Batch Exploration with Examples for Scalable Robotic Reinforcement Learning

Annie S. Chen*, Hyunji Nam*, Suraj Nair*, Chelsea Finn

Stanford Artificial Intelligence

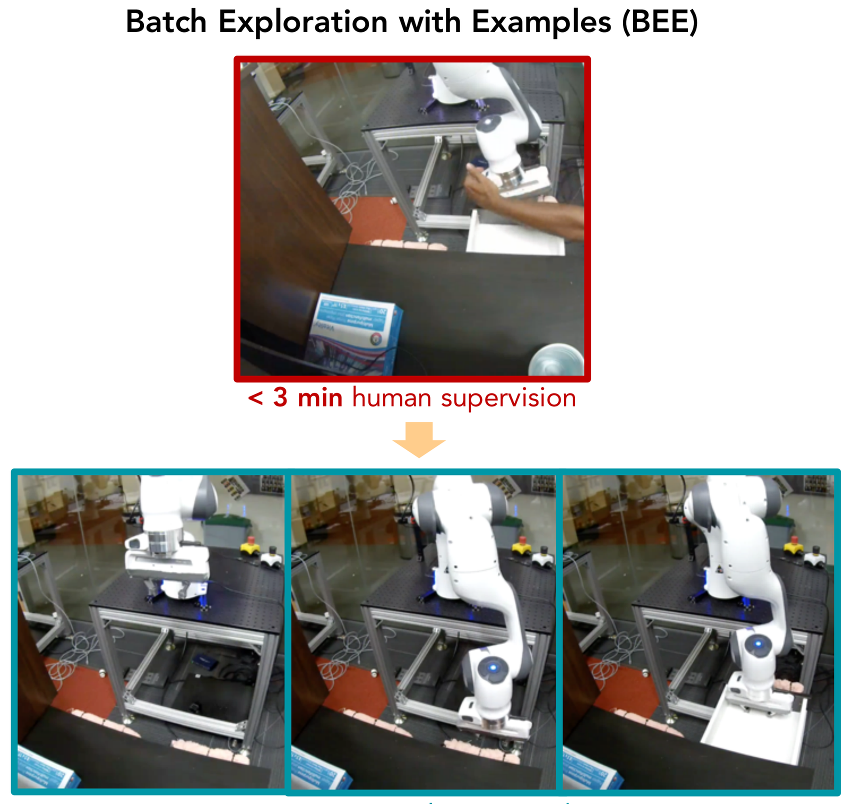
Introduction

Motivation: Leverage diverse offline datasets to learn general purpose robotic agents → requires large & meaningful robotic datasets

What makes robotic data collection challenging?

- Lack of scalability: Difficult to collect data at scale with a human in the loop
- Irrelevance of collected data: Prior novelty-seeking exploration methods are generally *task-agnostic*, so the agent tries to explore the entire state space when **only some subset of interactions are actually useful.**

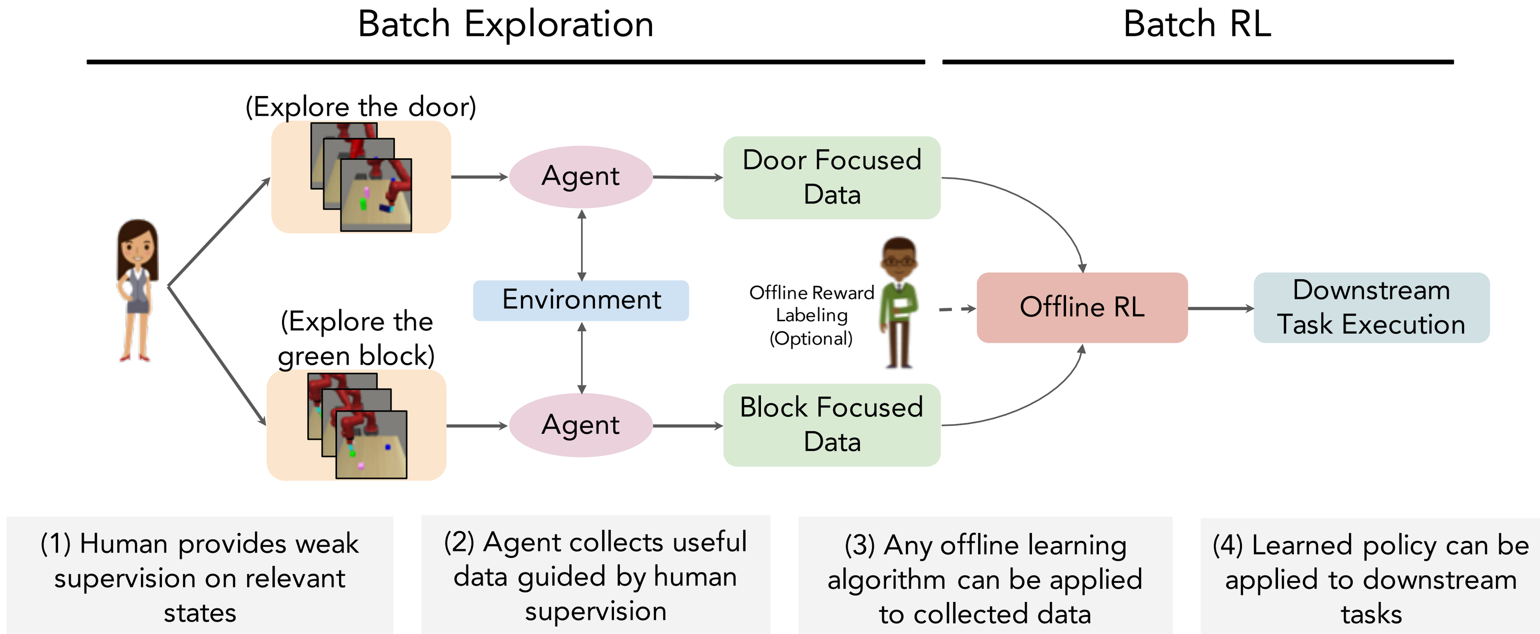
Batch Exploration with Examples (BEE)



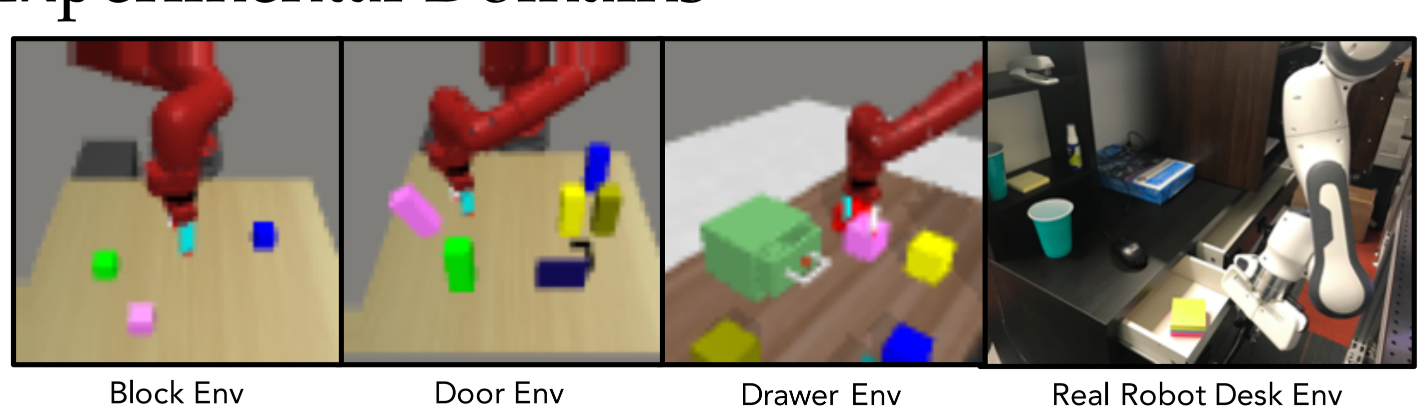
Objective: Collect large amounts of *meaningful* data for robotic manipulation without depending on a human in the loop

Key Insight: Use some *weak human supervision before data collection* to guide exploration

Batch Exploration + Batch RL Framework



Experimental Domains

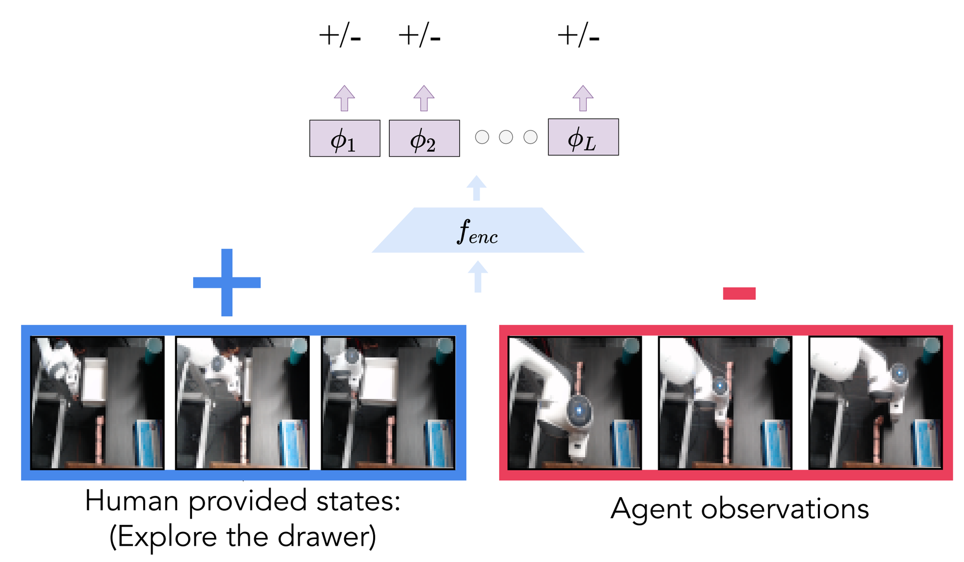


Experimental Comparisons:

- State Marginal Matching [1]: match a given target state distribution
- Model Disagreement [2], [3]: explore where an ensemble of dynamics models disagree the most

Exploration with BEE

Goal: Explore around states that the weak human supervision indicates are relevant

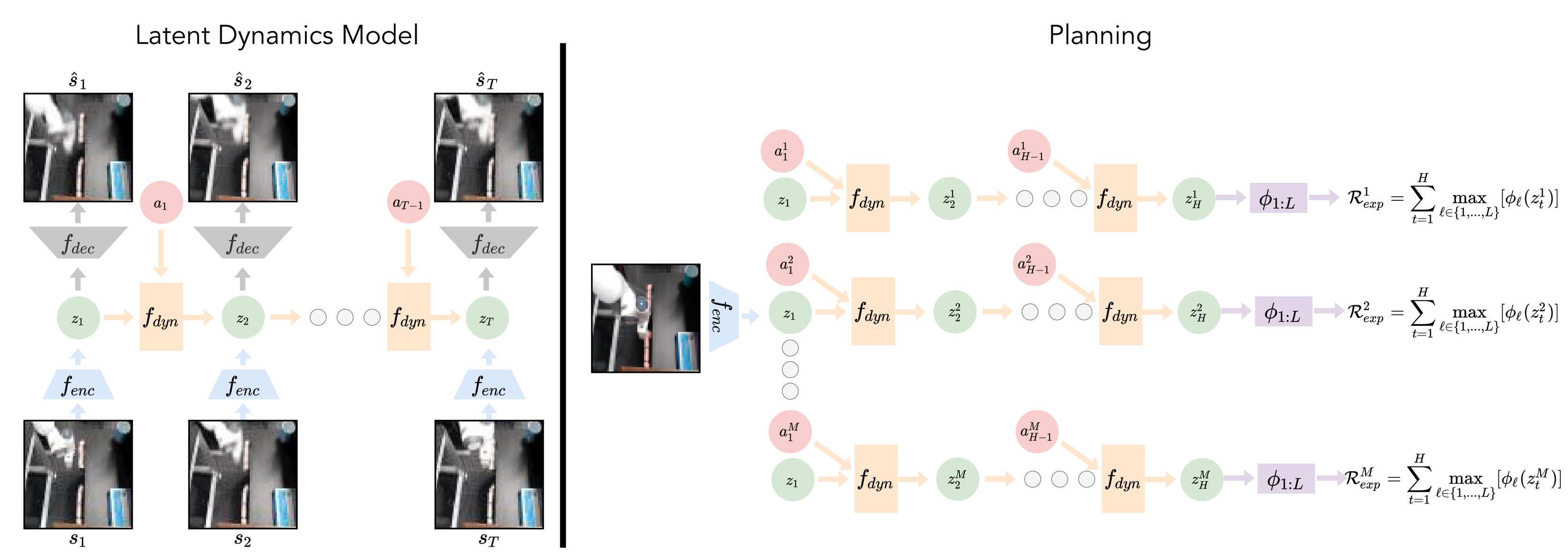


1. Identify what states are relevant:

- Learn an ensemble of relevance discriminators
- Differentiate between the human provided states (+ label) and the agents own experience (- label)

2. Plan actions to explore states identified as relevant or uncertain:

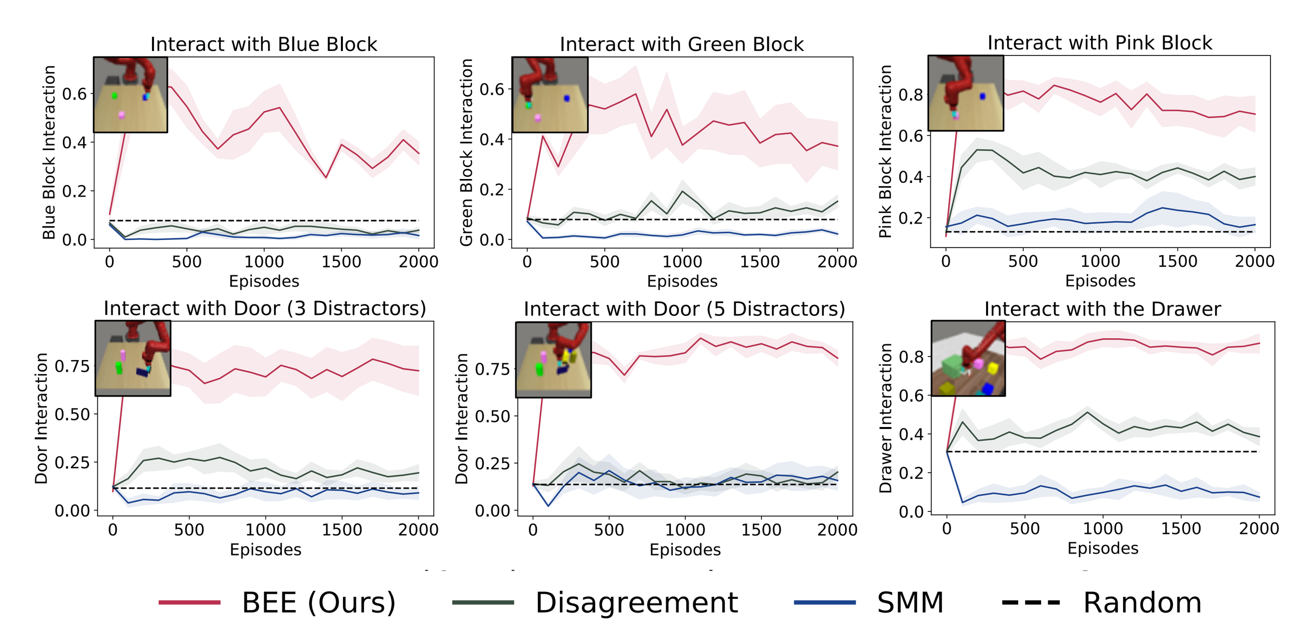
- Learn a latent dynamics model
- Perform sampling-based planning using this model, where the reward function is the max score over the ensemble of discriminators



Results in Simulated Environments

Q1: Does BEE interact more with relevant objects?

Across all settings, BEE interacts with the relevant object *more than twice as often* as the comparisons.



	Open Drawer	Push Door (3)	Push Door (5)	Push Green	Push Blue
BEE (Ours)	**0.42**	**0.63**	0.65	**0.47**	**0.50**
Disagreement	0.36	0.59	0.69	0.45	0.44
SMM	0.29	0.58	**0.70**	0.43	0.46
Random	0.31	0.60	0.65	0.45	0.45

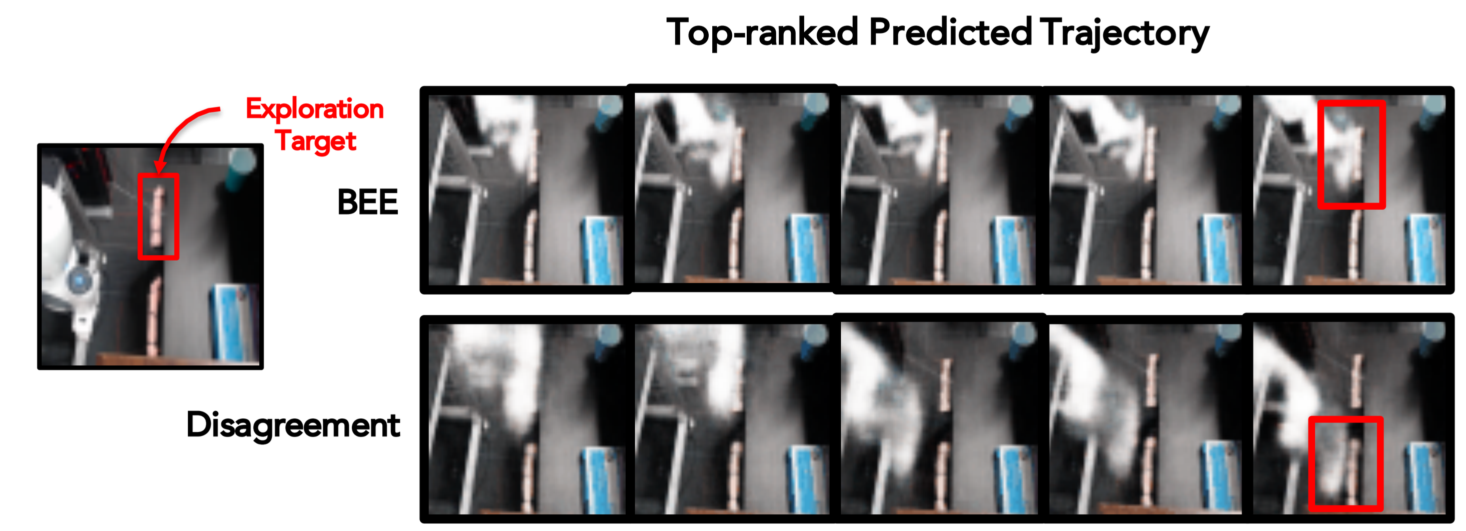
Q2: Does data from BEE enable better downstream performance?

Using data from BEE w/ visual foresight [4] for downstream tasks ***improves performance on 4 out of 5 tasks.***

Results on a Franka Robot

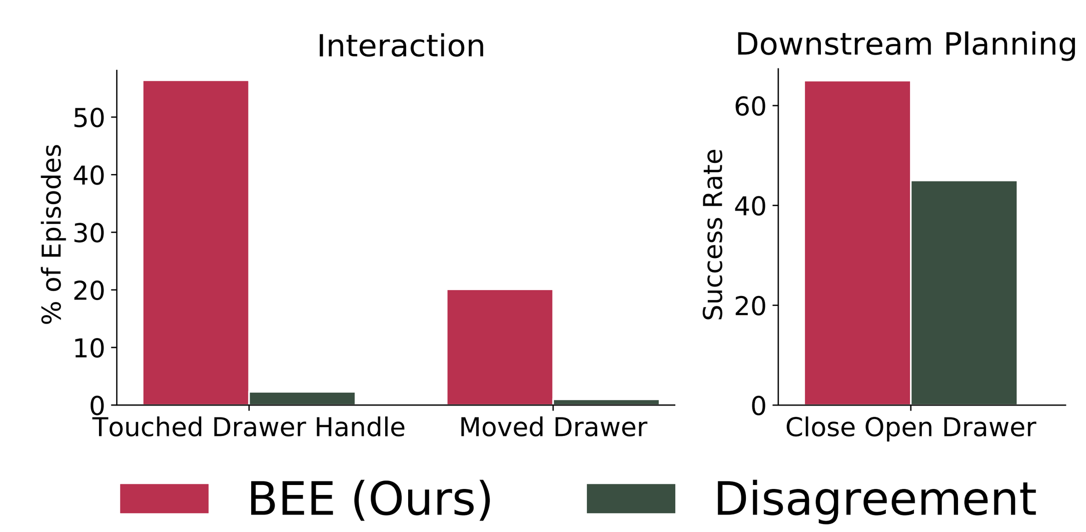
Qualitative Example

- During data collection, BEE rankings effectively discriminate target (top drawer) & non-target interaction
- High-ranked states under Disagreement do not involve the target



Q3: Is BEE effective on a real robot?

- BEE interacts w/ the target drawer ***an order of magnitude more*** than Disagreement
- Using data collected from BEE improves performance on the downstream task of closing an open drawer.



Key Takeaways

- Batch Exploration + Batch RL framework allows for a ***scalable, data-driven approach*** to robotic learning
- BEE
  (1) yields improved interaction with relevant objects
  (2) enables better offline RL on downstream tasks than using data collected via existing state-of-the-art exploration techniques
  (3) performs well on a real robotic system.

References

[1] L. Lee, B. Eysenbach, E. Parisotto, E. Xing, S. Levine, and R. Salakhutdinov. "Efficient exploration via state marginal matching." arXiv preprint arXiv:1906.05274, 2019.
[2] D. Pathak, D. Gandhi, and A. Gupta. "Self-supervised exploration via disagreement." arXiv preprint arXiv:1906.04161, 2019.
[3] R. Sekar, O. Rybkin, K. Daniilidis, P. Abbeel, D. Hafner, and D. Pathak.. "Planning to Explore via Self-Supervised World Models." arXiv preprint arXiv:2005.05960, 2020.
[4] C. Finn and S. Levine. "Deep visual foresight for planning robot motion." IEEE International Conference on Robotics and Automation (ICRA), 2017.