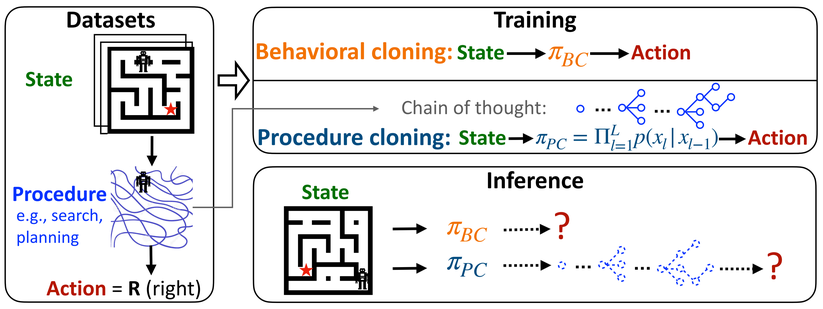
# Chain of Thought Imitation with Procedure Cloning

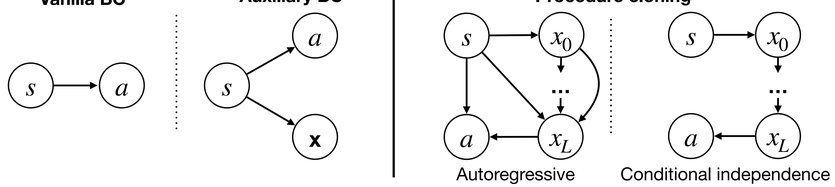
Sherry Yang, Dale Schuurmans, Pieter Abbeel, Ofir Nachum

## Imitation Learning

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




## Procedure Cloning




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

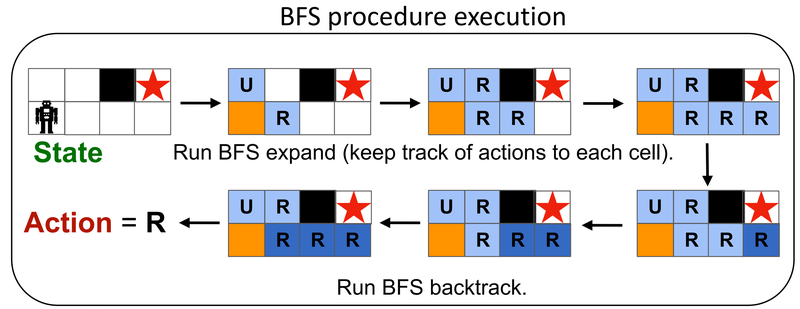
$J_{\mathrm{BC}}(\pi) := \hat{\mathbb{E}}_{(s,a) \sim \mathcal{D}_*}[-\log \pi(a|s)]$

Vanilla BC objective

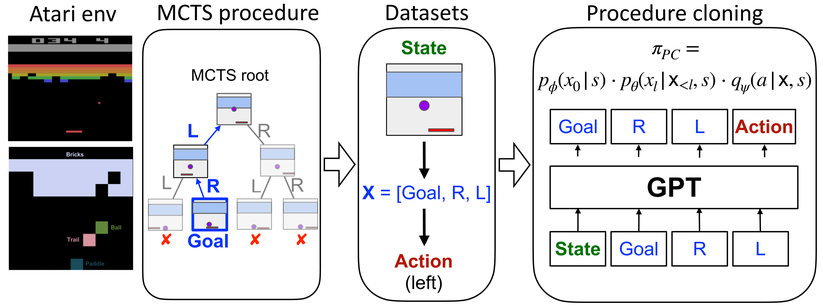
$$\min_{\phi,\theta,\psi} J_{\mathrm{PC}}(\phi,\theta,\psi) = \hat{\mathbb{E}}_{(s,\mathbf{x},a)\sim\mathcal{D}_\Pi}[-\log p(a,\mathbf{x}|s)]$$
$$= \mathbb{E}_{(s,\mathbf{x},a)\sim\mathcal{D}_\Pi}\Big[-\log q_\psi(a|\mathbf{x}, s) - \sum_{\ell=1}^{L} \log p_\theta(x_\ell|\mathbf{x}_{<\ell}, s) - \log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

PC objective

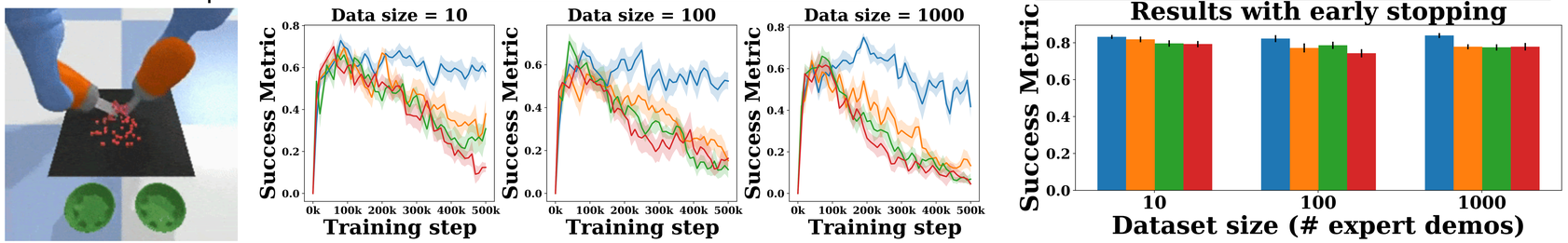
## Example: BFS



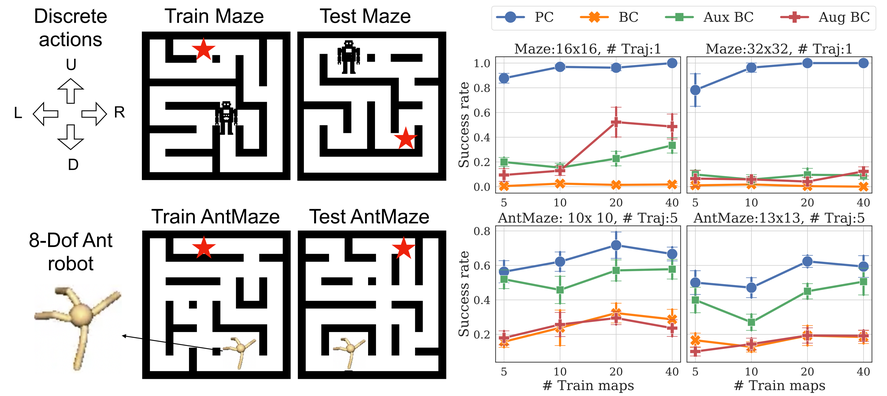

## Example: MCTS



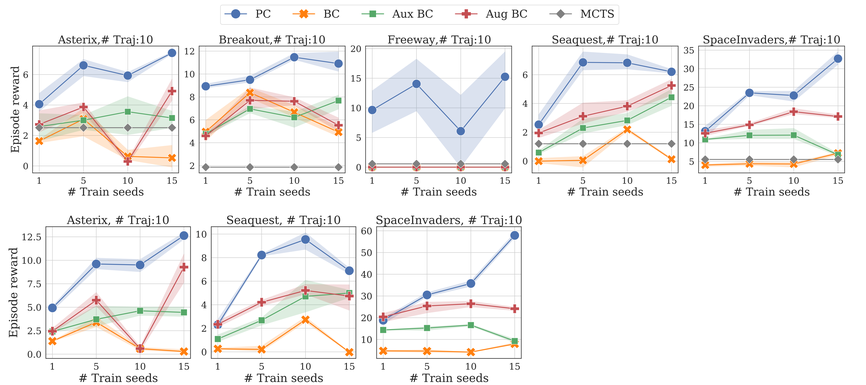

## Results: Simulated Robotic Manipulation




## Results: GridWorld & AntMaze




## Results: MinAtar




## Conclusion

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

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