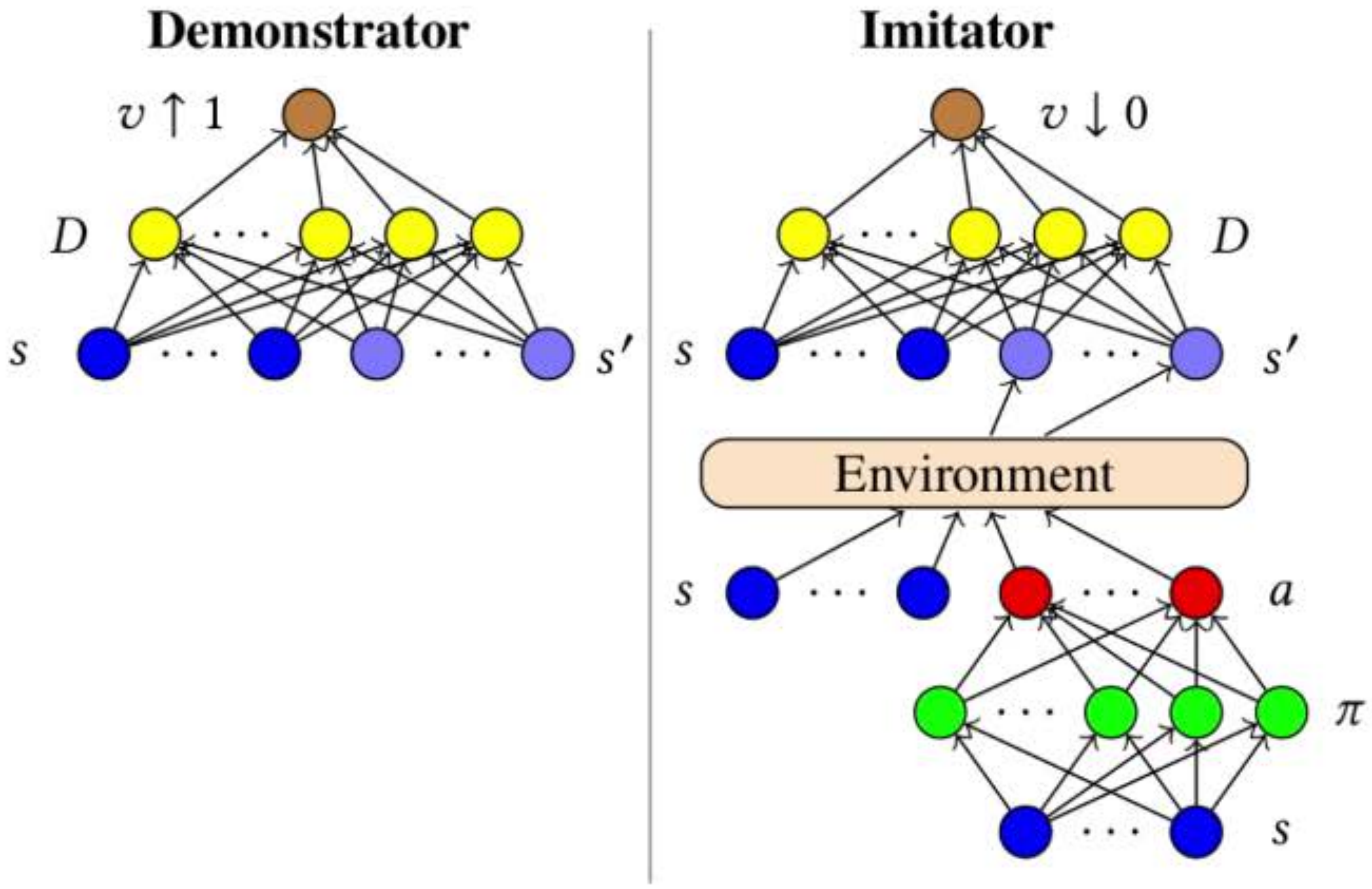


(b) Demonstration

Figure: State transition distribution in Hopper domain.

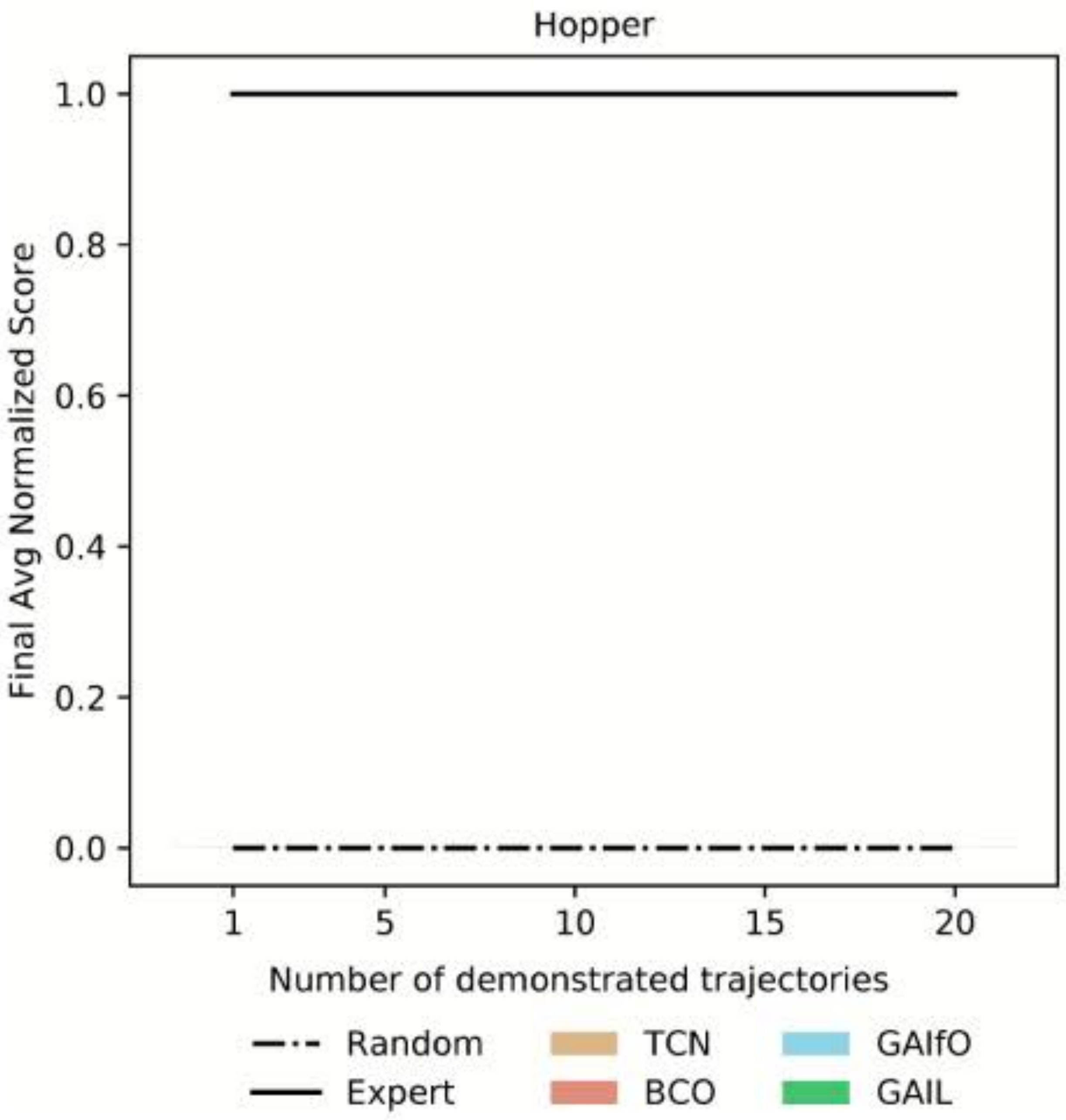
Gen. Adversarial Imitation from Observation (GAIfO)

Algorithm:



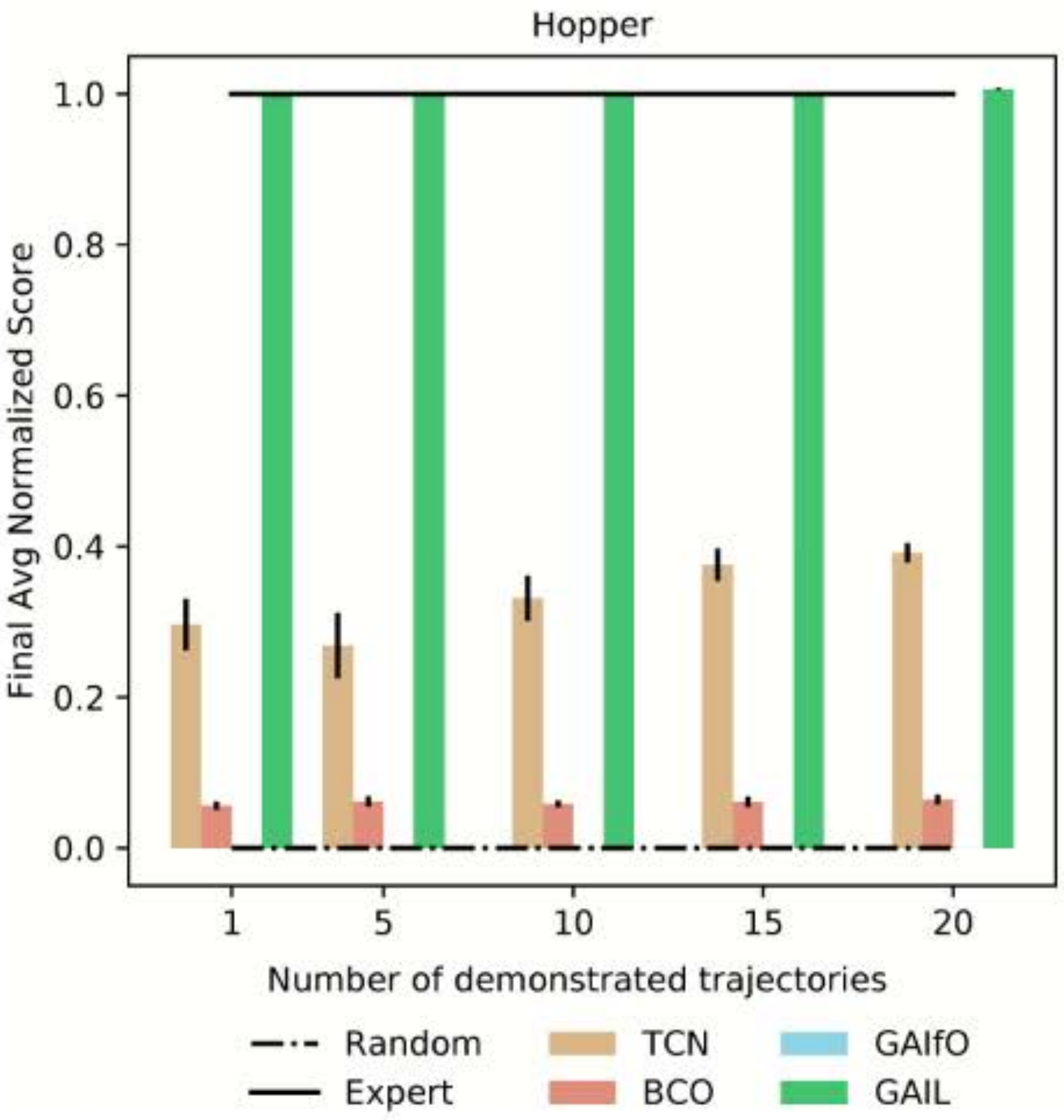
Gen. Adversarial Imitation from Observation (GAIfO)

Comparison against other IfO approaches and GAIL:



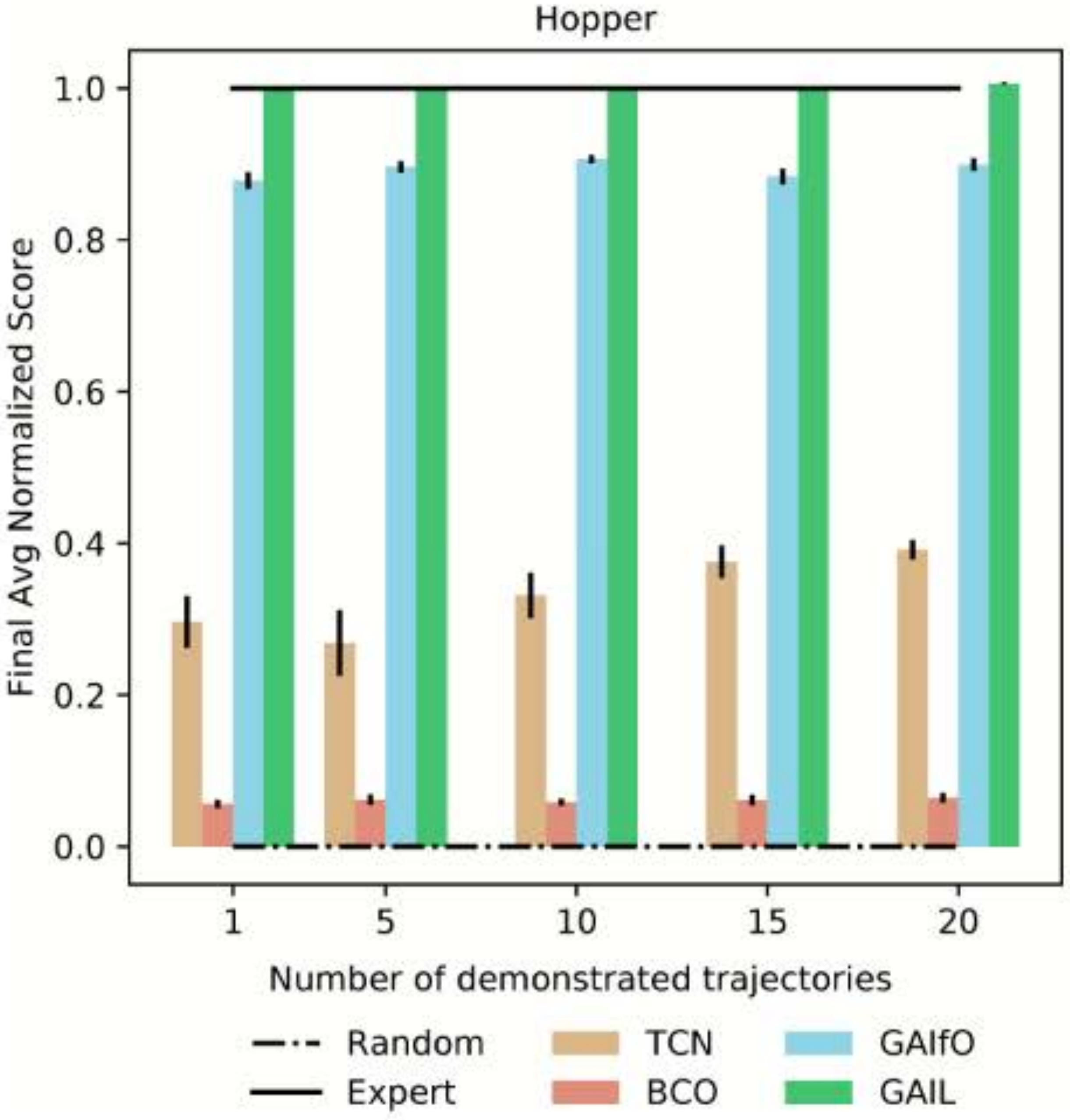
Gen. Adversarial Imitation from Observation (GAIfO)

Comparison against other IfO approaches and GAIL:



Gen. Adversarial Imitation from Observation (GAIfO)

Comparison against other IfO approaches and GAIL:



Gen. Adversarial Imitation from Observation (GAIfO)

Challenges:

Gen. Adversarial Imitation from Observation (GAIfO)

Challenges:

- States are not fully-observable.

Gen. Adversarial Imitation from Observation (GAIfO)

Challenges:

- States are not fully-observable.
- States are not Markovian.

Gen. Adversarial Imitation from Observation (GAIfO)

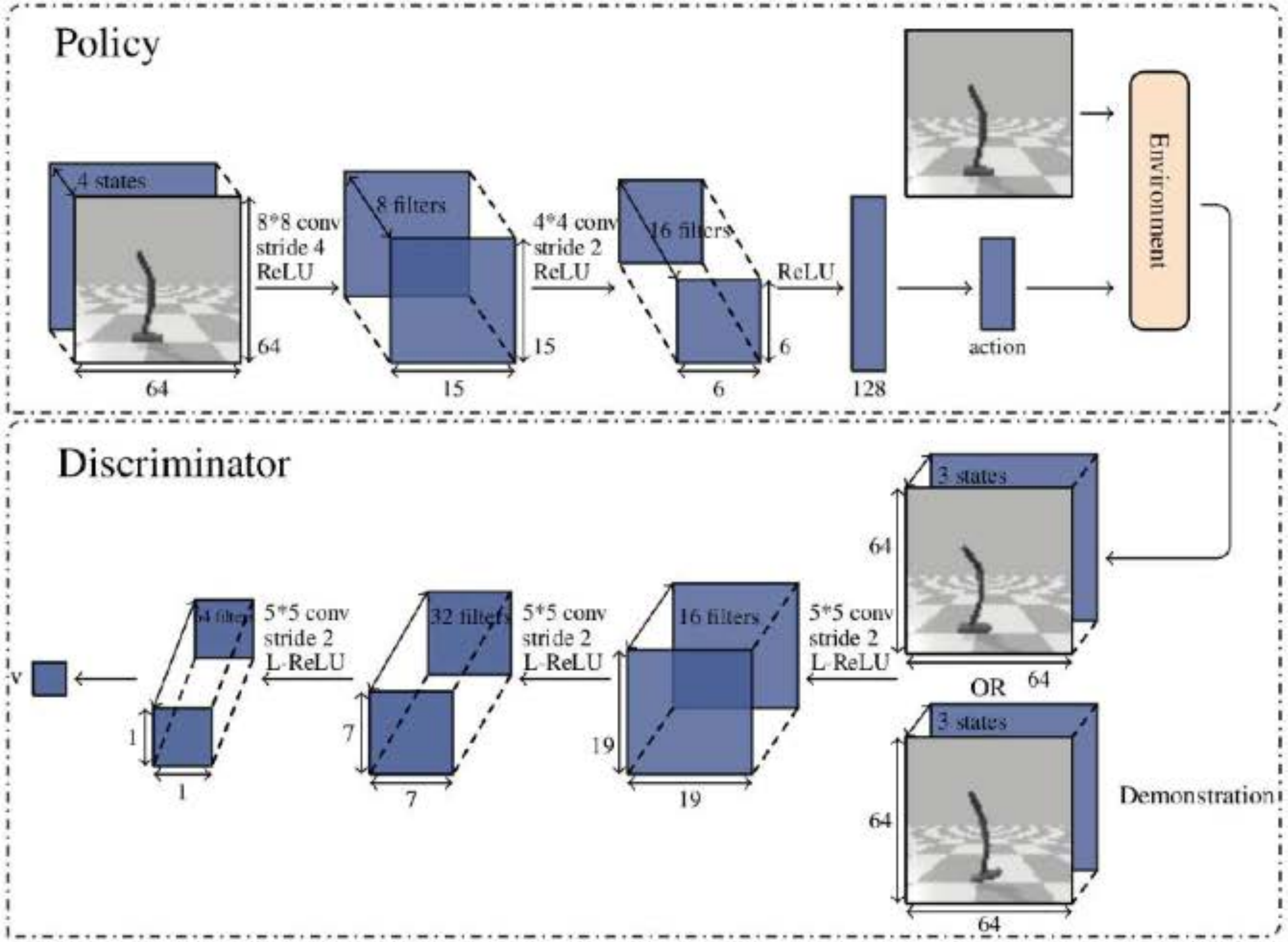
Challenges:

- States are not fully-observable.
- States are not Markovian.

Solution:

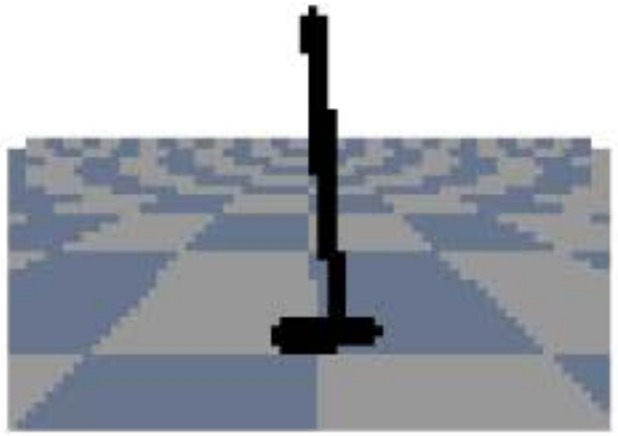
Gen. Adversarial Imitation from Observation (GAIfO)

Algorithm:



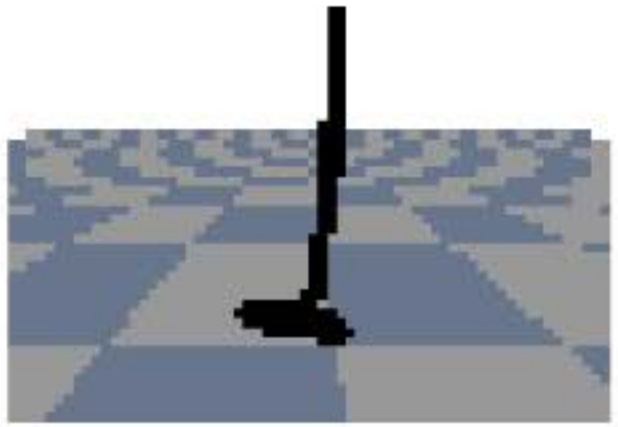
Gen. Adversarial Imitation from Observation (GAIfO)

Demonstration:



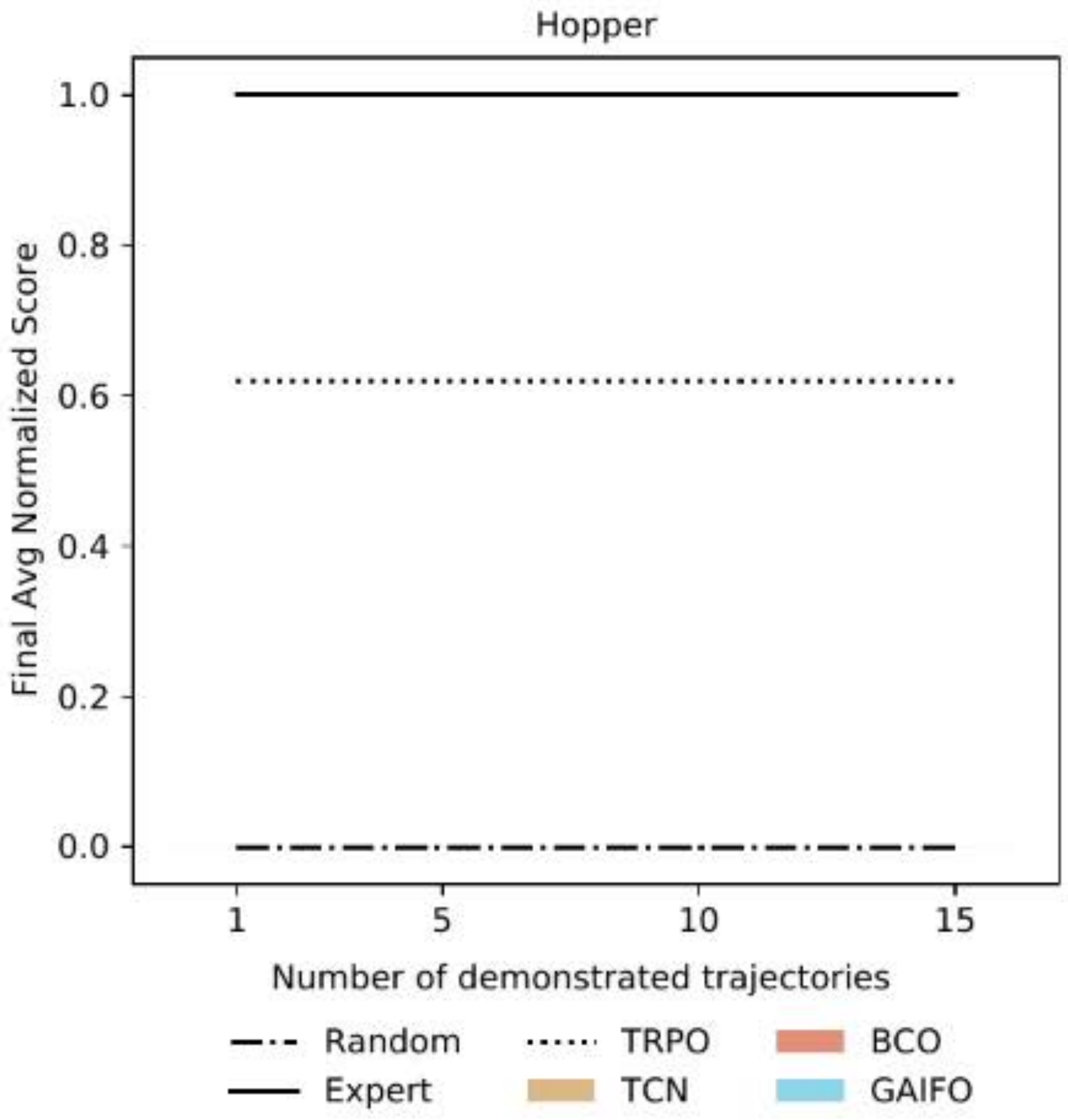
Gen. Adversarial Imitation from Observation (GAIfO)

Learned Policy:



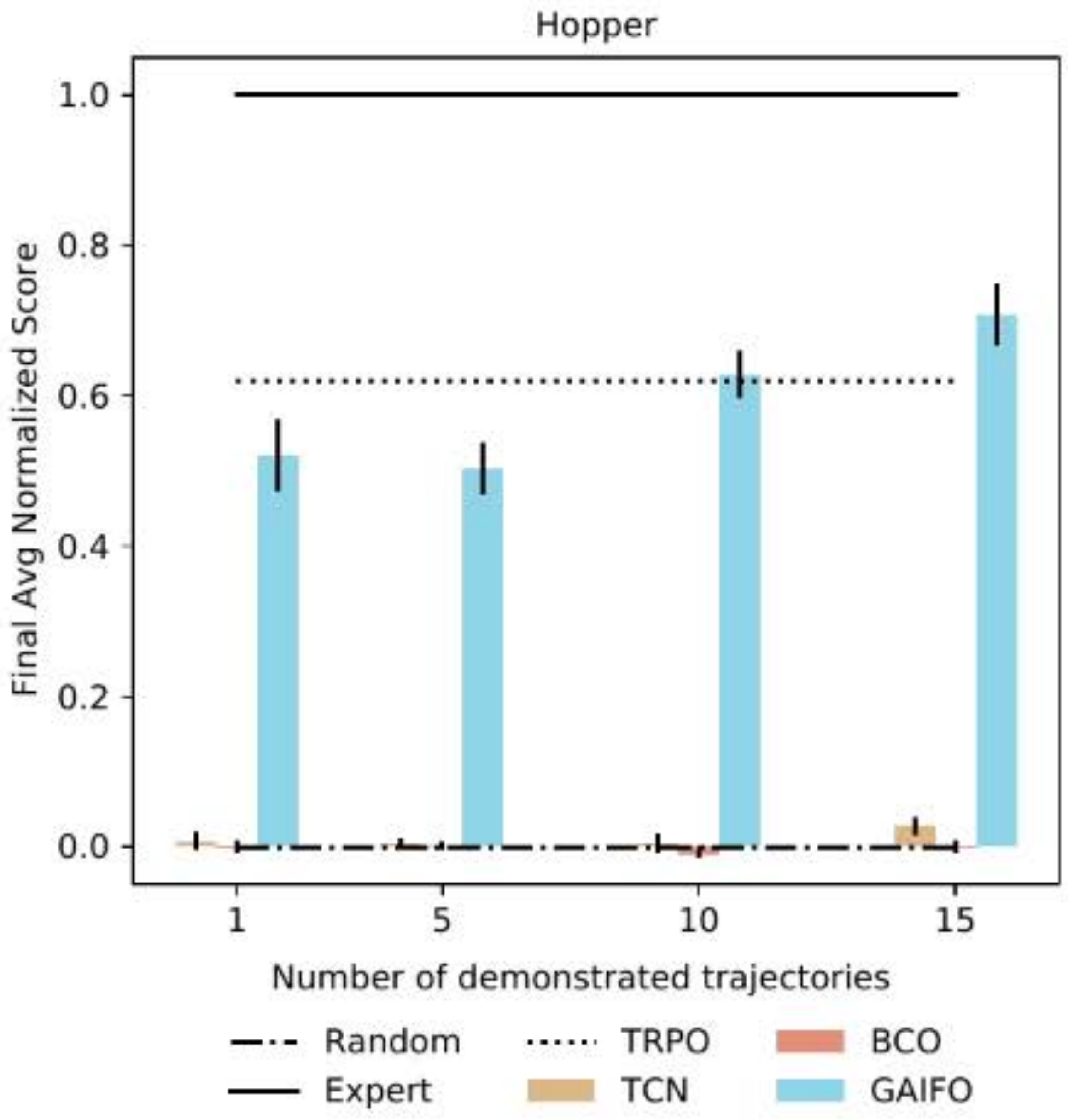
Gen. Adversarial Imitation from Observation (GAIfO)

Comparison against other IfO approaches:



Gen. Adversarial Imitation from Observation (GAIfO)

Comparison against other IfO approaches:



Ongoing Work

Ongoing Work

- Testing algorithms on more domains.

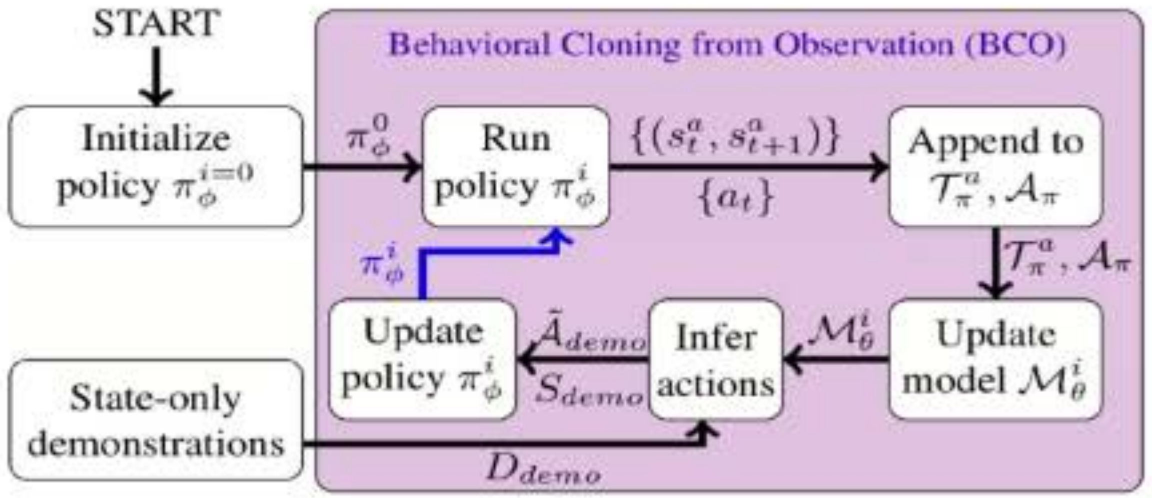
Ongoing Work

- Testing algorithms on more domains.
- Adapt algorithms for physical robots.

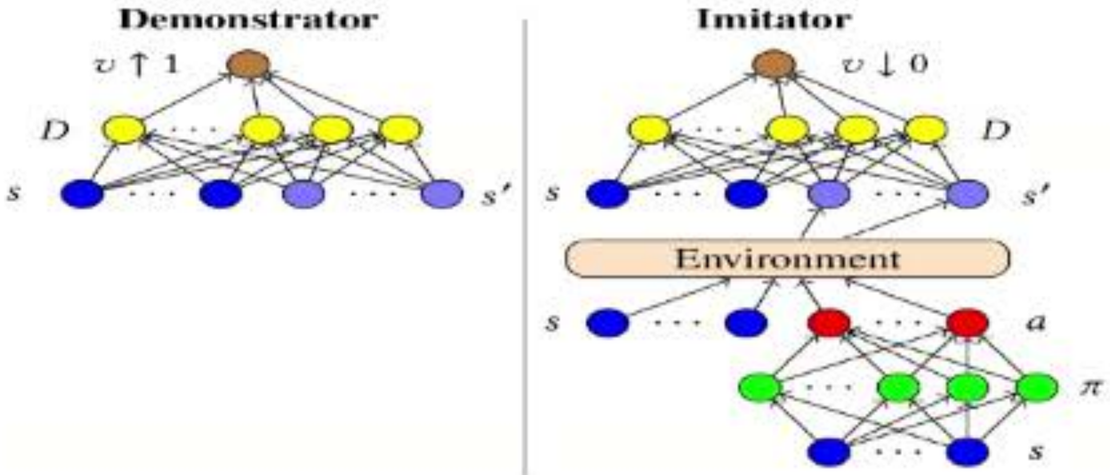
Ongoing Work

- Testing algorithms on more domains.
- Adapt algorithms for physical robots.
- Sim-to-real transfer using the algorithms.

Imitation Learning Summary



(a) BCO



(b) GAIfO

Faraz Torabi

Garrett Warnell

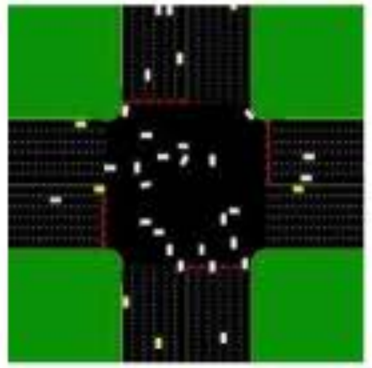
Research Question

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?

Research Areas

- Autonomous agents
- Multiagent systems
- Machine learning
  - **Reinforcement learning**
- **Robotics**



Selected other RL Contributions

- Human interaction
  - Advice, Demonstration
  - Positive/Negative Feedback

[Knox & Stone, '09]

- Transfer learning for RL [Taylor & Stone, '07]
- Curriculum Learning [Narvekar et al., '16]
- RL for musical playlist recommendation [Liebman et al., '15]
- TEXPLORE for Robot RL [Hester & Stone, '13]
  - Sample efficient; real-time
  - Continuous state; delayed effects
- Deep RL in continuous action spaces [Hausknecht & Stone, '16]

Selected MAS Contributions

- Autonomous traffic management
- Trading Agent Competition (PowerTAC)
- Ad Hoc Teamwork

Ad Hoc Teams

- Ad hoc team player is an individual
  - Unknown teammates (programmed by others)
- Teammates likely sub-optimal: no control

Challenge: Create a good team player

- Introduced as AAAI Challenge Problem [AAAI'10]
  - Theory: repeated games, bandits [AIJ'13]
  - Experiments: **pursuit**, **flocking** [Genter & Stone, '12]
  - **RoboCup experiments** [Genter et al., '15]

Benchmarking Robot Cooperation without Pre-Coordination in the RoboCup Standard Platform League Drop-In Player Competition

Katie Genter*, Tim Laue°, Peter Stone*

* University of Texas at Austin, Austin, TX, USA

° University of Bremen, Germany

Ad Hoc Teams

- Ad hoc team player is an individual
  - Unknown teammates (programmed by others)
- Teammates likely sub-optimal: no control

Challenge: Create a good team player

- Introduced as AAAI Challenge Problem [AAAI'10]
  - Theory: repeated games, bandits [AIJ'13]
  - Experiments: **pursuit**, **flocking** [Genter & Stone, '12]
  - **RoboCup experiments** [Genter et al., '15]

Ad Hoc Teams

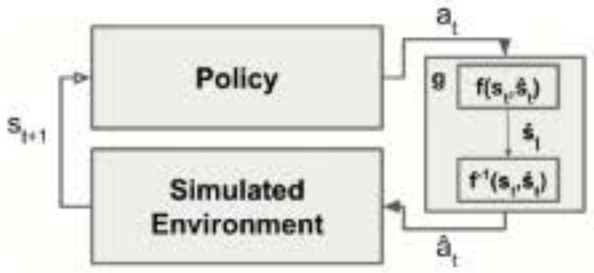
- Ad hoc team player is an individual
  - Unknown teammates (programmed by others)
- Teammates likely sub-optimal: no control

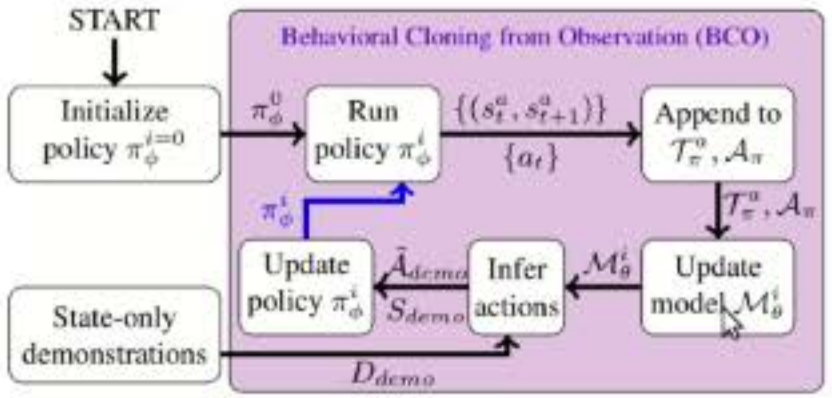
Challenge: Create a good team player

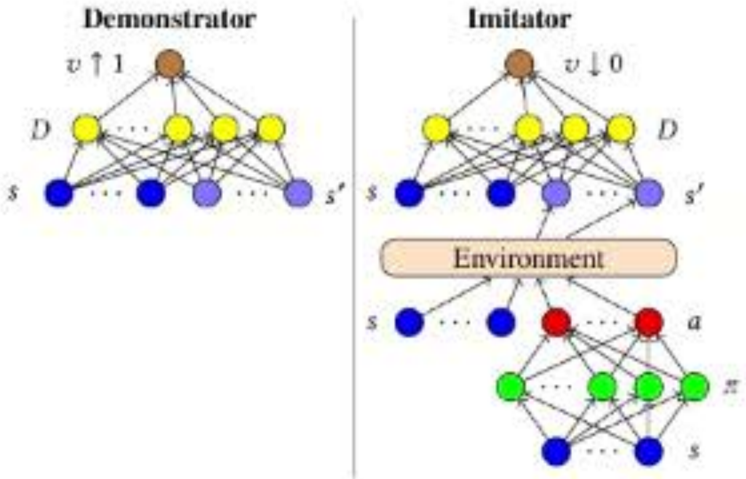
- Introduced as AAAI Challenge Problem [AAAI'10]
  - Theory: repeated games, bandits [AIJ'13]
  - Experiments: **pursuit**, **flocking** [Genter & Stone, '12]
  - **RoboCup experiments** [Genter et al., '15]
  - Community: MIPC Workshops, JAAMAS issue

Efficient Robot Skill Learning: GSL and IfO

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?







- **Motivation:** RoboCup
- **Sim2Real:** Grounded Simulation Learning
- **Imitation Learning from Observation:** BCO and GAIfO

Selected other RL Contributions

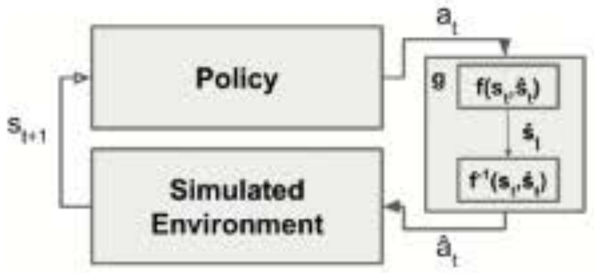
- Human interaction
  - Advice, Demonstration
  - Positive/Negative Feedback

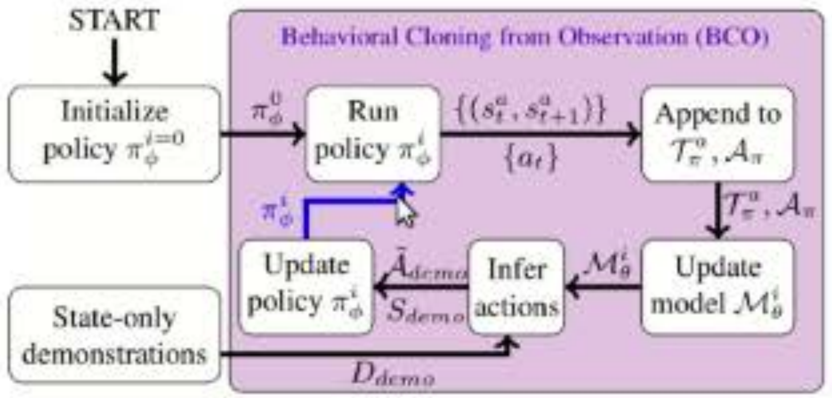
[Knox & Stone, '09]

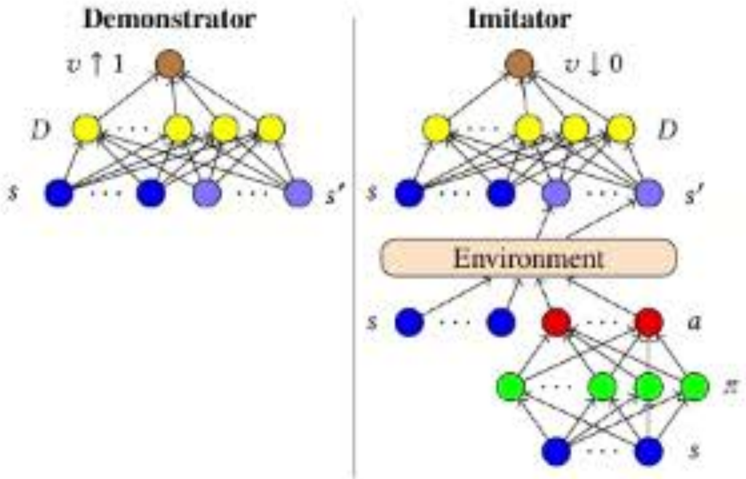
- Transfer learning for RL [Taylor & Stone, '07]
- Curriculum Learning [Narvekar et al., '16]
- RL for musical playlist recommendation [Liebman et al., '15]
- TEXPLORE for Robot RL [Hester & Stone, '13]
  - Sample efficient; real-time
  - Continuous state; delayed effects
- Deep RL in continuous action spaces [Hausknecht & Stone, '16]

Efficient Robot Skill Learning: GSL and IfO

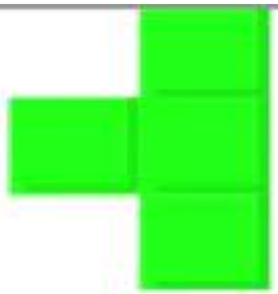
To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?







- **Motivation:** RoboCup
- **Sim2Real:** Grounded Simulation Learning
- **Imitation Learning from Observation:** BCO and GAIfO



Selected other RL Contributions

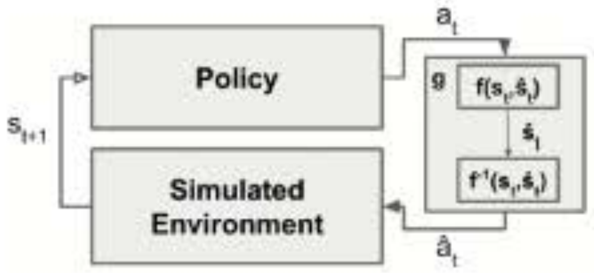
- Human interaction
  - Advice, Demonstration
  - Positive/Negative Feedback

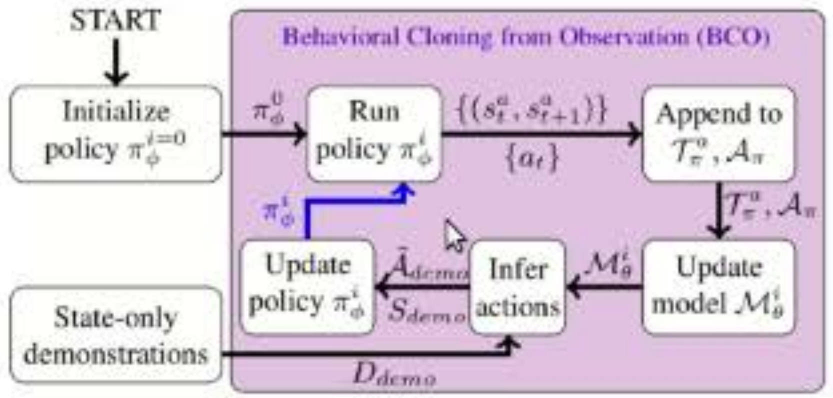
[Knox & Stone, '09]

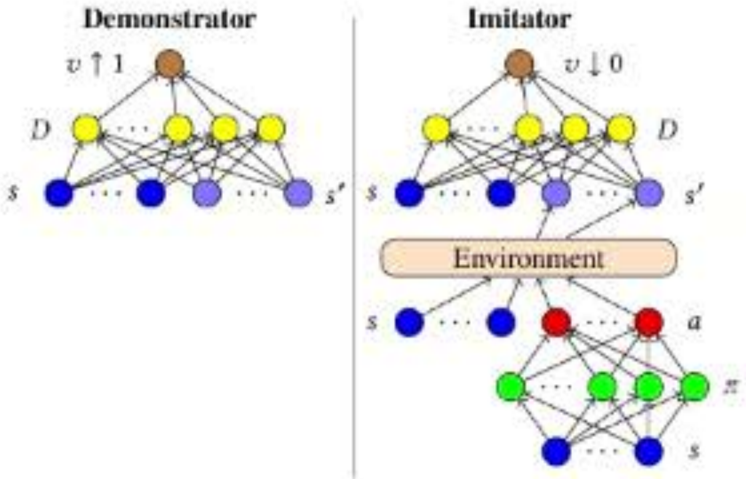
- Transfer learning for RL [Taylor & Stone, '07]
- Curriculum Learning [Narvekar et al., '16]
- RL for musical playlist recommendation [Liebman et al., '15]
- TEXPLORE for Robot RL [Hester & Stone, '13]
  - Sample efficient; real-time
  - Continuous state; delayed effects

Efficient Robot Skill Learning: GSL and IfO

To what degree can autonomous intelligent agents learn in the presence of teammates and/or adversaries in real-time, dynamic domains?







- **Motivation:** RoboCup
- **Sim2Real:** Grounded Simulation Learning
- **Imitation Learning from Observation:** BCO and GAIfO

Ad Hoc Teams

- Ad hoc team player is an individual
  - Unknown teammates (programmed by others)
- Teammates likely sub-optimal: no control

Challenge: Create a good team player

- Introduced as AAAI Challenge Problem [AAAI'10]
  - Theory: repeated games, bandits [AIJ'13]
  - Experiments: **pursuit**, **flocking** [Genter & Stone, '12]
  - **RoboCup experiments** [Genter et al., '15]
  - Community: MIPC Workshops, JAAMAS issue