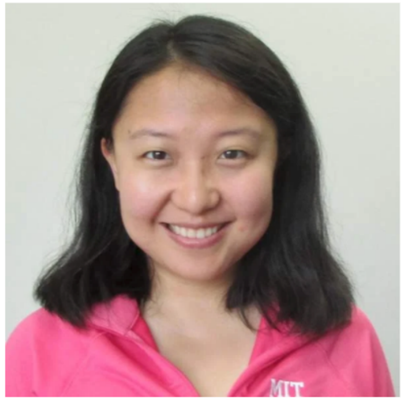
Chain of Thought Imitation with Procedure Cloning

Sherry Yang



Dale Schuurmans

Pieter Abeel

Ofir Nachum

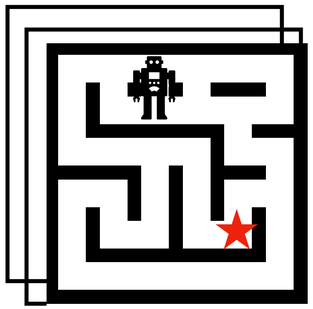
Paper: arxiv.org/abs/2205.10816?
Code: github.com/google-research/google-research/tree/master/procedure_cloning
Website: sites.google.com/corp/view/procedure-cloning

Background: Imitation Learning

- Datasets: expert state-action pairs

Datasets

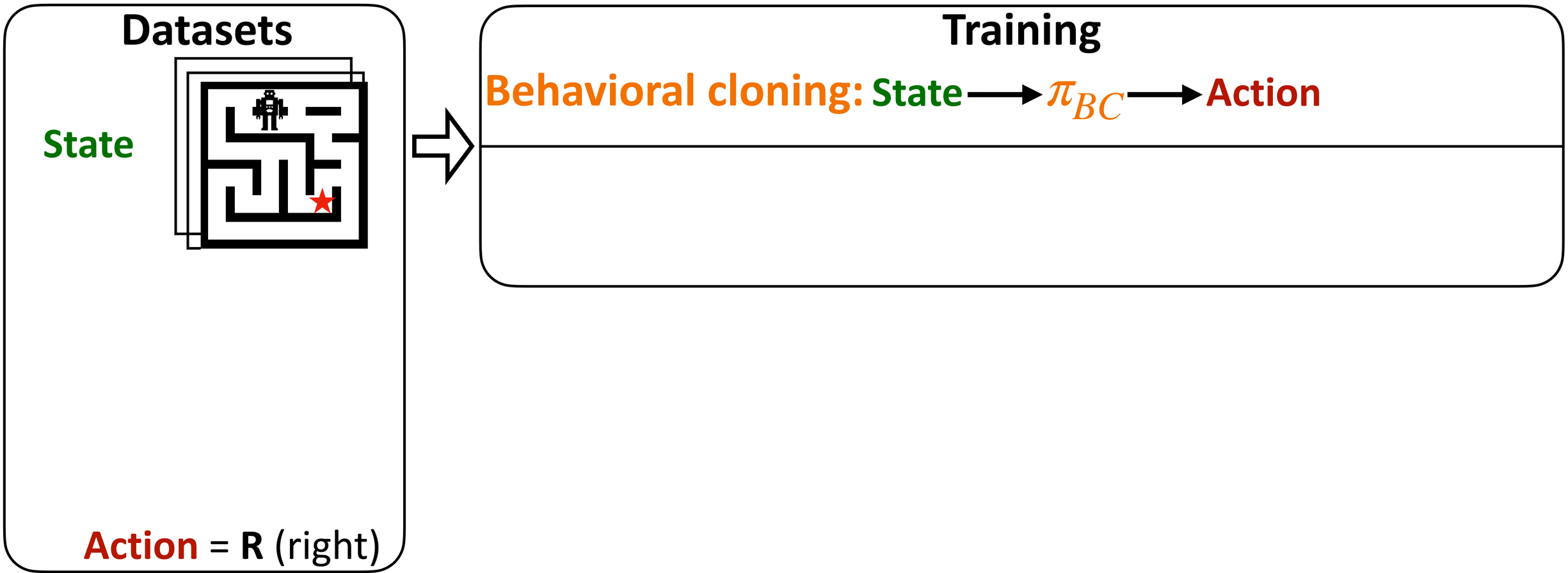
State



Action = **R** (right)

Background: Imitation Learning

- Datasets: expert state-action pairs
- Behavioral cloning: learns mapping from state to action

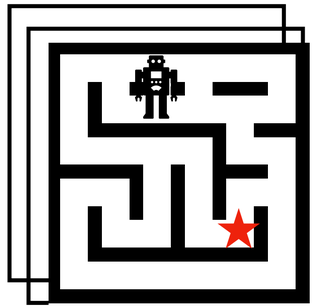


Background: Imitation Learning

- Datasets: expert state-action pairs
- Behavioral cloning: learns mapping from state to action

Datasets

State



Action = **R** (right)

Training

Behavioral cloning: **State** → π_{BC} → **Action**

Inference

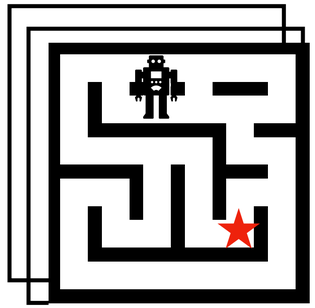
State

Background: Imitation Learning

- Datasets: expert state-action pairs
- Behavioral cloning: learns mapping from state to action

Datasets

State



Action = **R** (right)

Training

Behavioral cloning: **State** → π_{BC} → **Action**

Inference

State

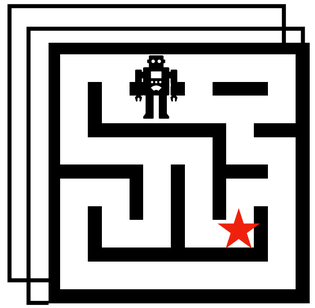
→ π_{BC}

Background: Imitation Learning

- Datasets: expert state-action pairs
- Behavioral cloning: learns mapping from state to action

Datasets

State



Action = **R** (right)

Training

Behavioral cloning: **State** → π_{BC} → **Action**

Inference

State

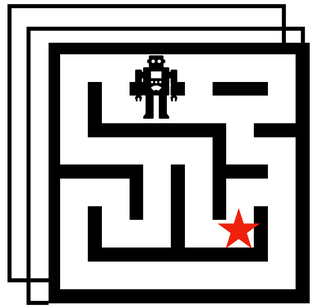
→ π_{BC} ⋯→ ?

Background: Imitation Learning

- Datasets: expert state-action pairs
- Behavioral cloning: learns mapping from state to action
- What's missing?

Datasets

State



Action = **R** (right)

Training

Behavioral cloning: **State** → π_{BC} → **Action**

Inference

State

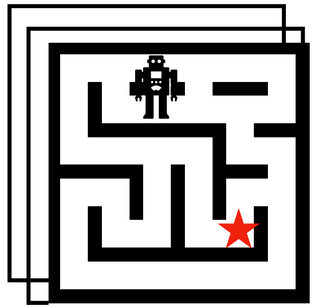
→ π_{BC} ⋯→ ?

Background: Imitation Learning

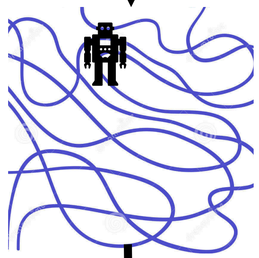
- Expert demos might provide more info!

Datasets

State



Procedure
e.g., search, planning



Action = **R** (right)

Training

Behavioral cloning: **State** → π_{BC} → **Action**

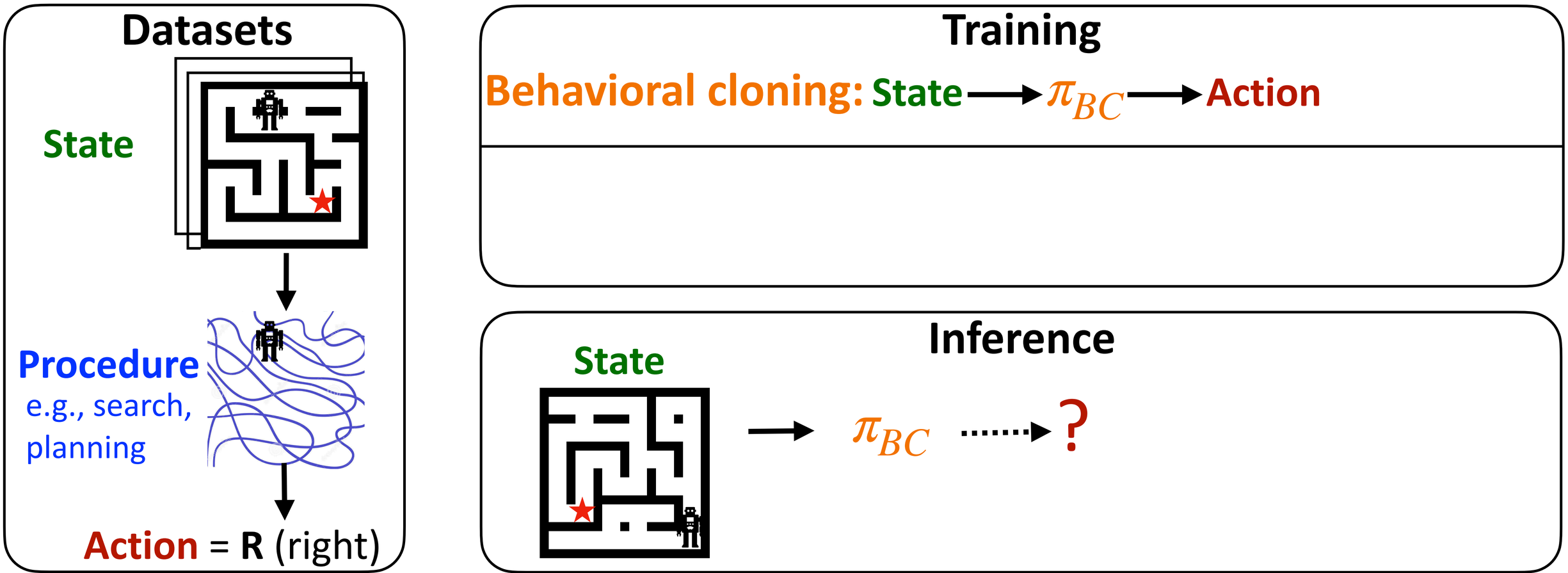
Inference

State

→ π_{BC} ⋯→ ?

Background: Imitation Learning

- Expert demos might provide more info!
  - e.g., planning, search, multi-step algo

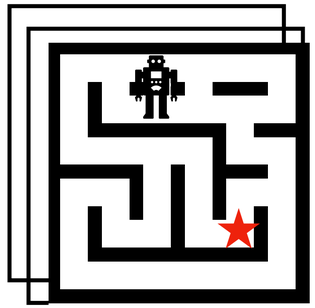


Background: Imitation Learning

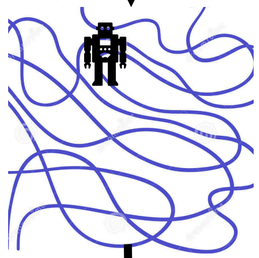
- Expert demos might provide more info!
  - e.g., planning, search, multi-step algo
  - Can't run procedures during inference (simulators, annotations)

Datasets

State



Procedure
e.g., search, planning



Action = **R** (right)

Training

Behavioral cloning: **State** → π_{BC} → **Action**

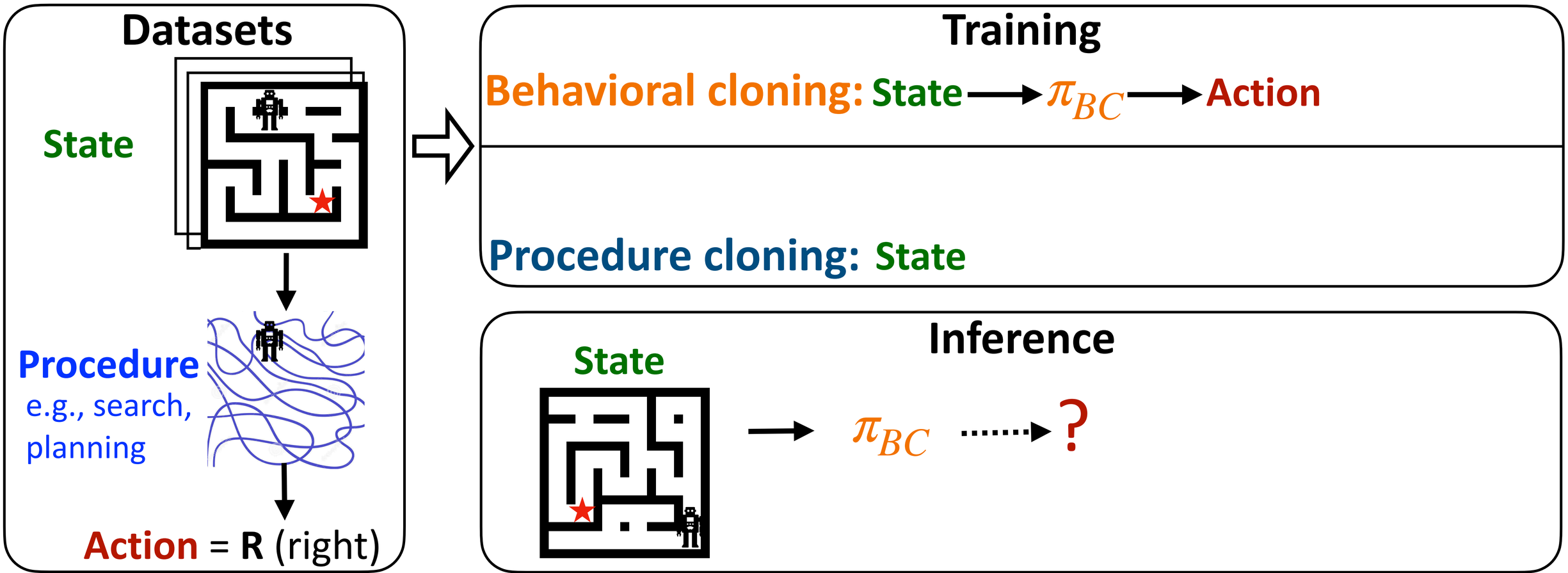
Inference

State

→ π_{BC} ⋯→ ?

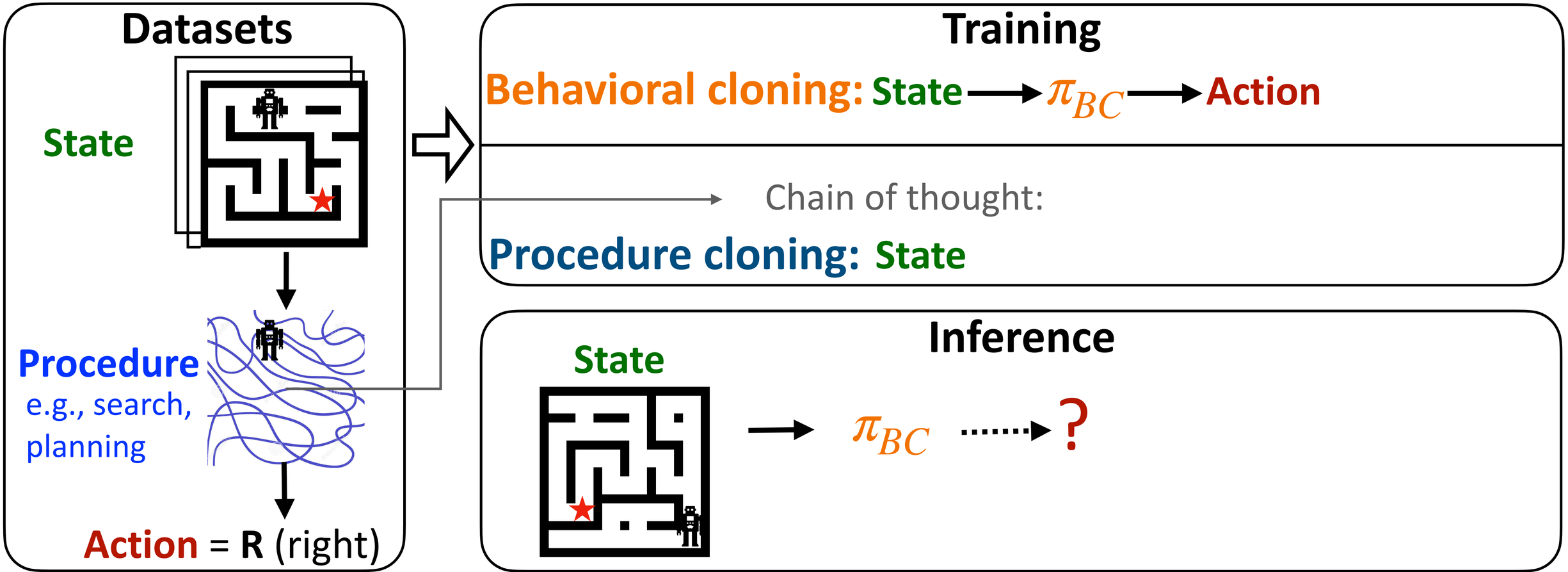
Proposal: Procedure Cloning

- Expert demos might provide more info!
- Imitate the whole expert procedure



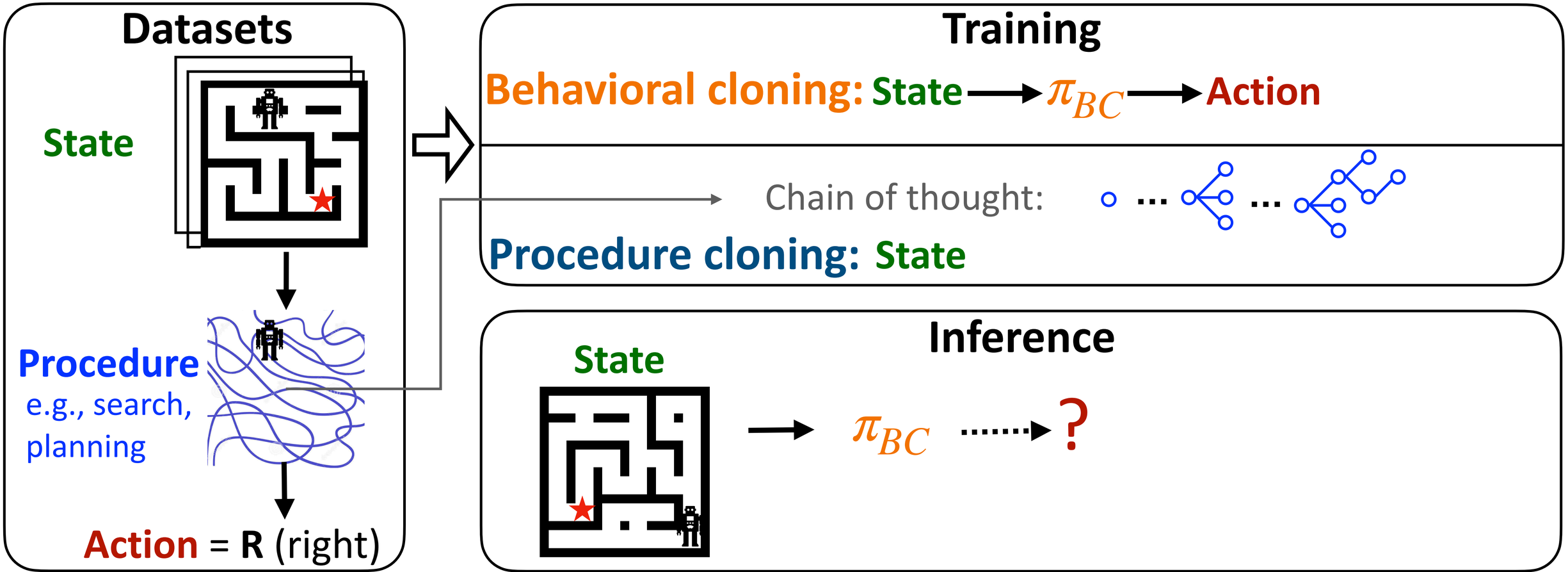
Proposal: Procedure Cloning

- Expert demos might provide more info!
- Imitate the whole expert procedure



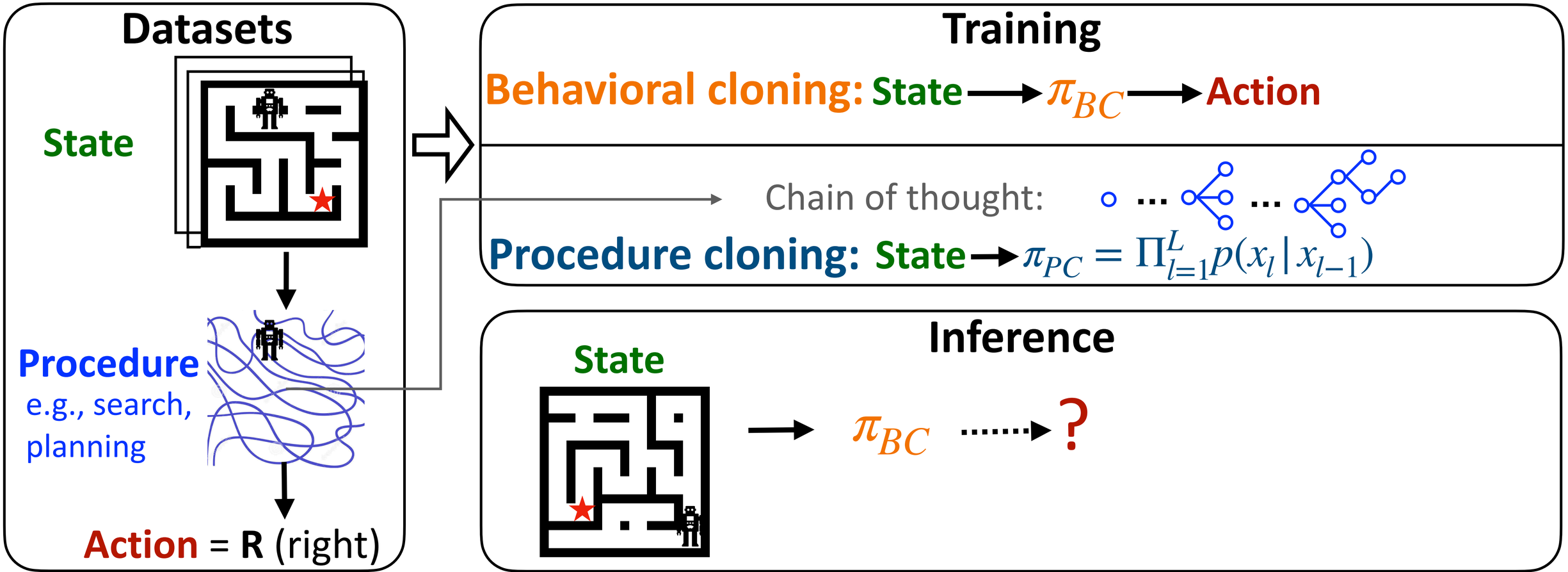
Proposal: Procedure Cloning

- Expert demos might provide more info!
- Imitate the whole expert procedure



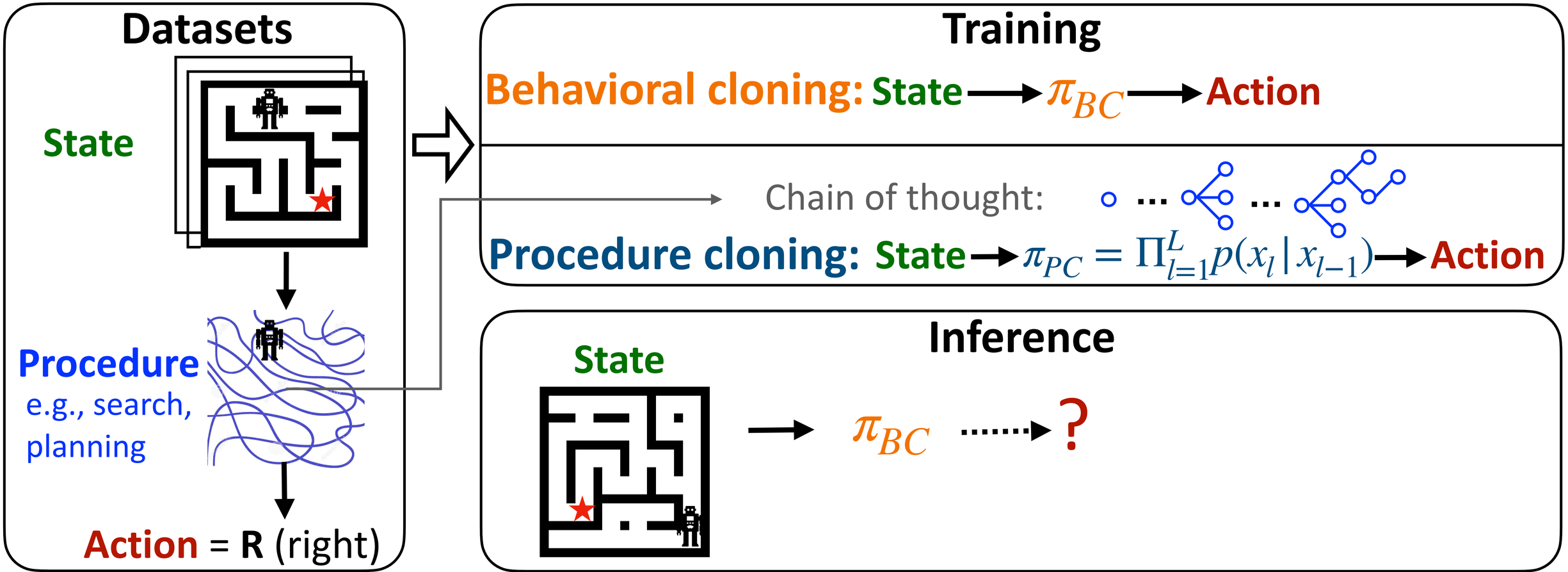
Proposal: Procedure Cloning

- Expert demos might provide more info!
- Imitate the whole expert procedure



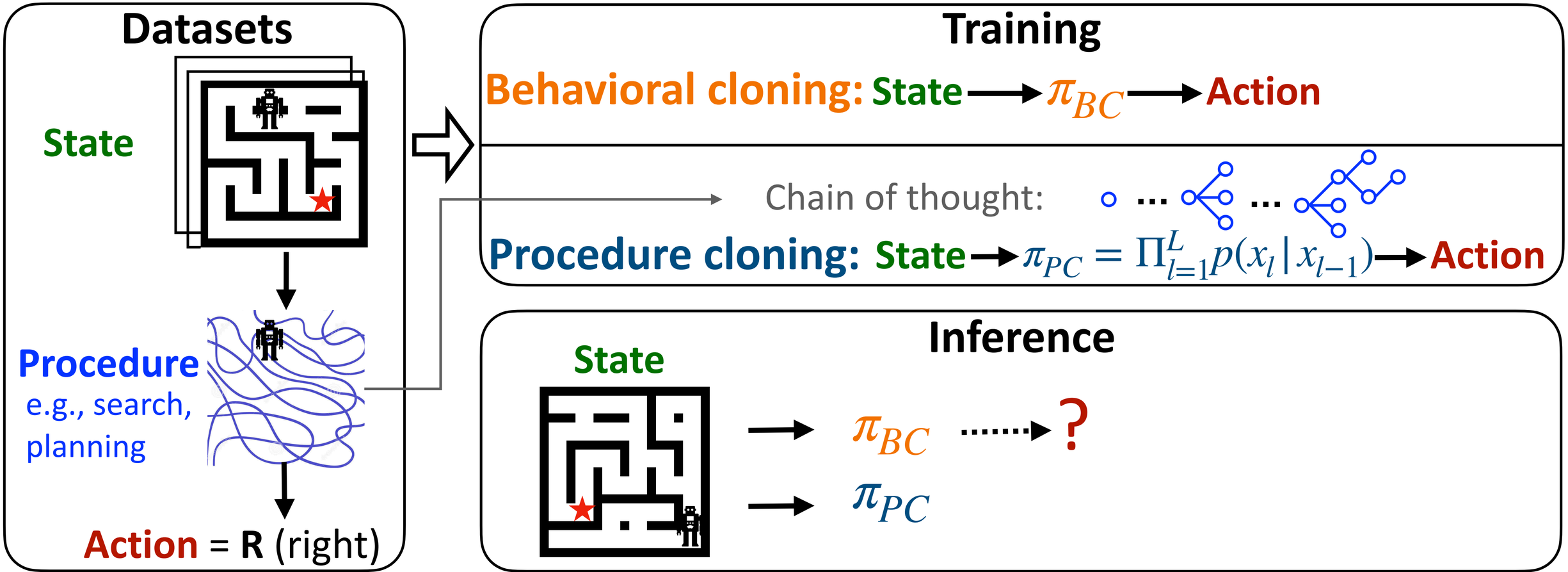
Proposal: Procedure Cloning

- Expert demos might provide more info!
- Imitate the whole expert procedure



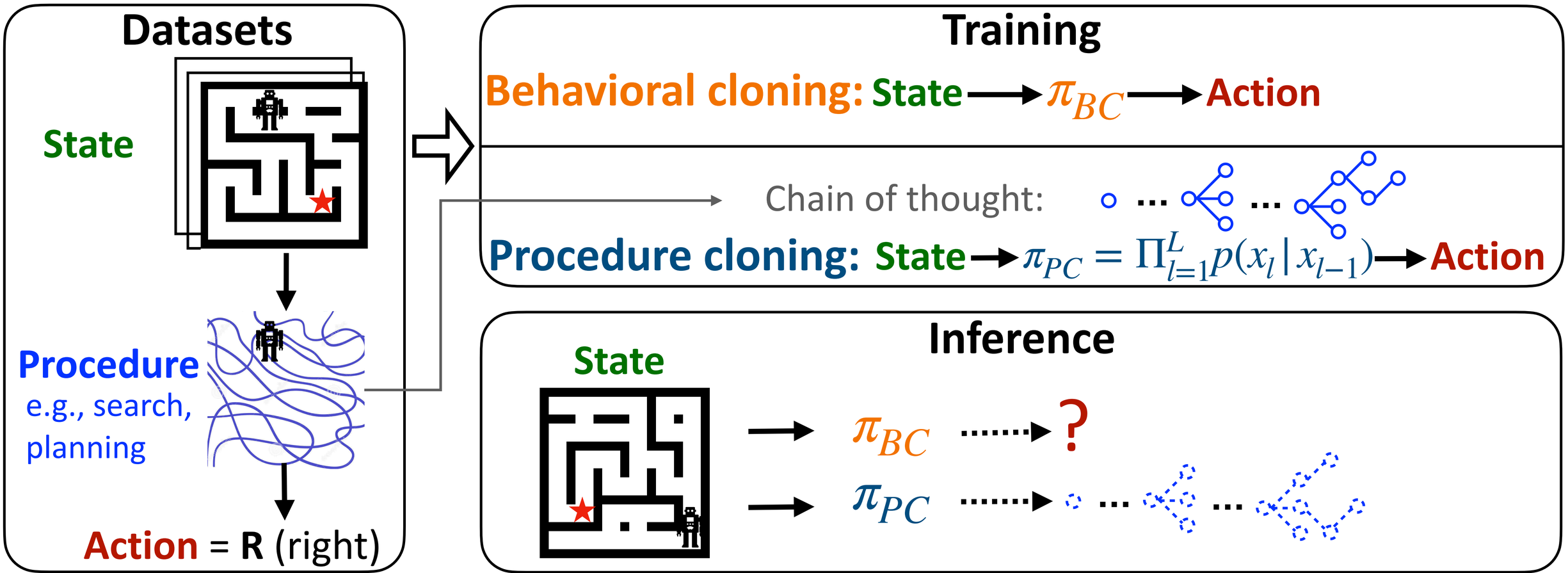
Proposal: Procedure Cloning

- Expert demos might provide more info!
- Imitate the whole expert procedure



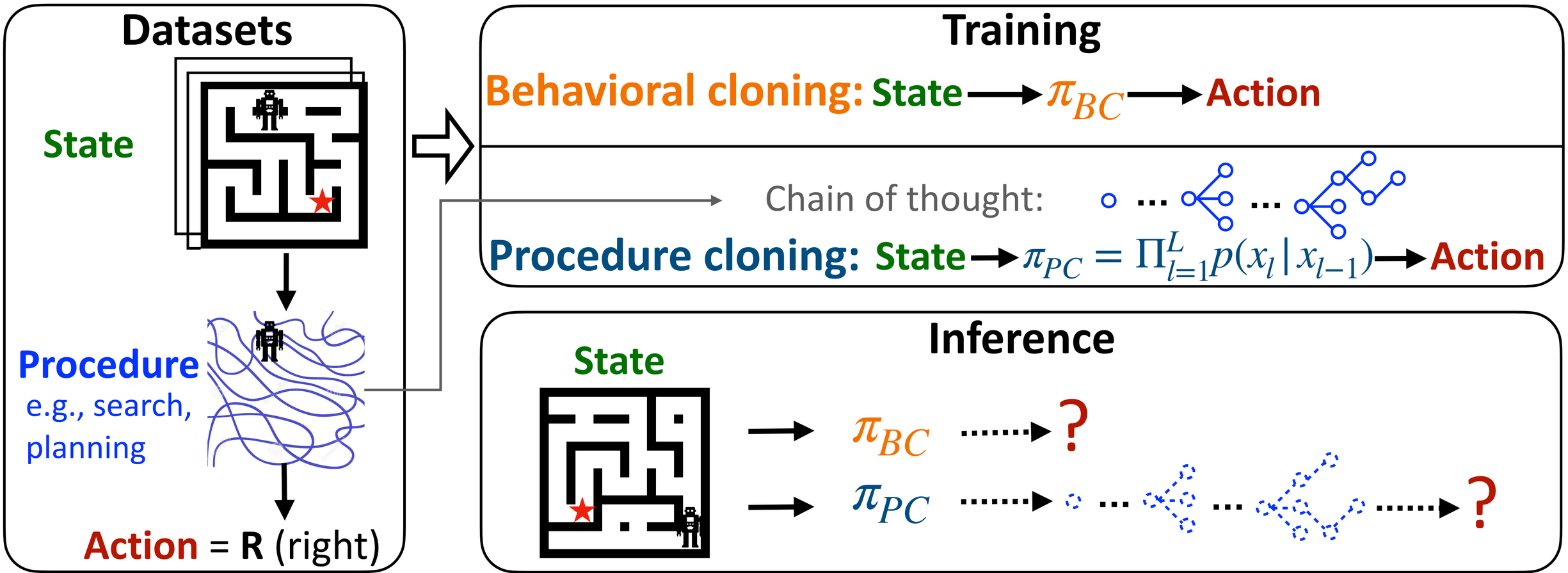
Proposal: Procedure Cloning

- Expert demos might provide more info!
- Imitate the whole expert procedure



Proposal: Procedure Cloning

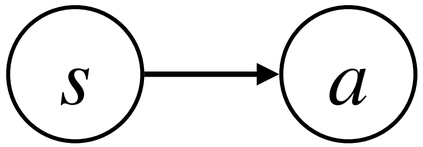
- Expert demos might provide more info!
- Imitate the whole expert procedure



Procedure Cloning

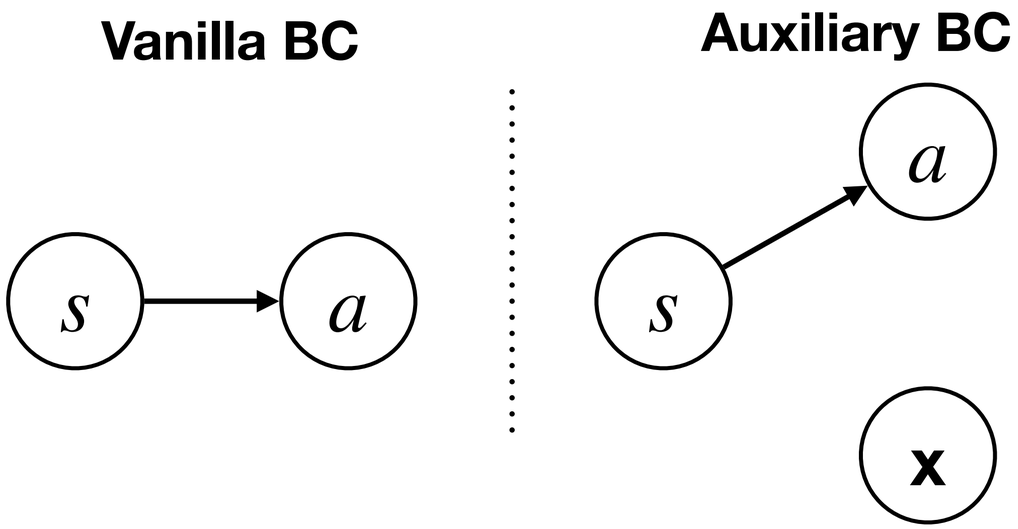
- Graphical models view

Vanilla BC



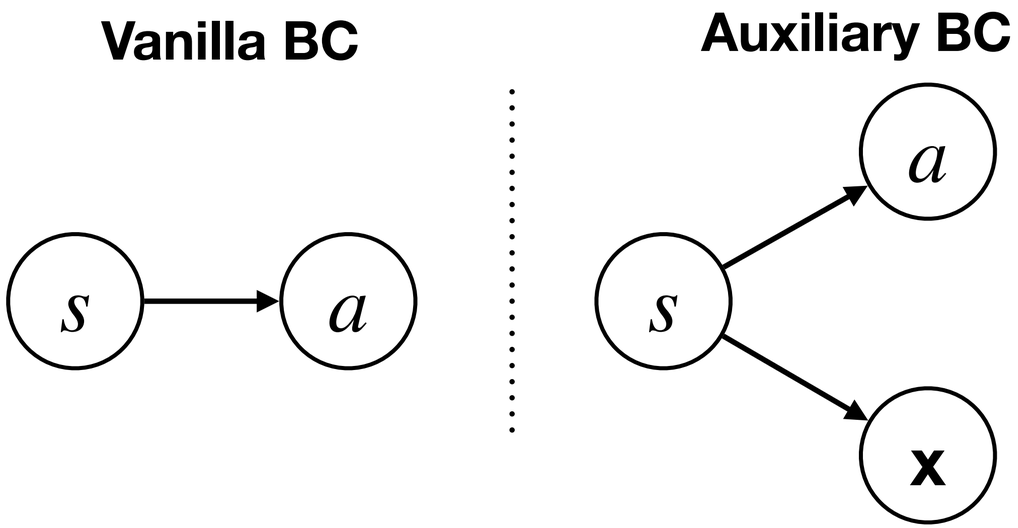
Procedure Cloning

- Graphical models view



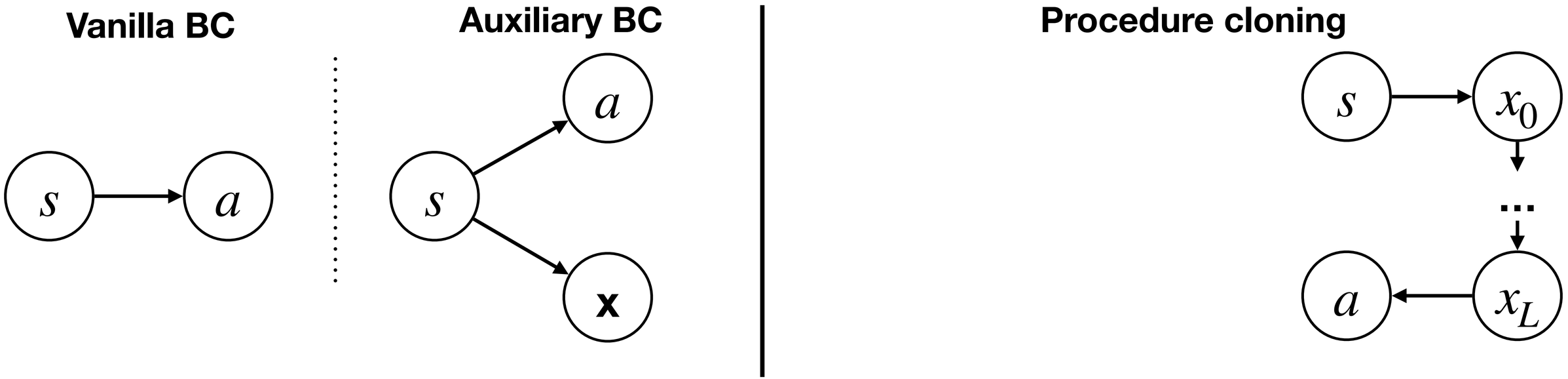
Procedure Cloning

- Graphical models view



Procedure Cloning

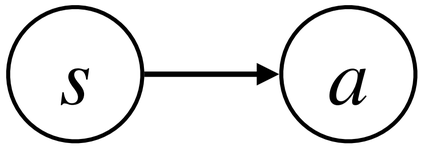
- Graphical models view



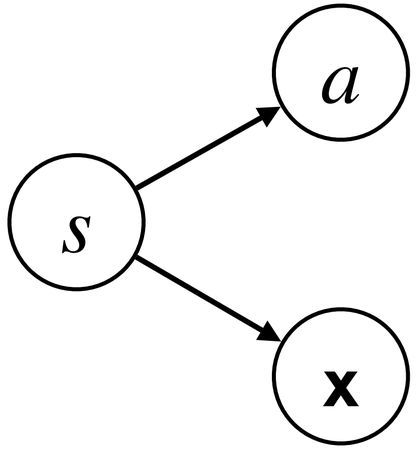
Procedure Cloning

- Graphical models view

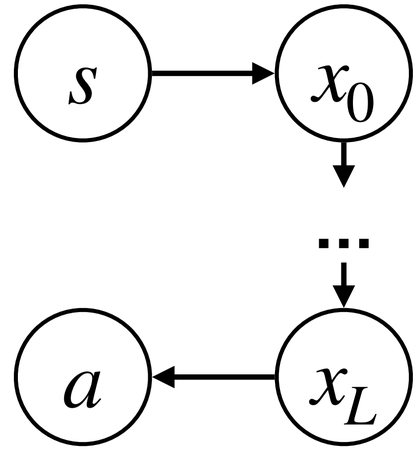
Vanilla BC



Auxiliary BC



Procedure cloning



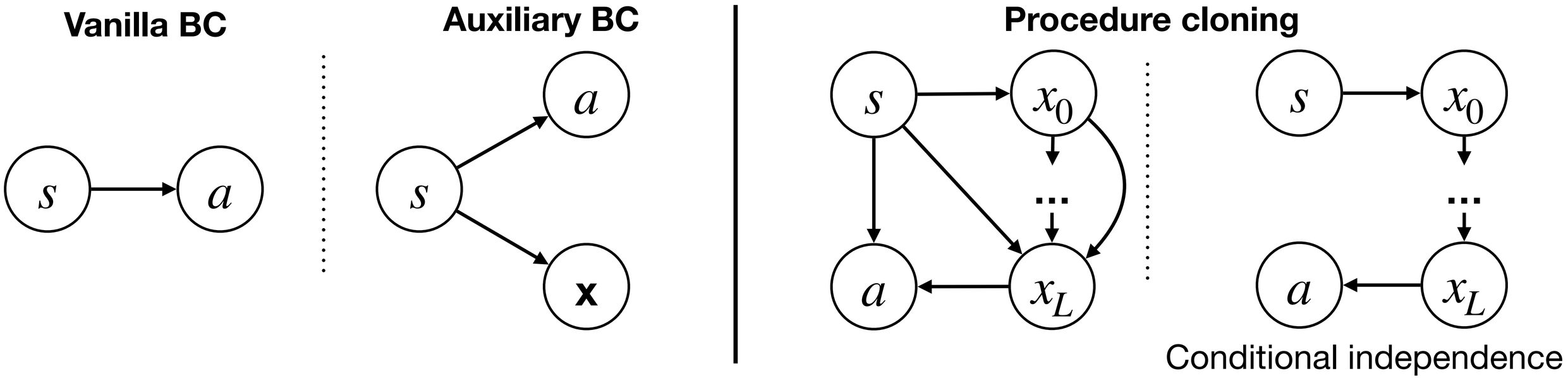
Conditional independence

Conditional independence:

$$p(a, \mathbf{x}|s) = p(a|x_L) \cdot \Pi_{l=1}^{L} p(x_\ell|x_{\ell-1}) \cdot p(x_0|s)$$

Procedure Cloning

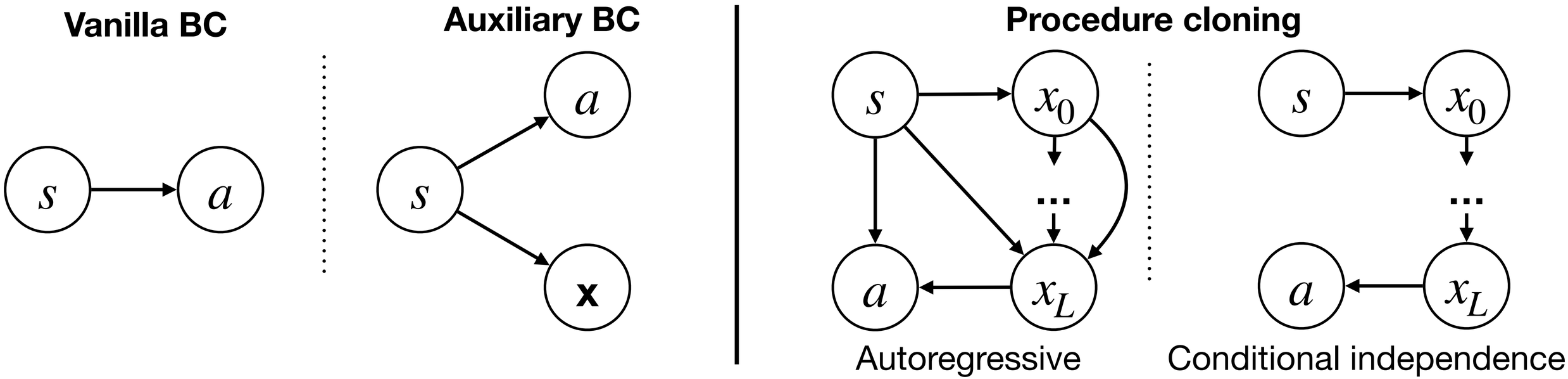
- Graphical models view



Conditional independence: $p(a, \mathbf{x}|s) = p(a|x_L) \cdot \Pi_{l=1}^{L} p(x_\ell|x_{\ell-1}) \cdot p(x_0|s)$

Procedure Cloning

- Graphical models view

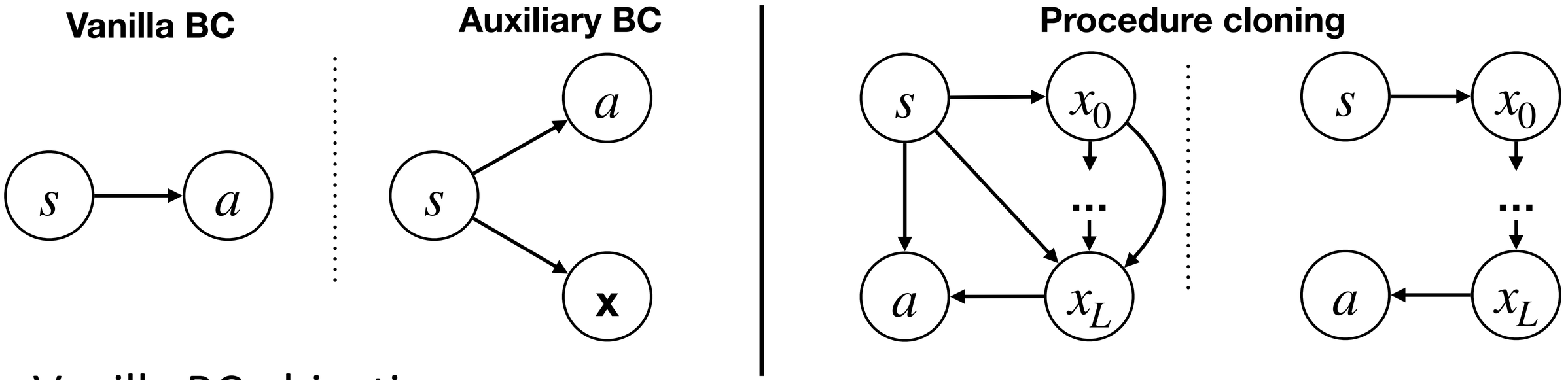


Autoregressive: $p(a, \mathbf{x}|s) = p(a|\mathbf{x}, s) \cdot \Pi_{l=1}^{L} p(x_\ell|\mathbf{x}_{<\ell}, s) \cdot p(x_0|s)$

Conditional independence: $p(a, \mathbf{x}|s) = p(a|x_L) \cdot \Pi_{l=1}^{L} p(x_\ell|x_{\ell-1}) \cdot p(x_0|s)$

Procedure Cloning

- Graphical models view

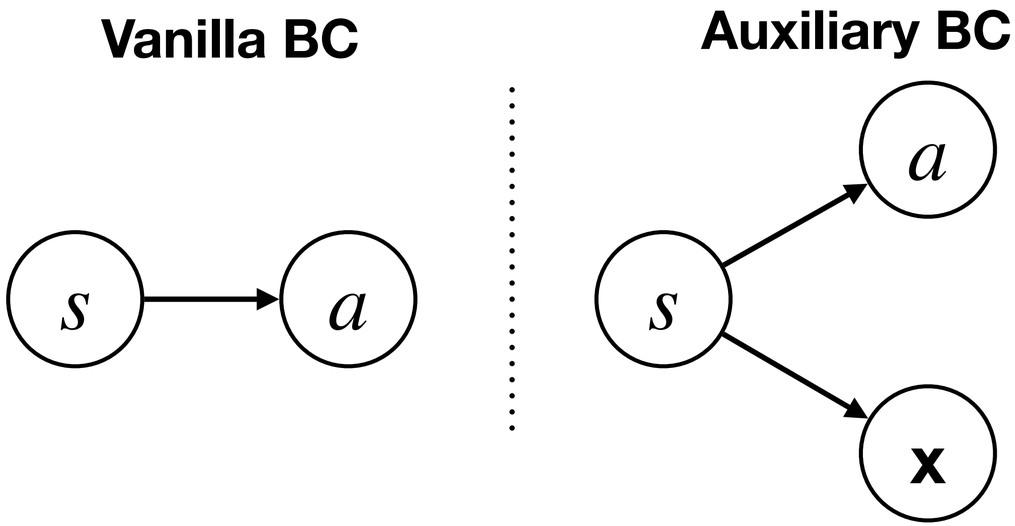


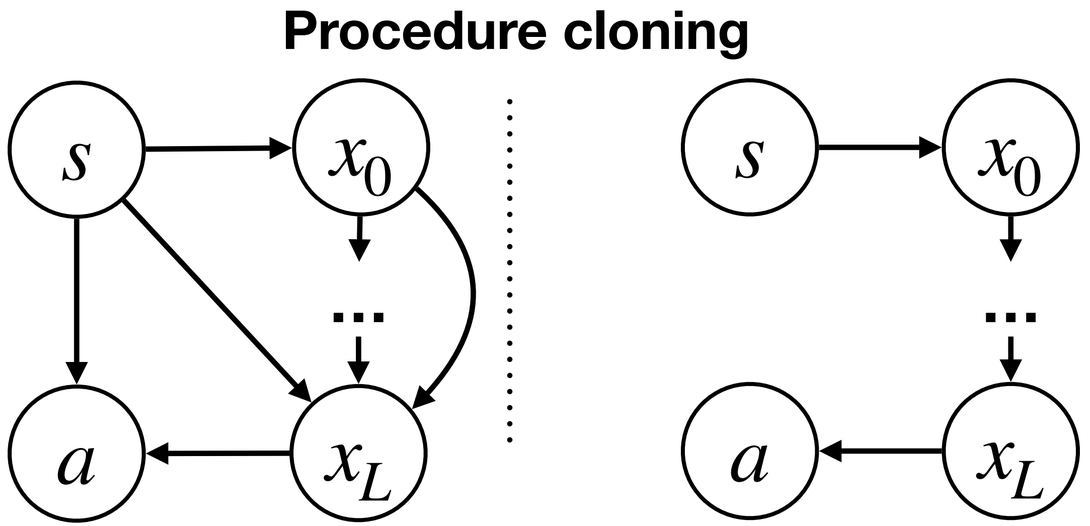
- Vanilla BC objective

$$J_{\mathrm{BC}}(\pi) := \hat{\mathbb{E}}_{(s,a)\ \sim\ \mathcal{D}_*}[-\log \pi(a|s)]$$

Procedure Cloning

- Graphical models view





- Vanilla BC objective

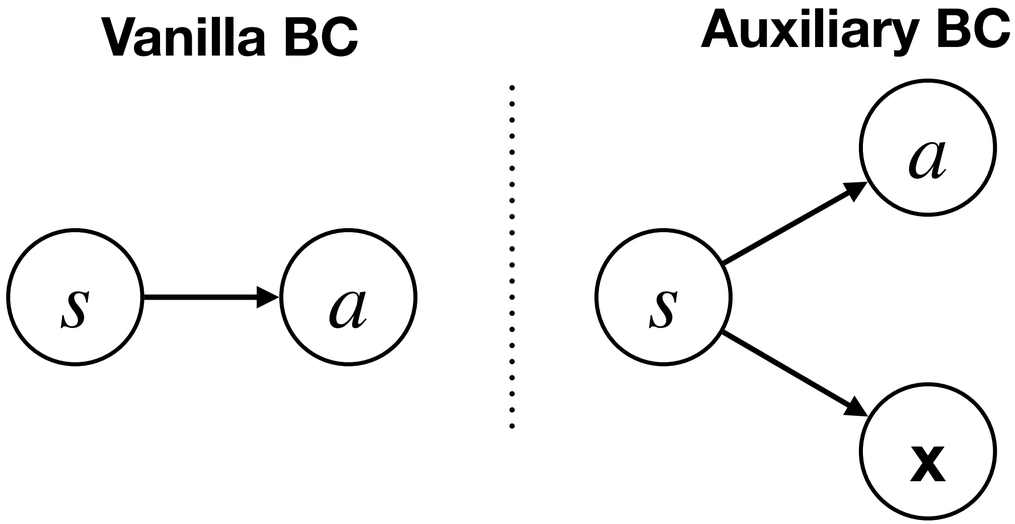
$$J_{\mathrm{BC}}(\pi) := \hat{\mathbb{E}}_{(s,a) \sim \mathcal{D}_*}[-\log \pi(a|s)]$$

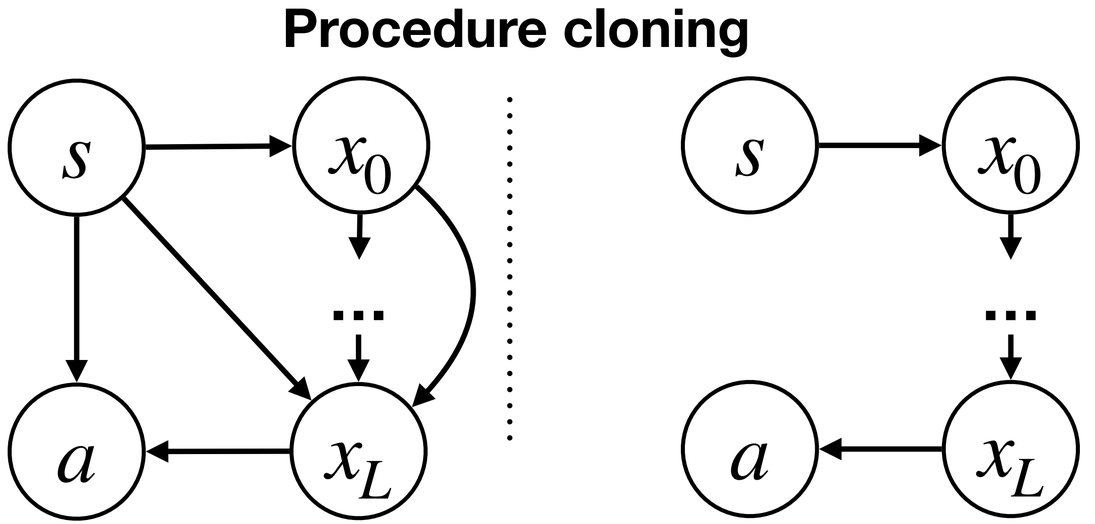
- PC objective

$$\min_{\phi,\theta,\psi} J_{\mathrm{PC}}(\phi,\theta,\psi) = \hat{\mathbb{E}}_{(s,\mathbf{x},a)\sim\mathcal{D}_\Pi}[-\log p(a,\mathbf{x}|s)]$$

Procedure Cloning

- Graphical models view





- Vanilla BC objective

$$J_{\mathrm{BC}}(\pi) := \hat{\mathbb{E}}_{(s,a)\ \sim\ \mathcal{D}_*}[-\log \pi(a|s)]$$

- PC objective

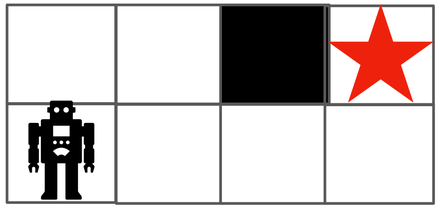
$$\min_{\phi,\theta,\psi} J_{\mathrm{PC}}(\phi,\theta,\psi) = \hat{\mathbb{E}}_{(s,\mathbf{x},a)\sim\mathcal{D}_\Pi}[-\log p(a,\mathbf{x}|s)]$$

$$= \mathbb{E}_{(s,\mathbf{x},a)\sim\mathcal{D}_\Pi}\Big[-\log q_\psi(a|\mathbf{x},s) - \sum_{\ell=1}^{L} \log p_\theta(x_\ell|\mathbf{x}_{<\ell},s) - \log p_\phi(x_0|s)\Big]$$

Procedure Clone BFS

- Proof of concept: synthetic maze navigation

BFS procedure execution



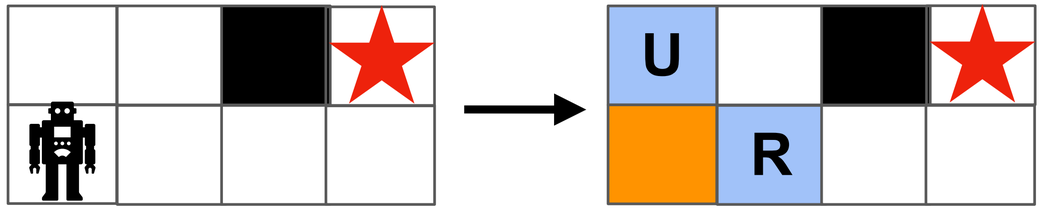
State

Run BFS expand (keep track of actions to each cell).

Procedure Clone BFS

- Proof of concept: synthetic maze navigation

BFS procedure execution



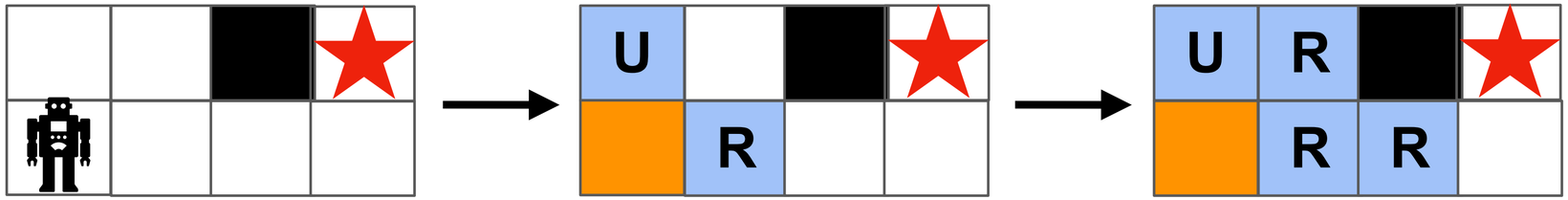
State

Run BFS expand (keep track of actions to each cell).

Procedure Clone BFS

- Proof of concept: synthetic maze navigation

BFS procedure execution



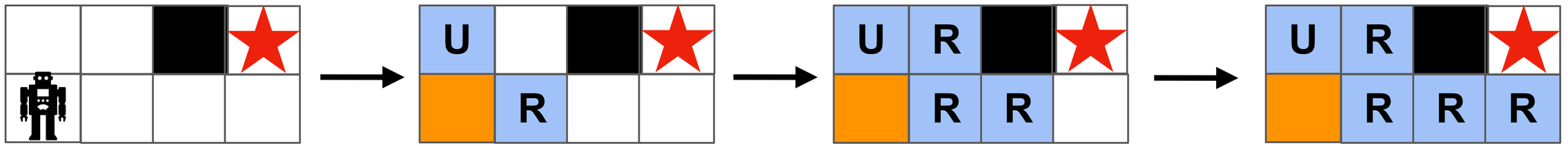
State

Run BFS expand (keep track of actions to each cell).

Procedure Clone BFS

- Proof of concept: synthetic maze navigation

BFS procedure execution



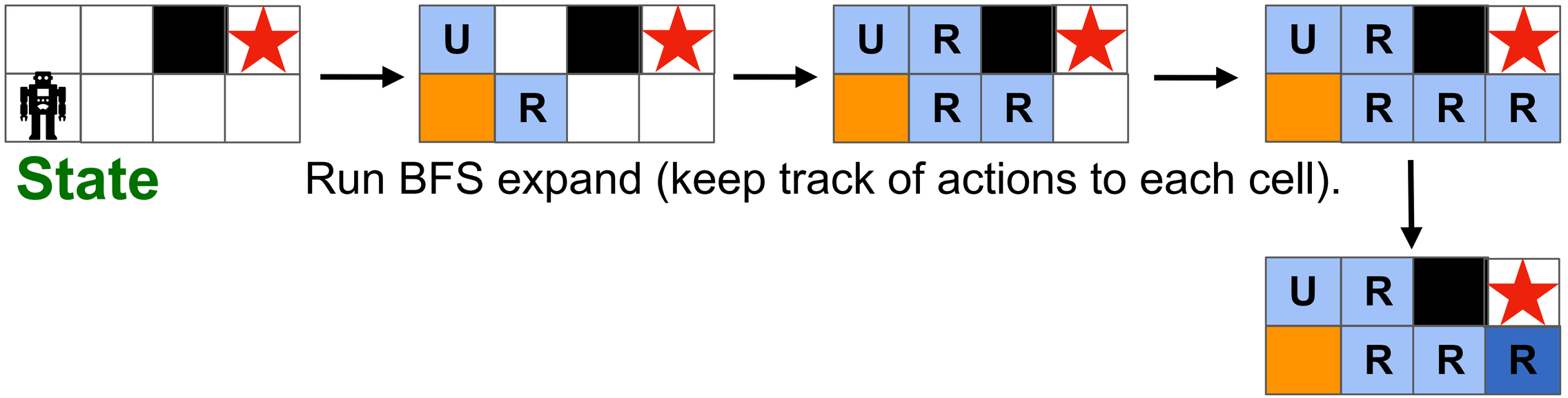
State

Run BFS expand (keep track of actions to each cell).

Procedure Clone BFS

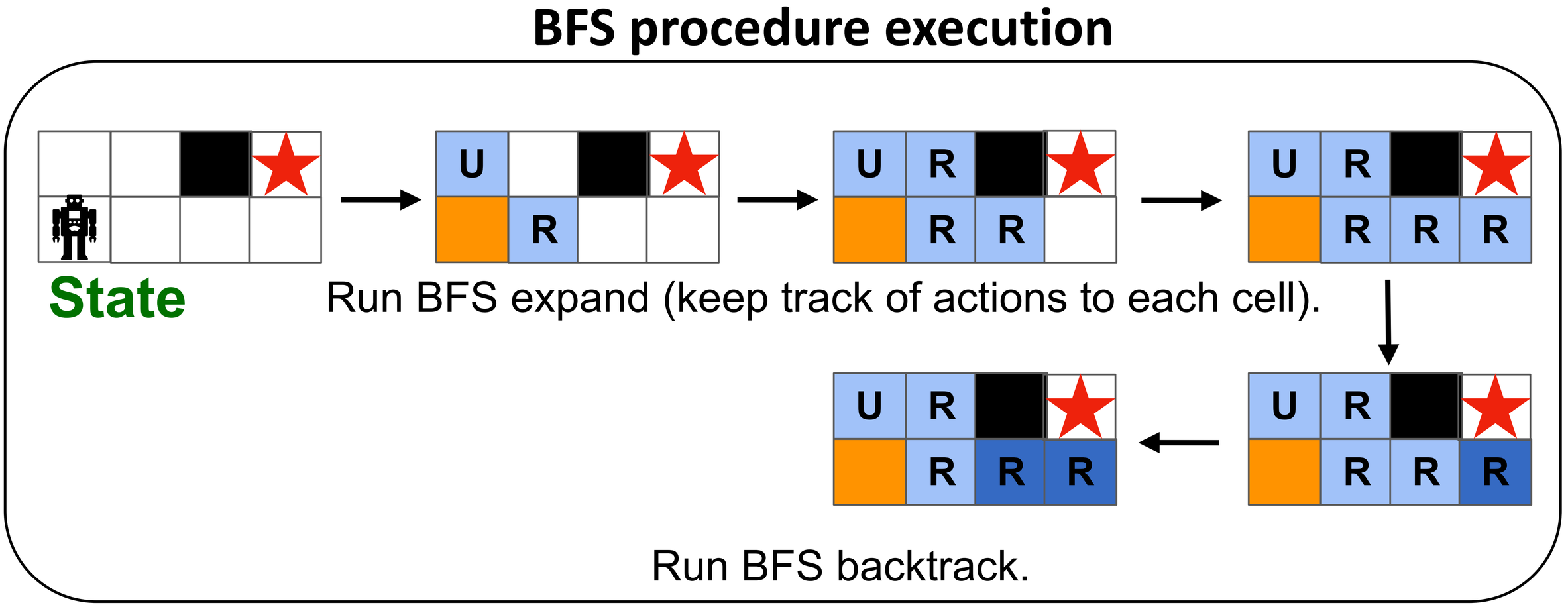
- Proof of concept: synthetic maze navigation

BFS procedure execution



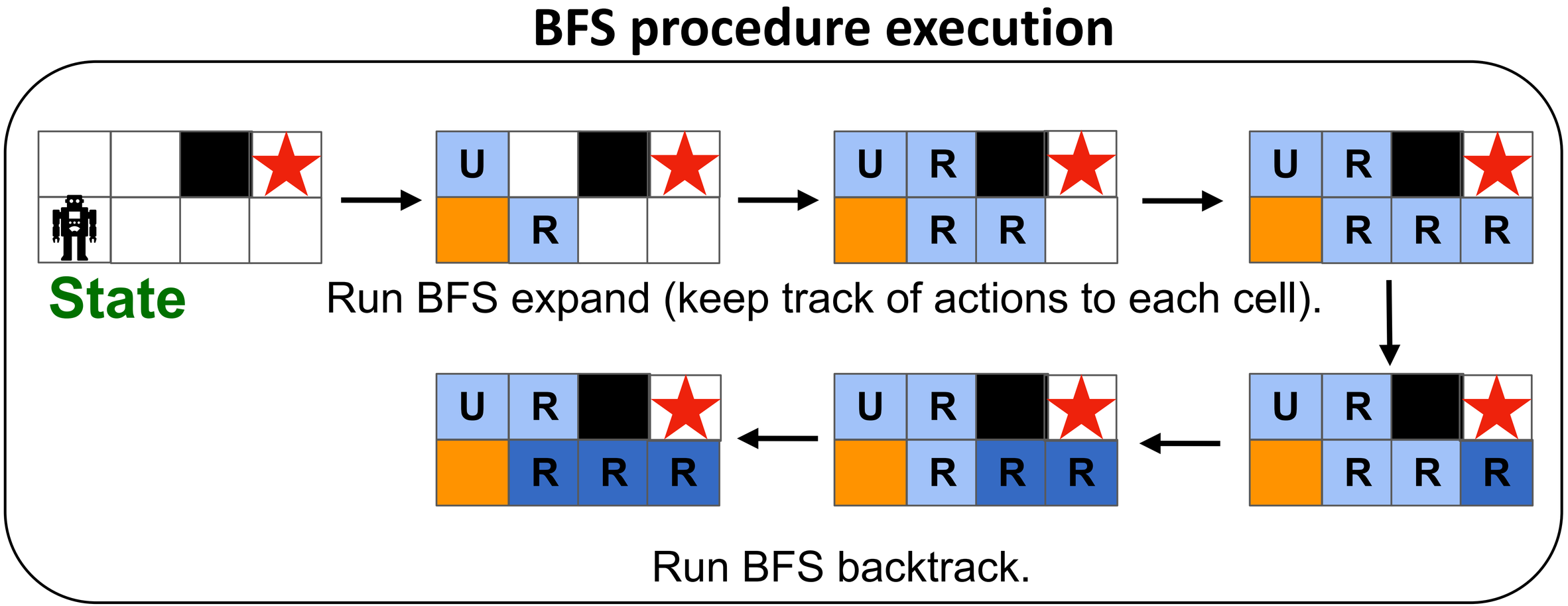
Procedure Clone BFS

- Proof of concept: synthetic maze navigation



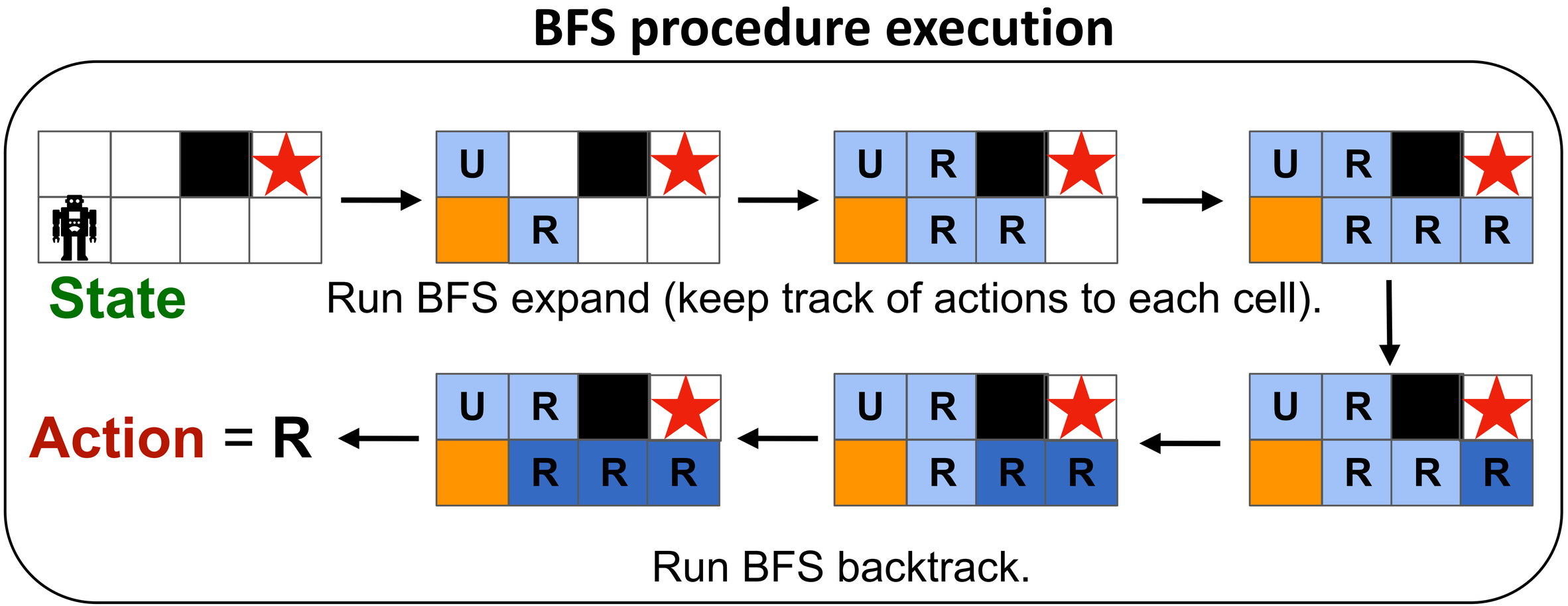
Procedure Clone BFS

- Proof of concept: synthetic maze navigation



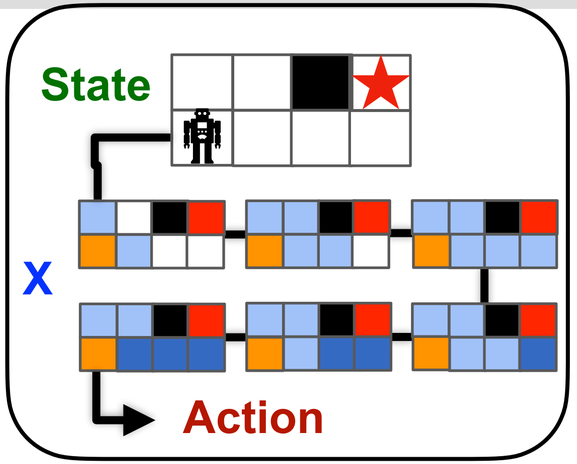
Procedure Clone BFS

- Proof of concept: synthetic maze navigation



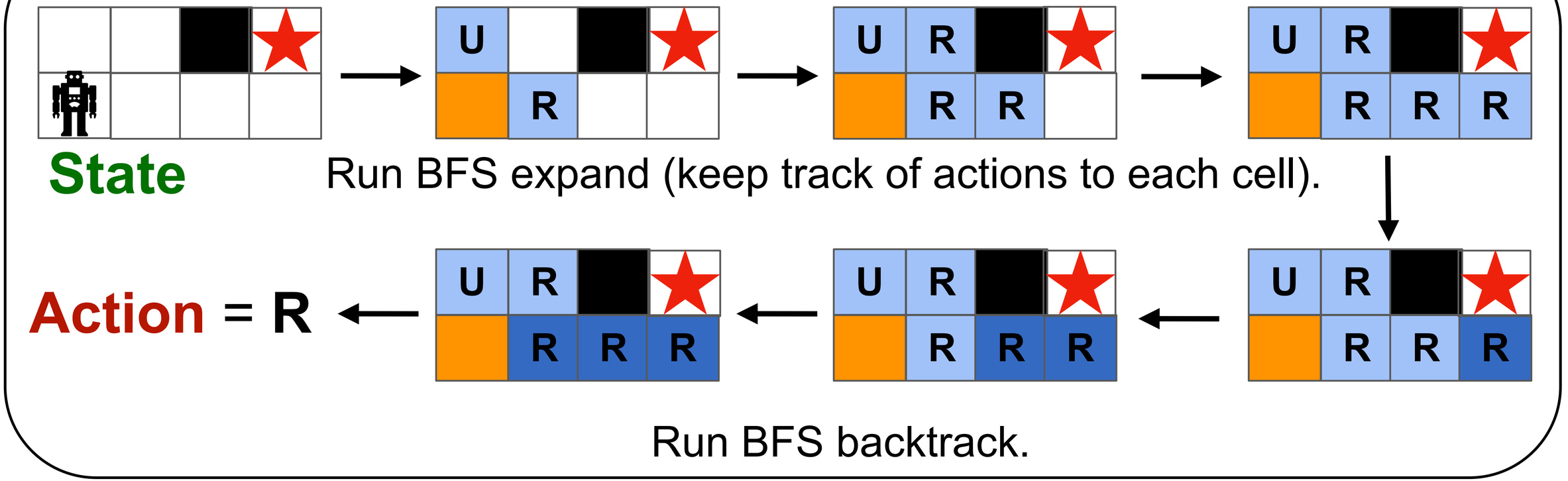
Procedure Clone BFS

- Proof of concept: synthetic maze navigation



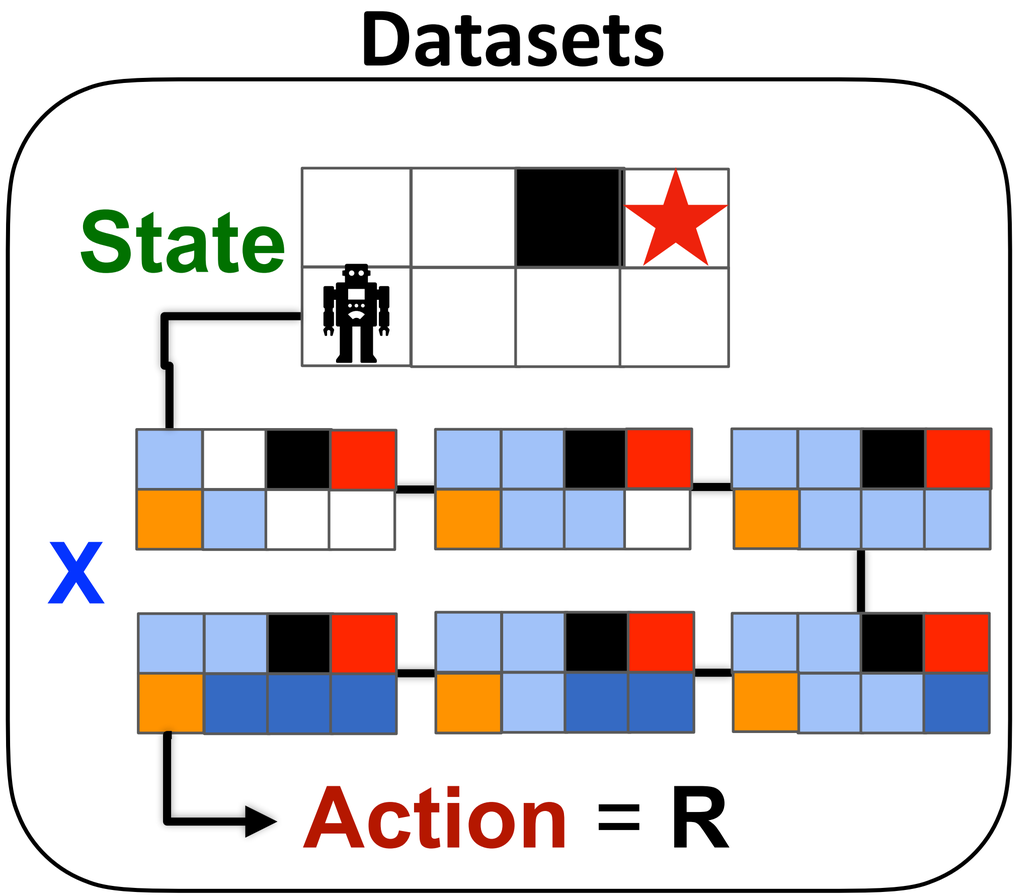
Datasets

BFS procedure execution



Procedure Clone BFS

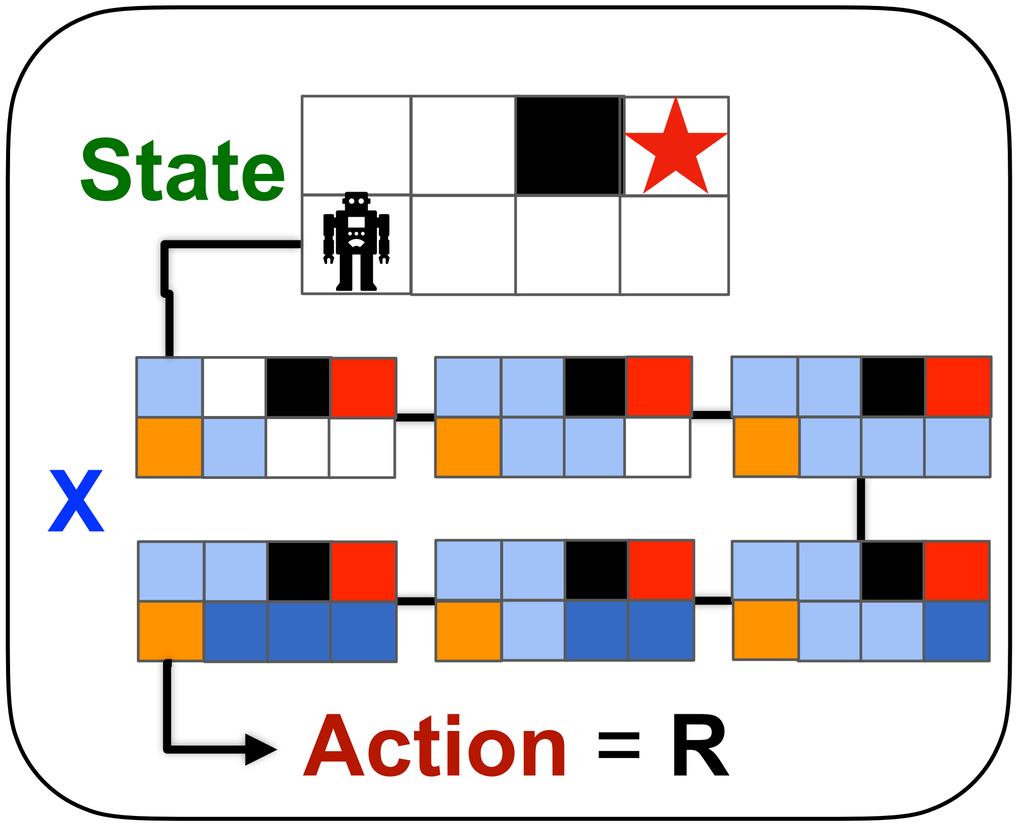
- Proof of concept: synthetic maze navigation



Procedure Clone BFS

- Proof of concept: synthetic maze navigation

Datasets



Procedure cloning

$$\pi_{PC} = p_\phi(x_0 \mid s) \bullet$$

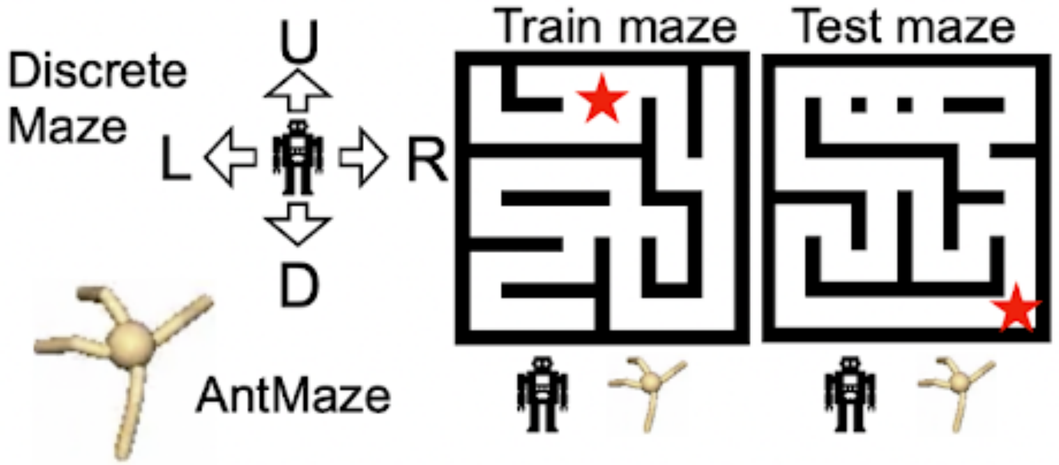
$$p_\theta(x_1 \mid x_0) \bullet p_\theta(x_2 \mid x_1) \bullet p_\theta(x_3 \mid x_2) \bullet$$

$$p_\theta(x_6 \mid x_5) \bullet p_\theta(x_5 \mid x_4) \bullet p_\theta(x_4 \mid x_3)$$

$$\bullet\, q_\psi(a \mid x_6)$$

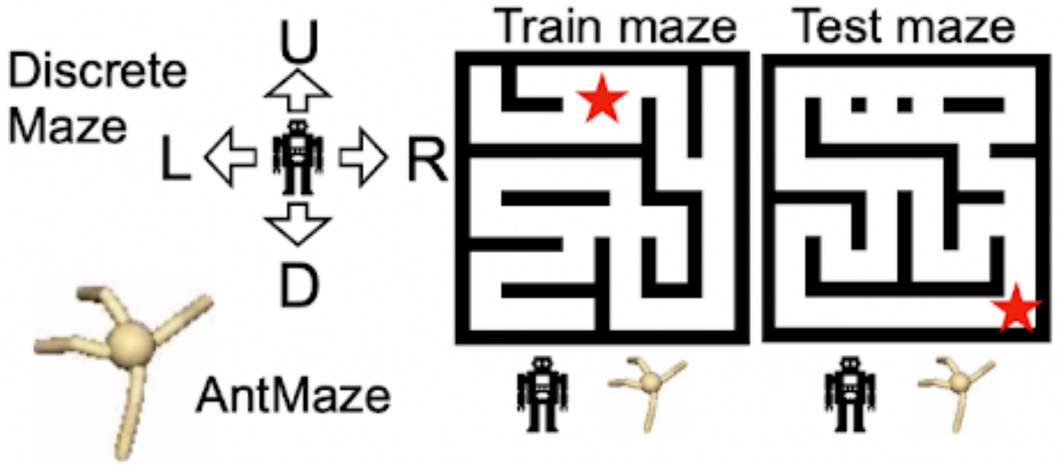
Procedure Clone BFS

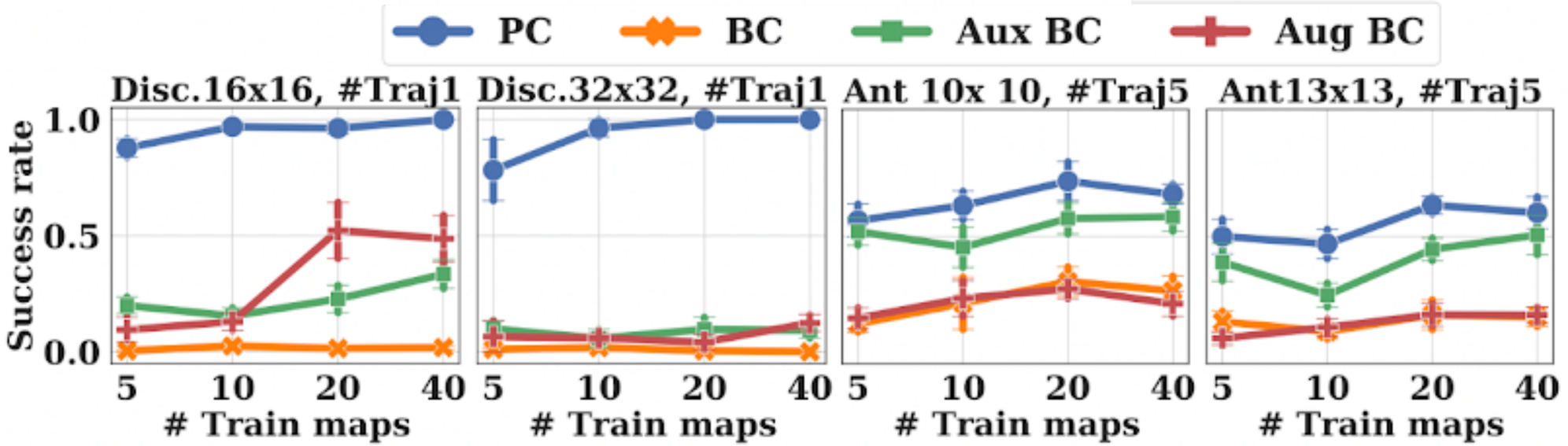
- Proof of concept: synthetic maze navigation
- Experiment:



Procedure Clone BFS

- Proof of concept: synthetic maze navigation
- Experiment:

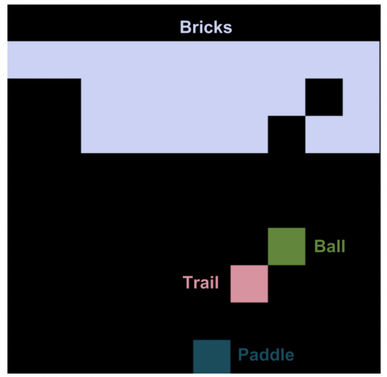




Procedure Clone MCTS

- Autoregressive procedure cloning

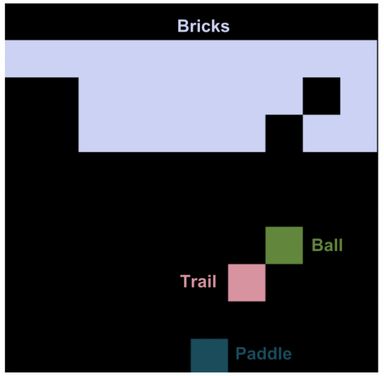
Atari env



Procedure Clone MCTS

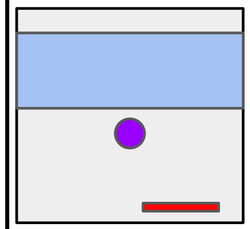
- Autoregressive procedure cloning

Atari env



MCTS Procedure execution

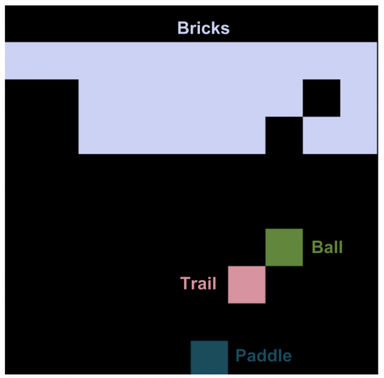
State



Procedure Clone MCTS

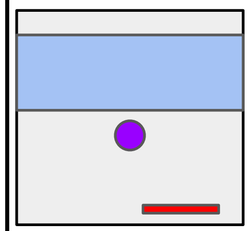
- Autoregressive procedure cloning

Atari env



MCTS Procedure execution

State

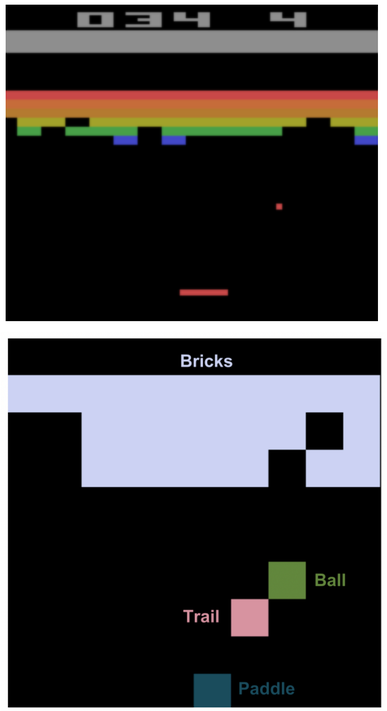


Run MCTS

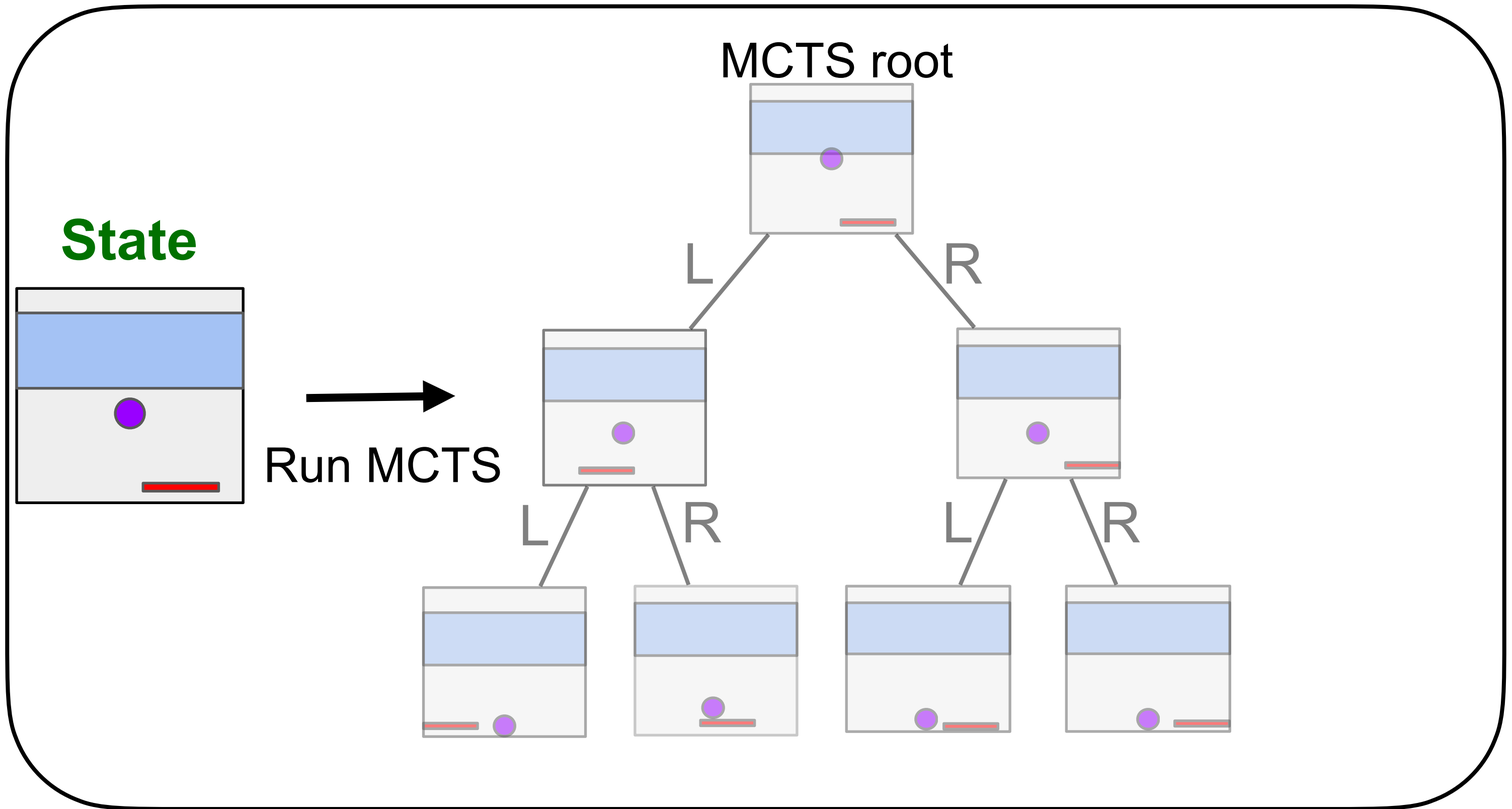
Procedure Clone MCTS

- Autoregressive procedure cloning

Atari env



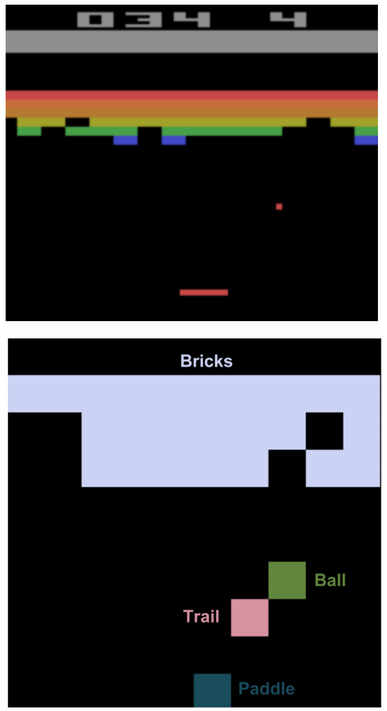
MCTS Procedure execution



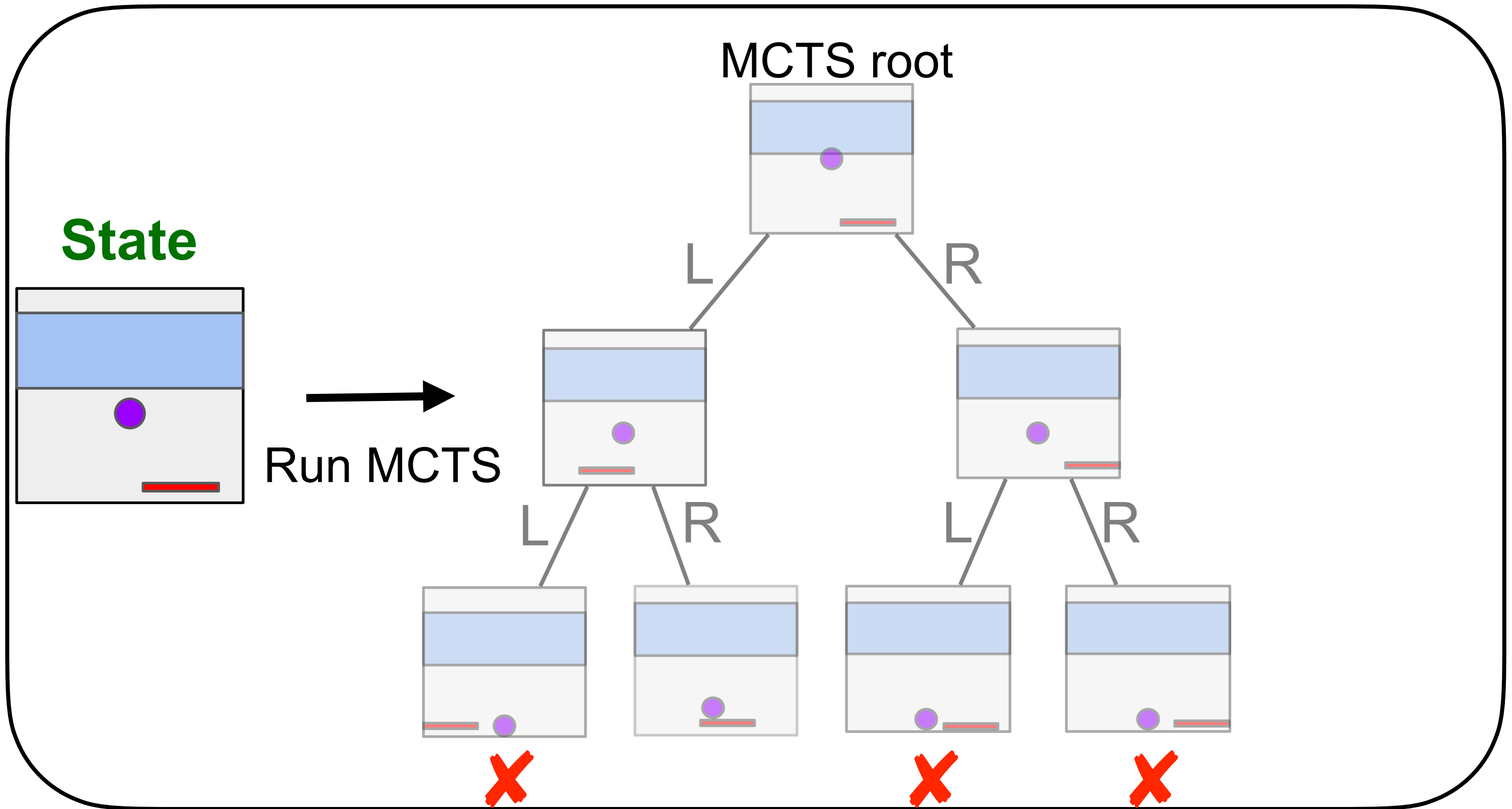
Procedure Clone MCTS

- Autoregressive procedure cloning

Atari env



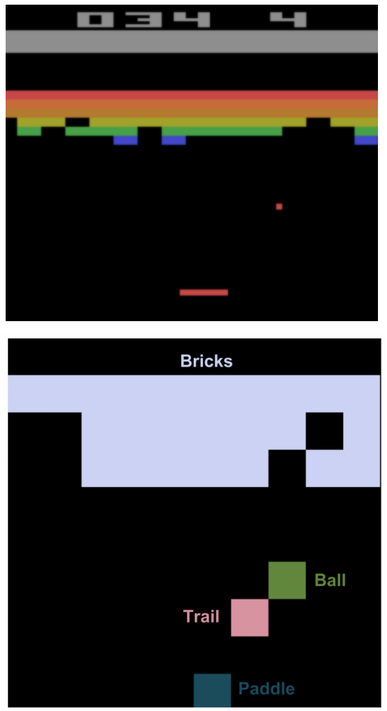
MCTS Procedure execution



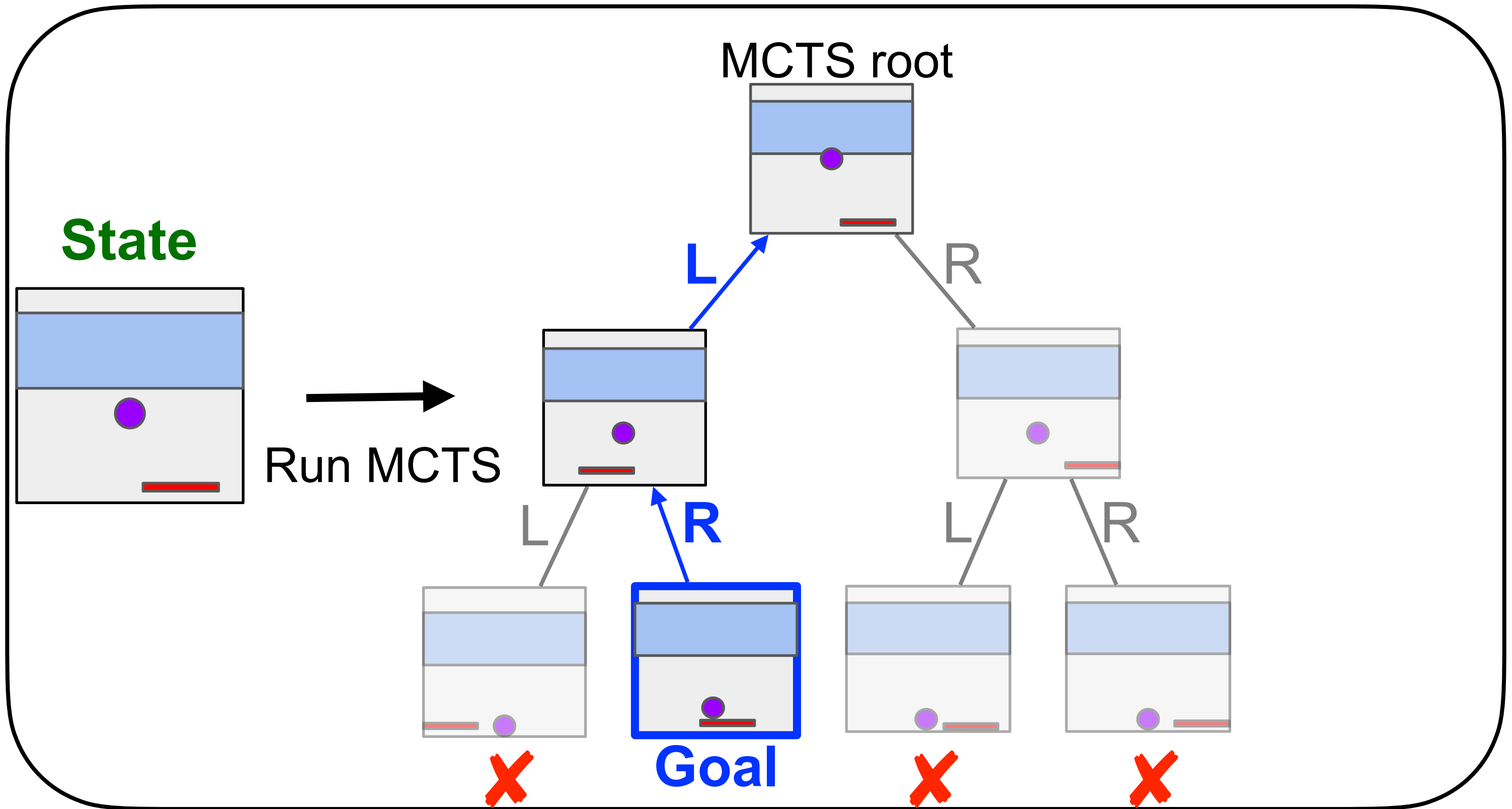
Procedure Clone MCTS

- Autoregressive procedure cloning

Atari env



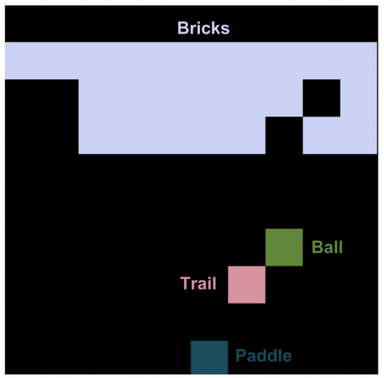
MCTS Procedure execution



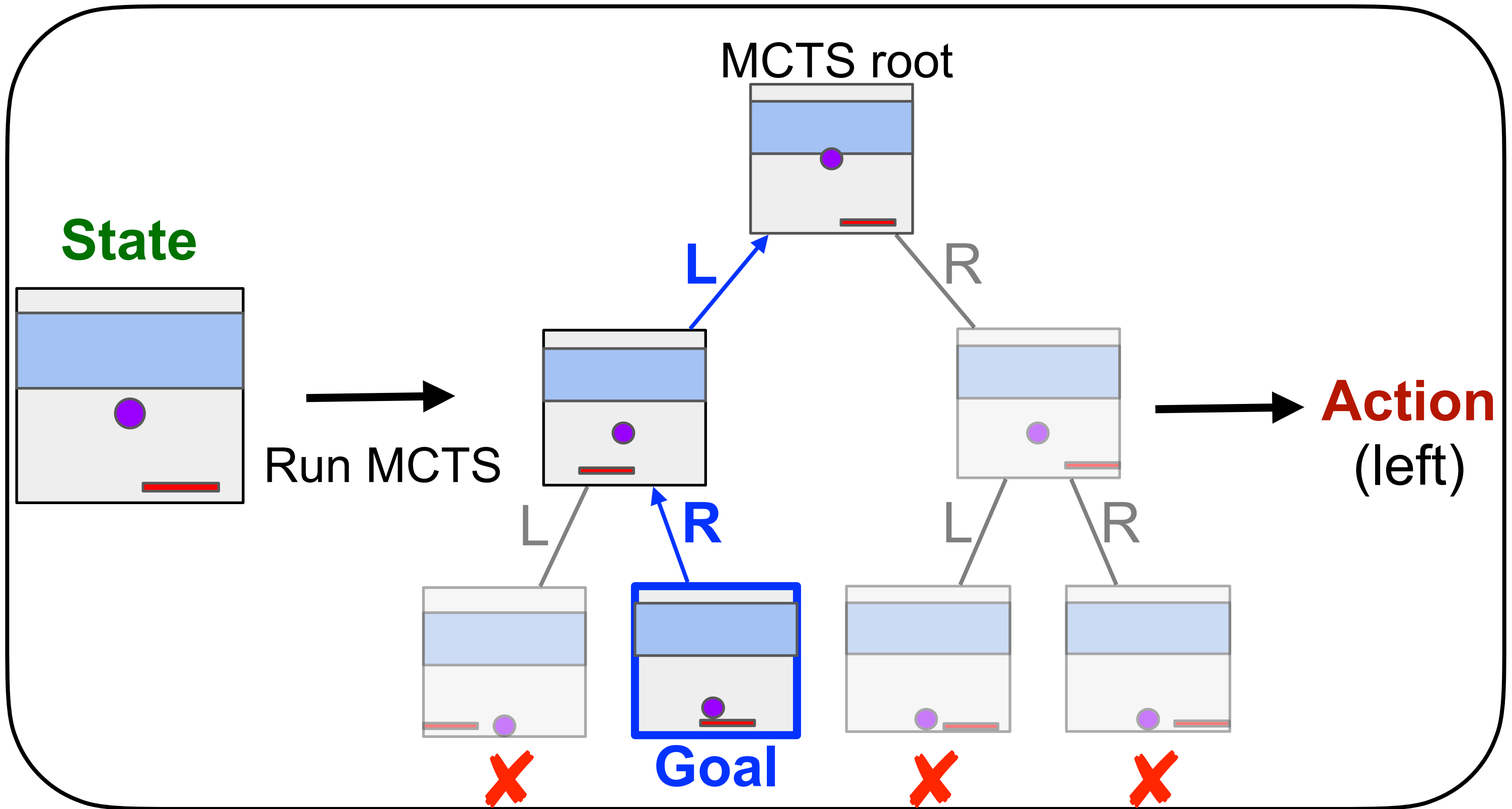
Procedure Clone MCTS

- Autoregressive procedure cloning

Atari env

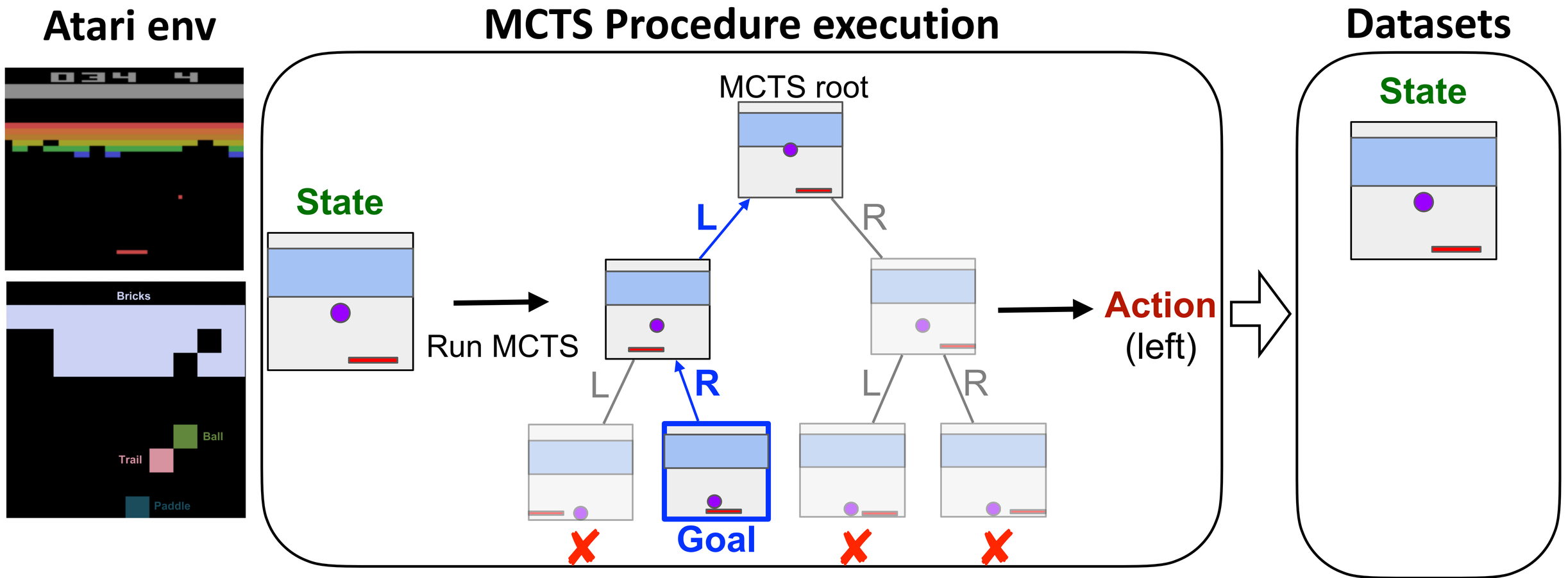


MCTS Procedure execution



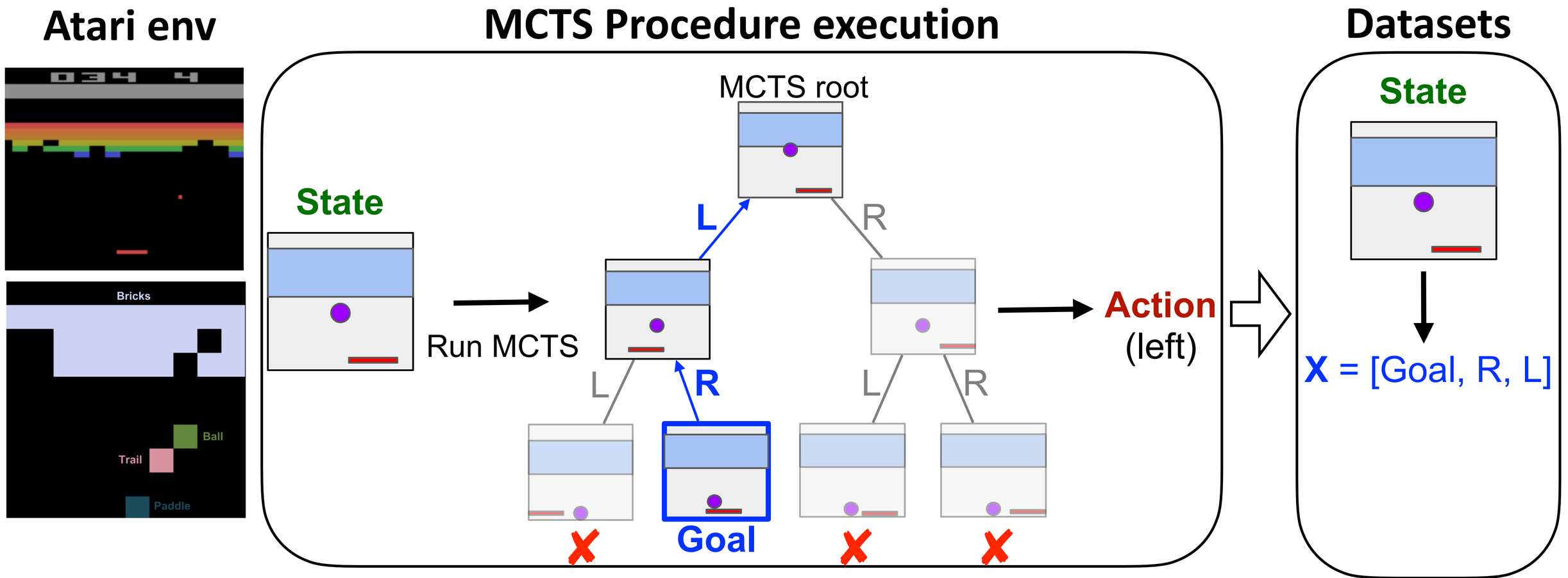
Procedure Clone MCTS

- Autoregressive procedure cloning



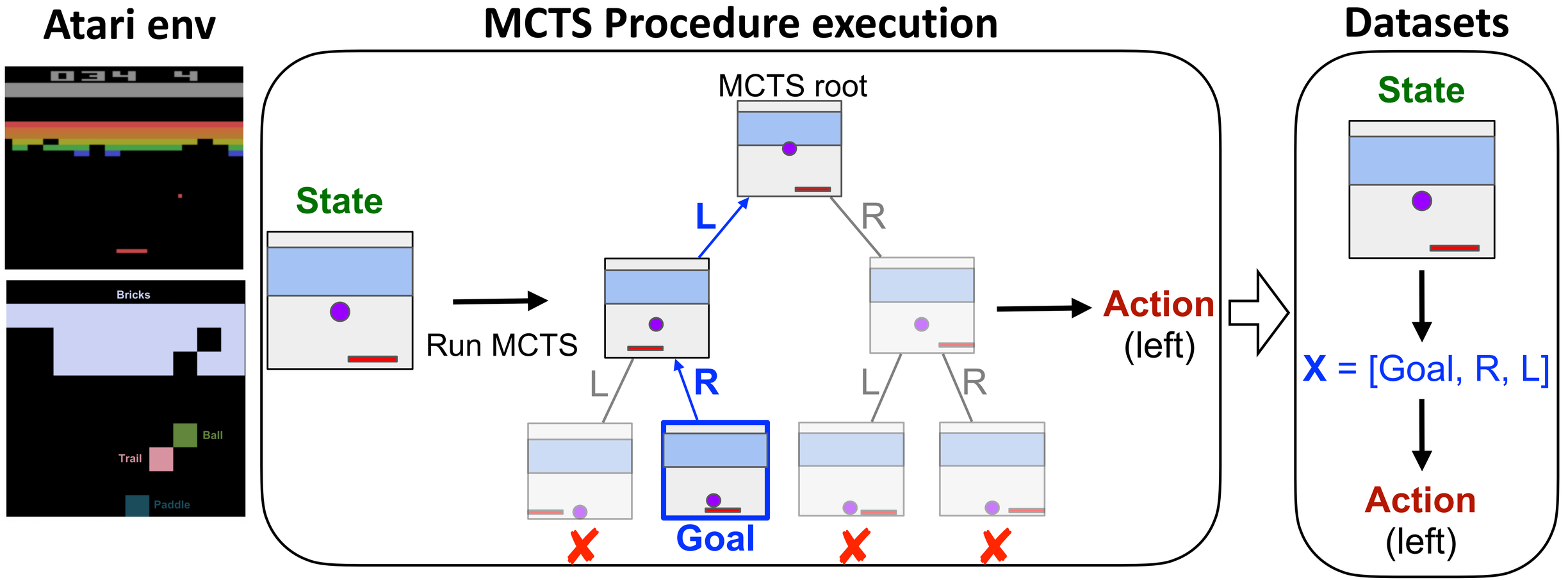
Procedure Clone MCTS

- Autoregressive procedure cloning



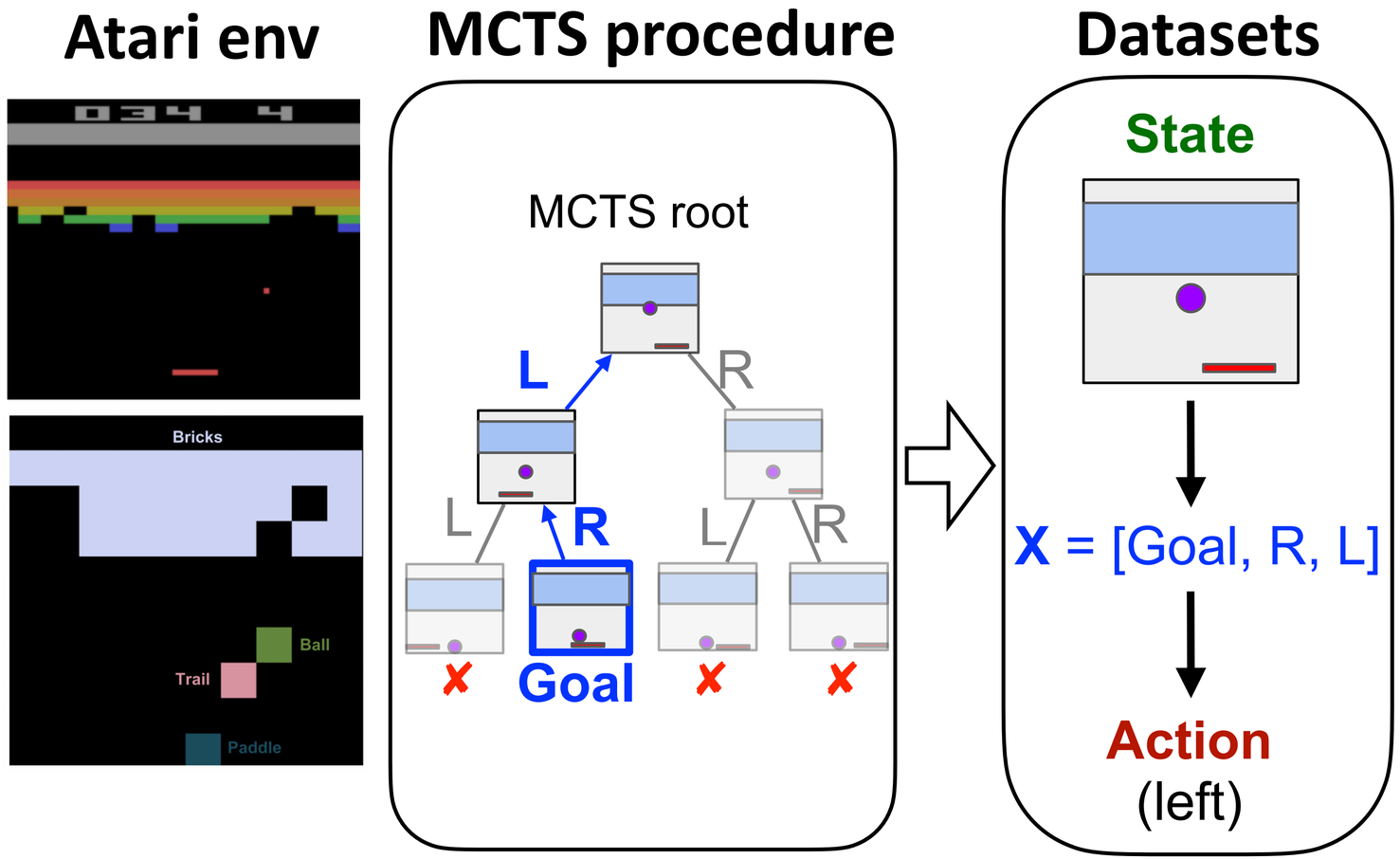
Procedure Clone MCTS

- Autoregressive procedure cloning



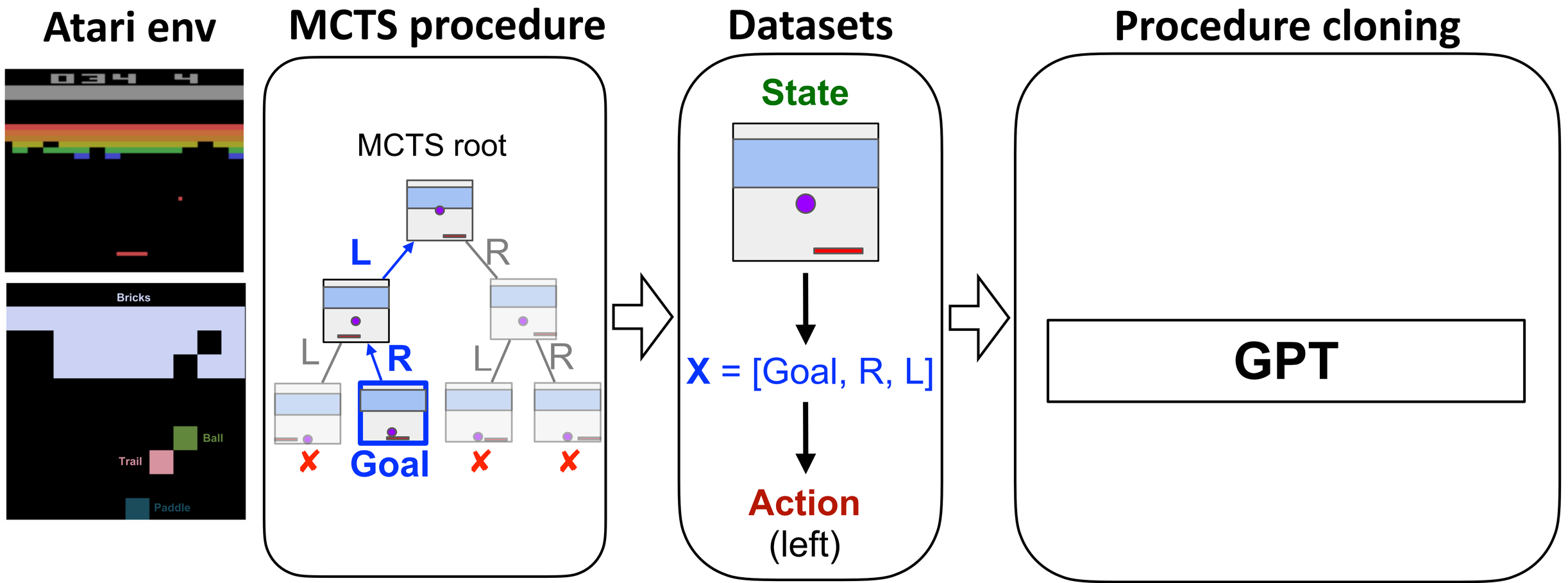
Procedure Clone MCTS

- Autoregressive procedure cloning



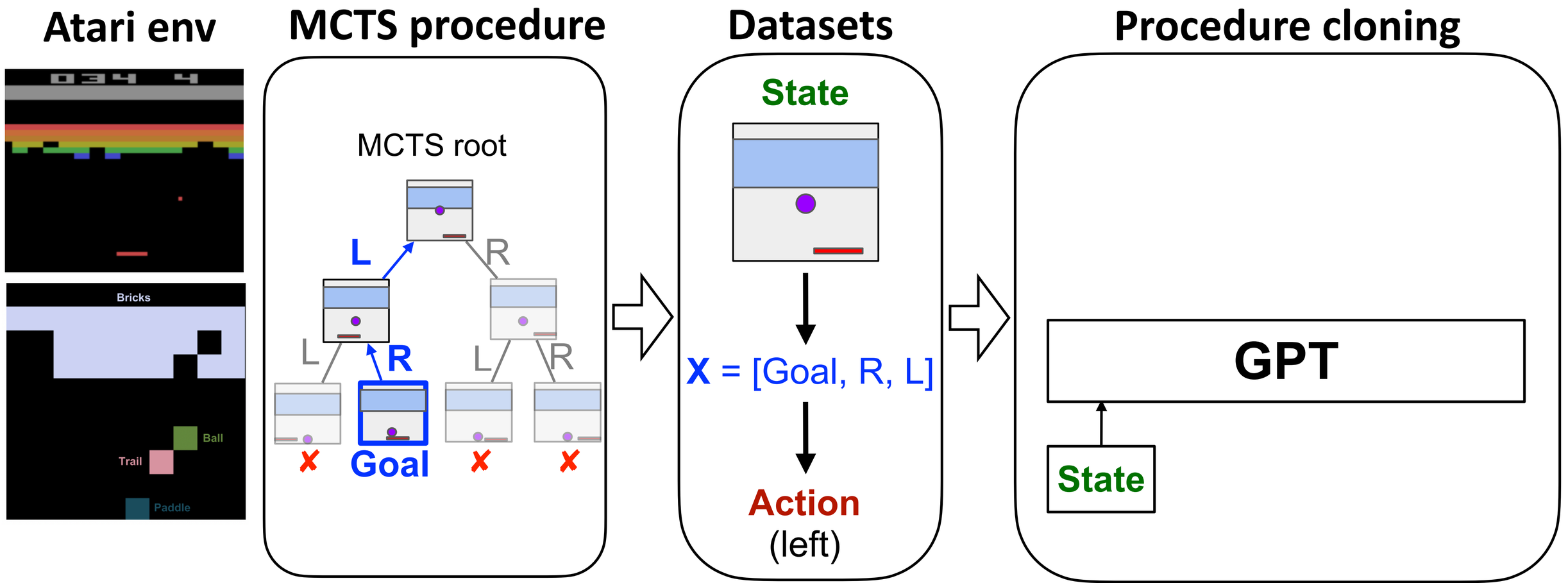
Procedure Clone MCTS

- Autoregressive procedure cloning



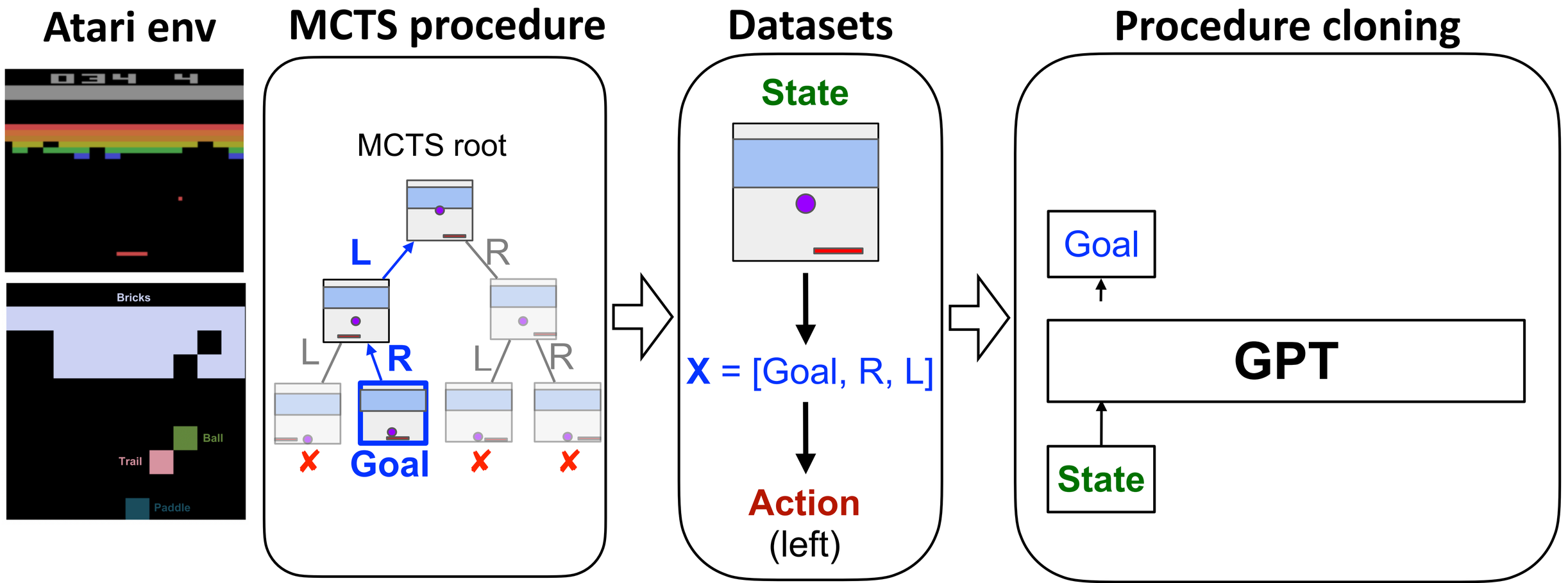
Procedure Clone MCTS

- Autoregressive procedure cloning



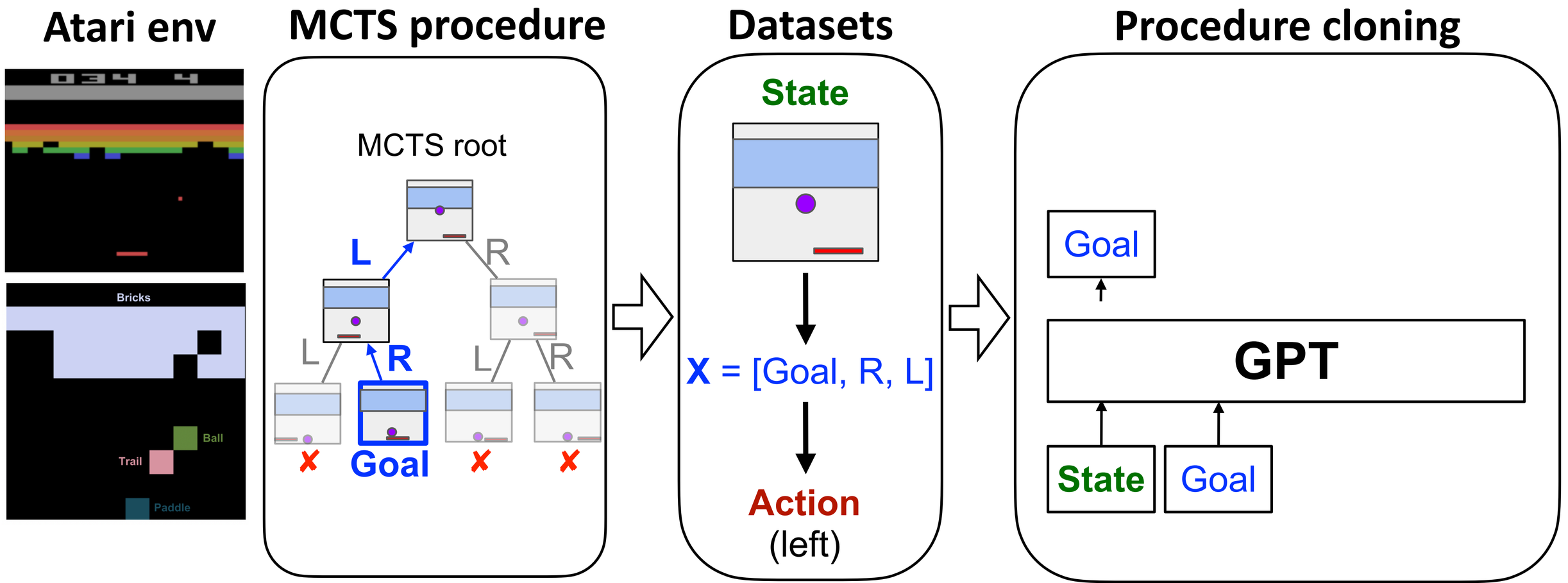
Procedure Clone MCTS

- Autoregressive procedure cloning



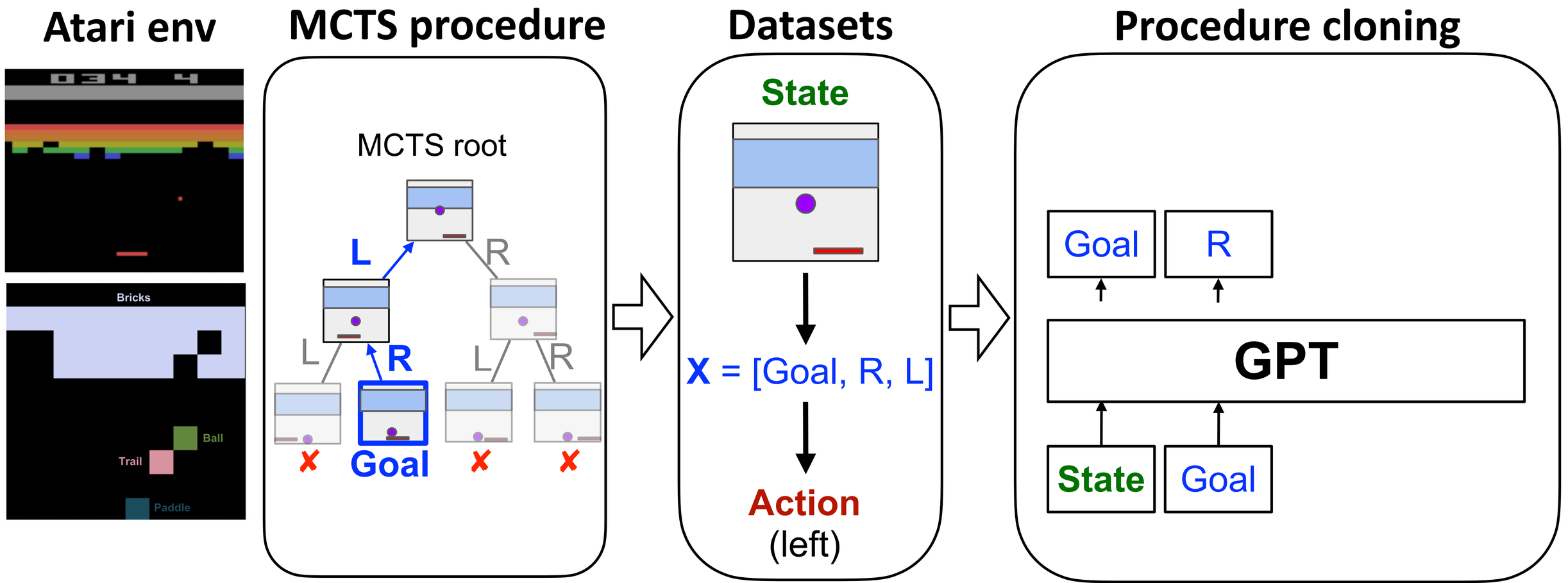
Procedure Clone MCTS

- Autoregressive procedure cloning



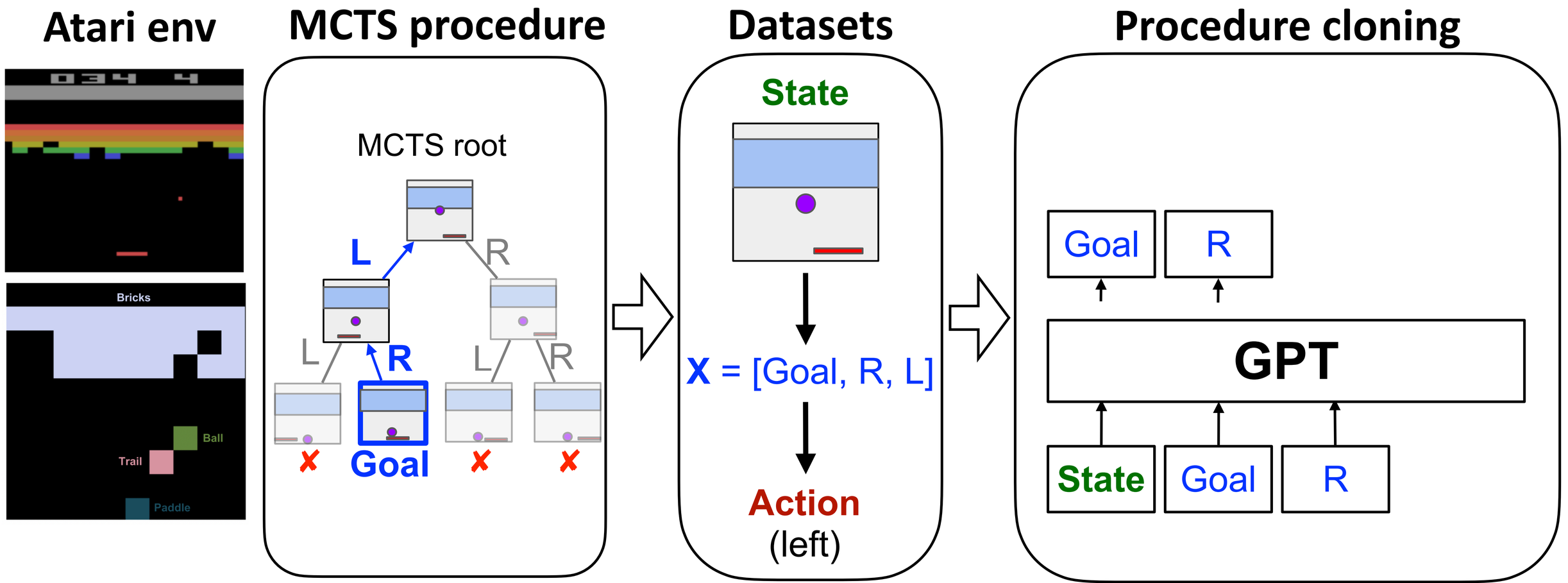
Procedure Clone MCTS

- Autoregressive procedure cloning



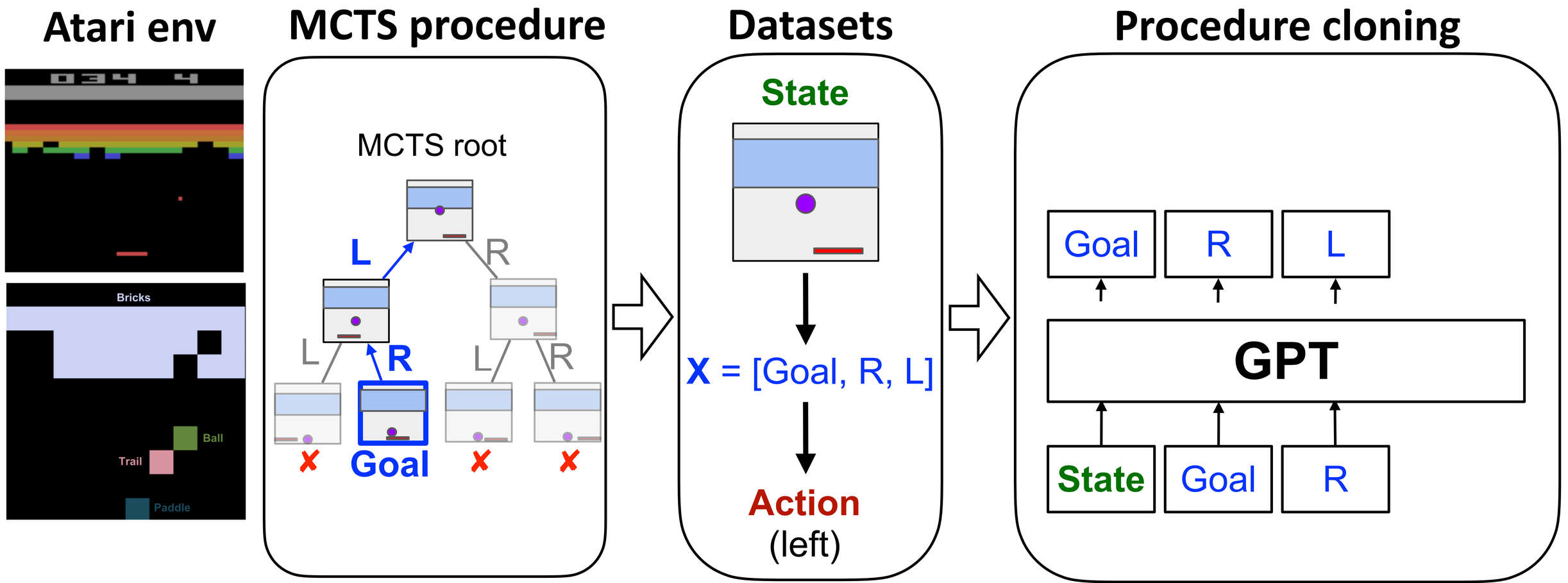
Procedure Clone MCTS

- Autoregressive procedure cloning



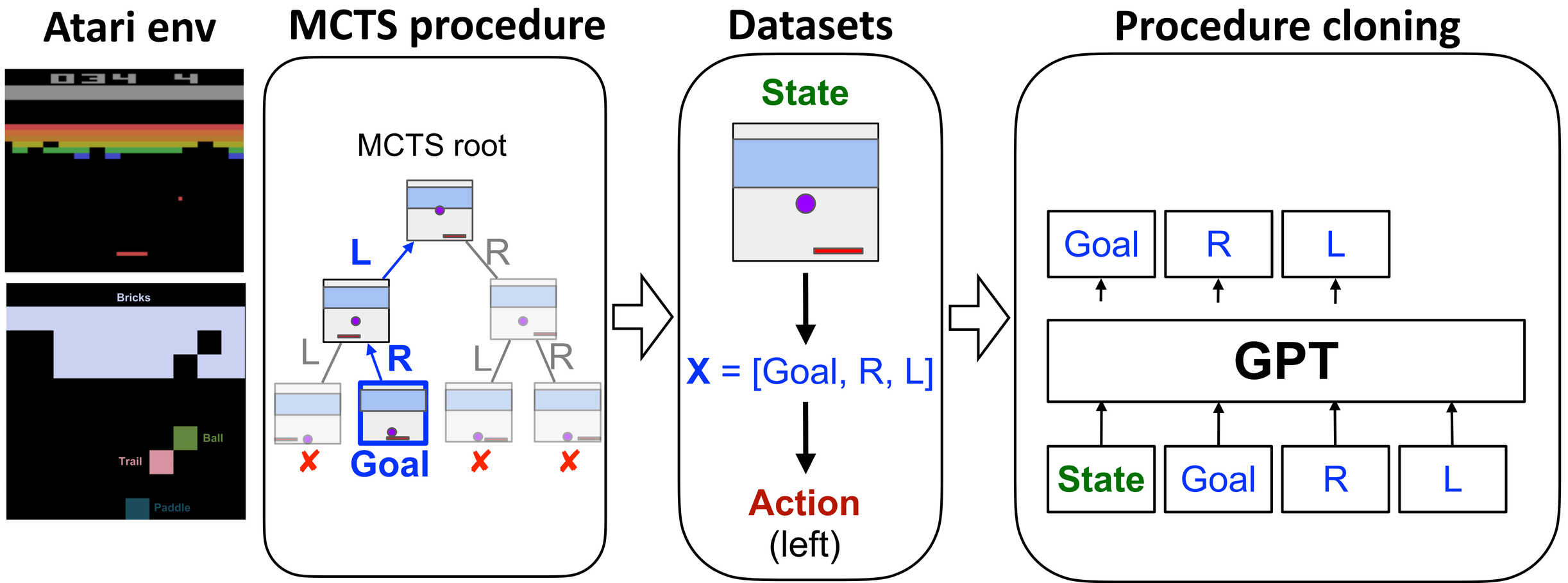
Procedure Clone MCTS

- Autoregressive procedure cloning



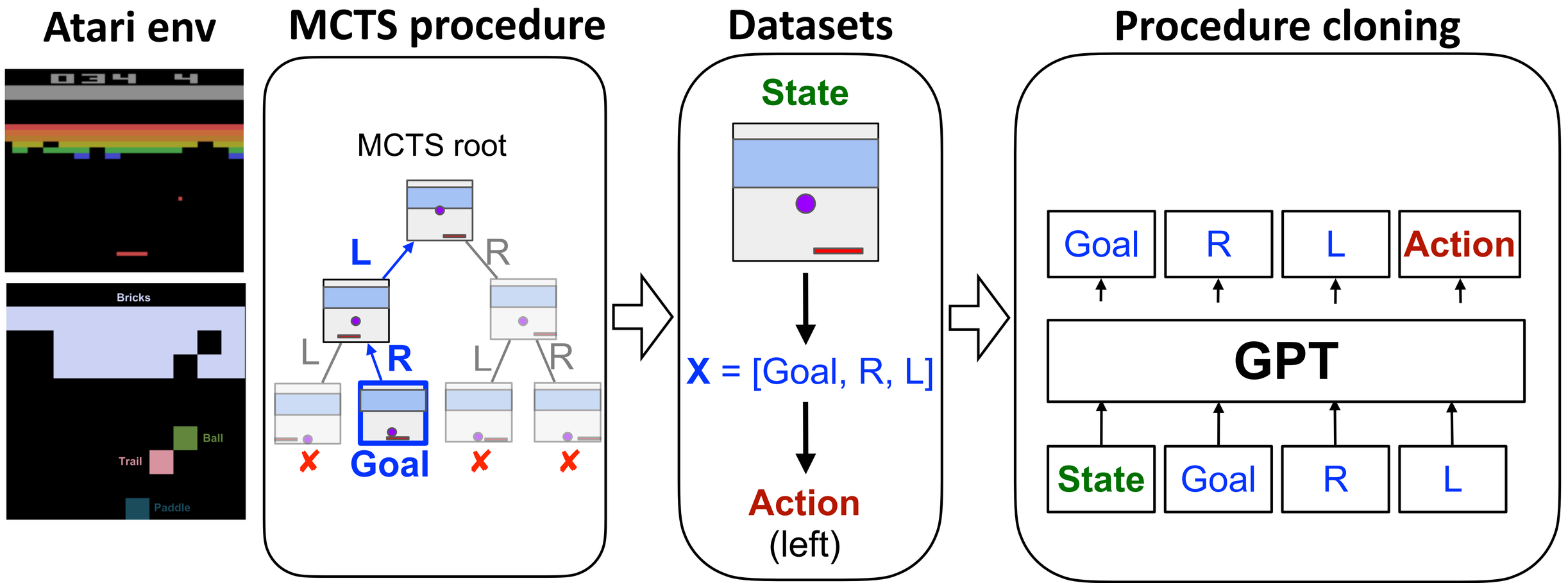
Procedure Clone MCTS

- Autoregressive procedure cloning



Procedure Clone MCTS

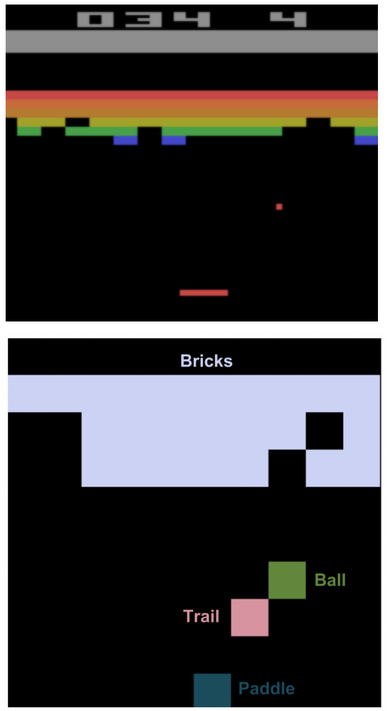
- Autoregressive procedure cloning



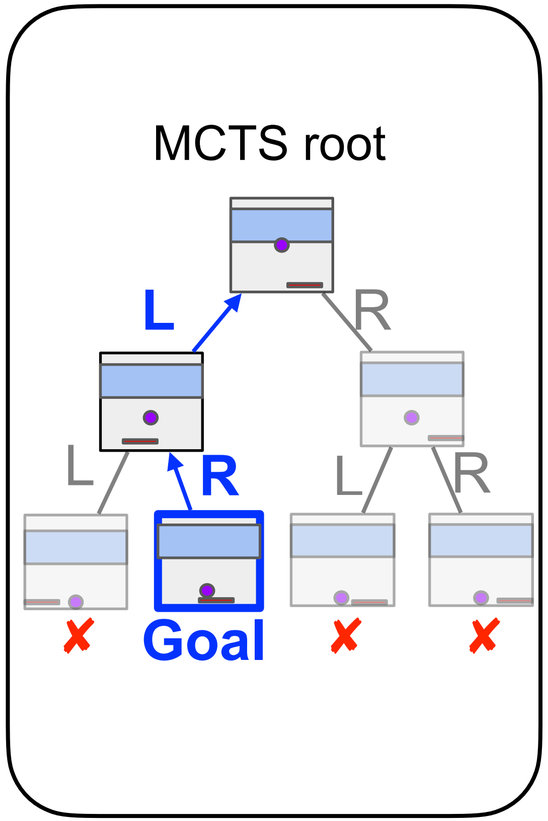
Procedure Clone MCTS

- Autoregressive procedure cloning

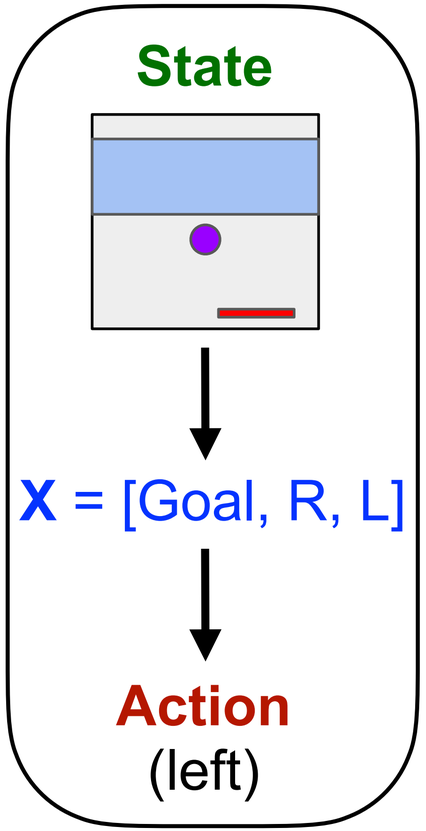
Atari env



MCTS procedure

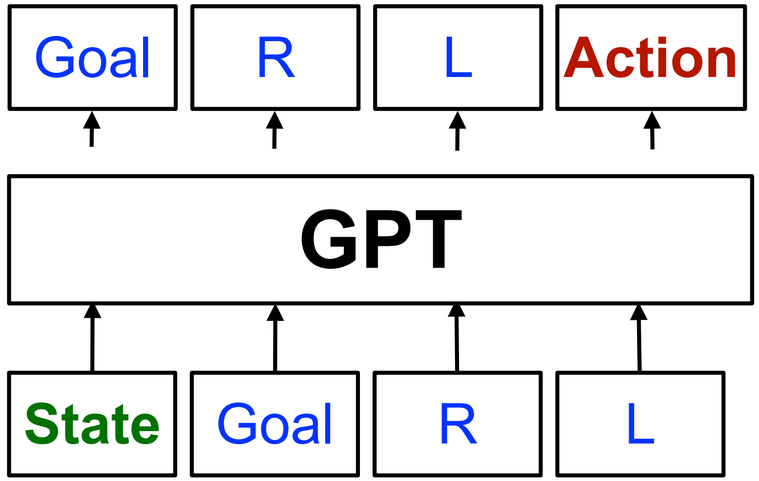


Datasets



Procedure cloning

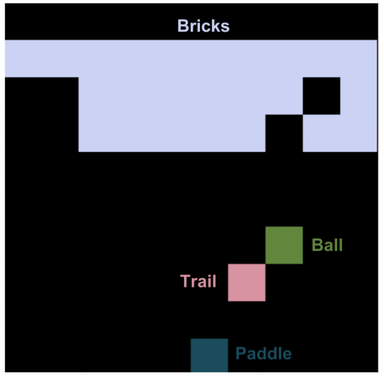
$$\pi_{PC} = p_\phi(x_0 \mid s) \cdot p_\theta(x_l \mid \mathbf{x}_{<l}, s) \cdot q_\psi(a \mid \mathbf{x}, s)$$

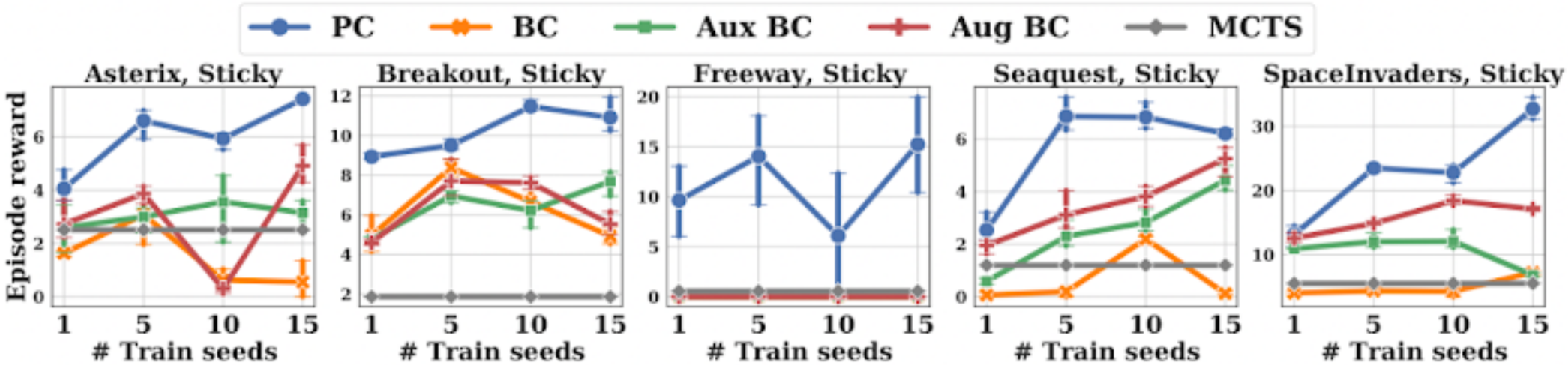


Procedure Clone MCTS

- Autoregressive procedure cloning
- Experiments:

Atari env

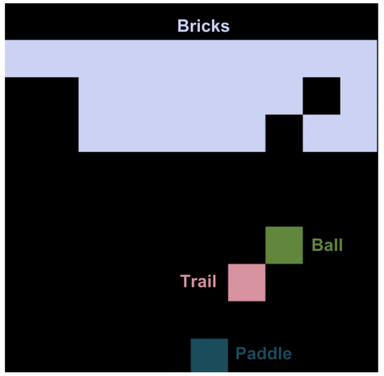


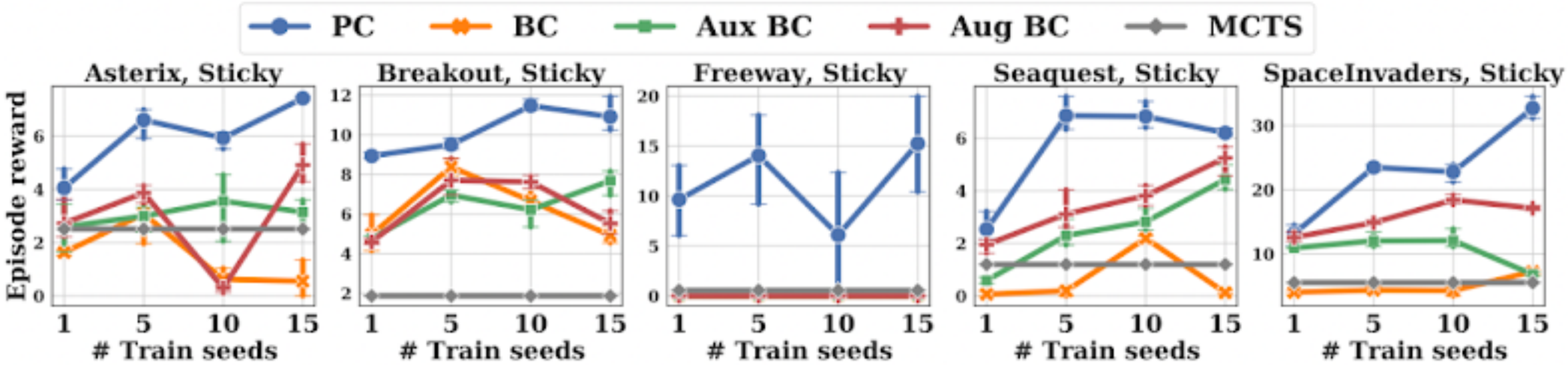


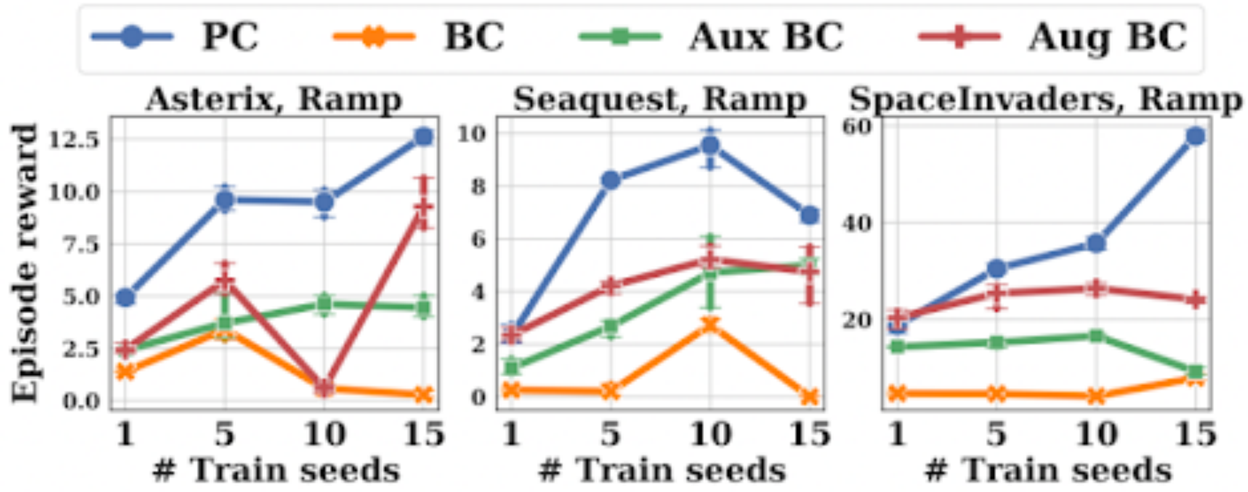
Procedure Clone MCTS

- Autoregressive procedure cloning
- Experiments:

Atari env

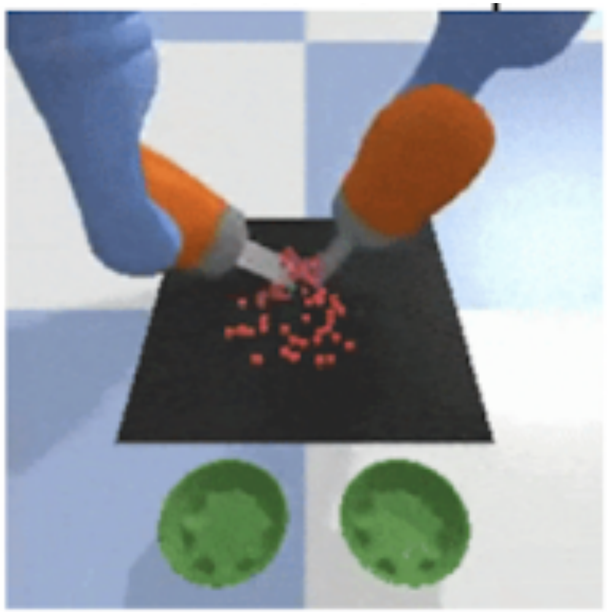






Procedure Clone Manipulation Script

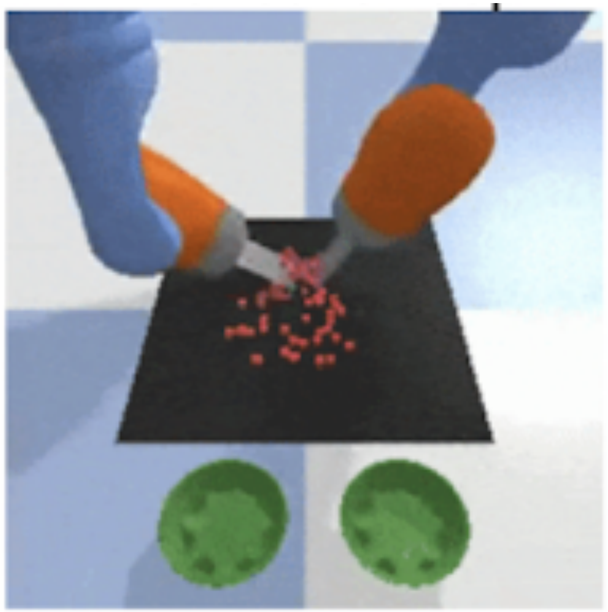
- Bimanual sweep task

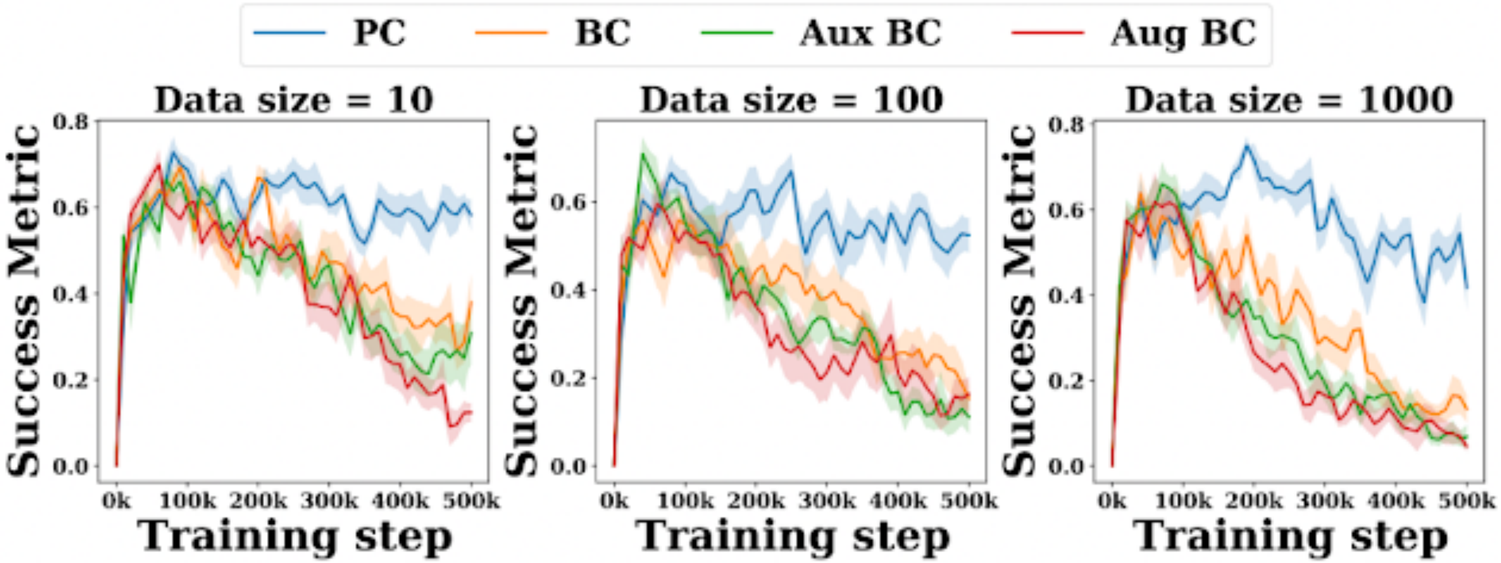


Pete Florence, Corey Lynch, Andy Zeng, Oscar A Ramirez, Ayzaan Wahid, Laura Downs, Adrian Wong, Johnny Lee, Igor Mordatch, and Jonathan Tompson. Implicit behavioral cloning. In Conference on Robot Learning, pages 158–168. PMLR, 2022.

Procedure Clone Manipulation Script

- Bimanual sweep task

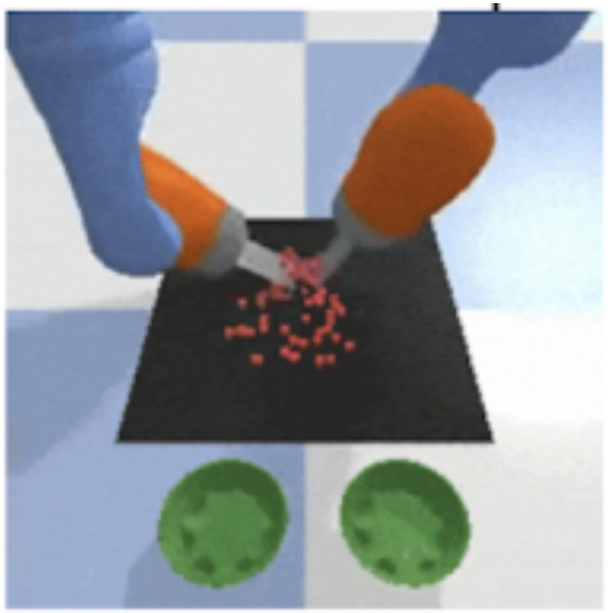


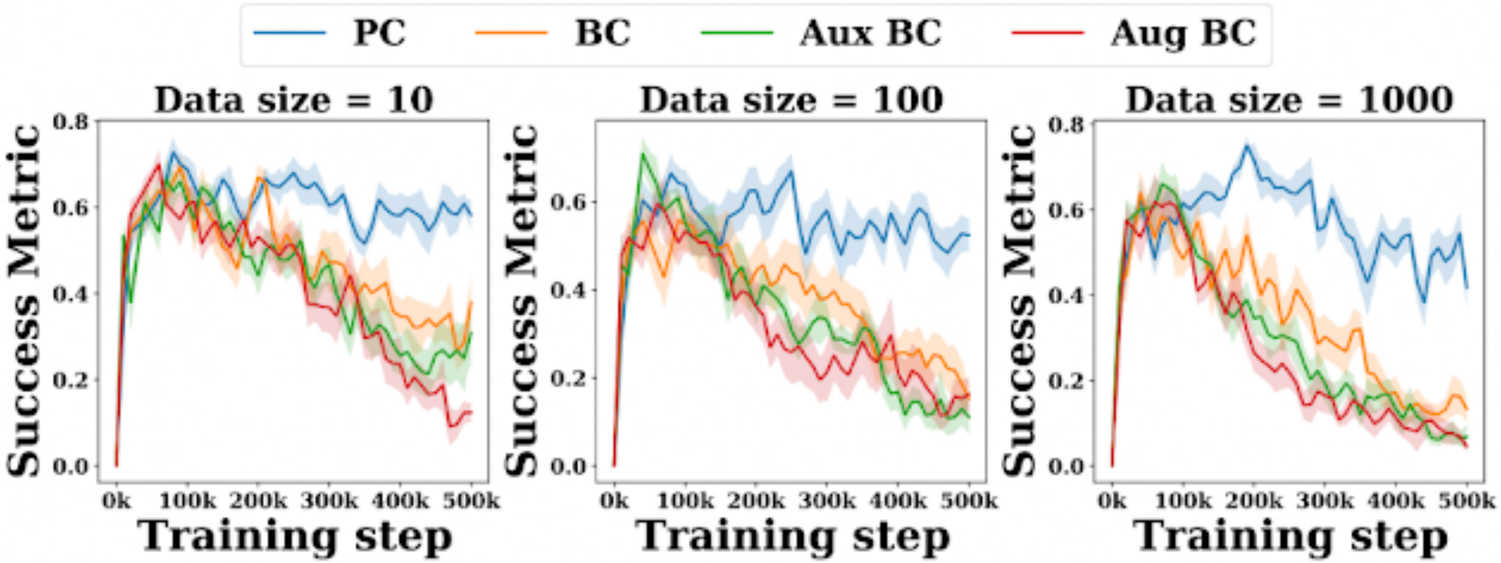


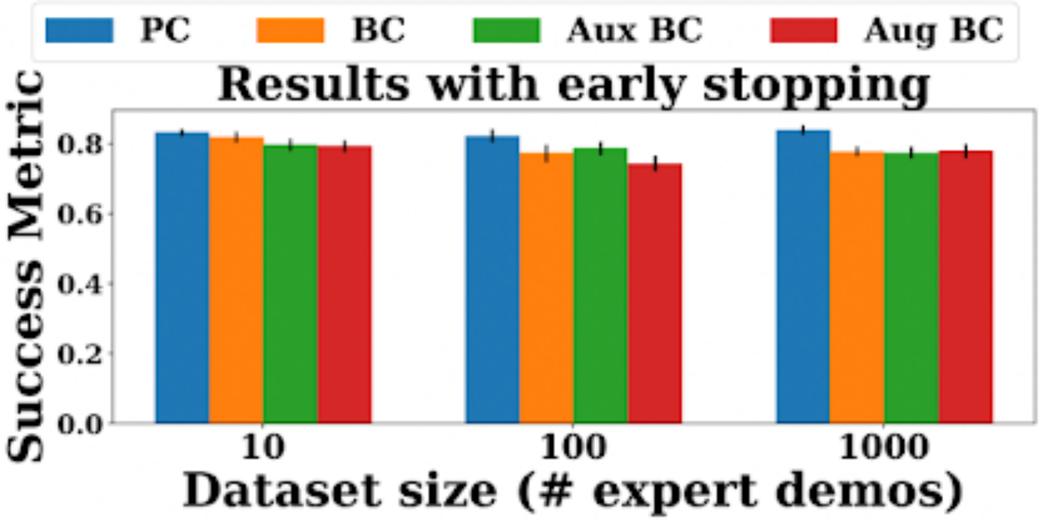
Pete Florence, Corey Lynch, Andy Zeng, Oscar A Ramirez, Ayzaan Wahid, Laura Downs, Adrian Wong, Johnny Lee, Igor Mordatch, and Jonathan Tompson. Implicit behavioral cloning. In Conference on Robot Learning, pages 158–168. PMLR, 2022.

Procedure Clone Manipulation Script

- Bimanual sweep task
- SoTA: 78.2% -> 83.9%







Pete Florence, Corey Lynch, Andy Zeng, Oscar A Ramirez, A lordatch, and Jonathan Tompson. Implicit behavioral cloning. In Conference on Rob

Broader Implications

- Connection to chain of thought prompting?

  Decomposing multi-step problems into intermediate steps and learning the intermediate steps using a sequence model.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou. Chain of thought prompting elicits reasoning in large language models.

Broader Implications

- Connection to chain of thought prompting?

  Decomposing multi-step problems into intermediate steps and learning the intermediate steps using a sequence model.

- Sequence modeling in Markovian environments? Why?

  E.g., Decision Transformer, Trajectory Transformer

  Procedure cloning models intermediate procedure computations as a sequence

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou. Chain of thought prompting elicits reasoning in large language models.
Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Misha Laskin, Pieter Abbeel, Aravind Srinivas, and Igor Mordatch. Decision transformer: Reinforcement learning via sequence modeling.
Michael Janner, Qiyang Li, and Sergey Levine. Offline reinforcement learning as one big sequence modeling problem.

Recap & Conclusion

- Gap in imitation learning: more expert info
  - Chain of thought imitation learning

Recap & Conclusion

- Gap in imitation learning: more expert info
  - Chain of thought imitation learning
- Expert computation relies on tools not available during inference
  - Procedure cloning

Recap & Conclusion

- Gap in imitation learning: more expert info
  - Chain of thought imitation learning
- Expert computation relies on tools not available during inference
  - Procedure cloning
- Results
  - Significant (zero-shot) generalization to new env configs
  - Better than expert

Recap & Conclusion

- Gap in imitation learning: more expert info
  - Chain of thought imitation learning
- Expert computation relies on tools not available during inference
  - Procedure cloning
- Results
  - Significant (zero-shot) generalization to new env configs
  - Better than expert

Thank you. Checkout

Paper: arxiv.org/abs/2205.10816?
Code: github.com/google-research/google-research/tree/master/procedure_cloning
Website: sites.google.com/corp/view/procedure-cloning