Planning, reasoning, and generalisation in deep learning

Jessica B. Hamrick
jhamrick@google.com
Google DeepMind

The Royal Society
“Beyond the symbols vs signals debate”
28 October 2024

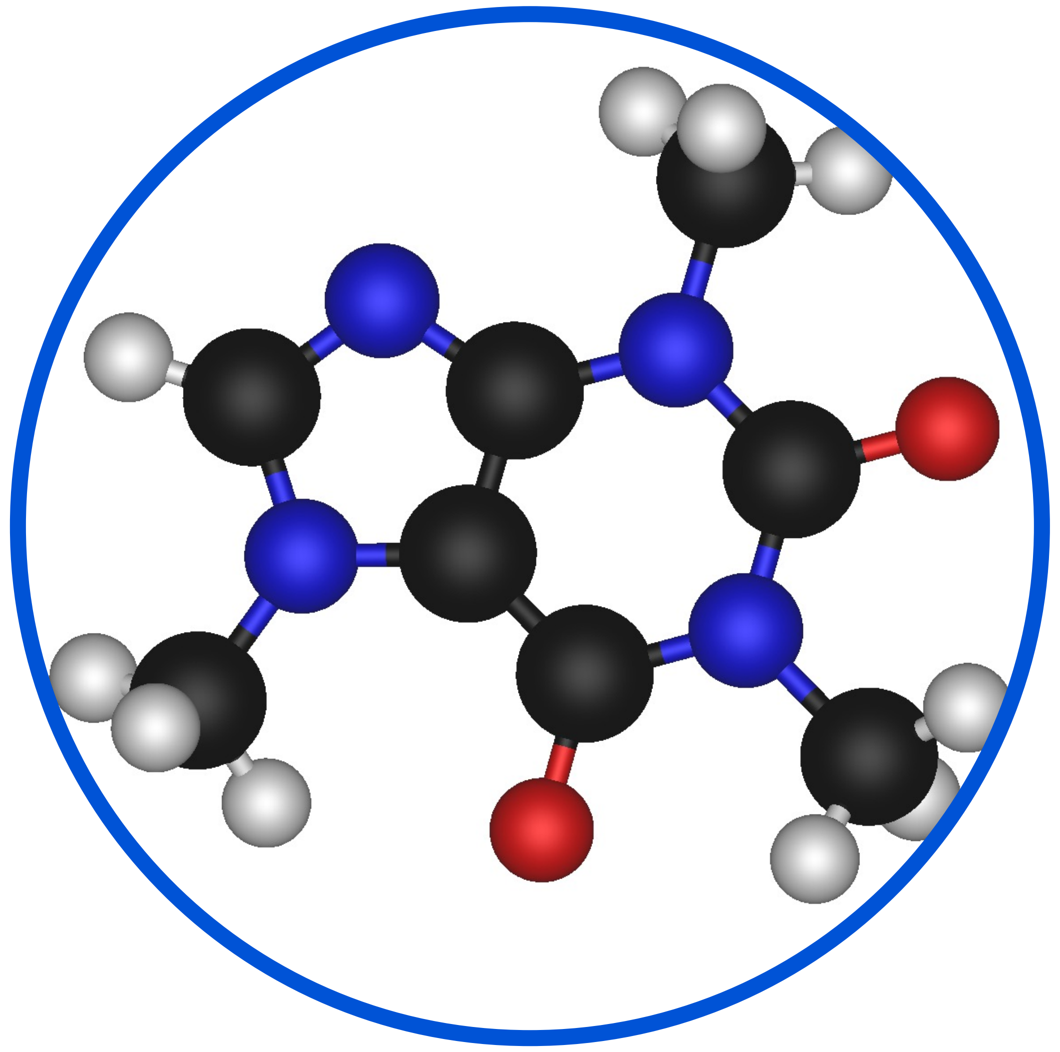
Reasoning with a world model

"If the organism carries a ***'small-scale model' of external reality*** *and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilise the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it."*

–Kenneth Craik, *The Nature of Explanation* (1943)

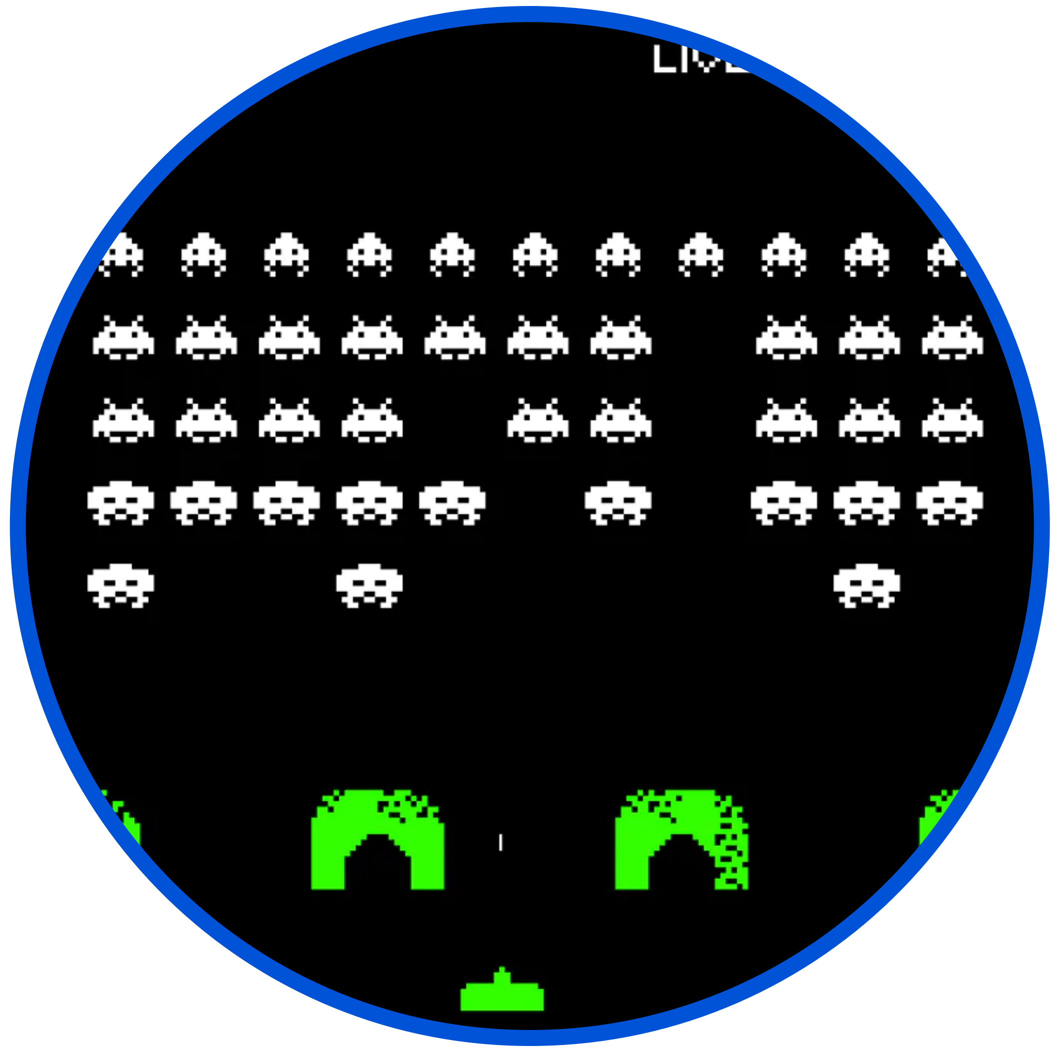
Silver et al. (2016)

OpenAI et al. (2019)



Segler et al. (2018)

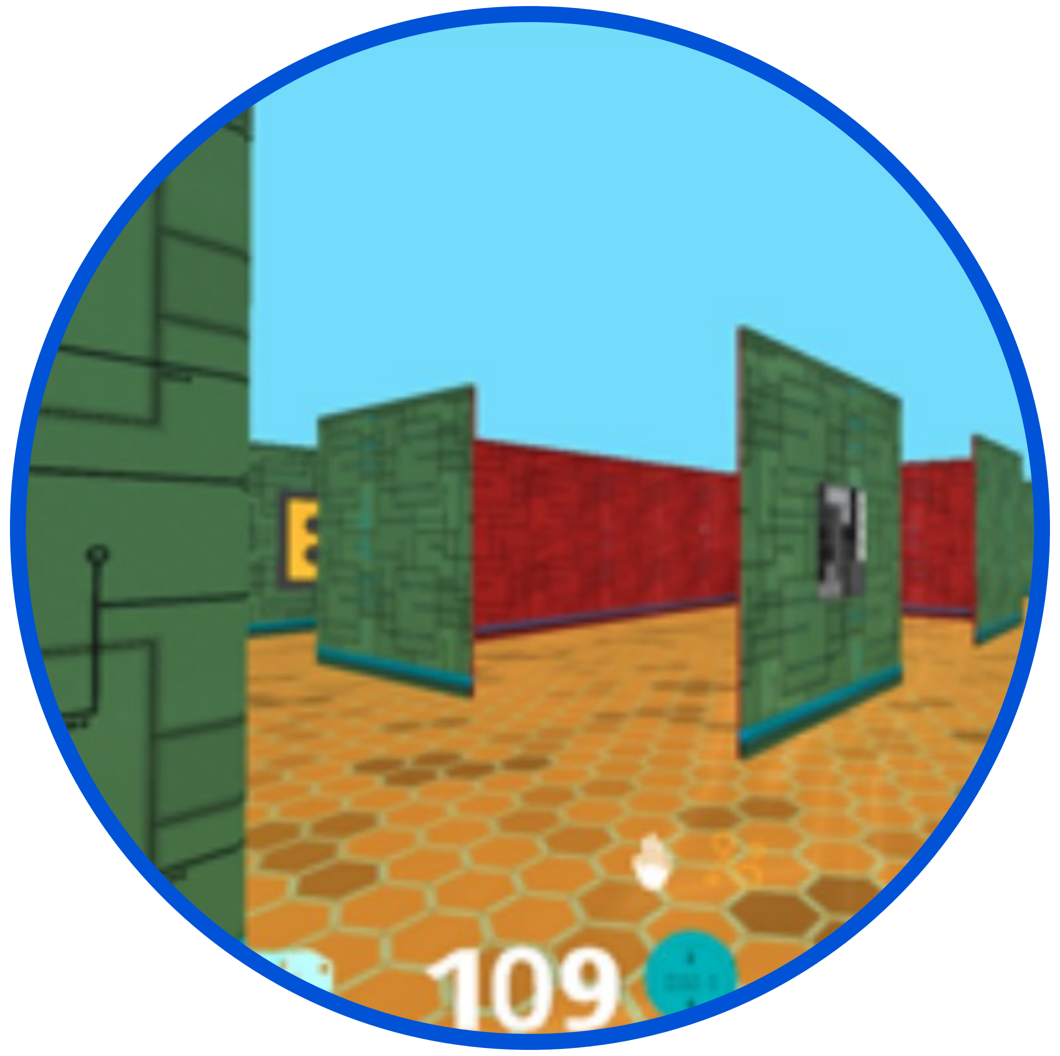
Finn et al. (2018)



Schrittwieser et al. (2020)

Luo et al. (2019)

Weber et al. (2017)



Hafner et al. (2019)

The promise of model-based RL

"Model-free algorithms are in turn far from the state of the art in domains that require ***precise and sophisticated lookahead***, such as chess and Go"
-Schrittwieser et al. (2019)

"Model-based planning is an essential ingredient of human intelligence, enabling ***flexible adaptation*** to new tasks and goals"
-Lake et al. (2016)

"By employing search, we can find strong move sequences potentially ***far away*** from the apprentice policy, accelerating learning in complex scenarios"
-Anthony et al. (2017)

"...a flexible and general strategy such as mental simulation allows us to reason about a wide range of scenarios, even ***novel*** ones..."
-Hamrick (2017)

"....predictive models can enable a real robot to manipulate ***previously unseen*** objects and solve new tasks"
-Ebert et al. (2018)

"...[models] enable better ***generalization*** across states, remain valid across tasks in the same environment, and exploit additional unsupervised learning signals..."
-Weber et al. (2017)

Generalization & transfer

Degree of transfer	Identical observations	Similar observations	Novel observations
Novel rewards	e.g., push rather than pull a familiar object	e.g., look for a different type of object in a new maze	e.g., figure out how to use a completely new object
Identical rewards	most of 'classic' RL (including model-based)	e.g., different maze layout	e.g., build a taller tower than during training

Degree of generalization

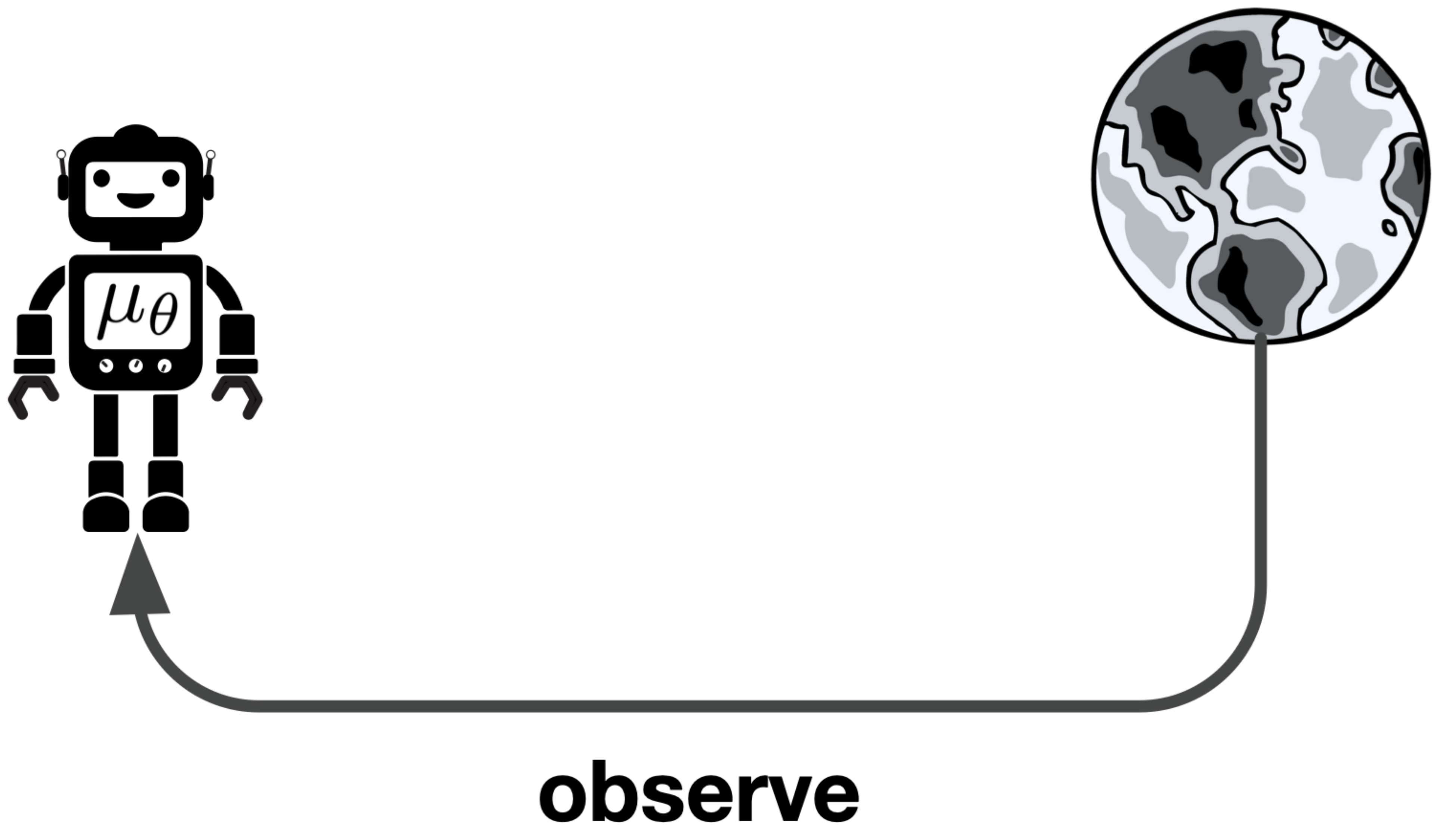
Plan for the talk

1. What is model-based RL?
2. Lessons from studying generalization & transfer in MBRL
3. The missing ingredient for neurosymbolic AI

What is model-based RL?

Model free RL: act according to a policy and update the policy from experience

Model based RL: construct **plans** using a model of the world, and use those to update the policy

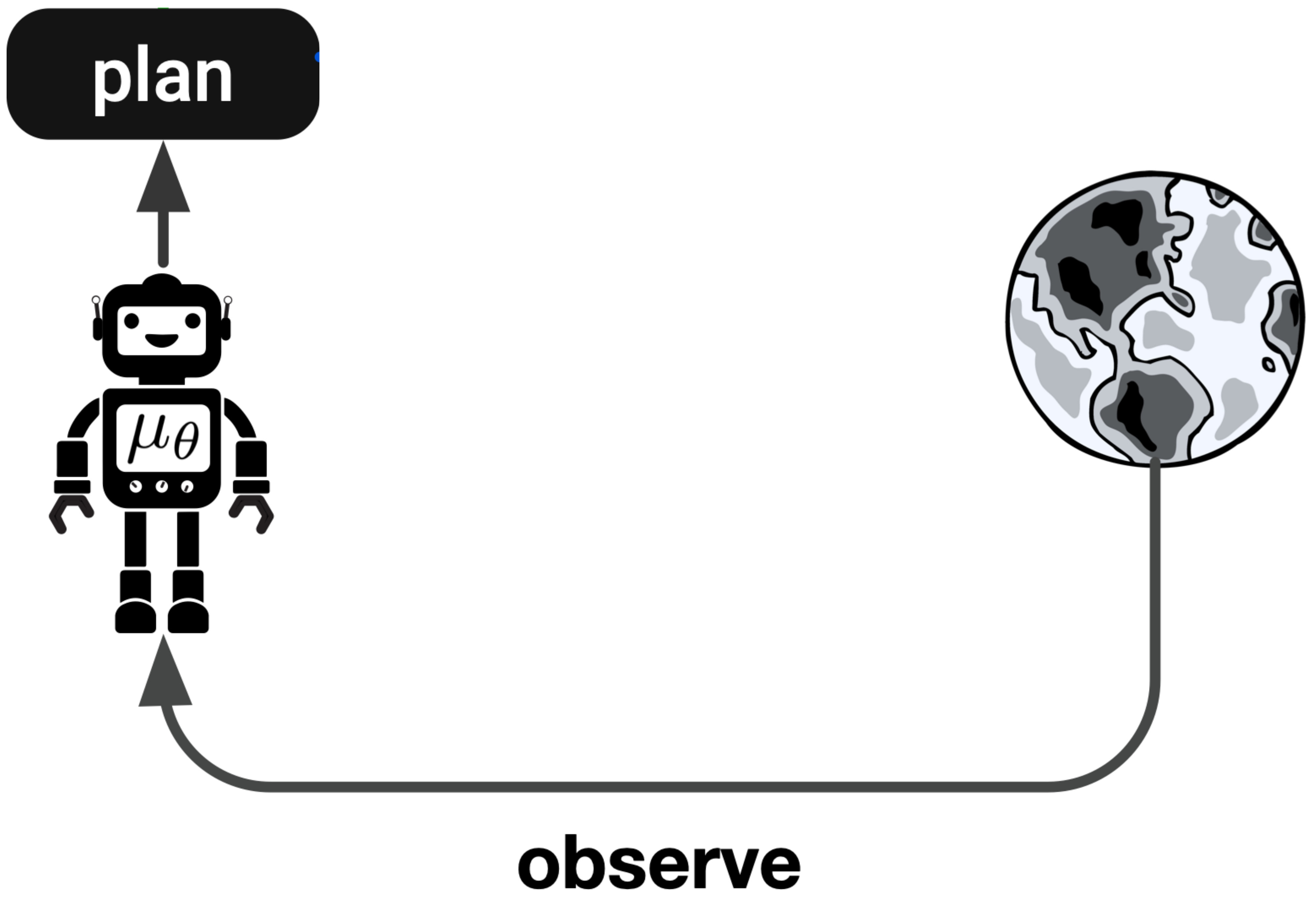


What is model-based RL?

Model free RL: act according to a policy and update the policy from experience

Model based RL: construct **plans** using a model of the world, and use those to update the policy

policy: where to search?
model: what will happen?

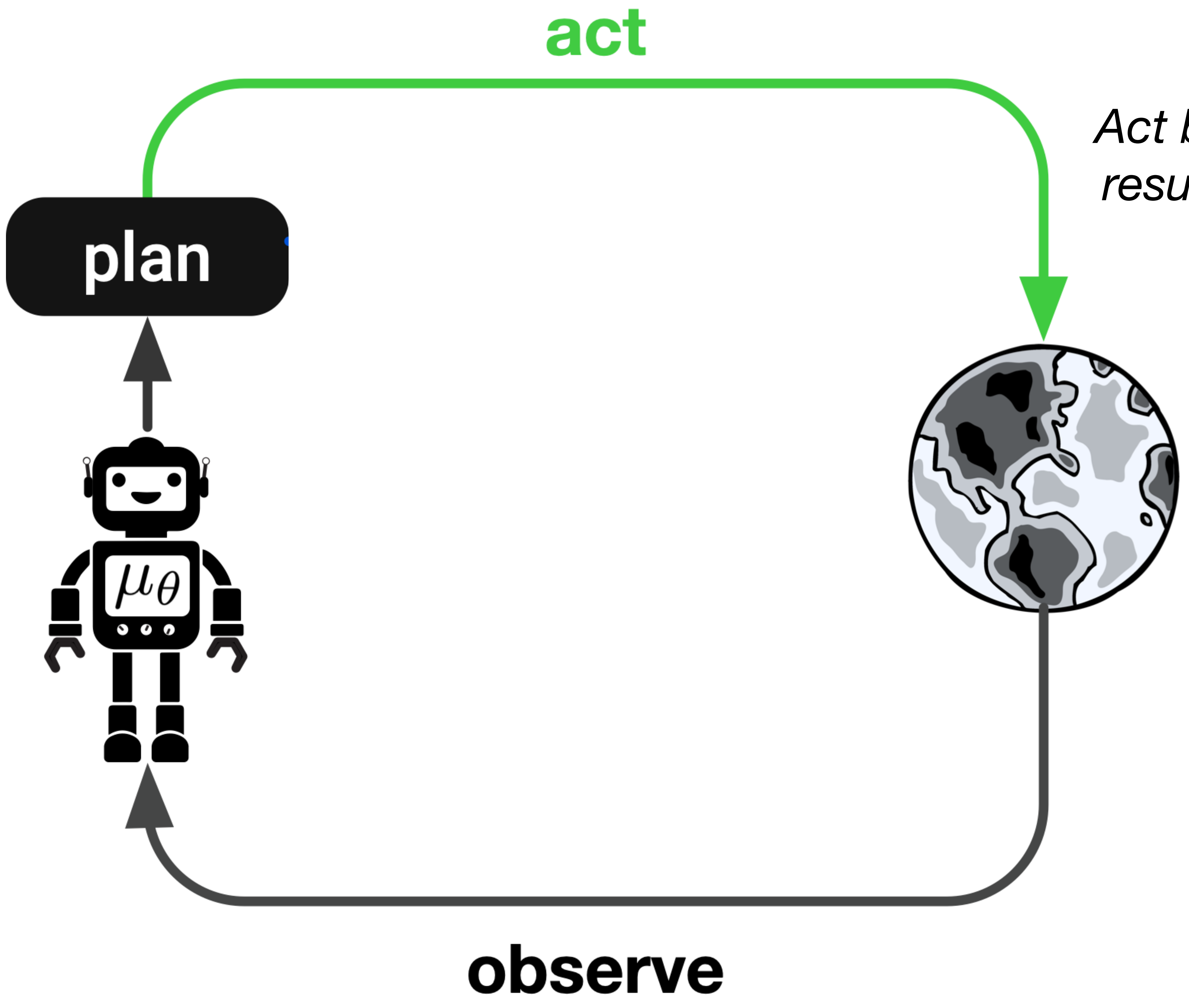


What is model-based RL?

Model free RL: act according to a policy and update the policy from experience

Model based RL: construct **plans** using a model of the world, and use those to update the policy

policy: where to search?
model: what will happen?



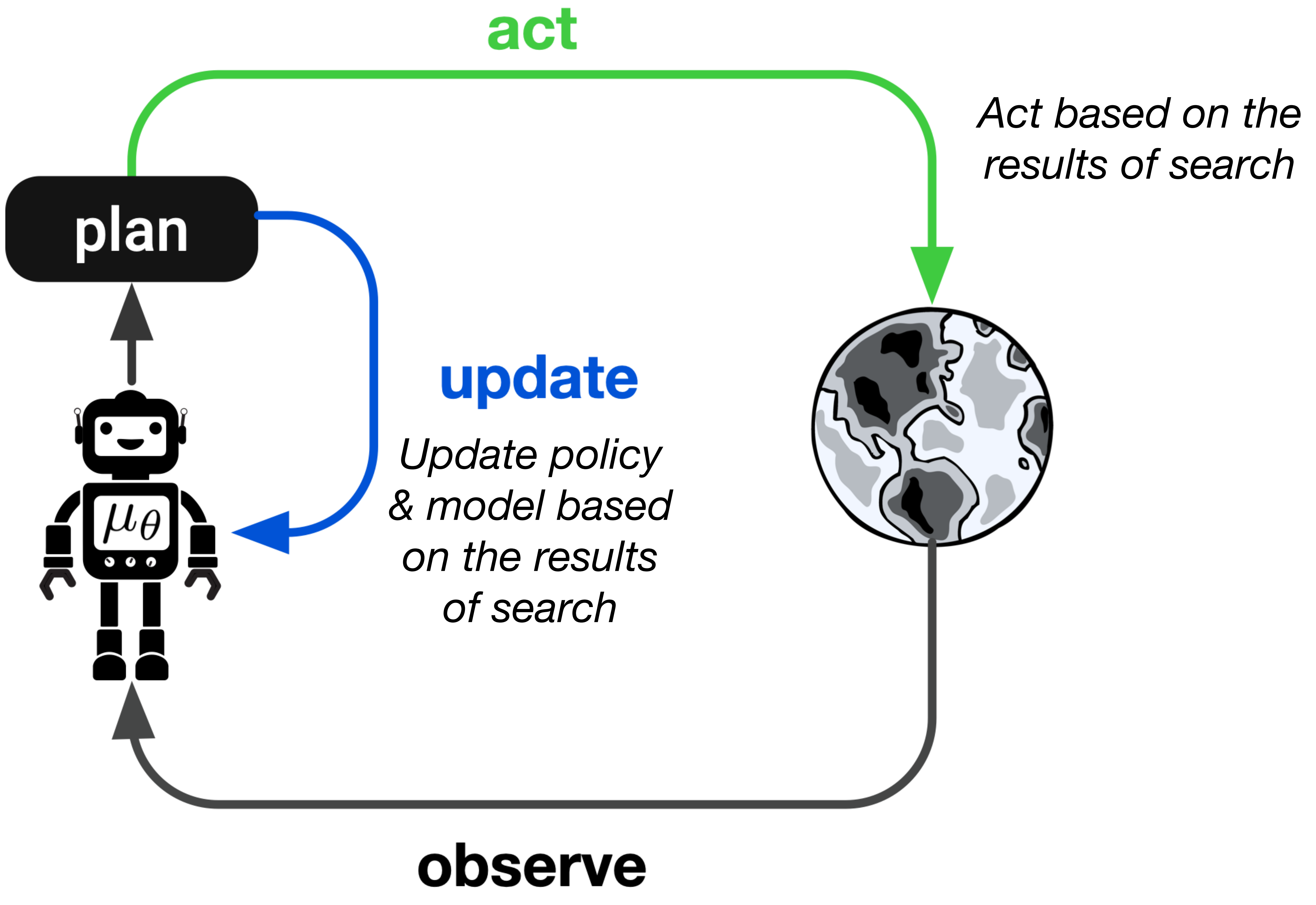
Act based on the results of search

What is model-based RL?

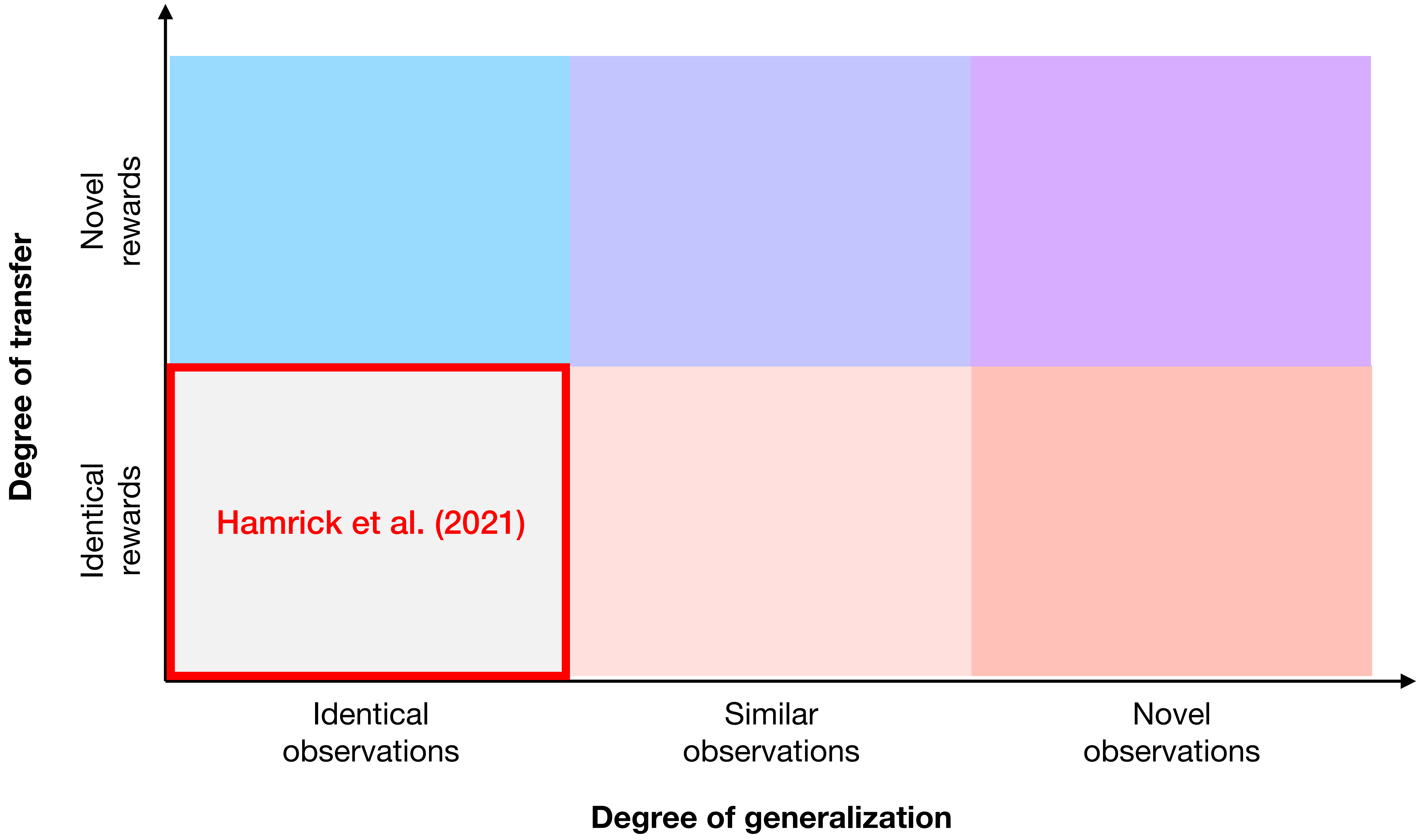
Model free RL: act according to a policy and update the policy from experience

Model based RL: construct **plans** using a model of the world, and use those to update the policy

policy: where to search?
model: what will happen?

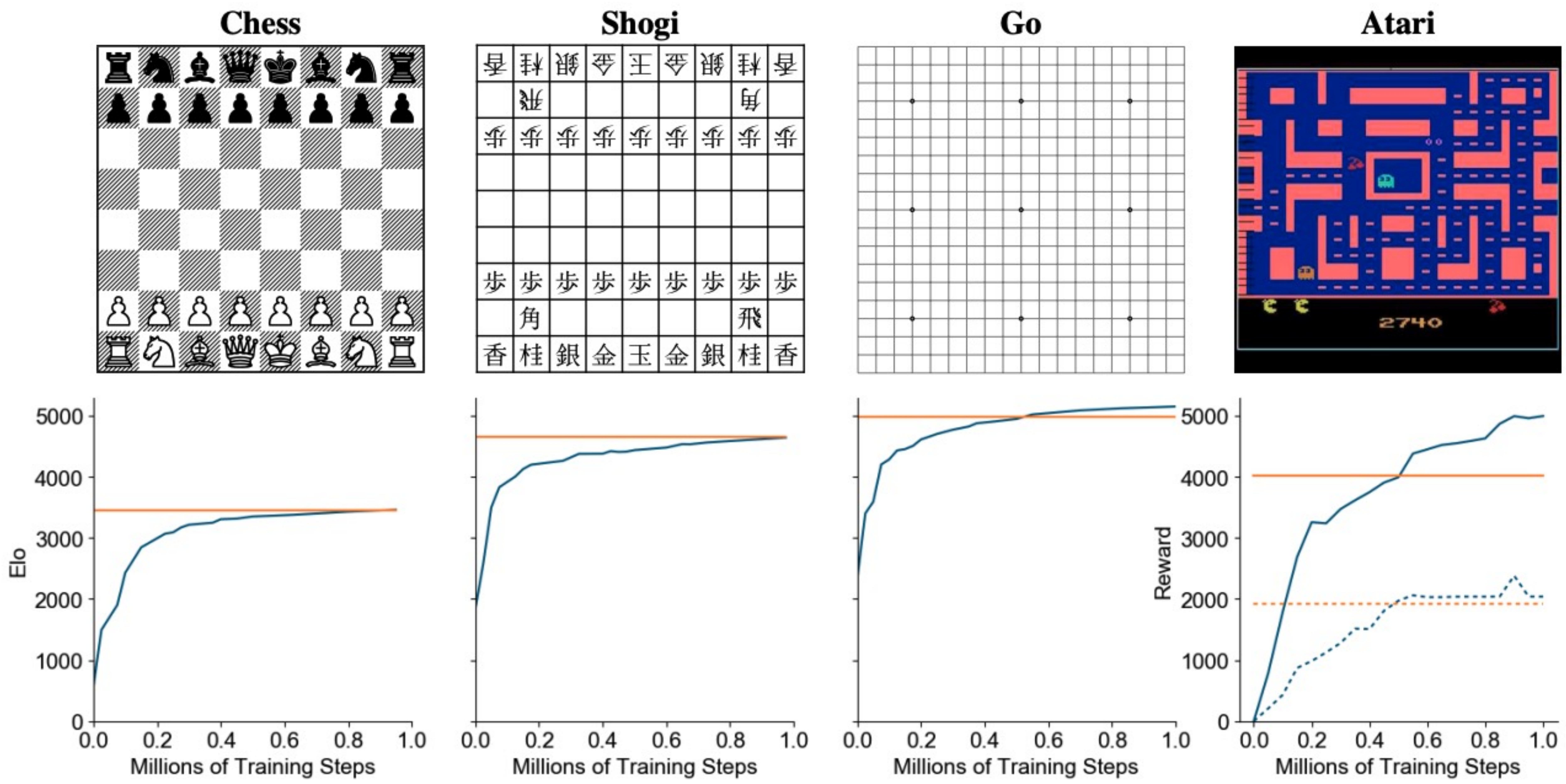


Lessons in generalization & transfer



MuZero

Schrittwieser et al. (2019)



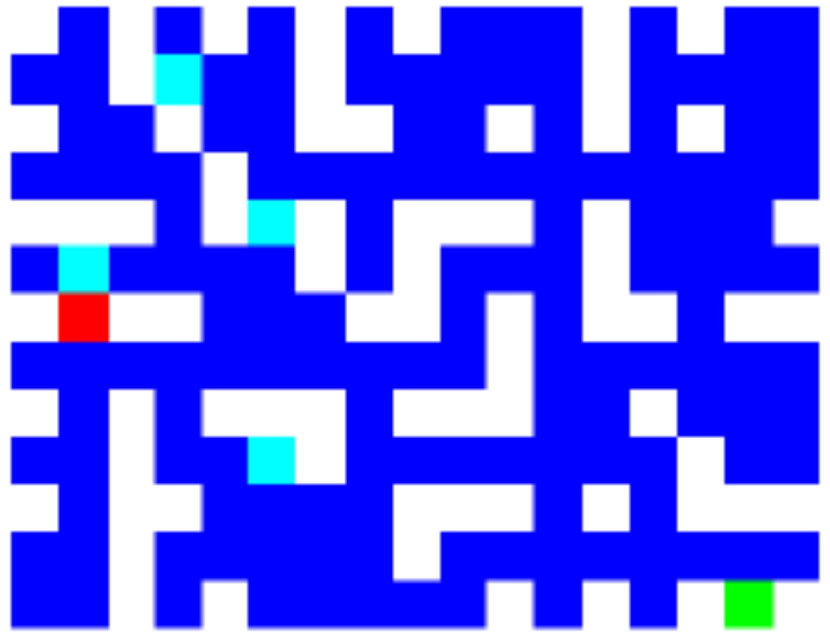
Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.

Environments

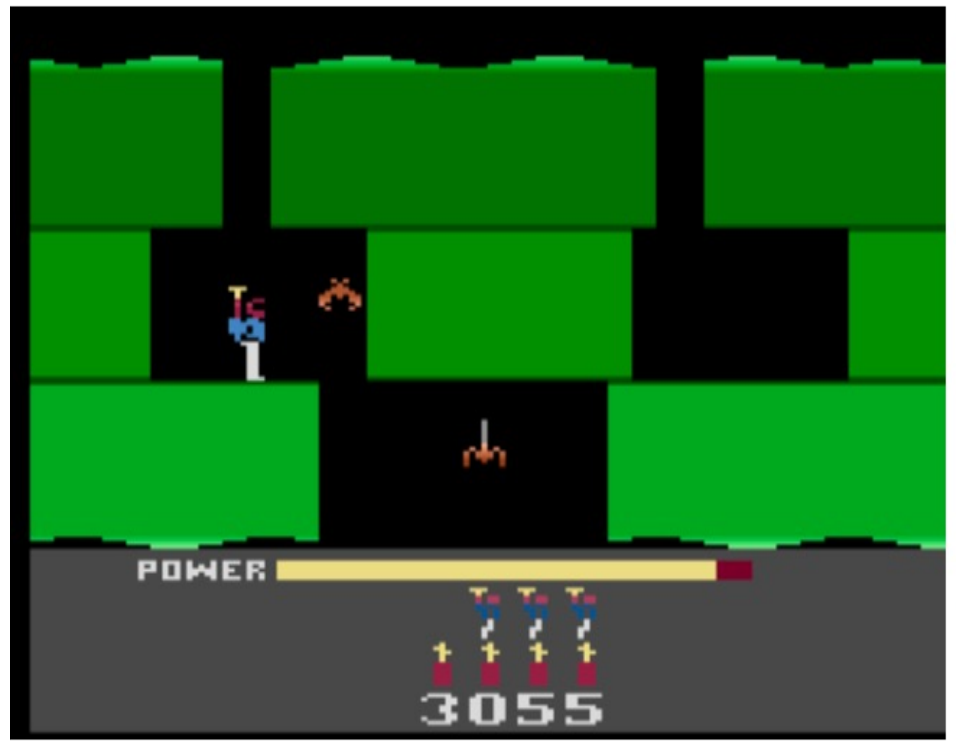
Acrobot
(Swingup Sparse)

Cheetah
(Run)

Humanoid
(Stand)

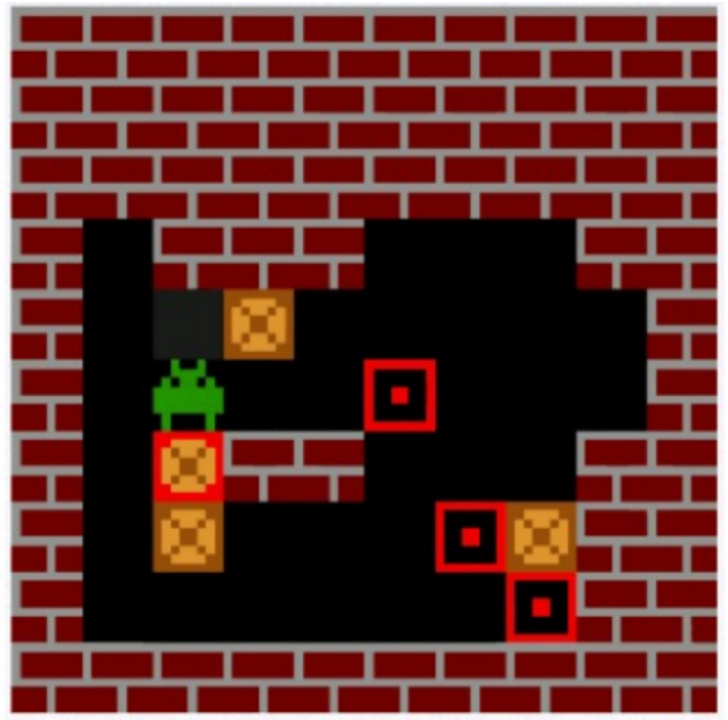


Minipacman
(Procedural)

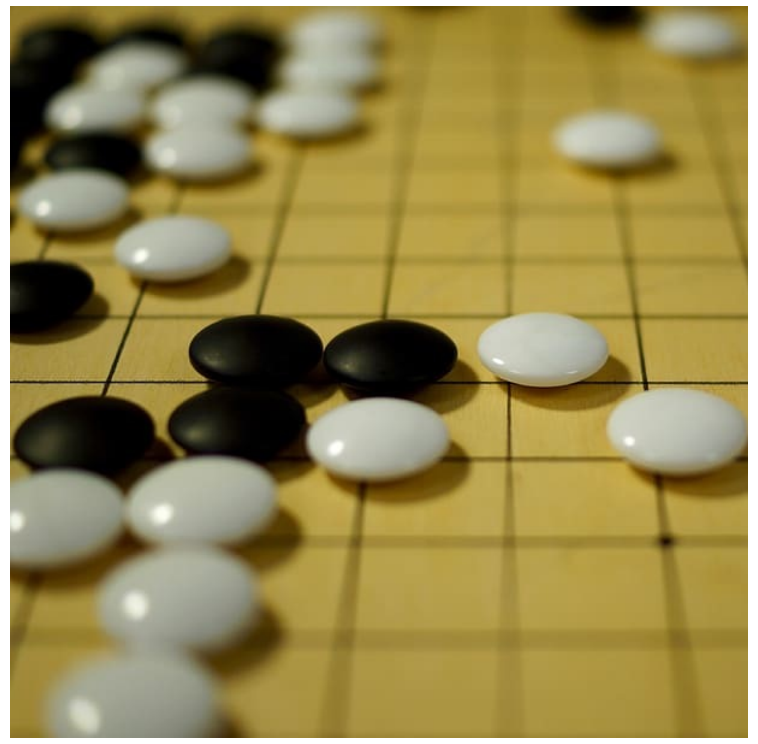


Hero

Ms. Pacman



Sokoban



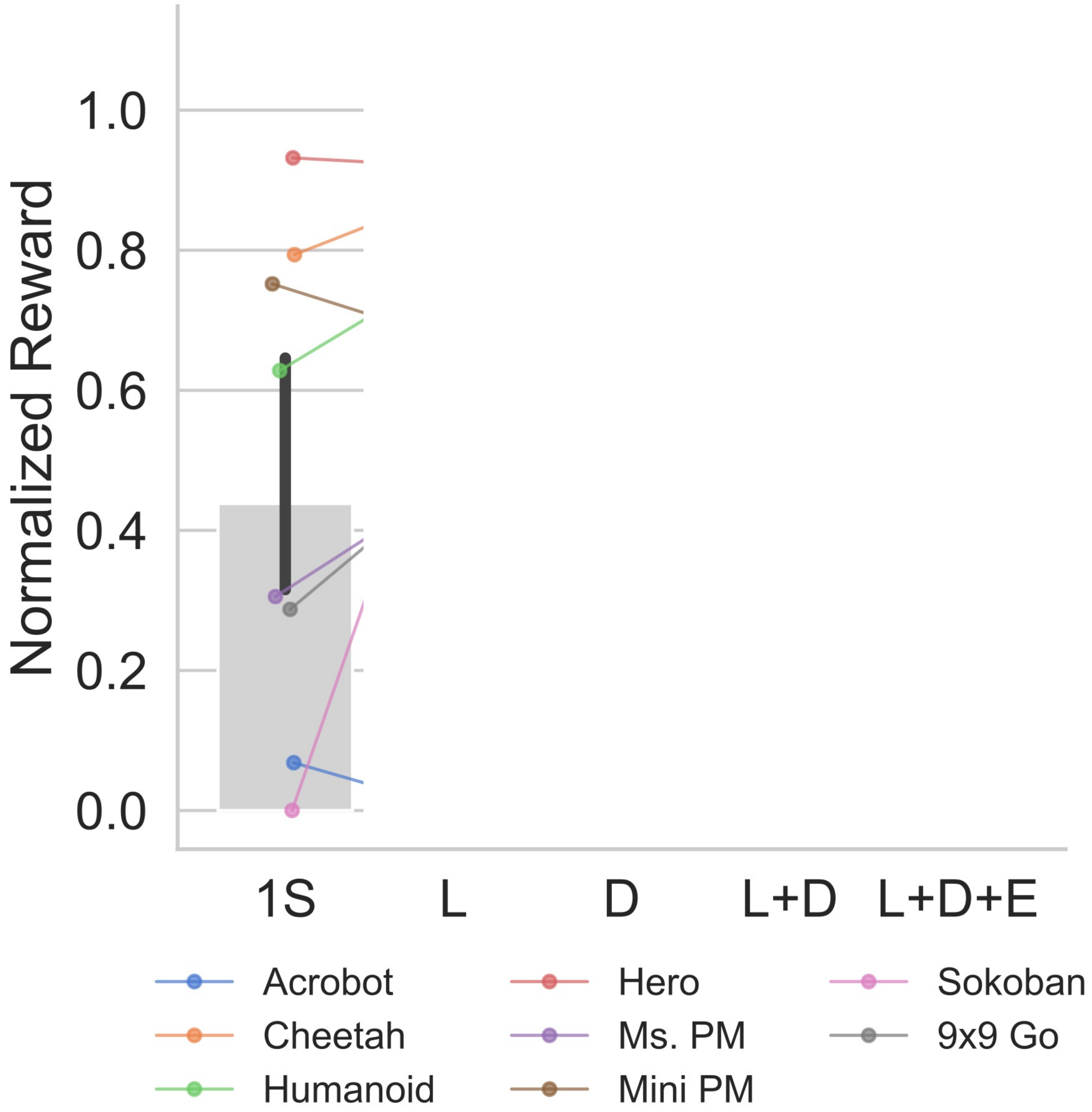
9x9 Go

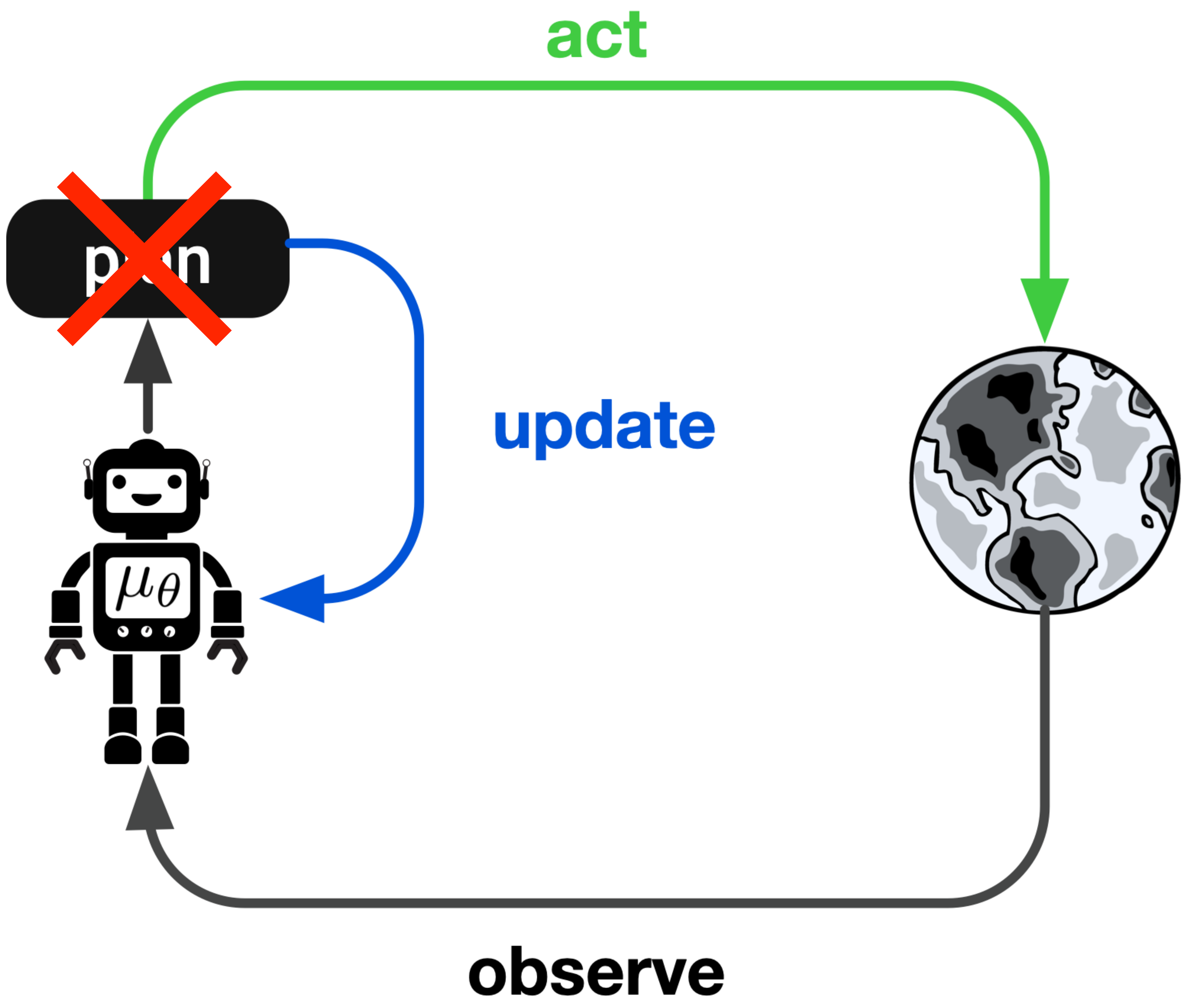
Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.

Using search in different ways

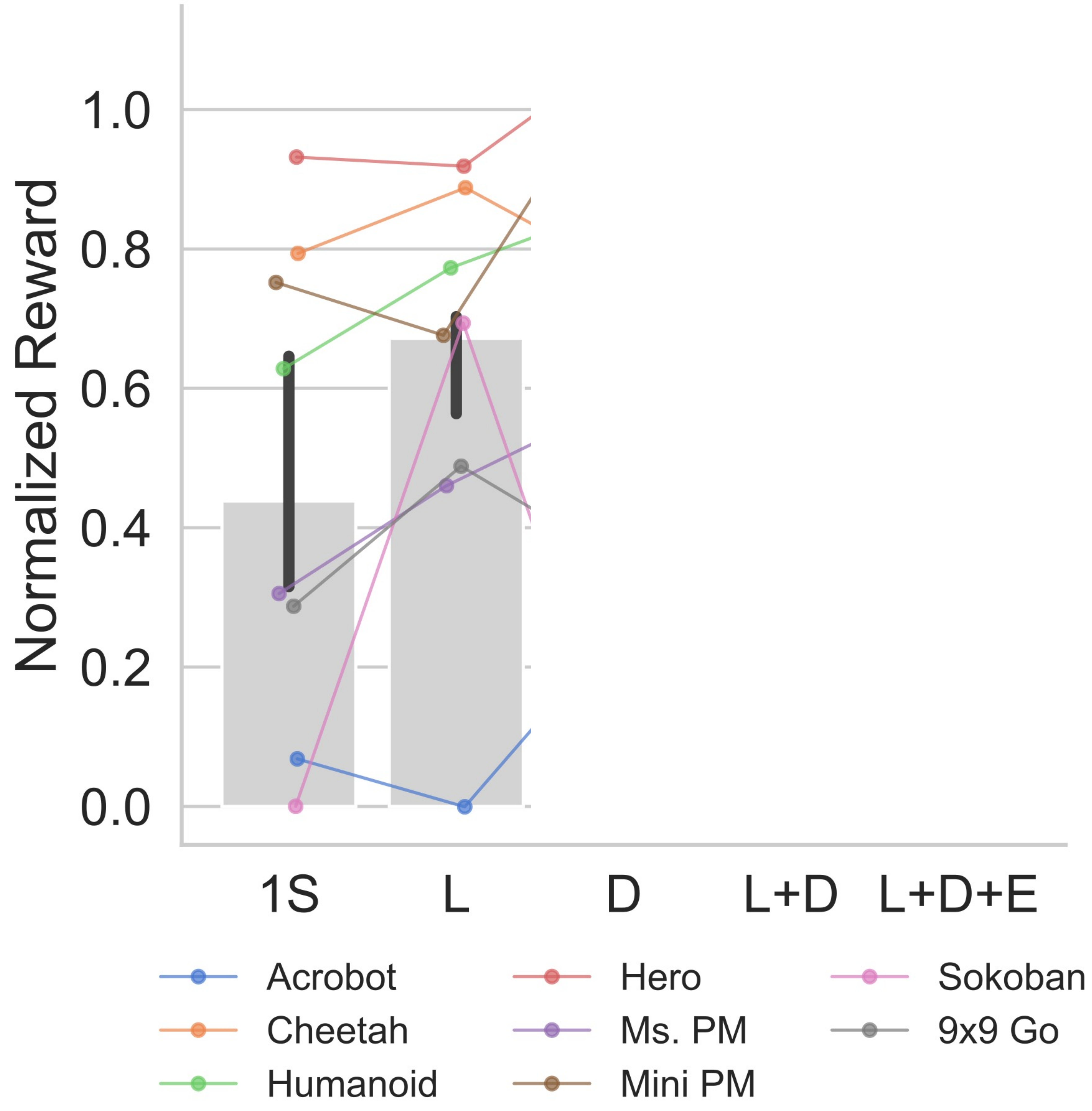
	Train Update	Train Act	Test Act
One-Step	Model-free	Model-free	Model-free
Learn	**Model-based**	Model-free	Model-free
Data	Model-free	**Model-based**	Model-free
Learn+Data	**Model-based**	**Model-based**	Model-free
Learn+Data+Eval (vanilla MuZero)	**Model-based**	**Model-based**	**Model-based**

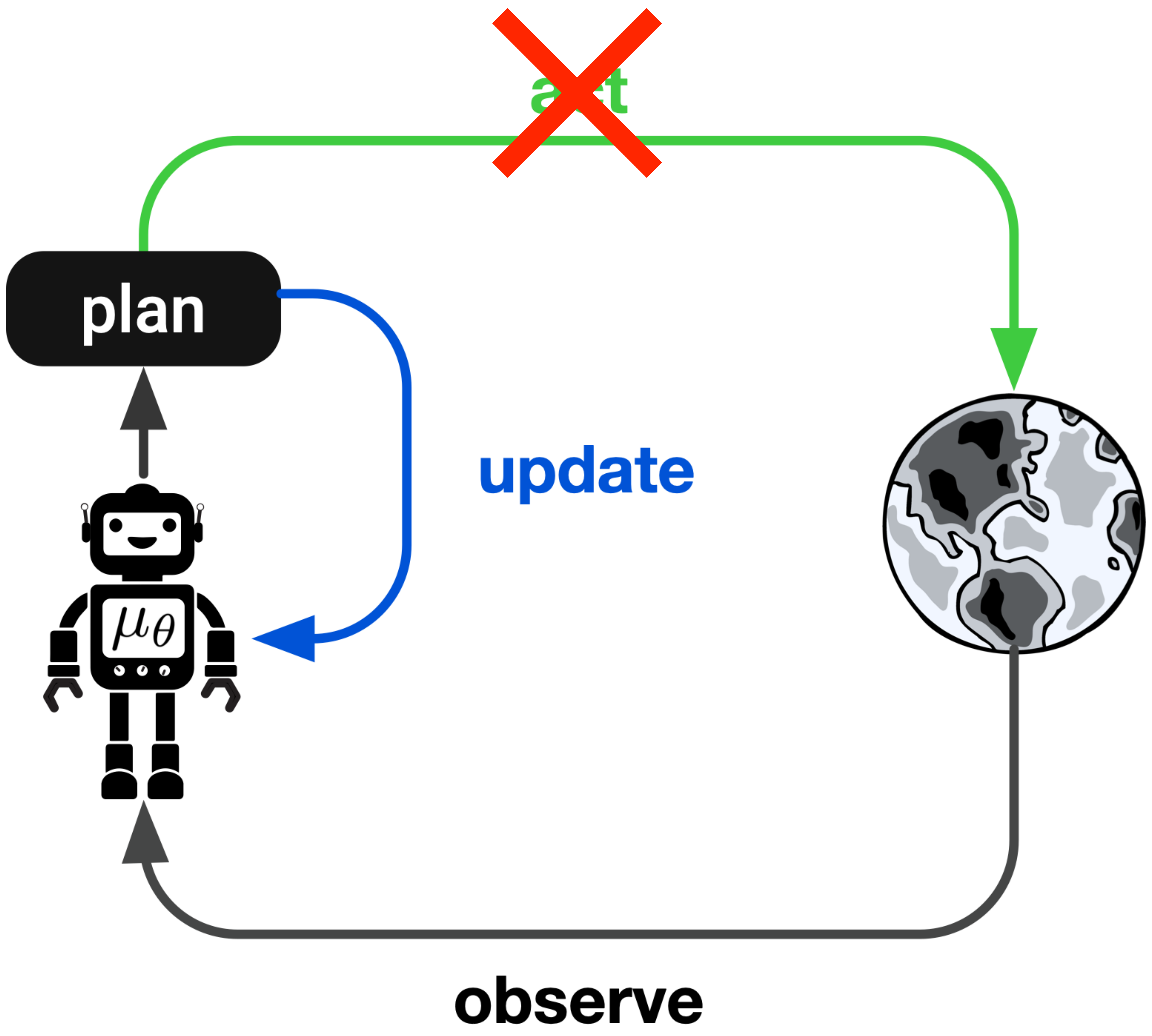
Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.



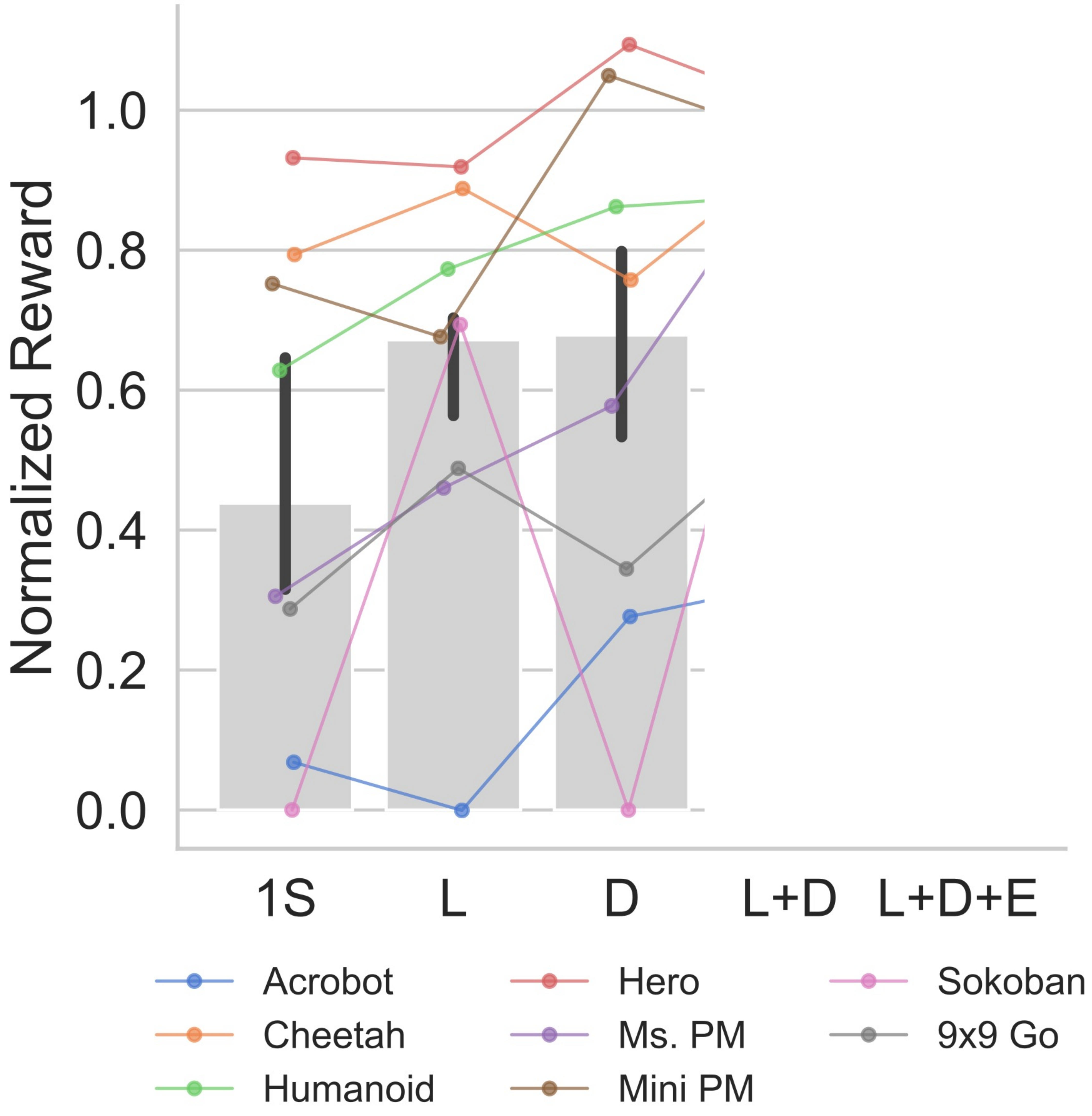


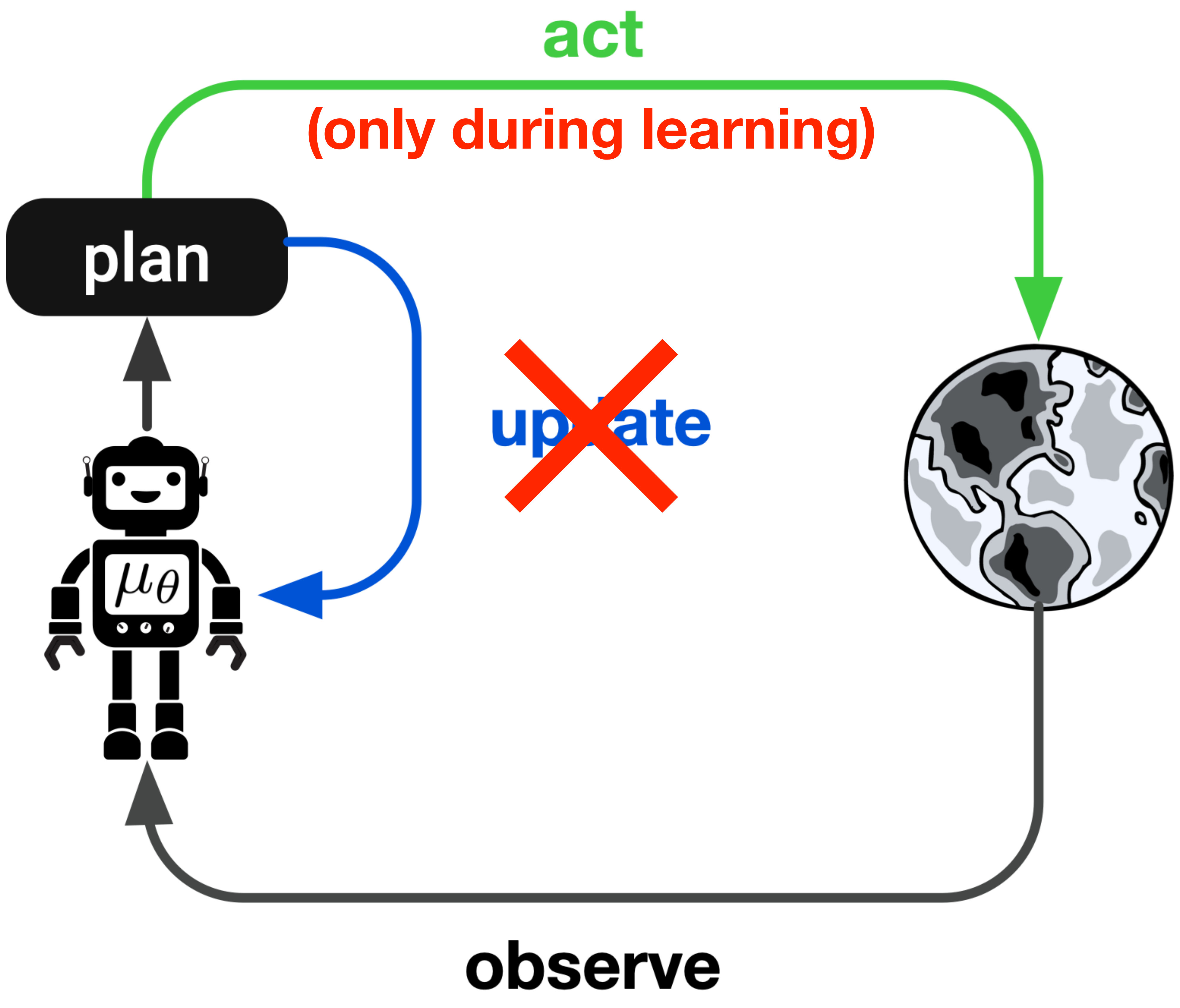
Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.



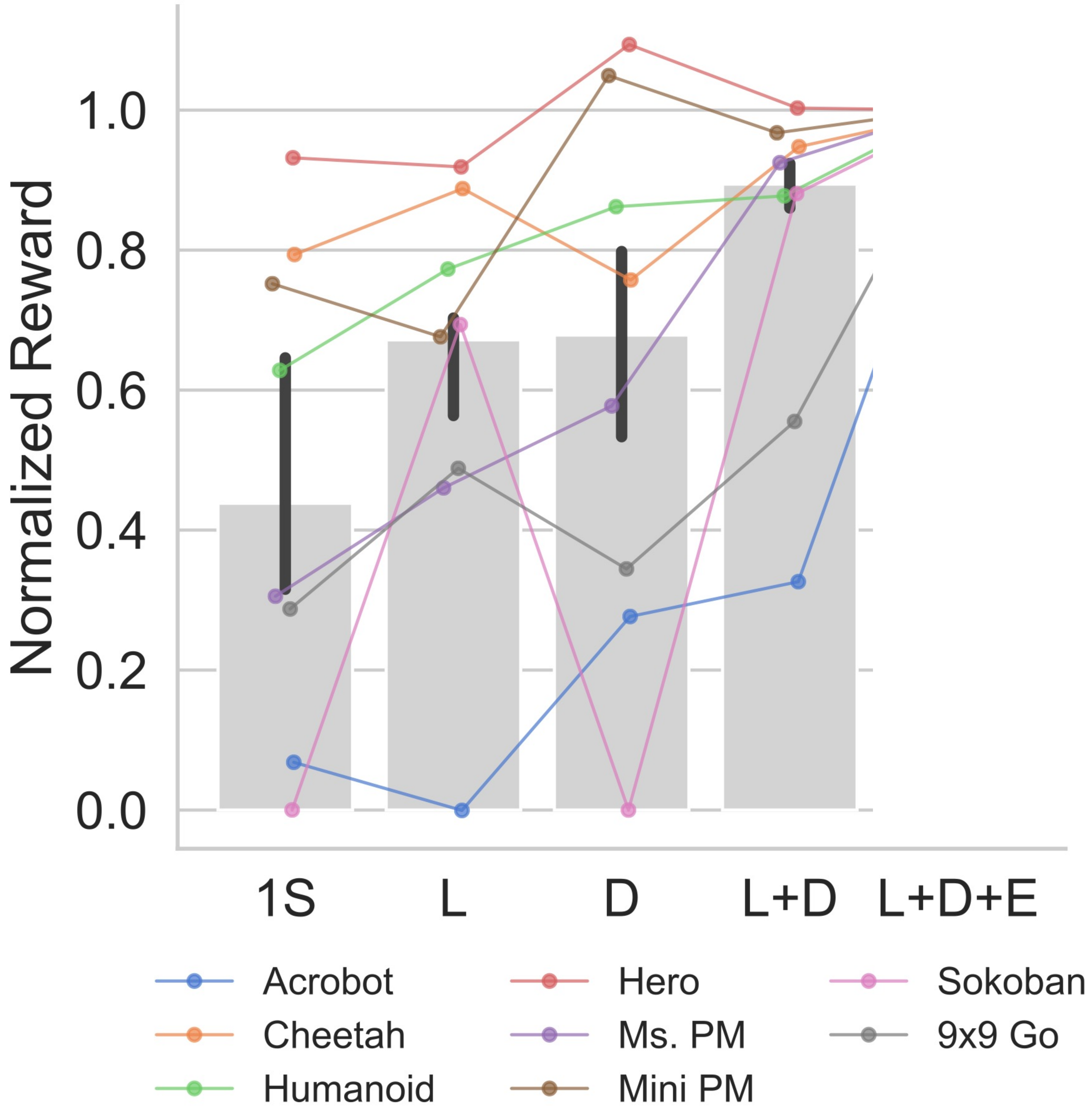


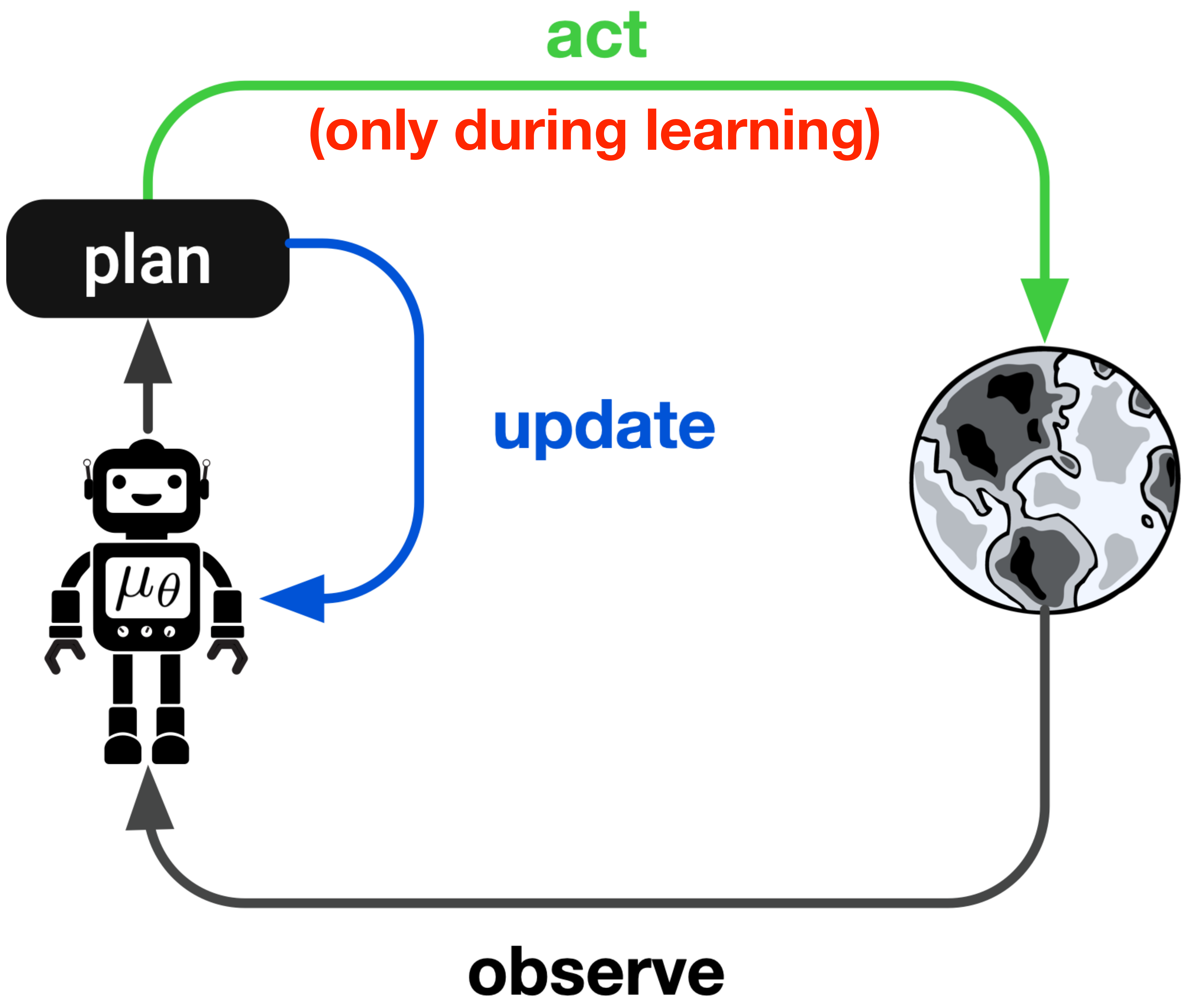
Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.



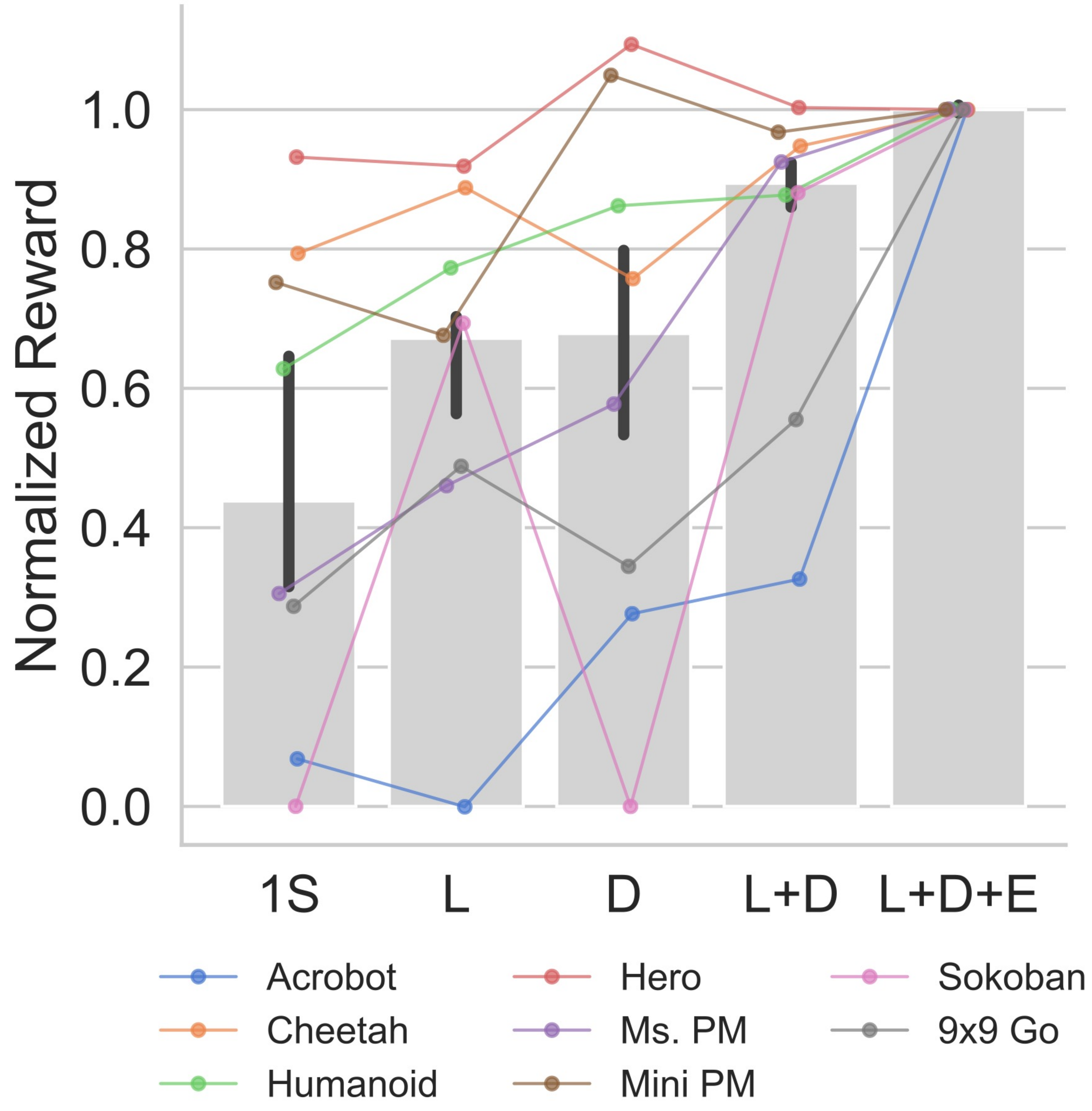


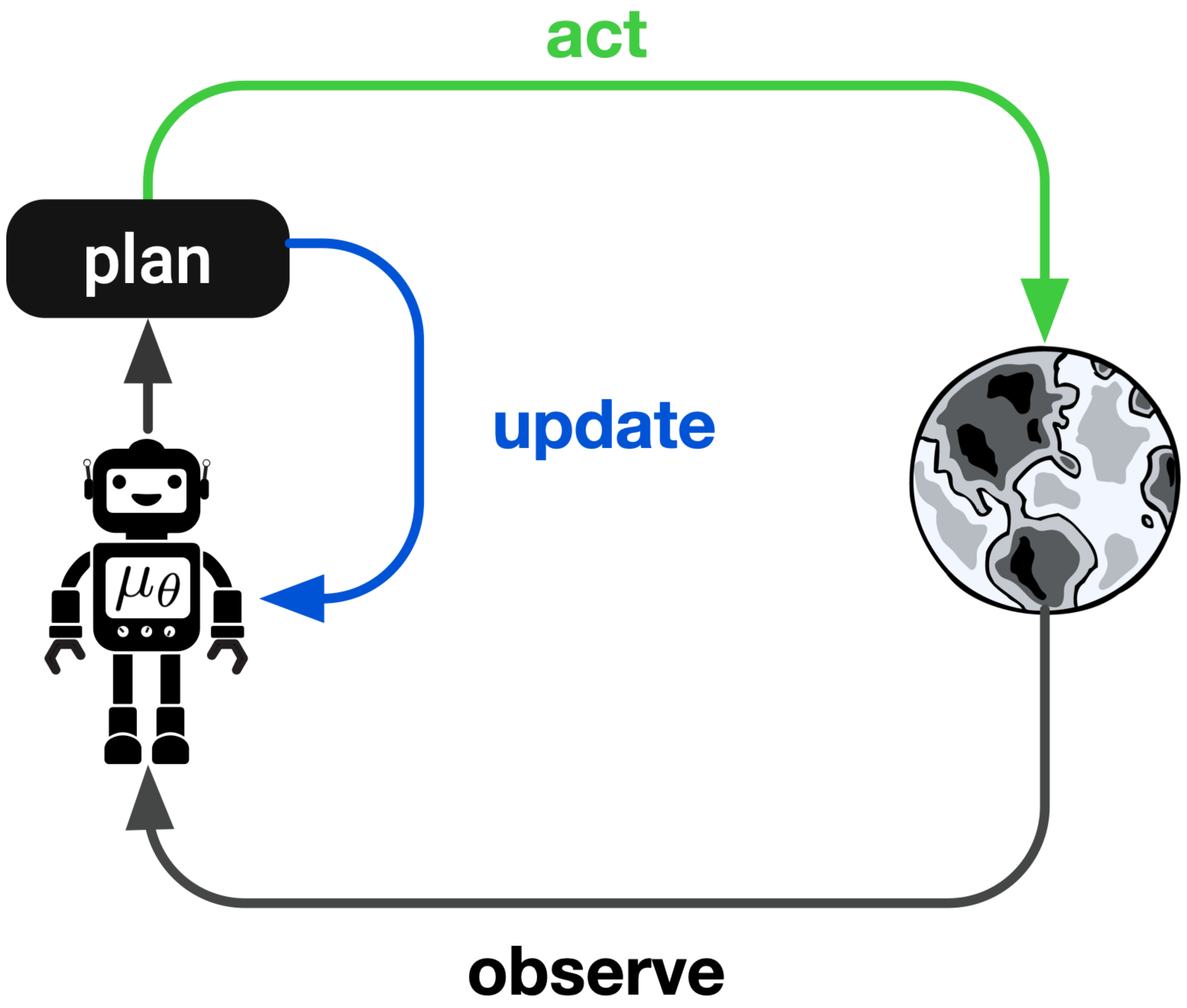
Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.



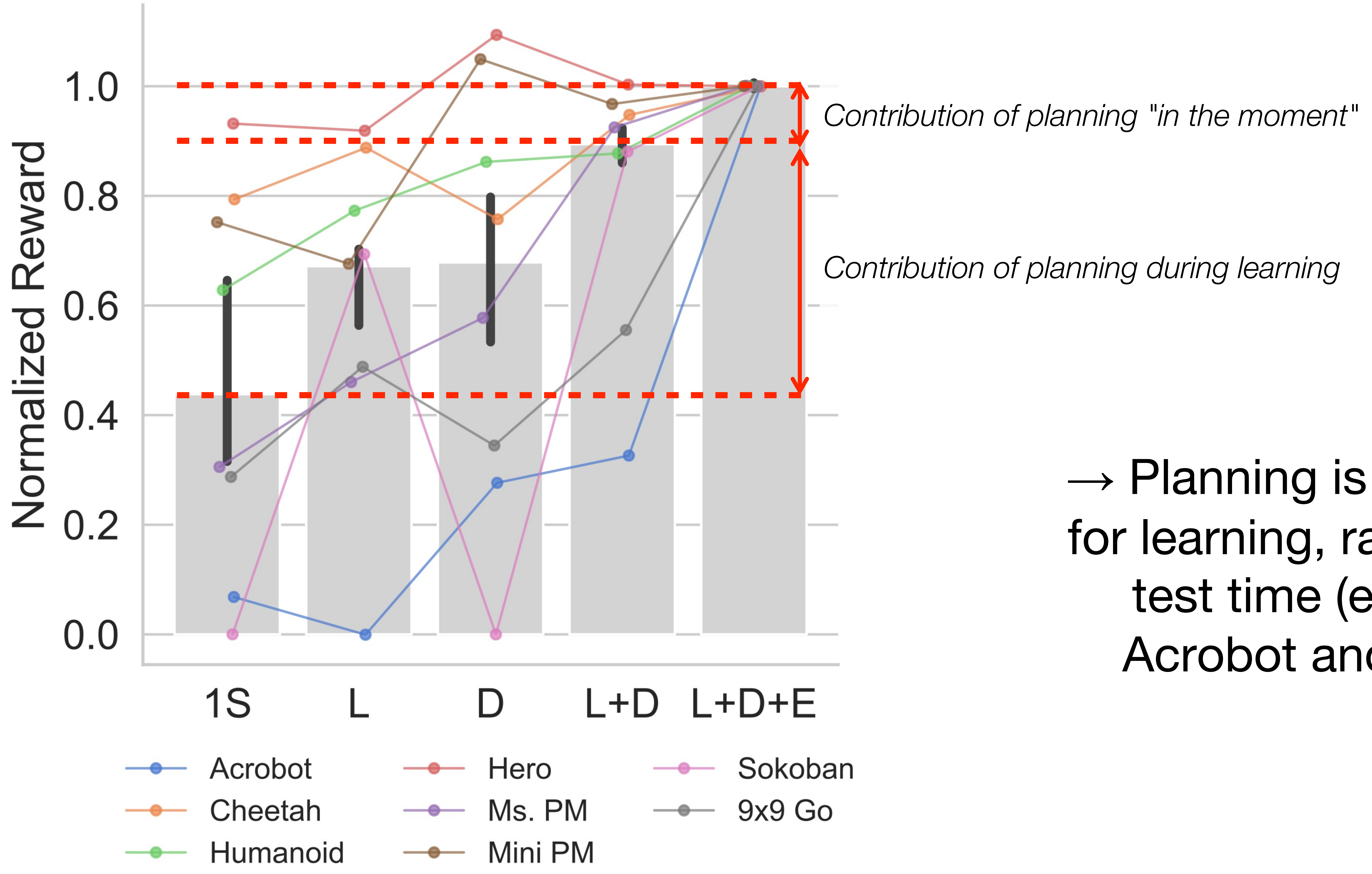


Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.





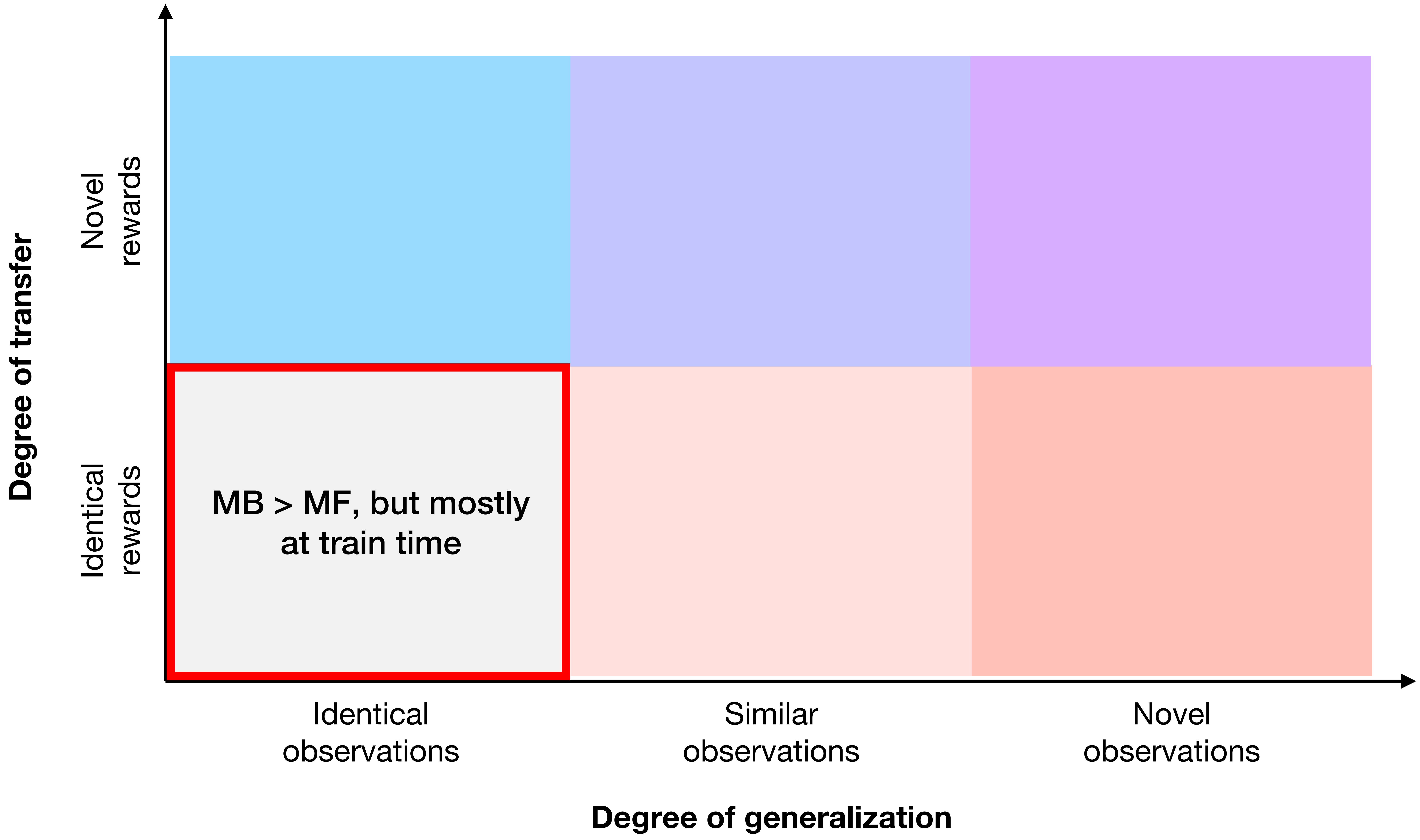
Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.



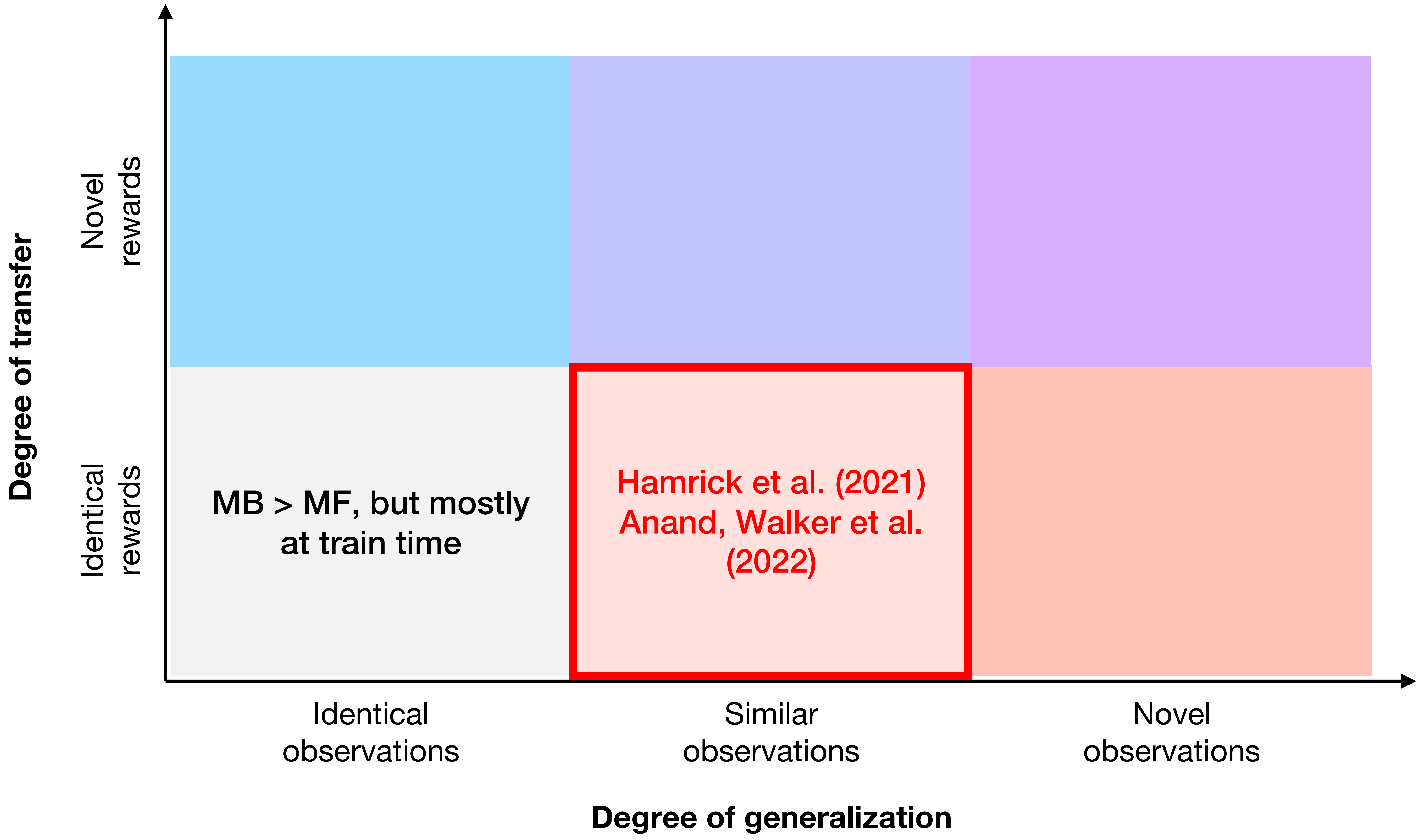
→ Planning is most useful for learning, rather than at test time (except for Acrobot and 9x9 Go)

Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.

Lessons in generalization & transfer

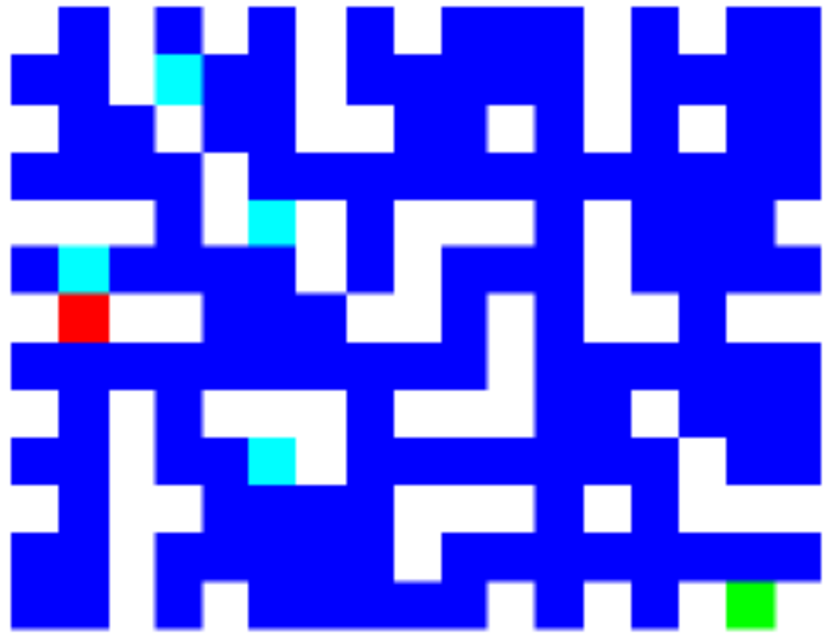


Lessons in generalization & transfer



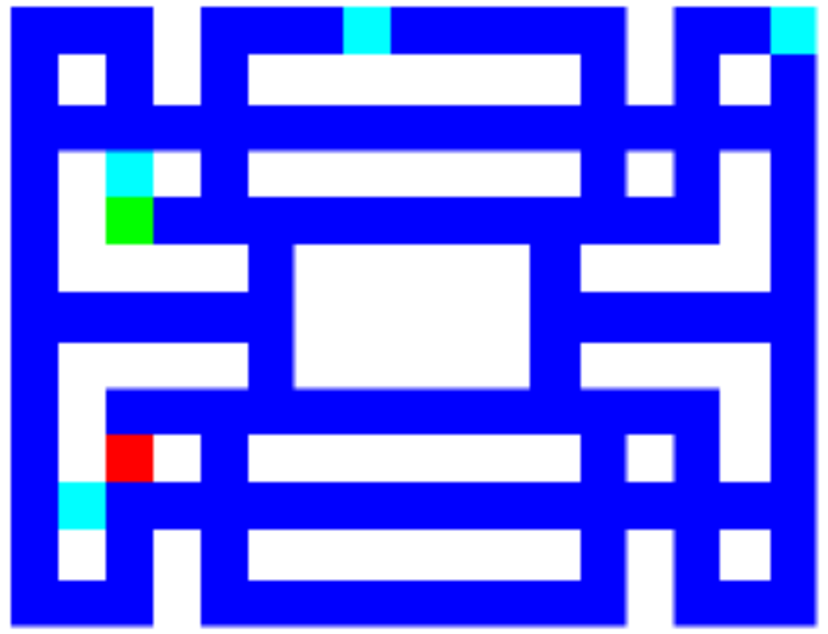
Procedural generalization

Train



Train on a **procedurally-generated** distribution of environments

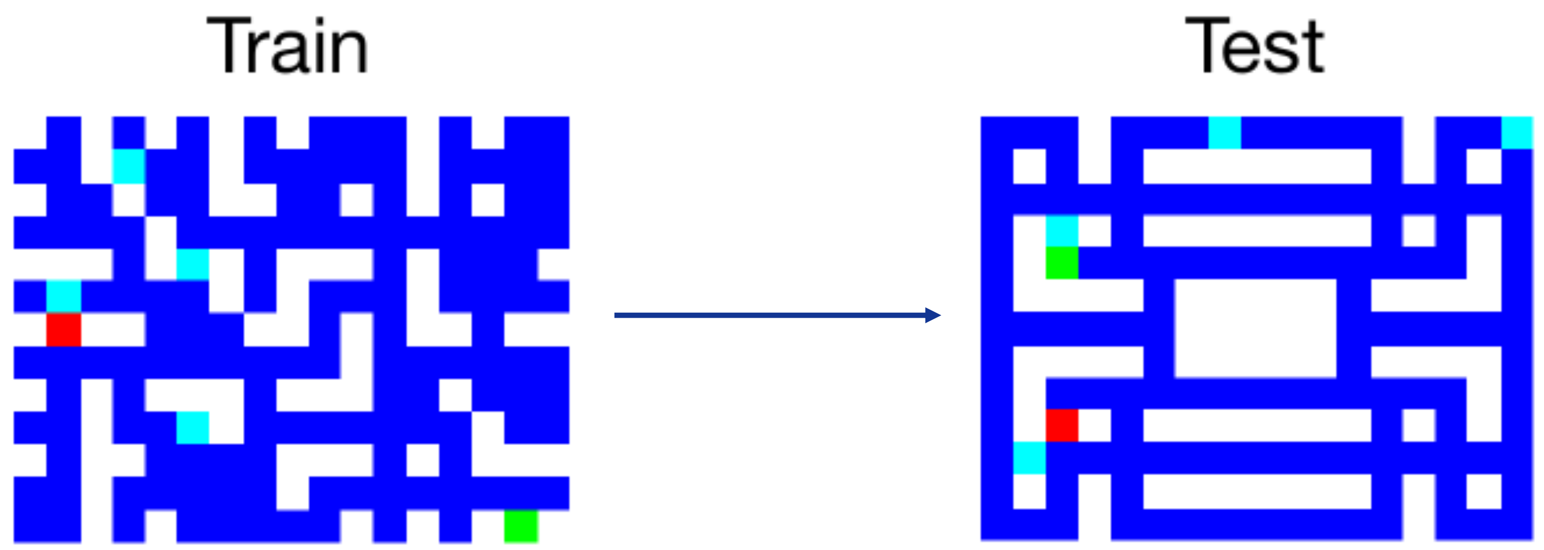
Test

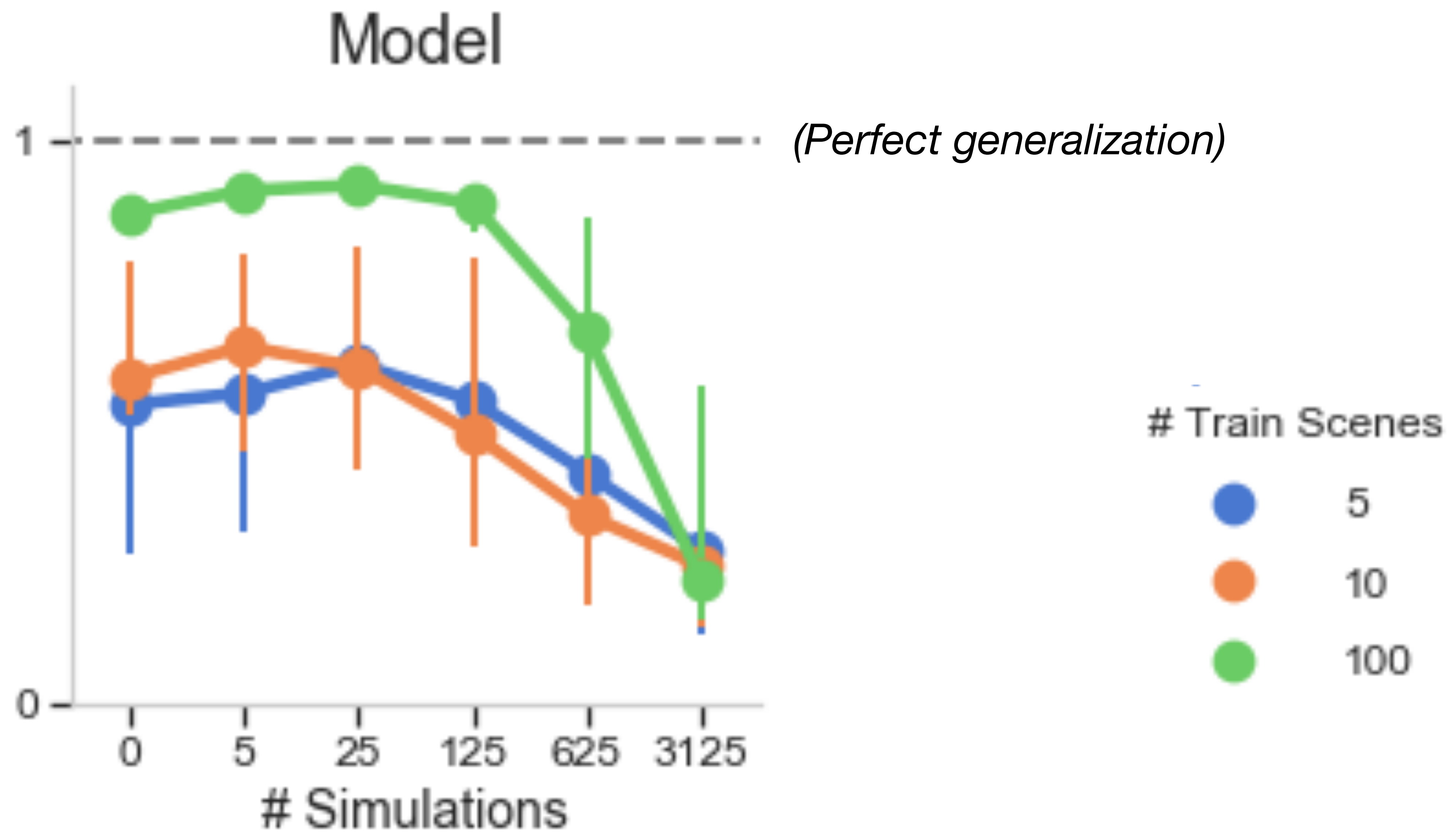


Zero-shot generalization to unseen environments

Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.

Generalizing to new mazes

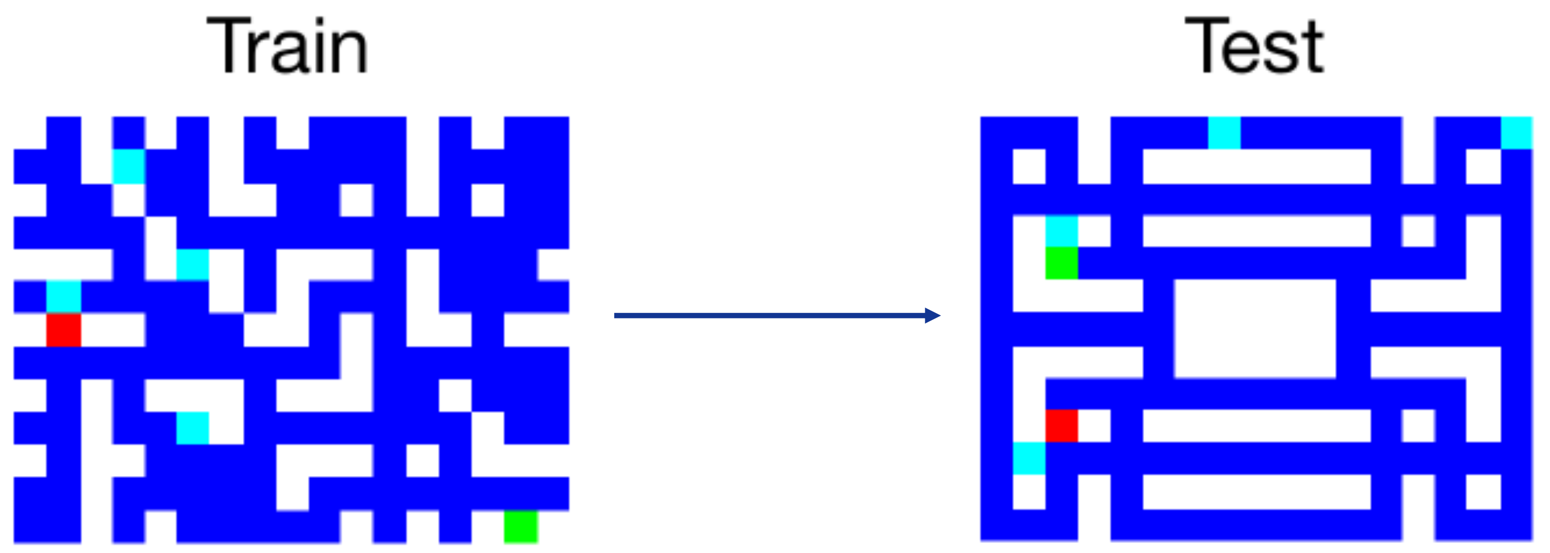


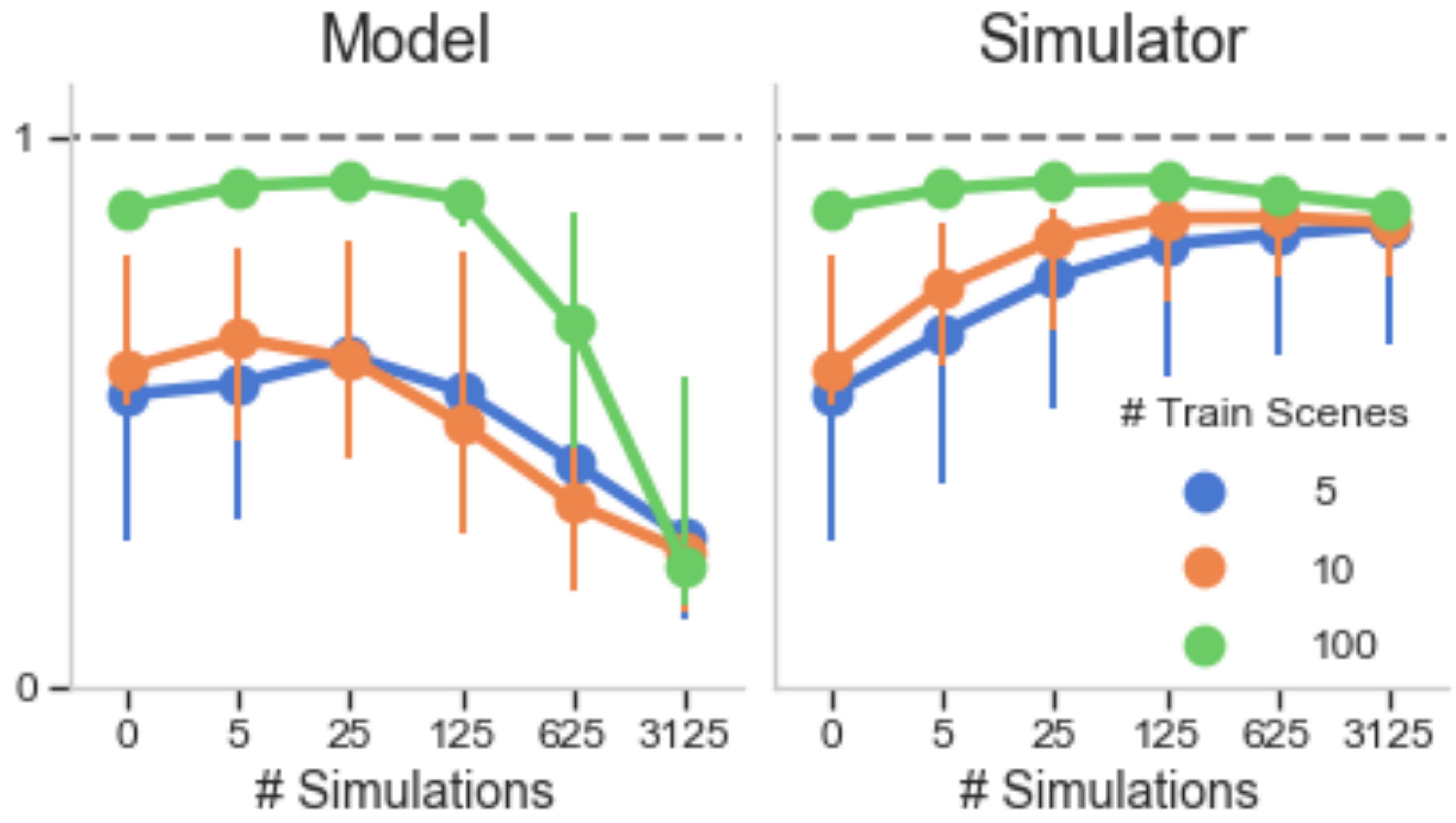


→ The model learned by MuZero is not very good

Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.

Generalizing to new mazes





(Perfect generalization)

→ The model learned by MuZero is not very good

→ But even a perfect model is not sufficient: we also need to know *where to search*

Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.

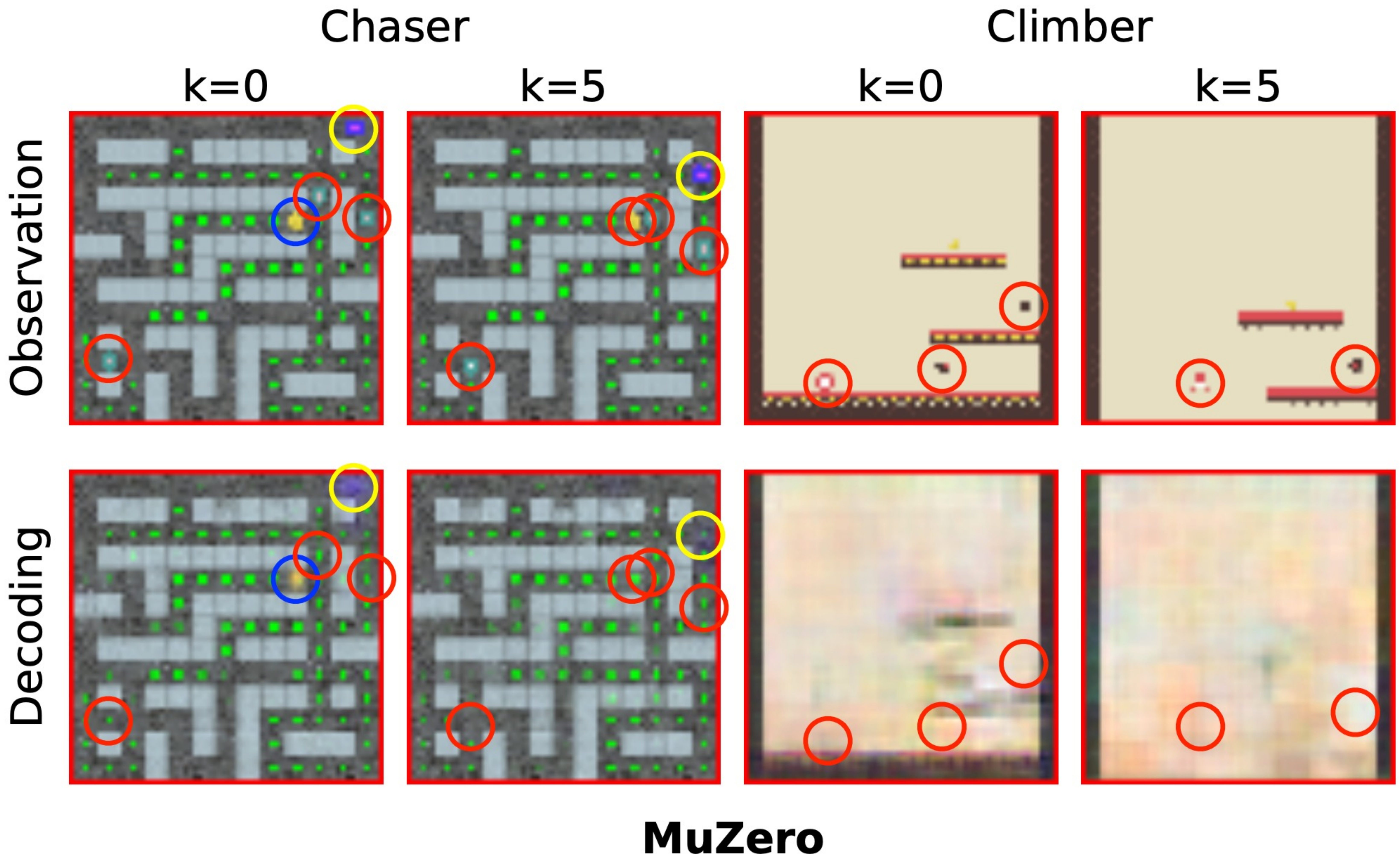
Procedural generalization

Procgen (Cobbe et al., 2020)

Train on a **procedurally-generated** distribution of environments → **Zero-shot generalization** to unseen environments

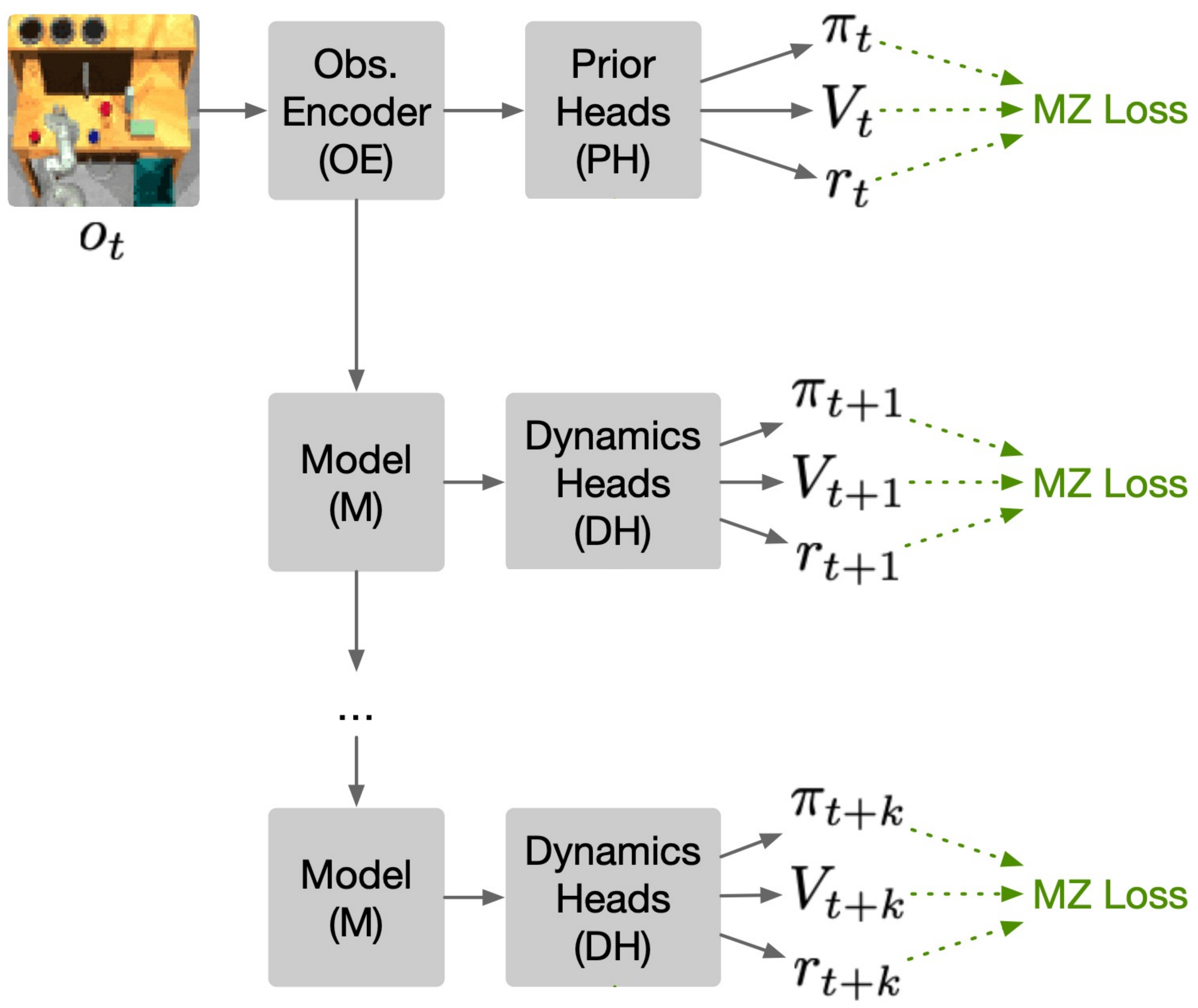
Anand, Walker et al. (2022). Procedural generalization by planning with self-supervised world models. ICLR.

Failure of representation



Anand, Walker et al. (2022). Procedural generalization by planning with self-supervised world models. ICLR.

Improving MuZero with self-supervision

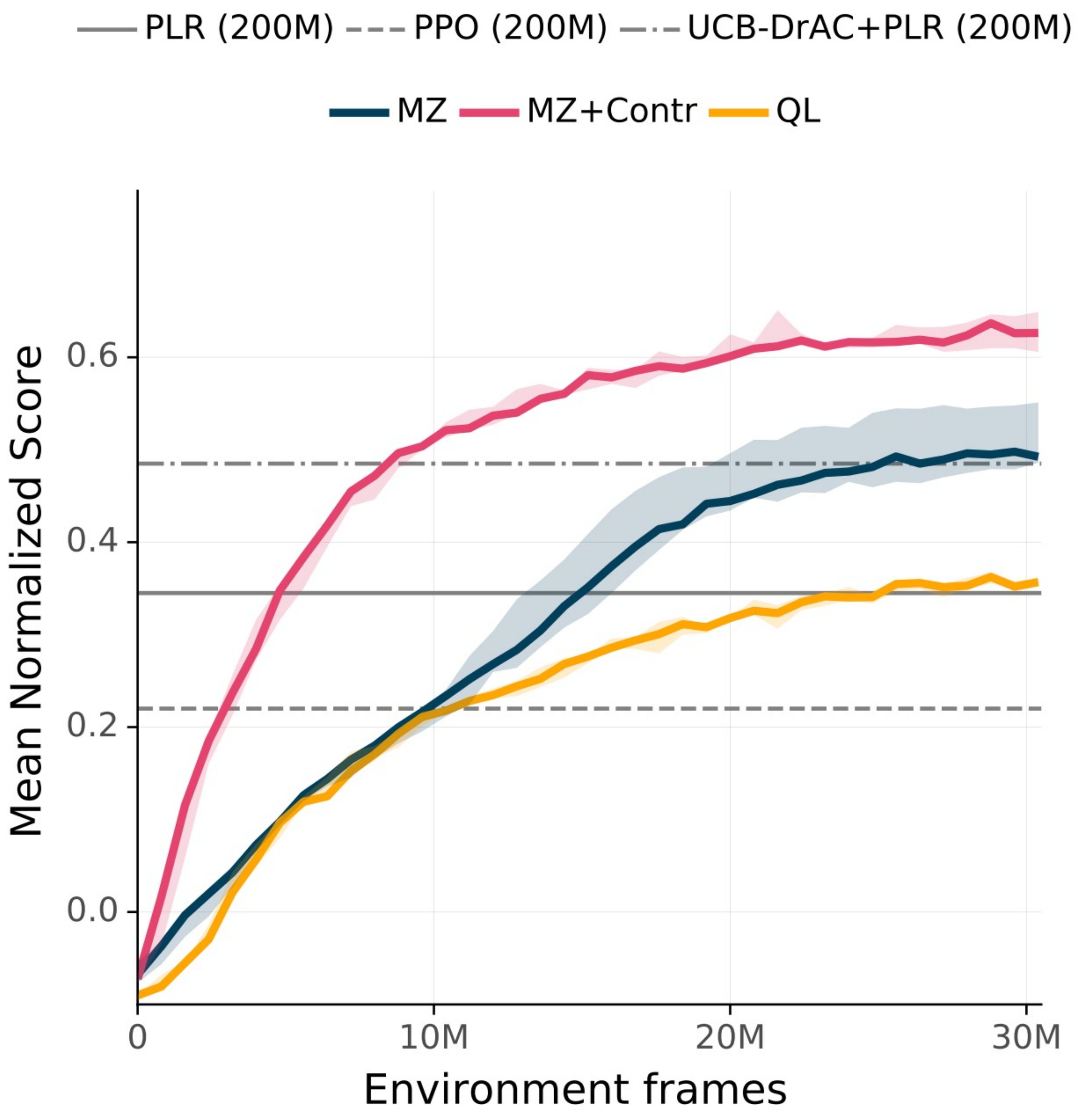


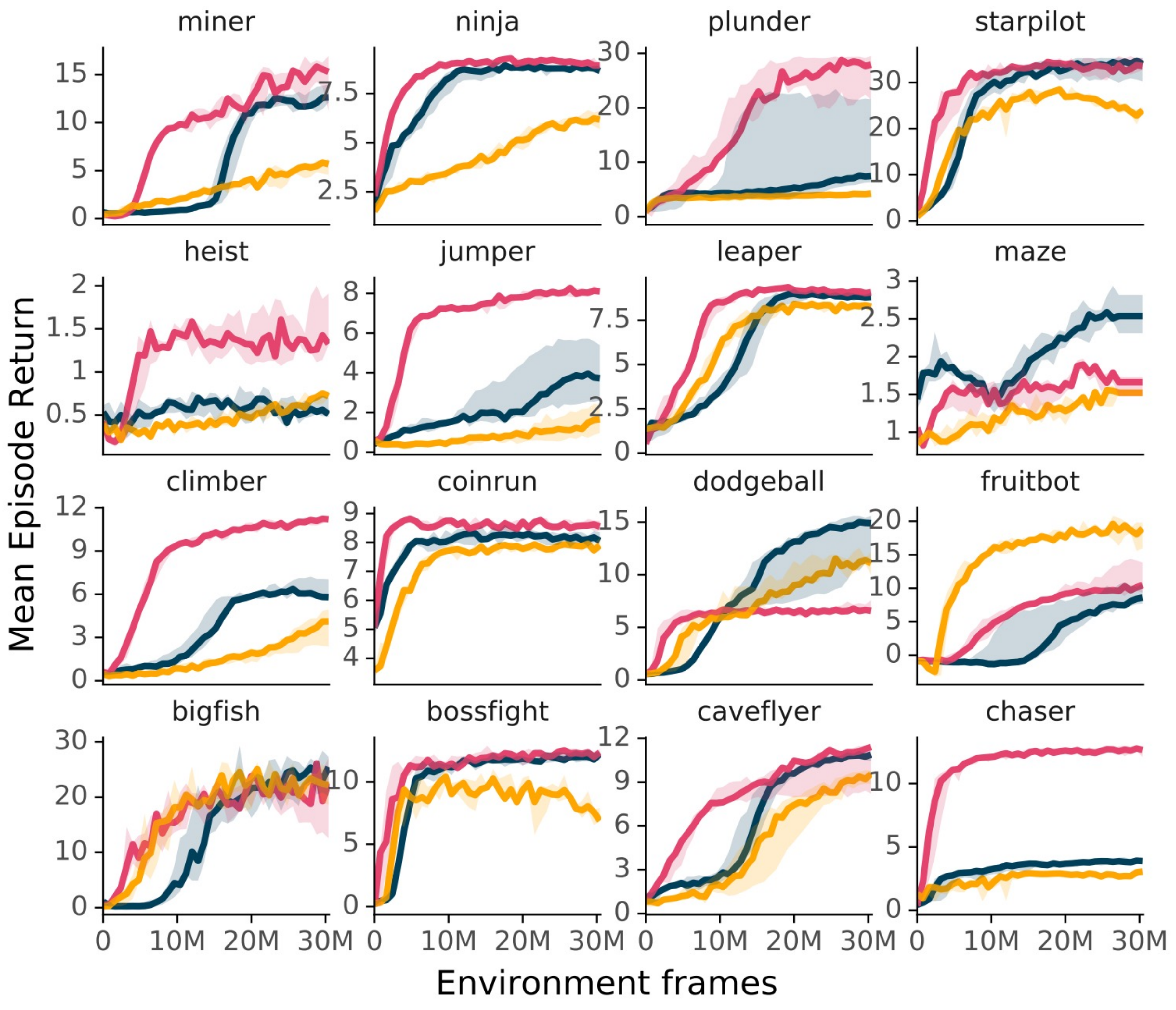
Self-supervised losses:

- *Reconstruction*: predict the obs. at time *t+k*
- *SPR:* predict the obs. embedding at time *t+k*
- *Contrastive:* classify whether a predicted obs. embedding at time *t+k* should correspond to the observation at time *t+i*

Anand, Walker et al. (2022). Procedural generalization by planning with self-supervised world models. ICLR.

Procgen results (train on 500 levels)

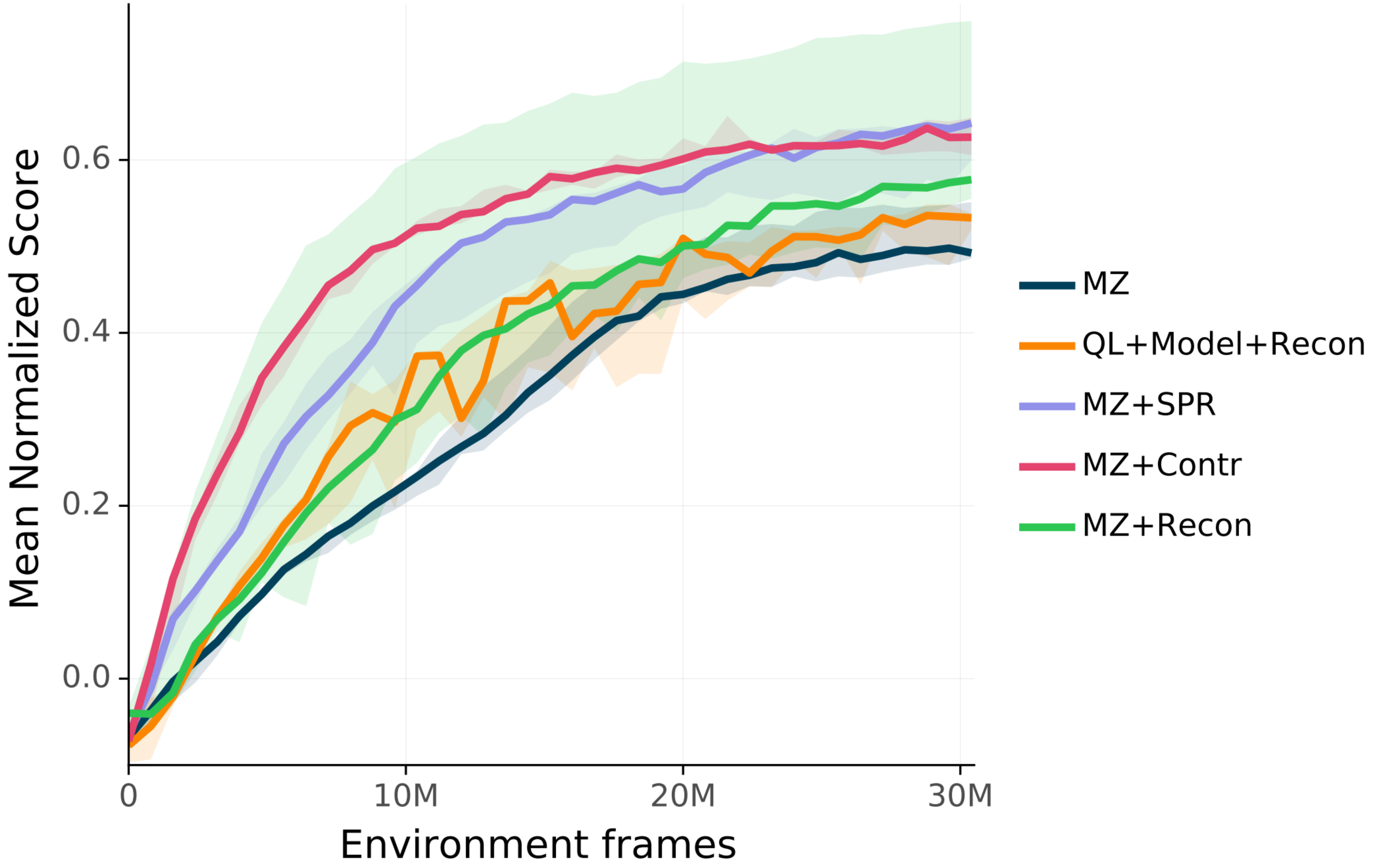




→ Self-supervision has a huge impact on generalization!

Anand, Walker et al. (2022). Procedural generalization by planning with self-supervised world models. ICLR.

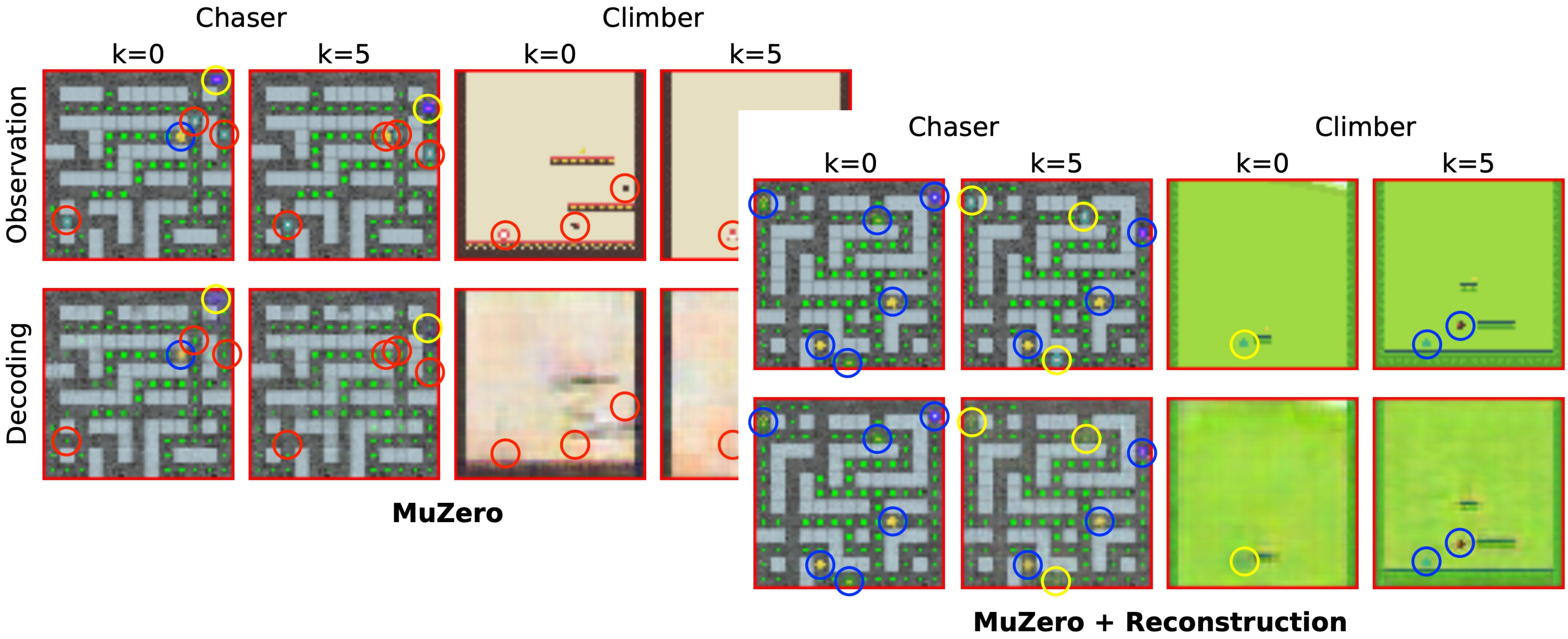
Comparing methods of self-supervision



→ All methods of self-supervision are roughly comparable

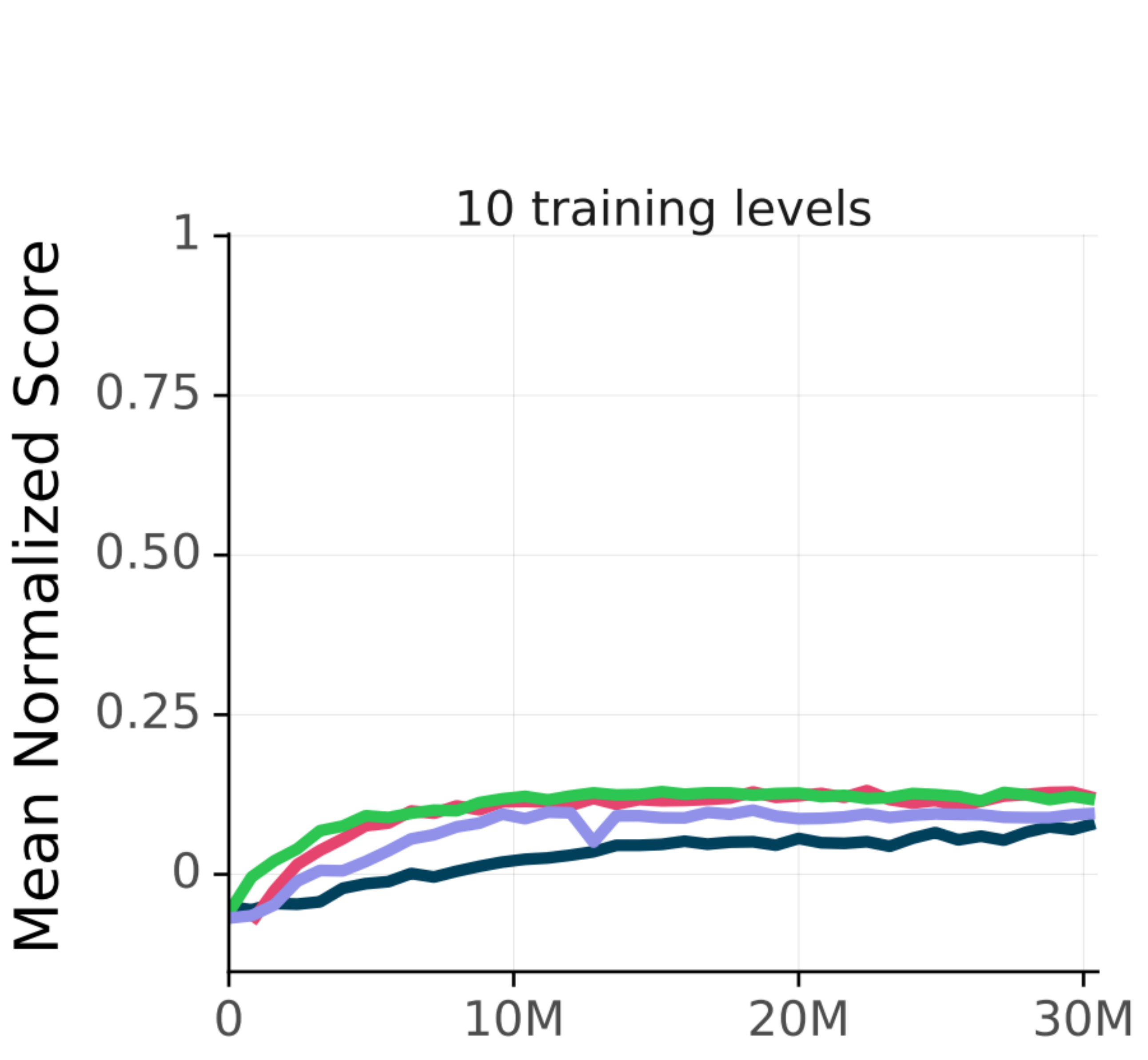
Anand, Walker et al. (2022). Procedural generalization by planning with self-supervised world models. ICLR.

Improved representations



Anand, Walker et al. (2022). Procedural generalization by planning with self-supervised world models. ICLR.

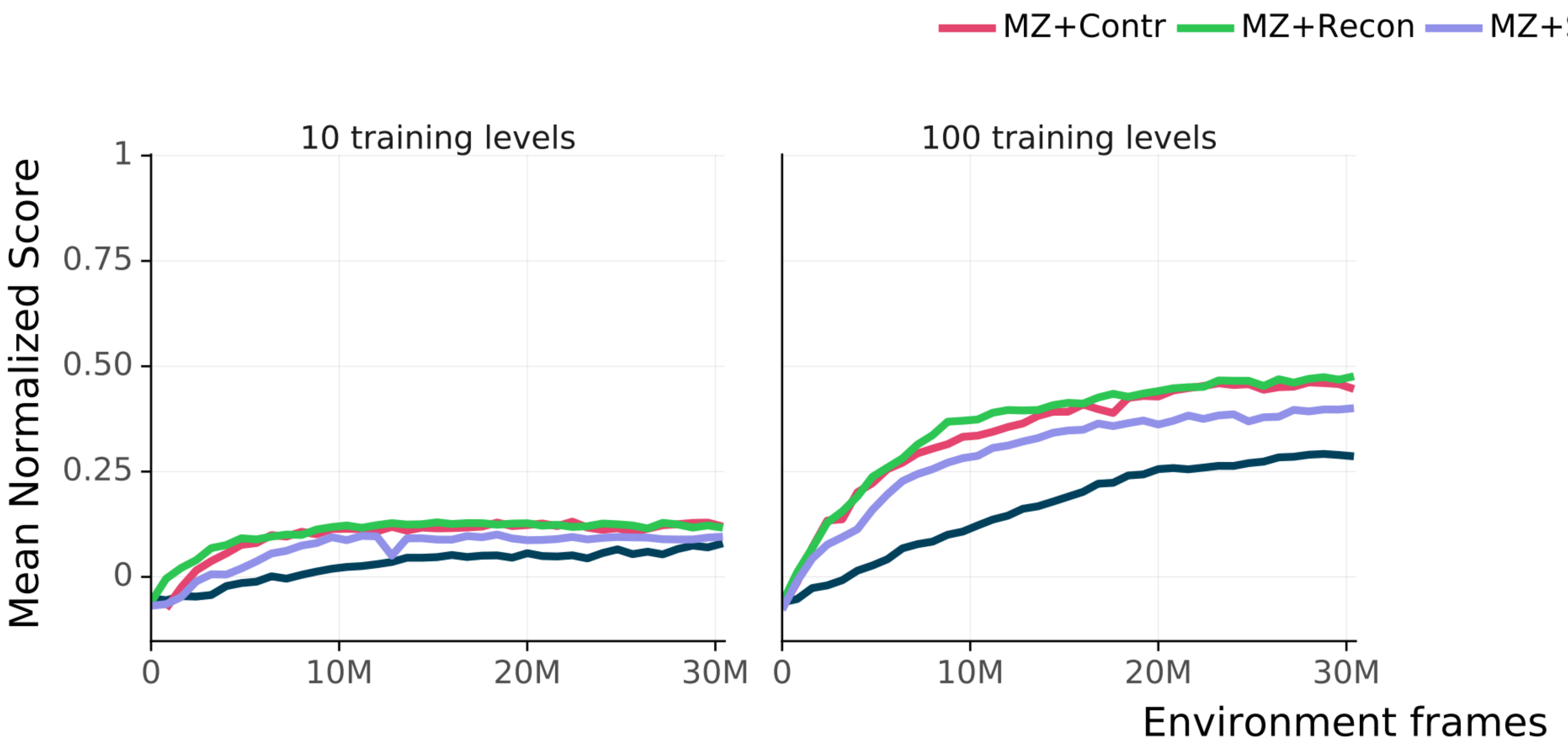
Interaction between self-supervision and dataset size



very little improvement w/ self-supervision

Anand, Walker et al. (2022). Procedural generalization by planning with self-supervised world models. ICLR.

Interaction between self-supervision and dataset size



very little improvement w/ self-supervision

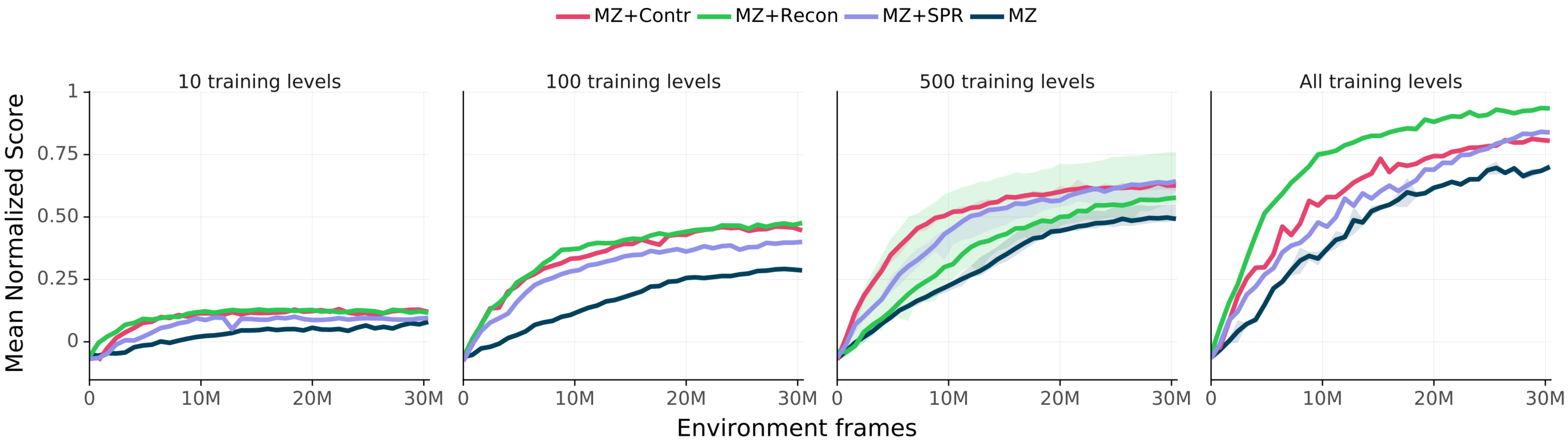
Anand, Walker et al. (2022). Procedural generalization by planning with self-supervised world models. ICLR.

Interaction between self-supervision and dataset size

very little improvement w/ self-supervision

Anand, Walker et al. (2022). Procedural generalization by planning with self-supervised world models. ICLR.

Interaction between self-supervision and dataset size



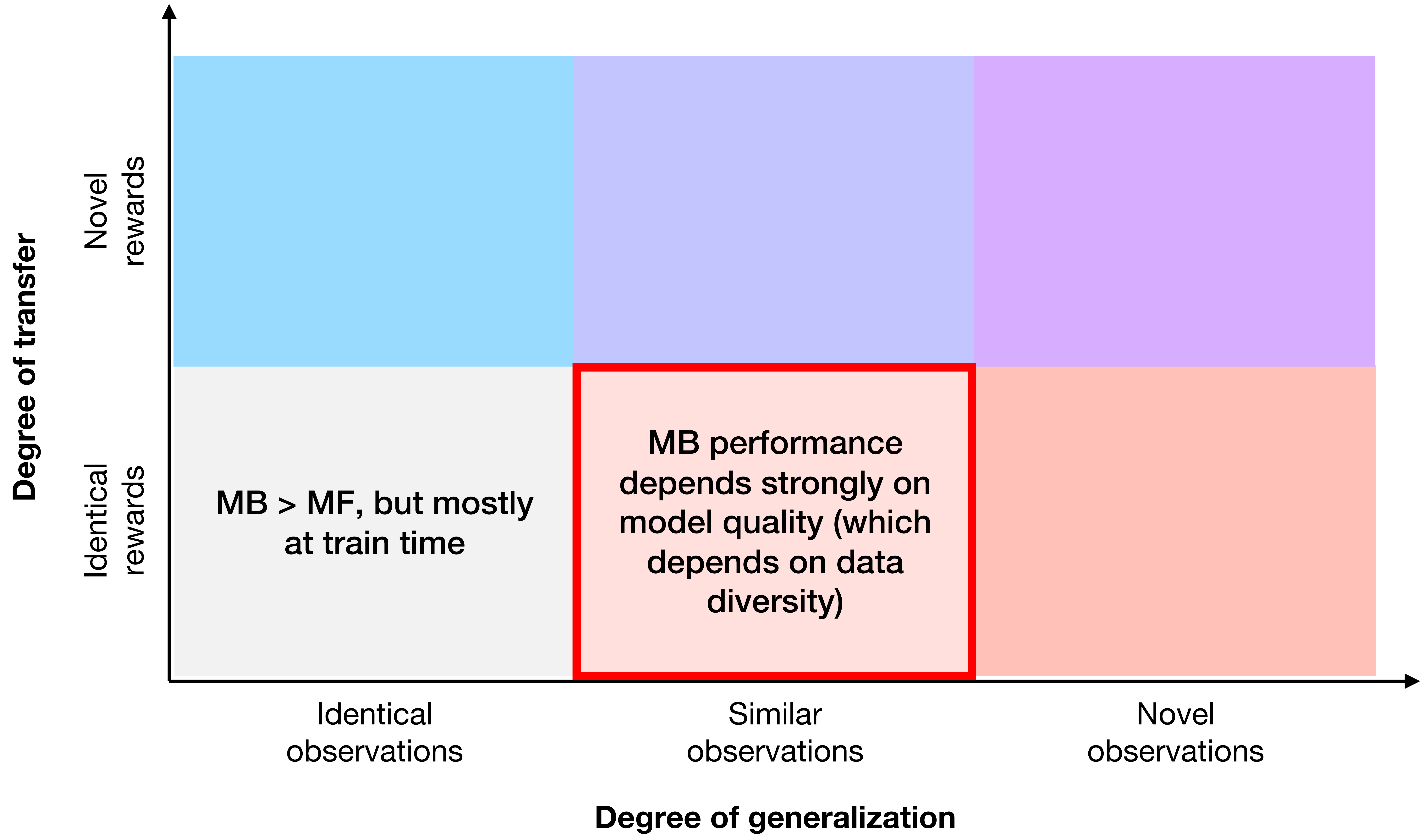
very little improvement w/ self-supervision

→ Self-supervision is more useful when training on more environments

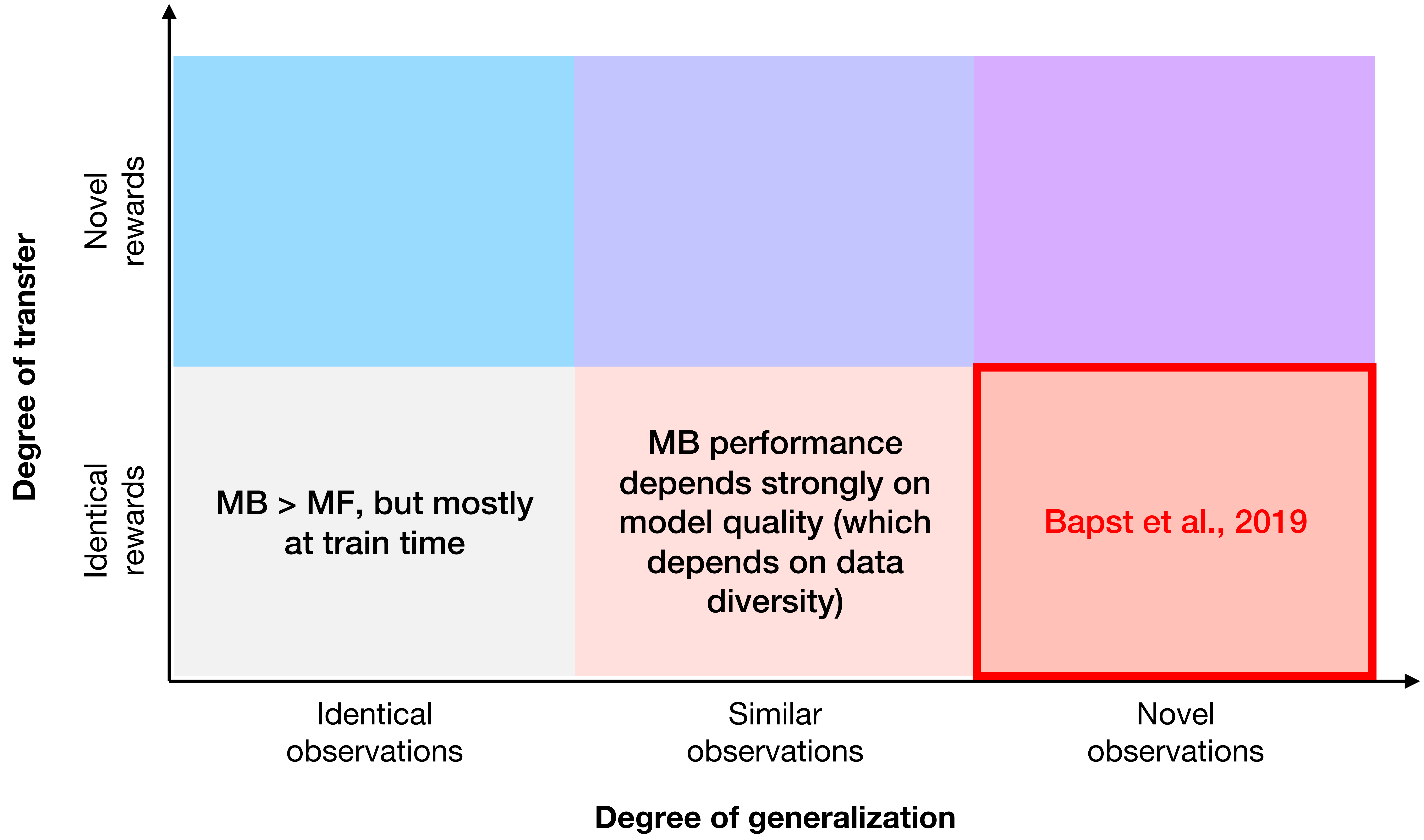
big improvement w/ self-supervision

Anand, Walker et al. (2022). Procedural generalization by planning with self-supervised world models. ICLR.

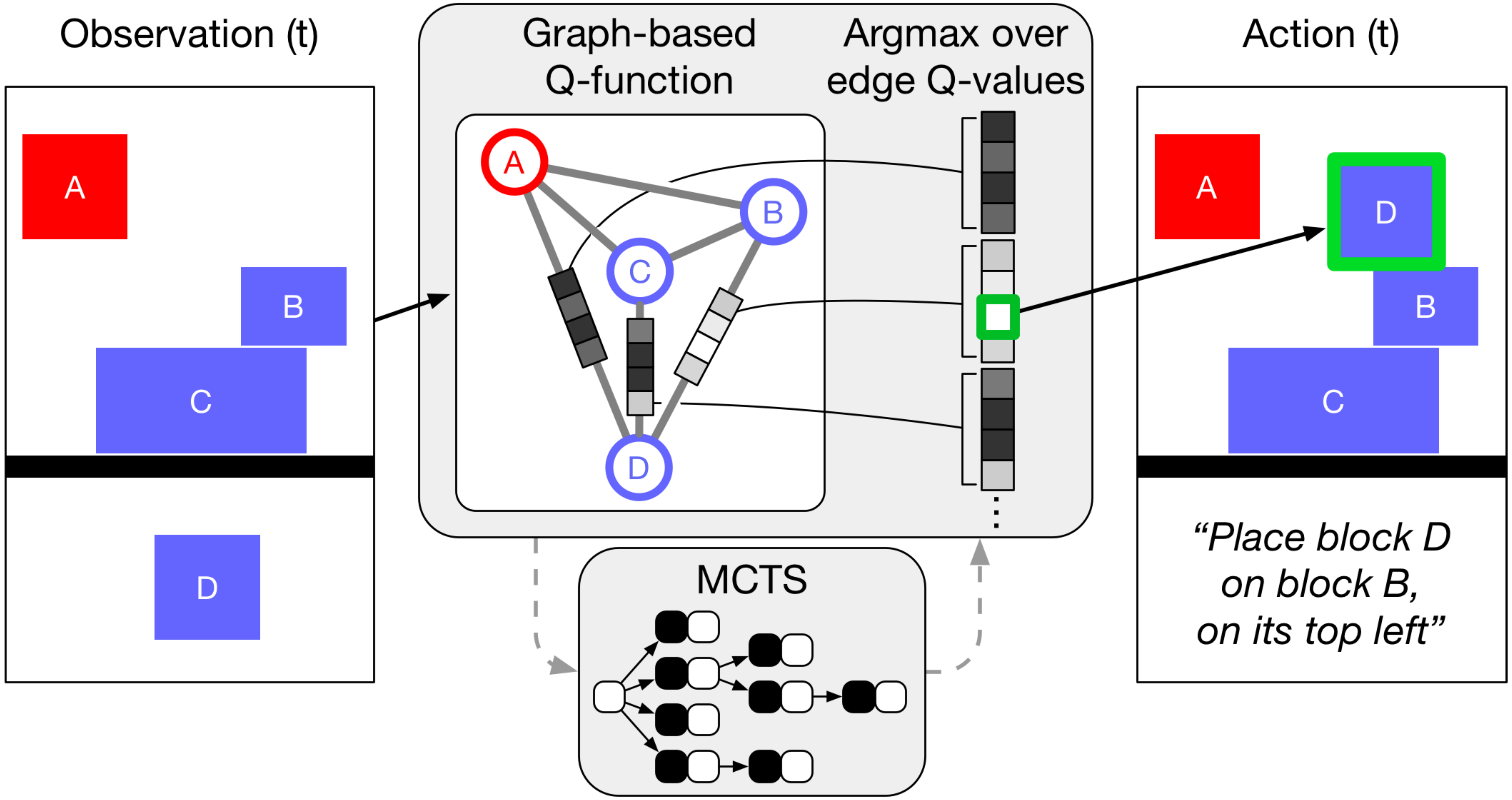
Lessons in generalization & transfer



Lessons in generalization & transfer

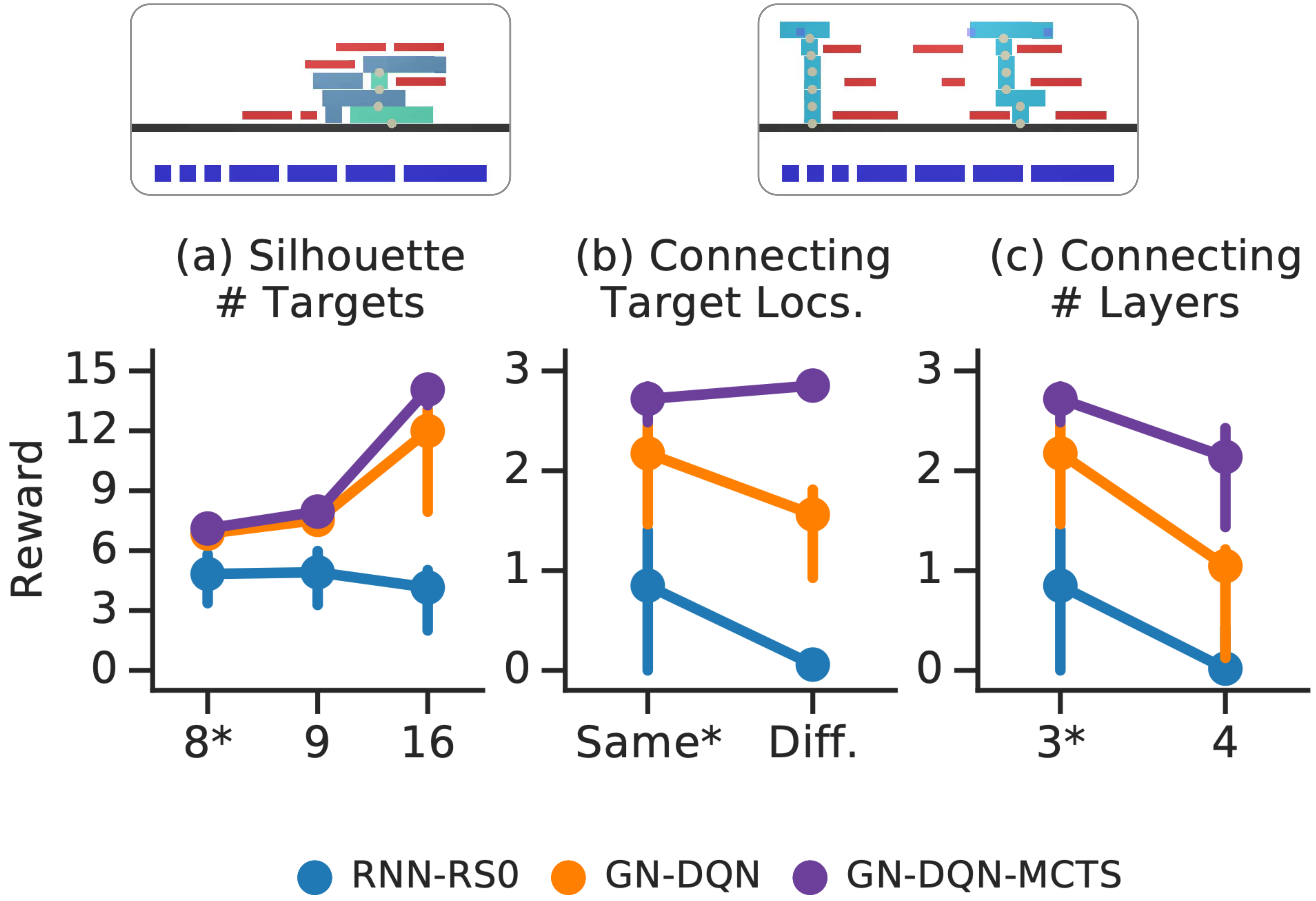


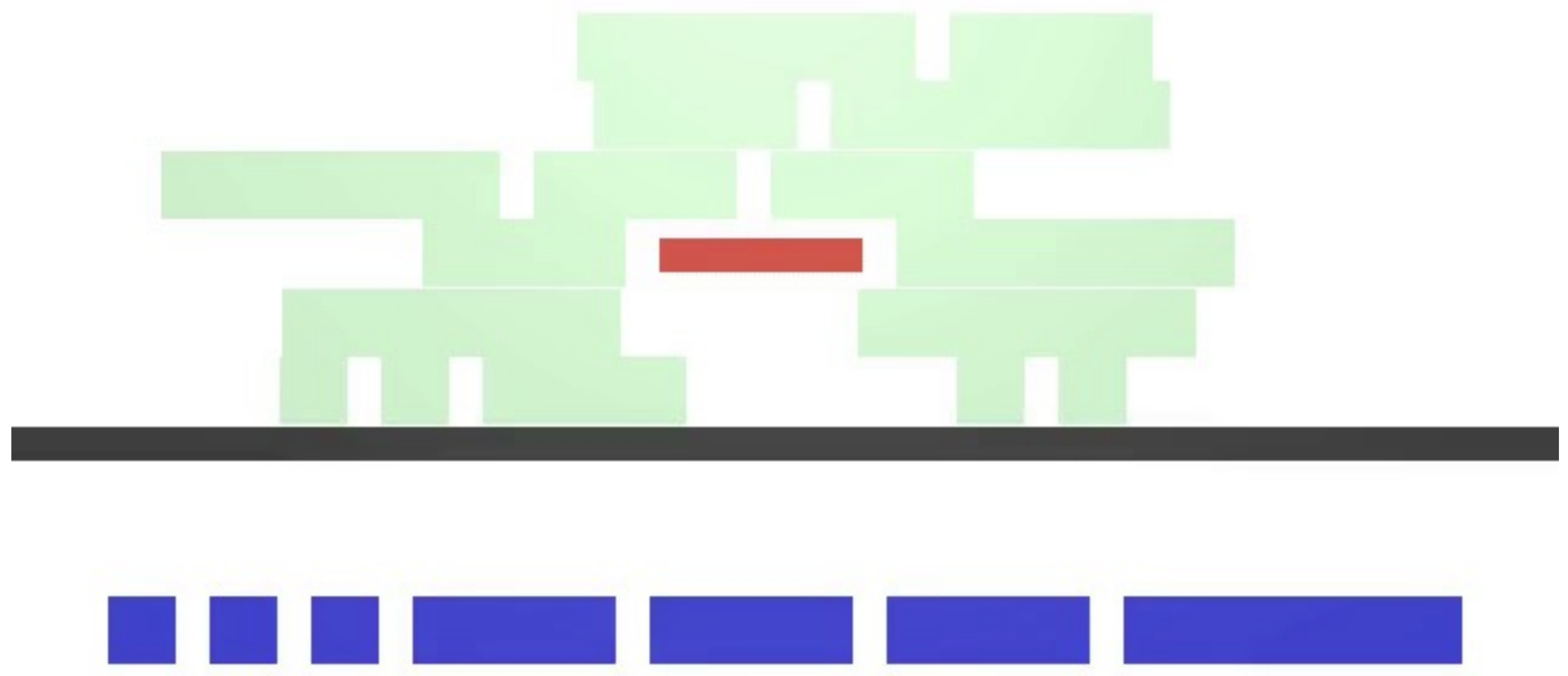
Generalizing to novel scenes



Bapst, Sanchez-Gonzalez et al. (2019). Structured agents for physical construction. ICML.

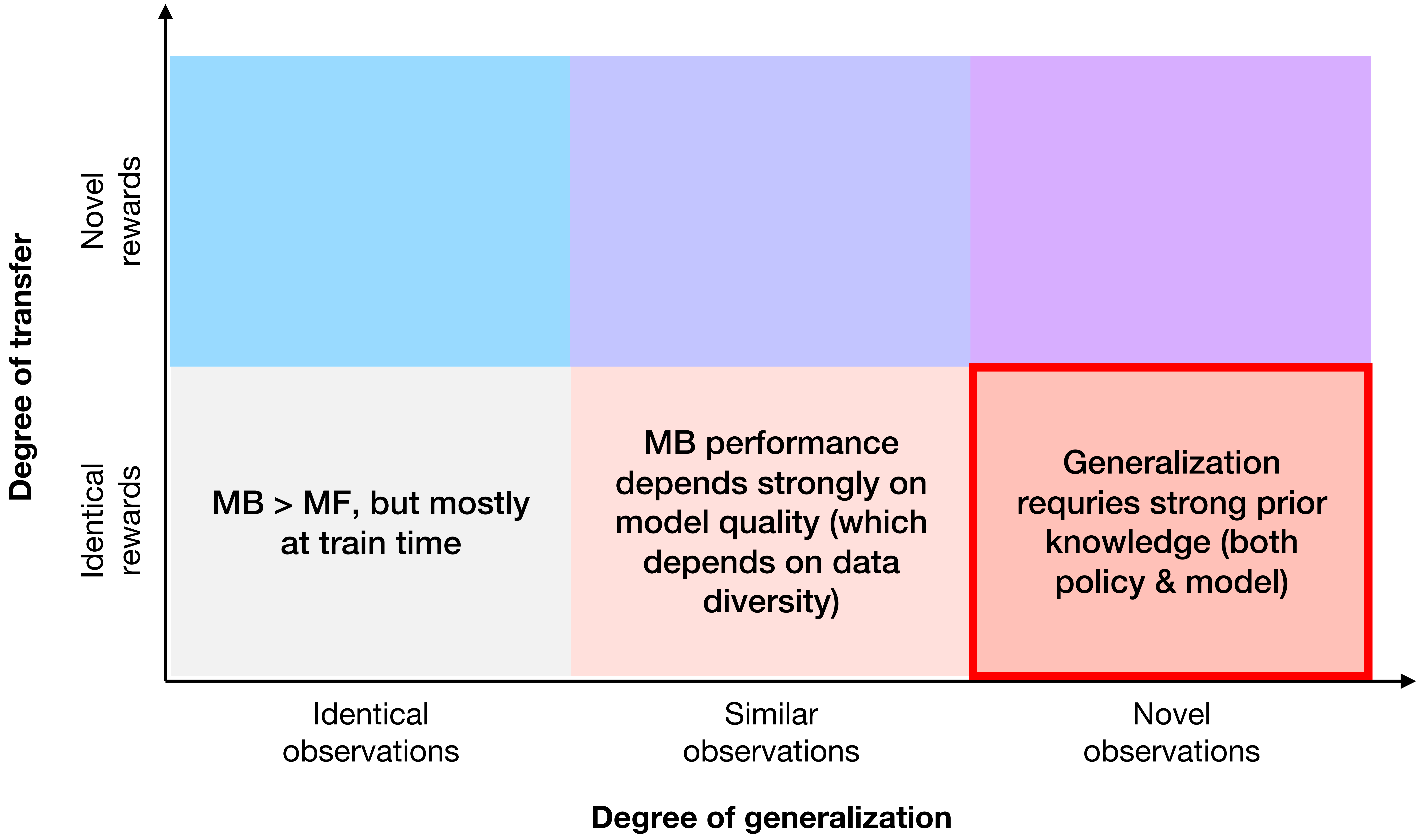
Generalizing to novel scenes



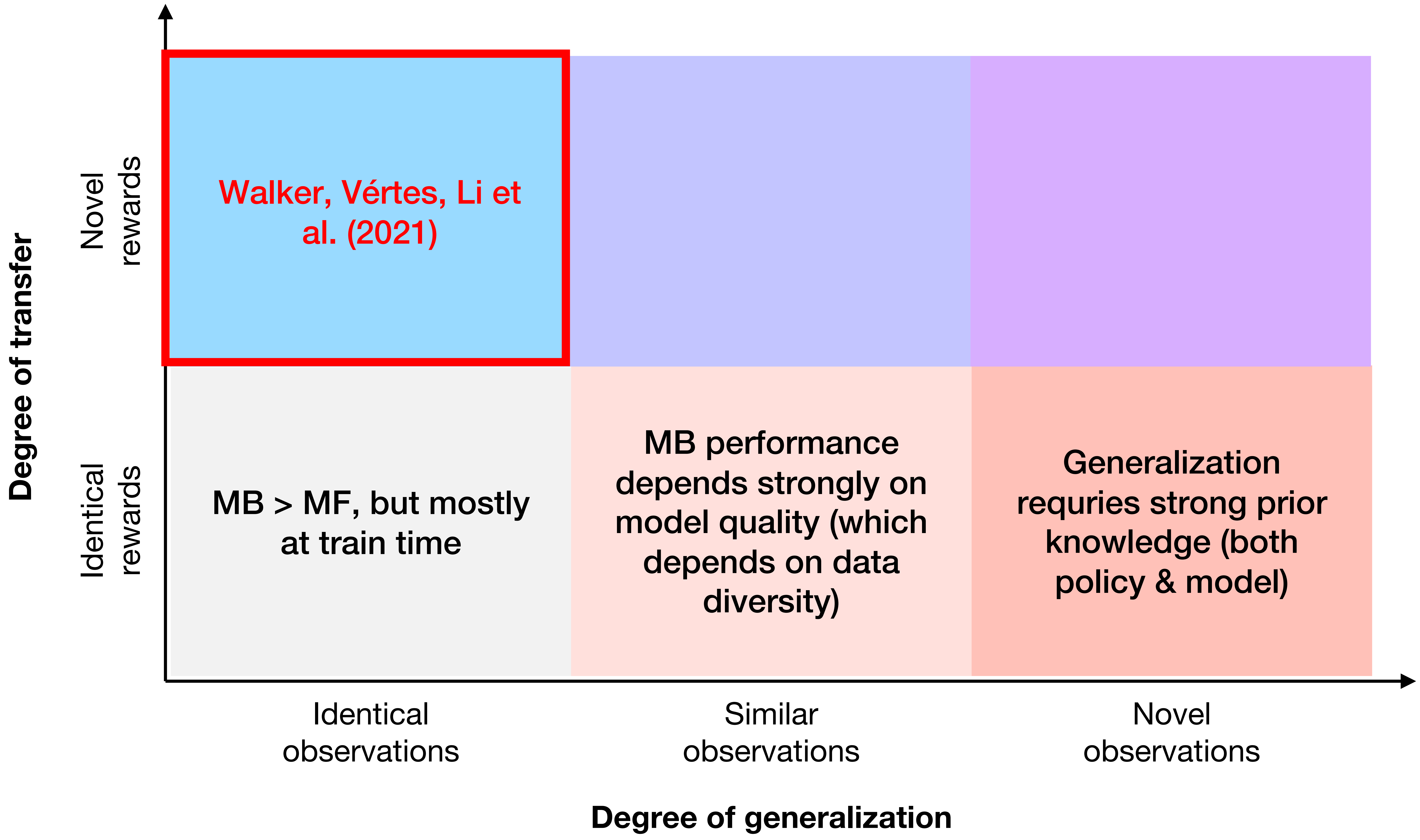


Bapst, Sanchez-Gonzalez et al. (2019). Structured agents for physical construction. ICML.

Lessons in generalization & transfer



Lessons in generalization & transfer

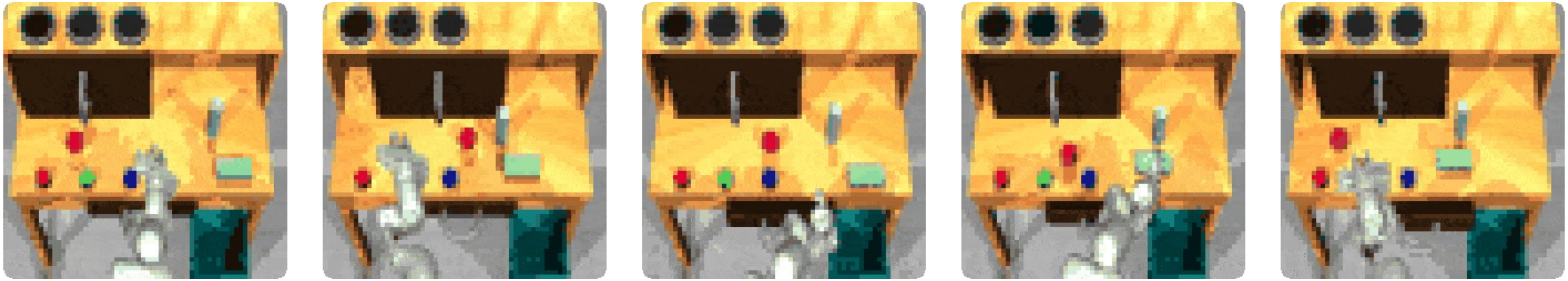


Experimental setup

Unsupervised exploration phase:
RL training with intrinsic rewards

Transfer phase:

Transfer policy and/or model and continue training with real rewards

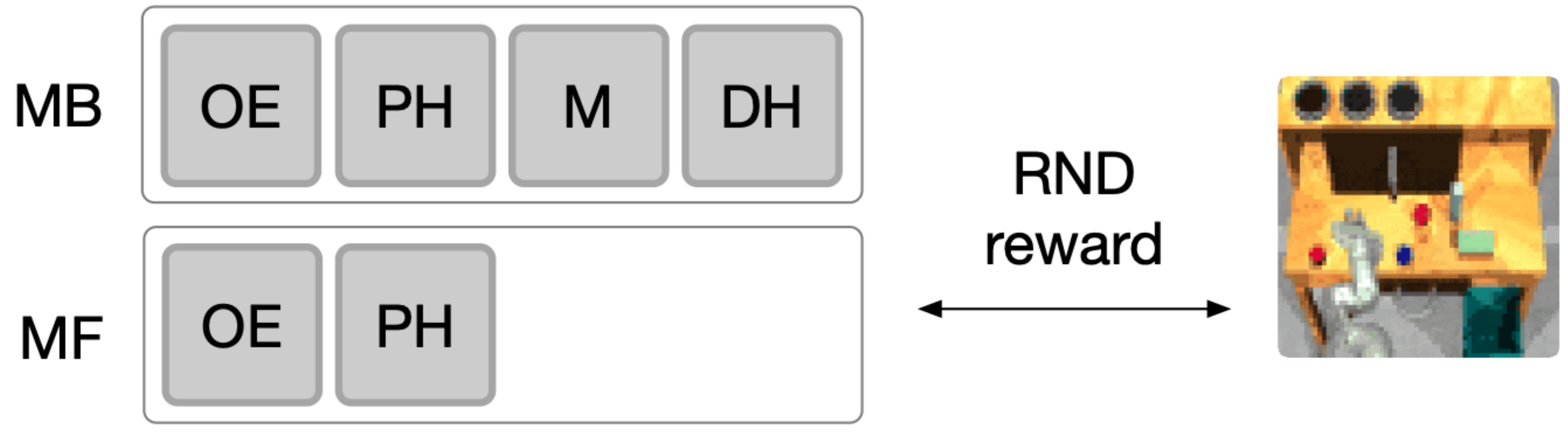


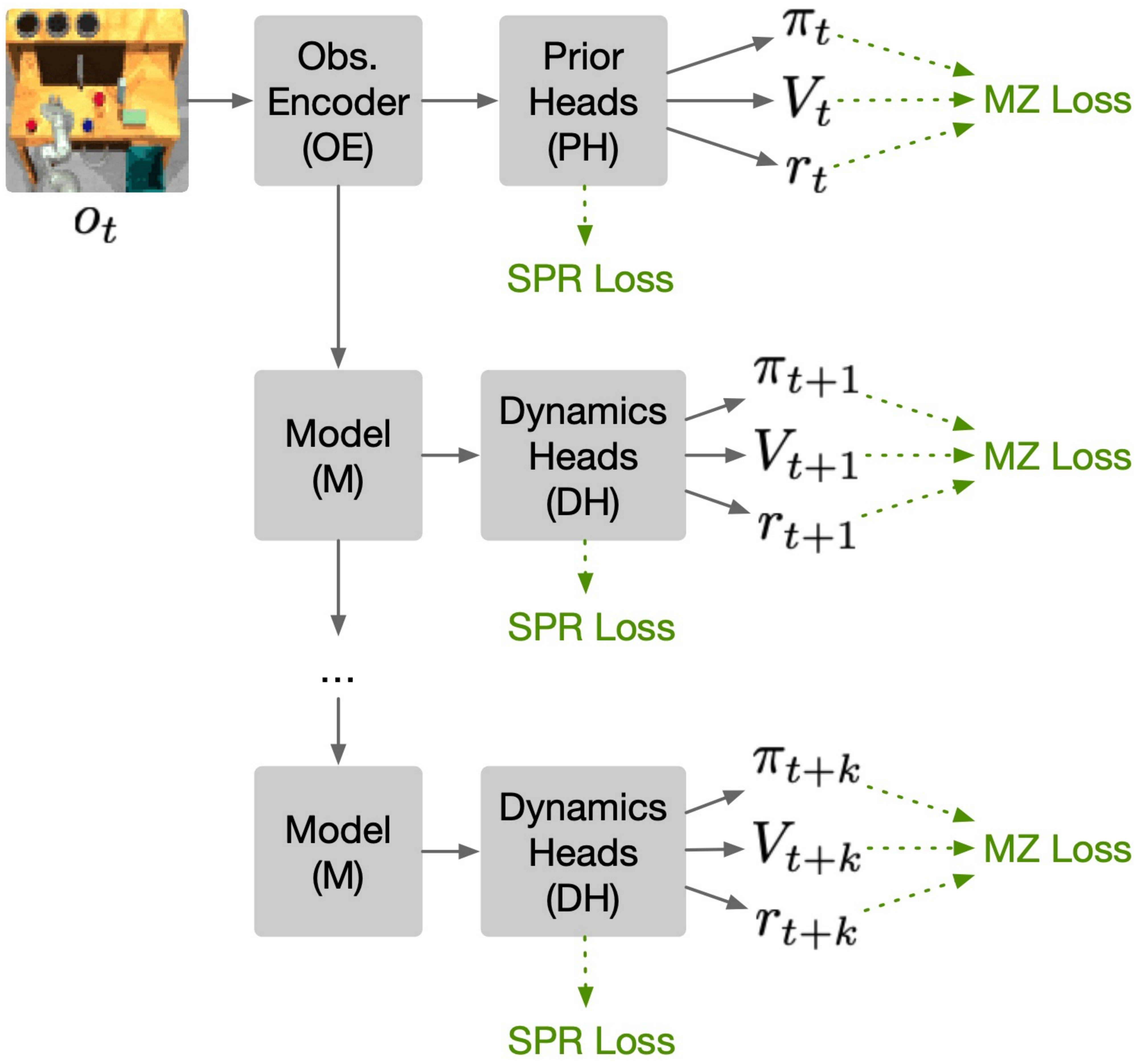
Robodesk (Kannan et al., 2021)

Walker, Vértes, Li et al. (2023). Investigating the role of model-based learning in exploration and transfer. ICML.

Experimental setup

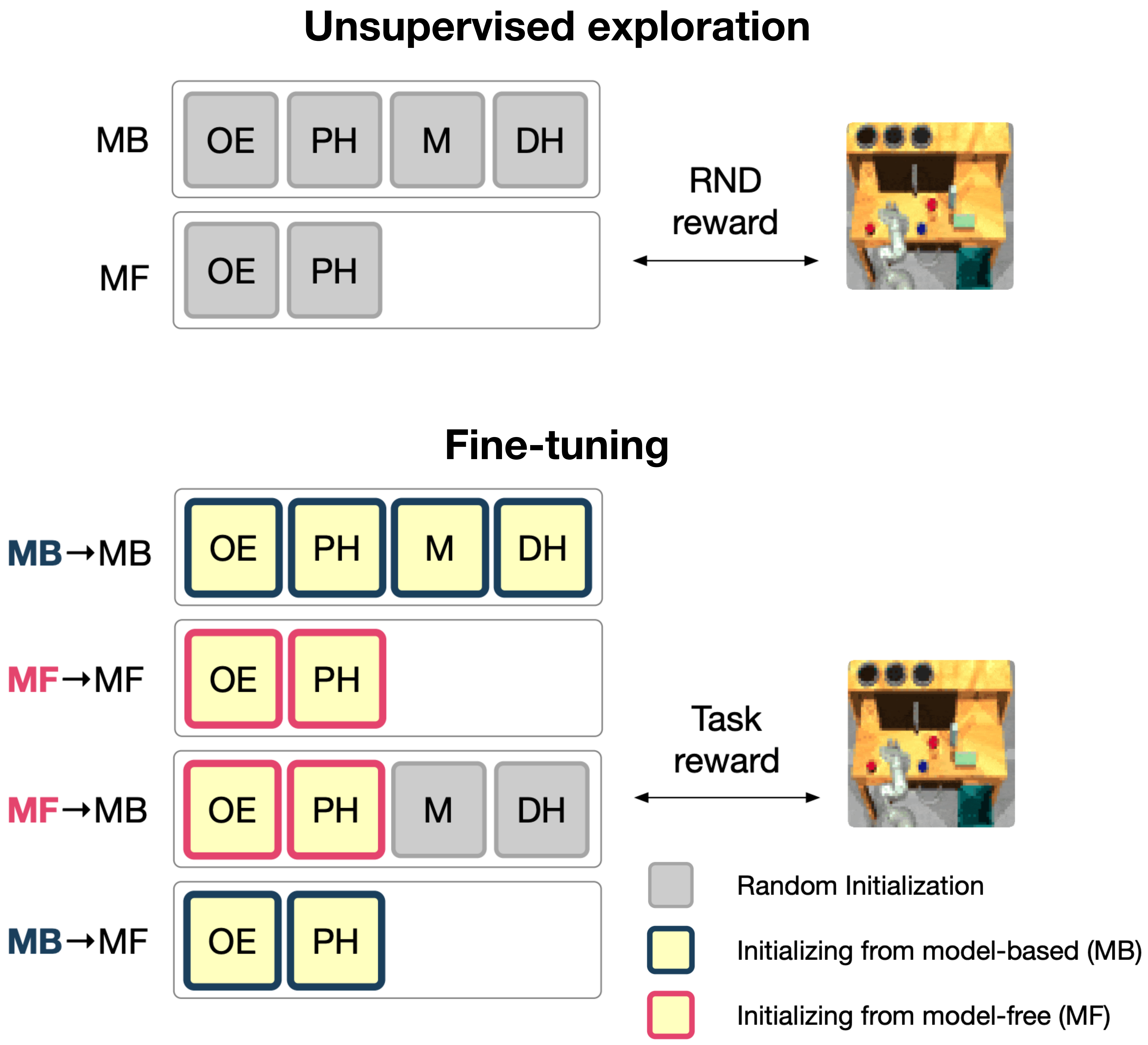
Unsupervised exploration

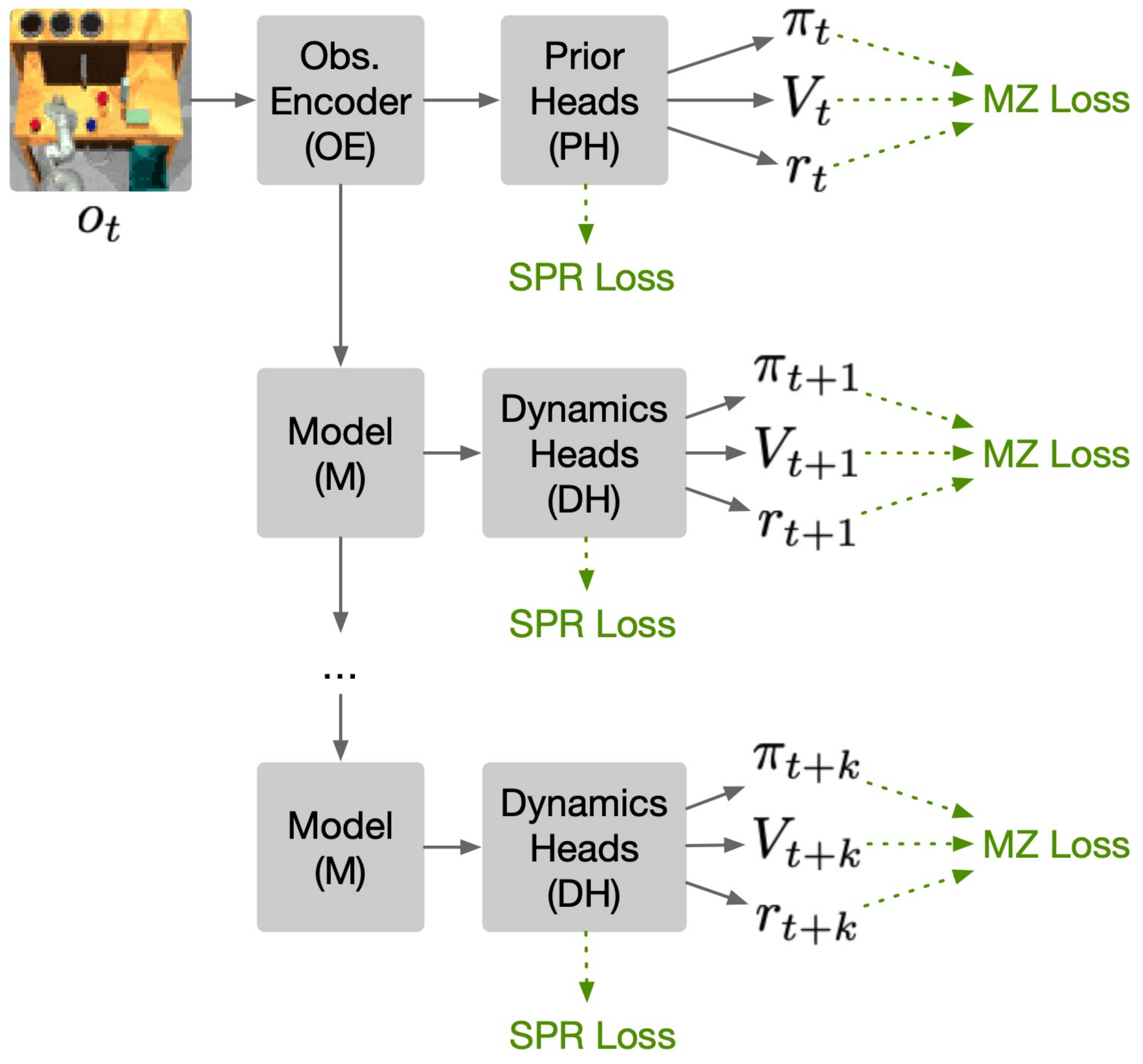




Walker, Vértes, Li et al. (2023). Investigating the role of model-based learning in exploration and transfer. ICML.

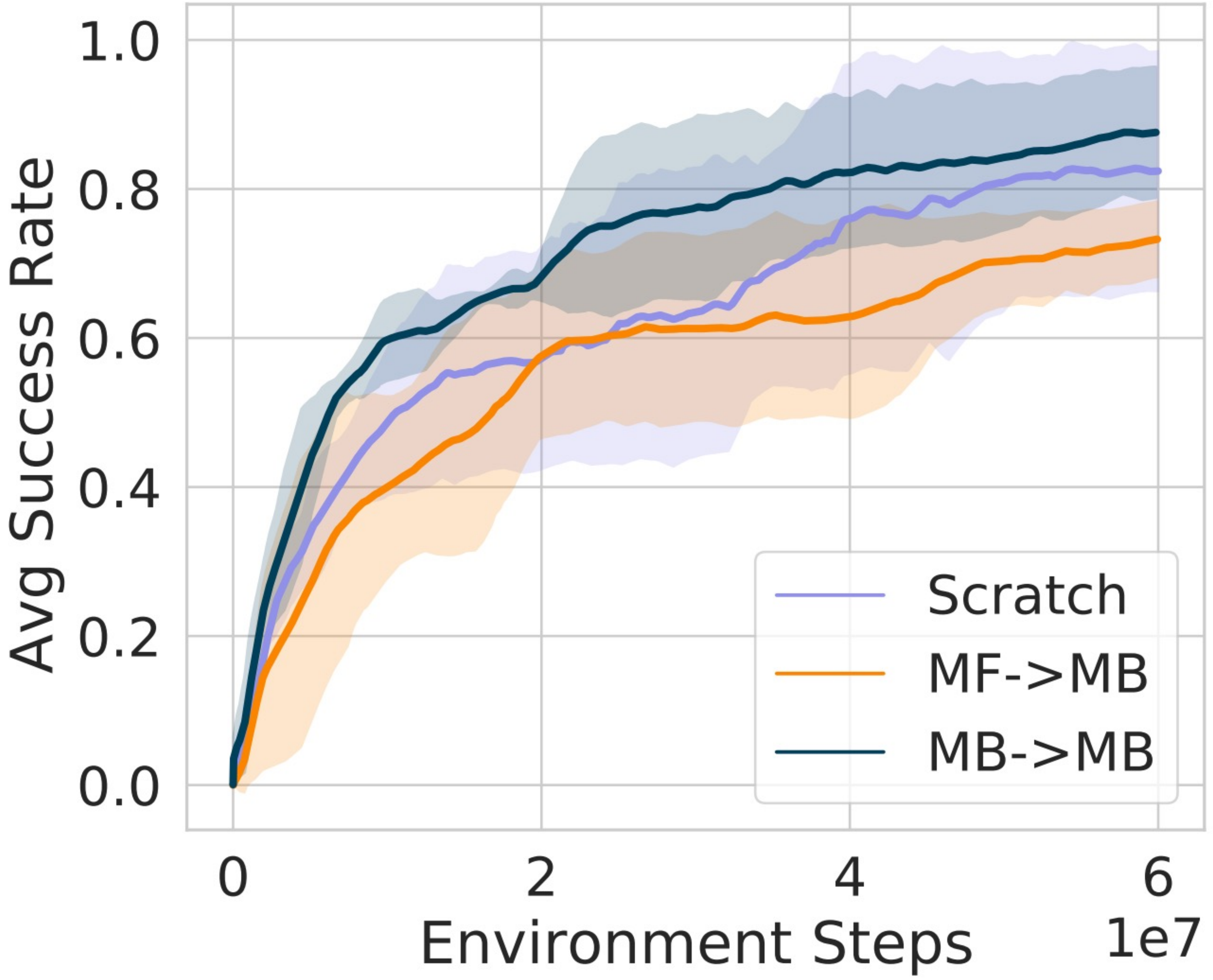
Experimental setup





Walker, Vértes, Li et al. (2023). Investigating the role of model-based learning in exploration and transfer. ICML.

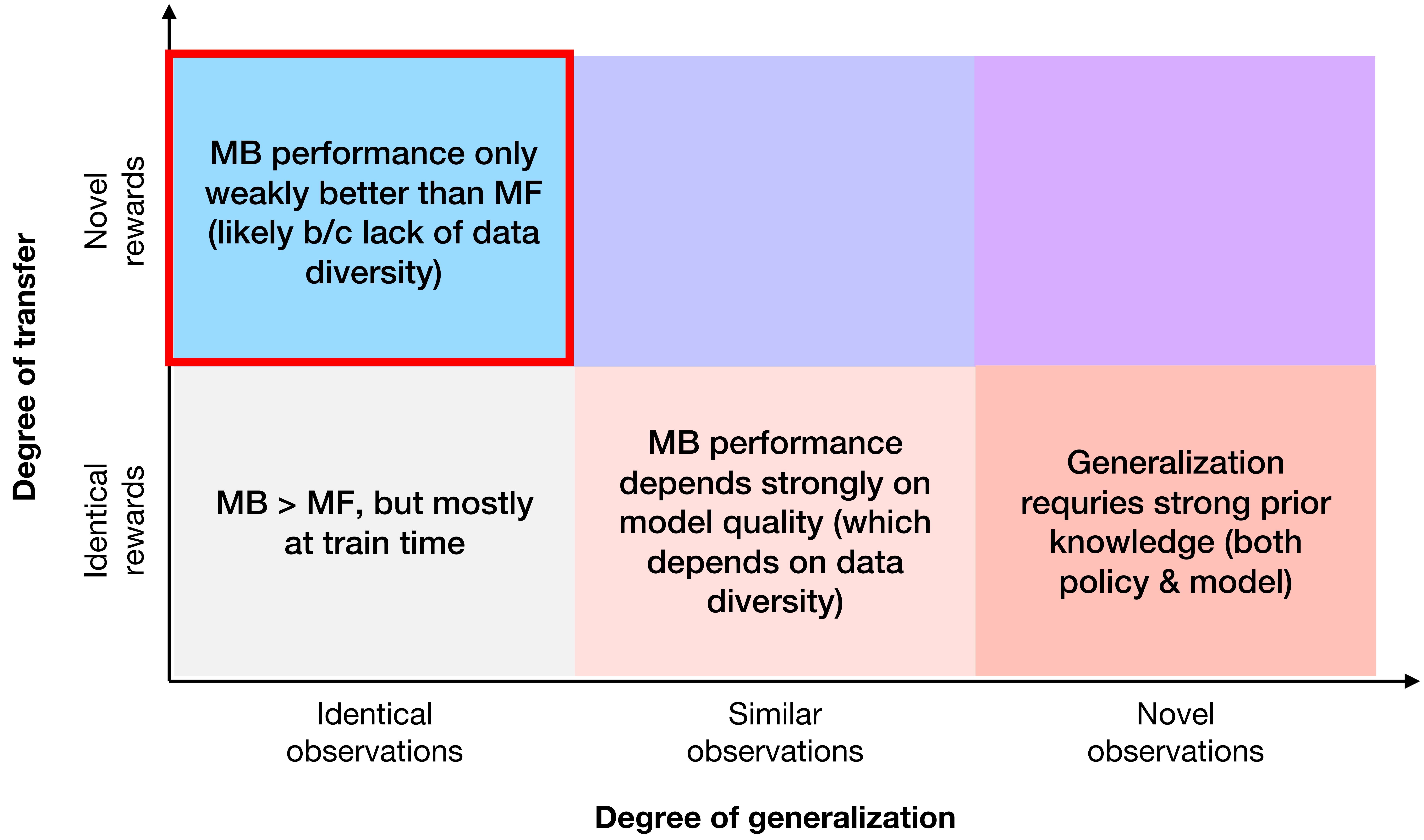
Transfer in Robodesk



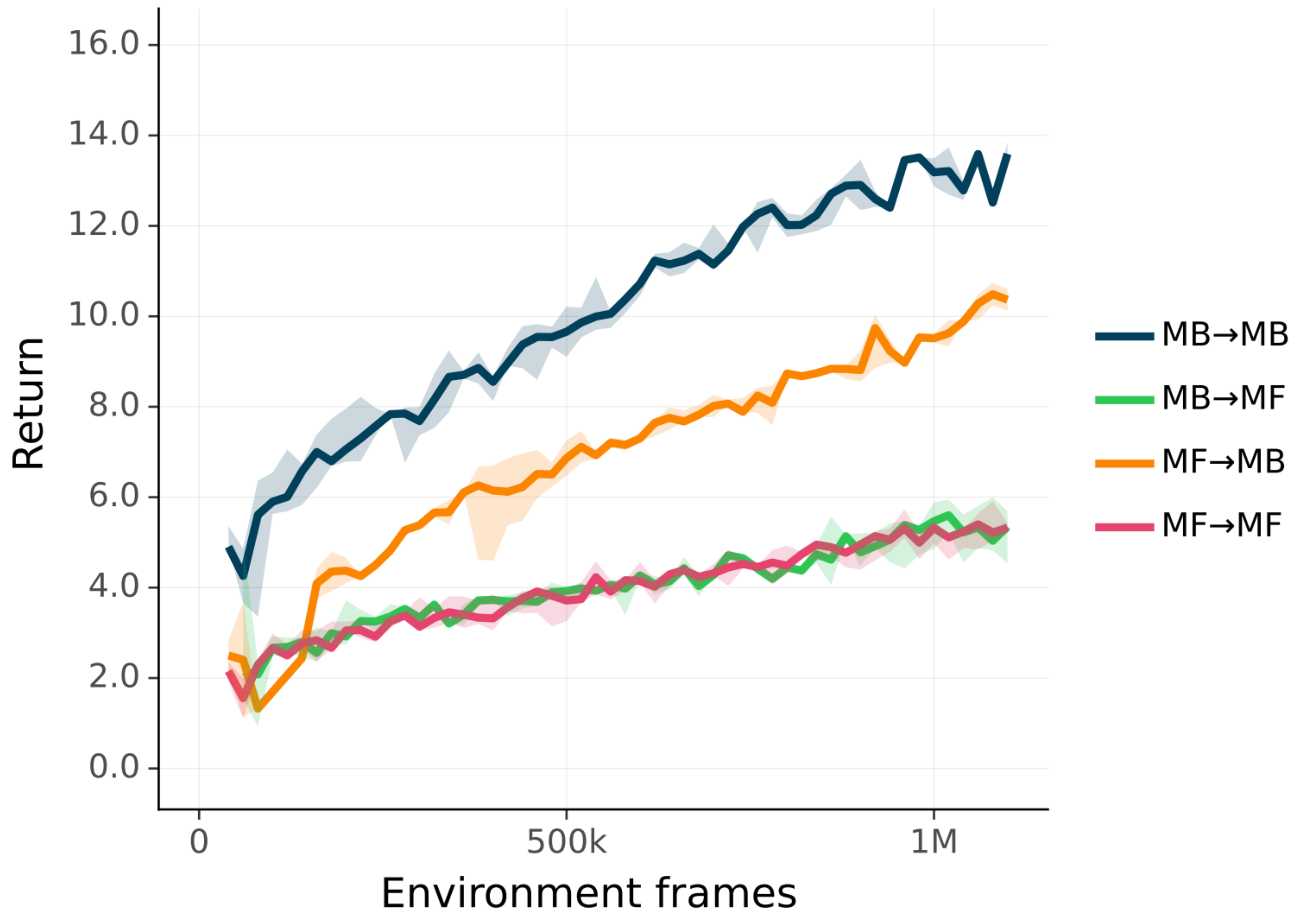
→ MB leads to slightly improved transfer performance, though the effect is weak

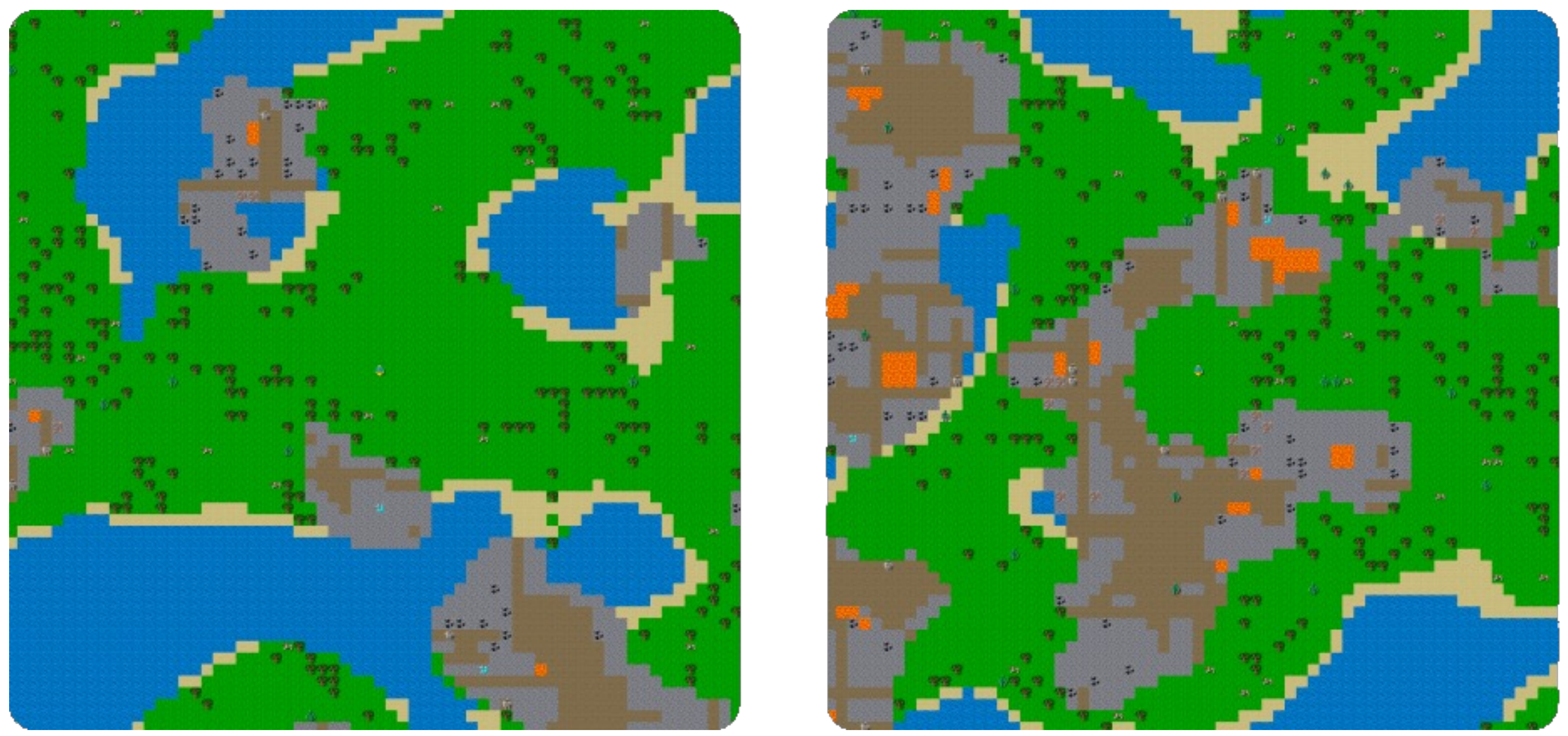
Walker, Vértes, Li et al. (2023). Investigating the role of model-based learning in exploration and transfer. ICML.

Lessons in generalization & transfer



Transfer in Crafter



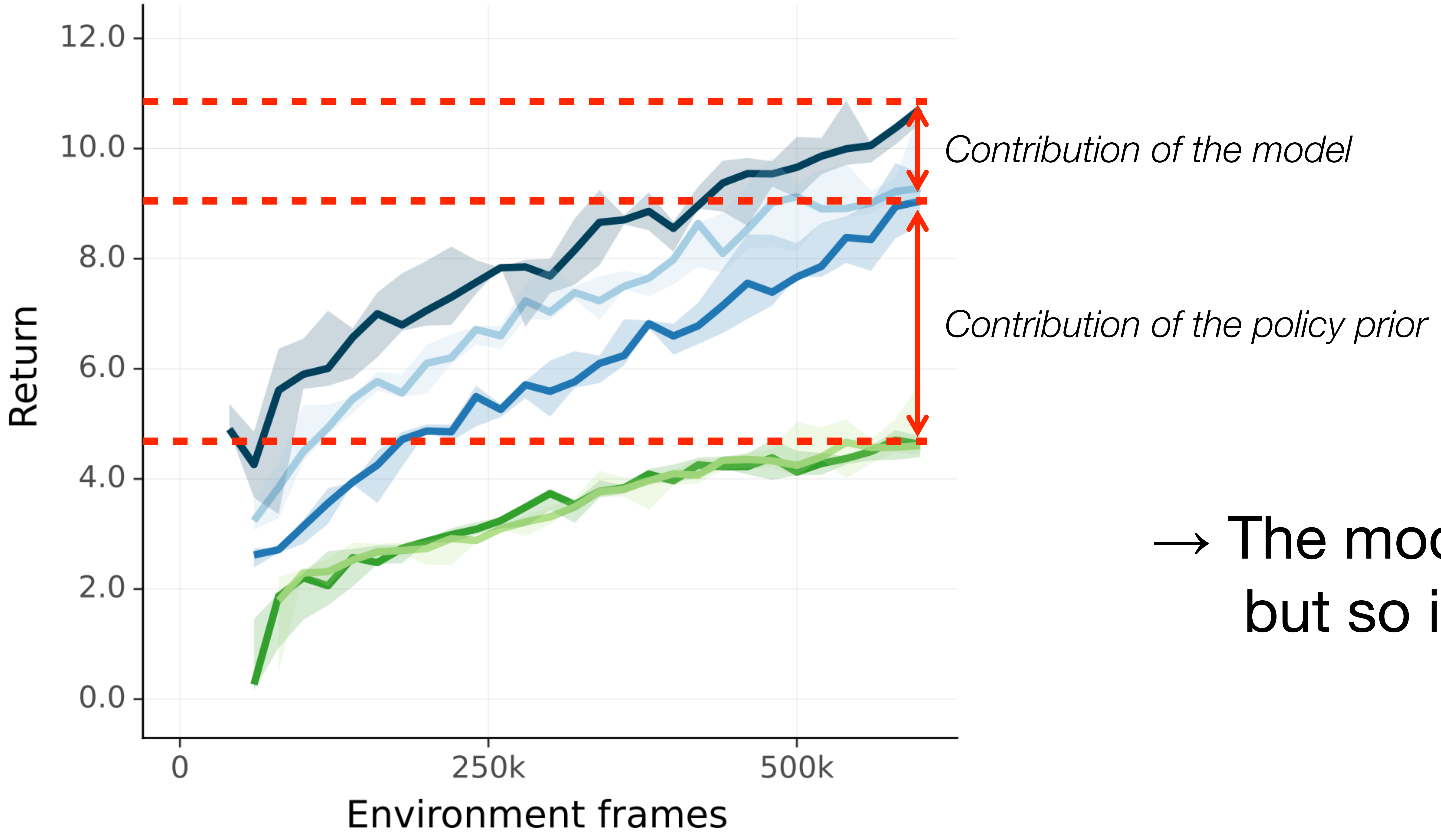


Crafter (Hafner, 2021)

→ MB leads to improved transfer performance, and matters a lot for finetuning

Walker, Vértes, Li et al. (2023). Investigating the role of model-based learning in exploration and transfer. ICML.

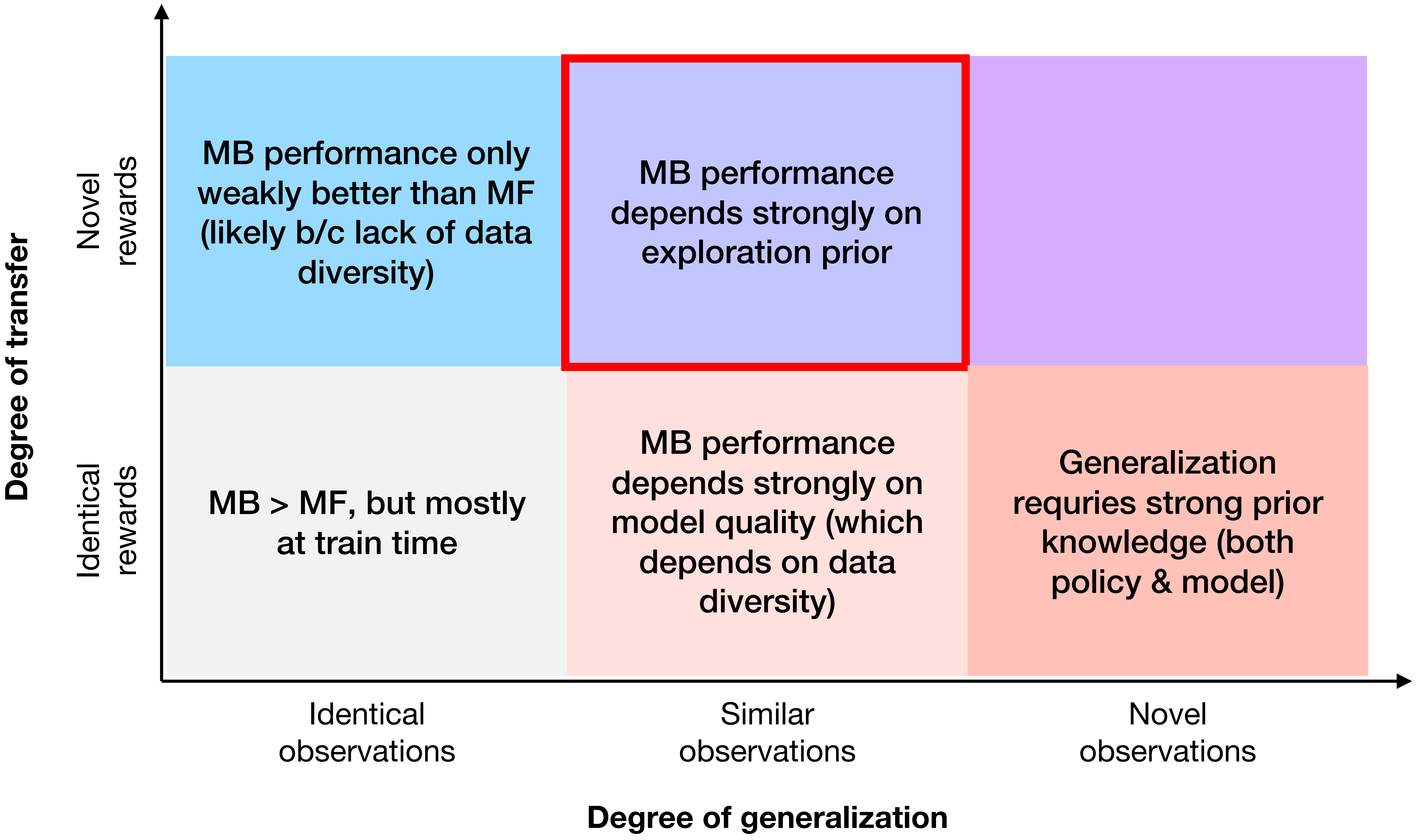
Contribution of different components



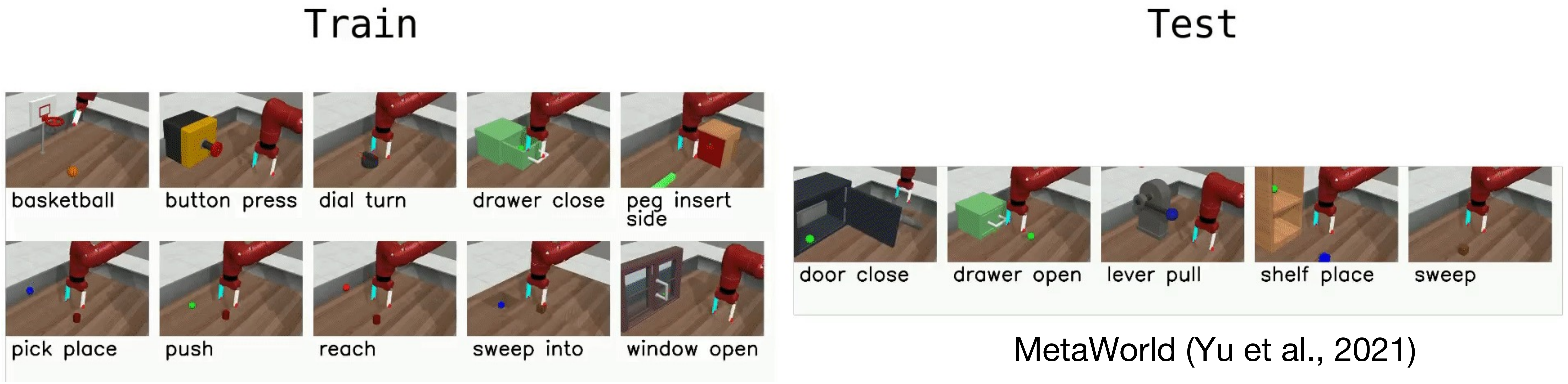
→ The model is important for transfer, but so is the exploration policy!

Walker, Vértes, Li et al. (2023). Investigating the role of model-based learning in exploration and transfer. ICML.

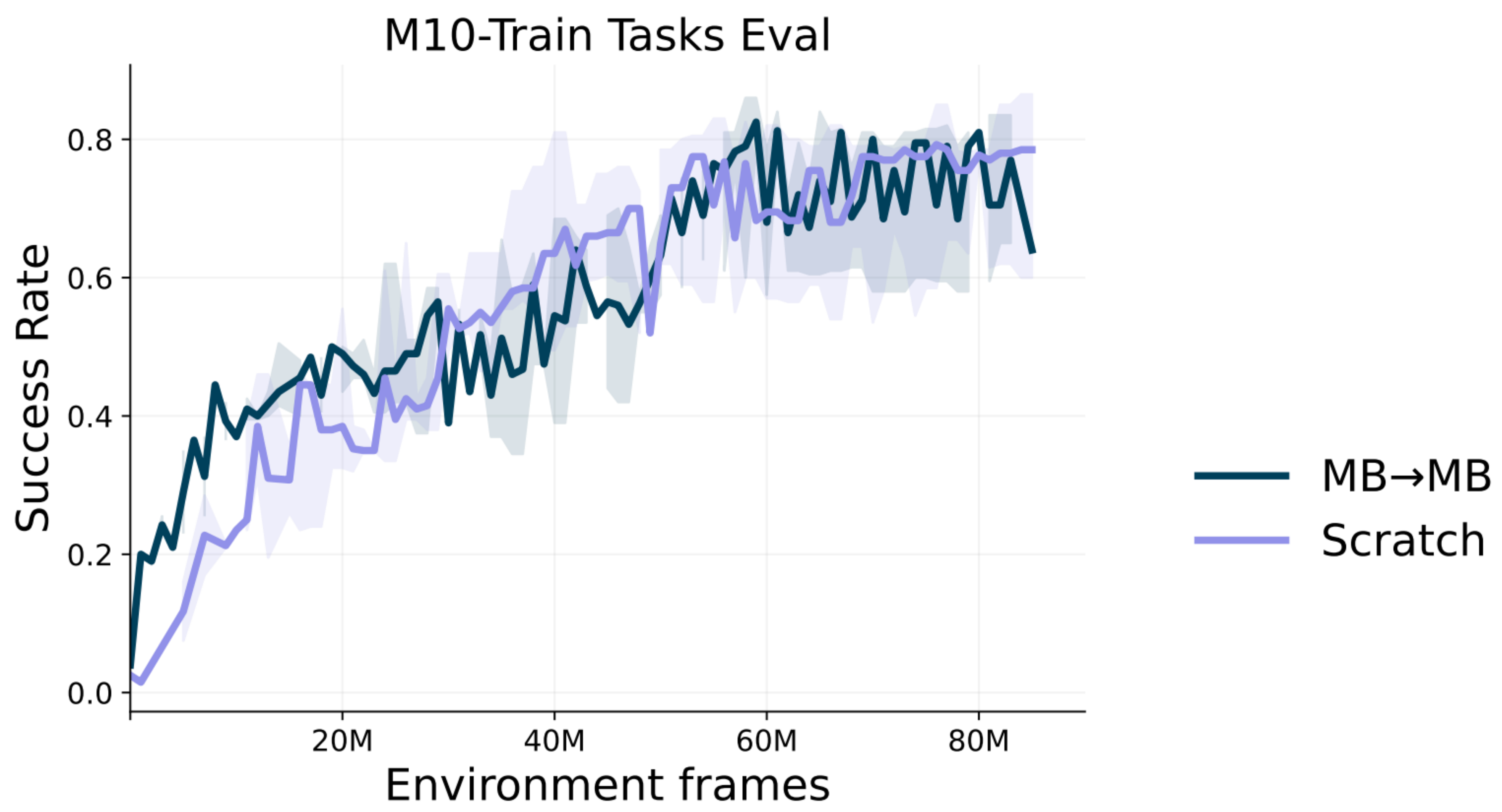
Lessons in generalization & transfer



Transfer in MetaWorld



MetaWorld (Yu et al., 2021)



→ MBRL may not substantially improve transfer performance if there is a large environment shift

Walker, Vértes, Li et al. (2023). Investigating the role of model-based learning in exploration and transfer. ICML.

Lessons in generalization & transfer

Degree of transfer	Identical observations	Similar observations	Novel observations
Novel rewards	MB performance only weakly better than MF (likely b/c lack of data diversity)	MB performance depends strongly on exploration prior	Need higher diversity & stronger prior knowledge for success
Identical rewards	MB > MF, but mostly at train time	MB performance depends strongly on model quality (which depends on data diversity)	Generalization requries strong prior knowledge (both policy & model)

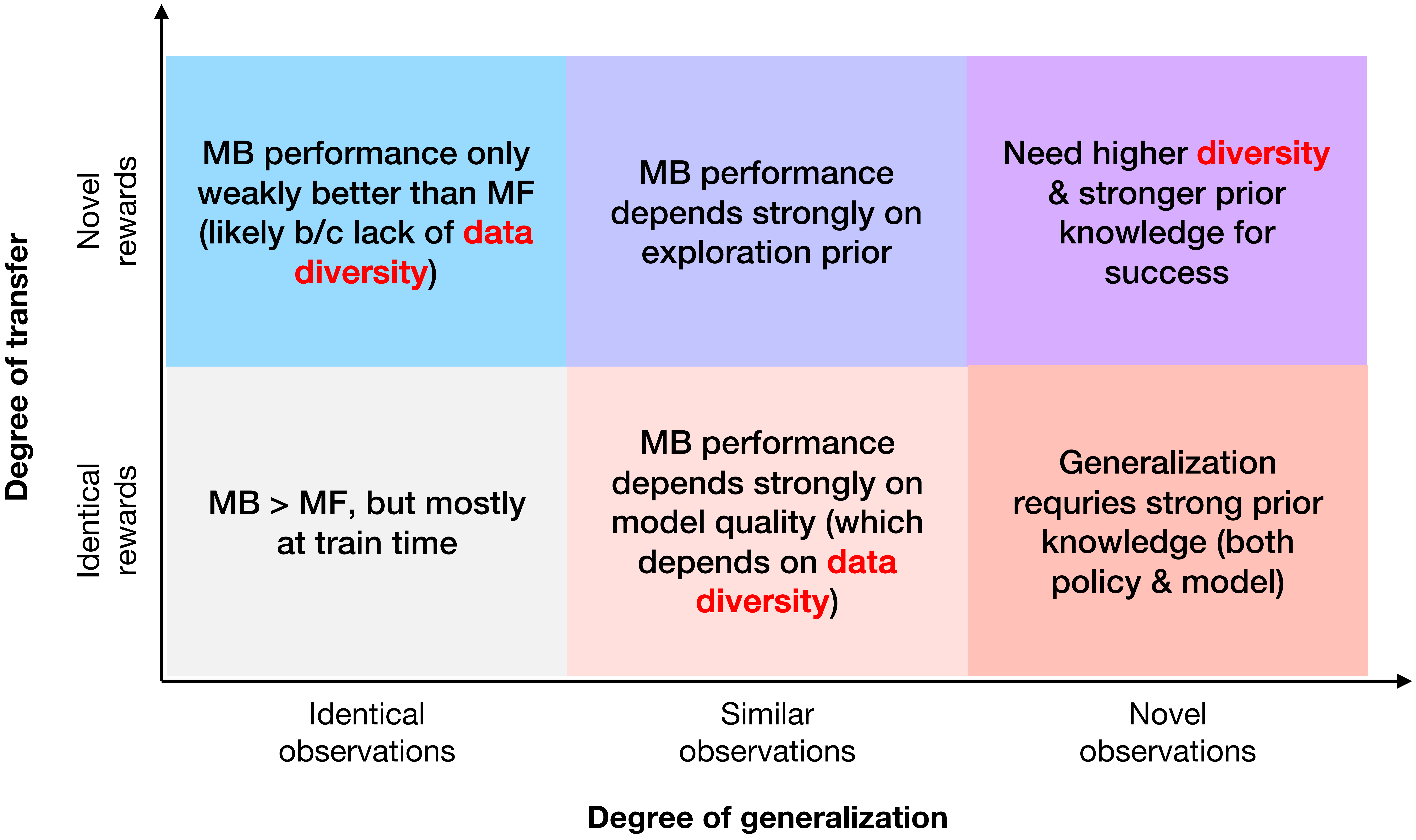
Degree of generalization

Lessons in generalization & transfer

Degree of transfer	Identical observations	Similar observations	Novel observations
Novel rewards	MB performance only weakly better than MF (likely b/c lack of data diversity)	MB performance depends strongly on **exploration prior**	Need higher diversity & stronger **prior knowledge** for success
Identical rewards	MB > MF, but mostly at train time	MB performance depends strongly on model quality (which depends on data diversity)	Generalization requries strong **prior knowledge** (both policy & model)

Degree of generalization

Lessons in generalization & transfer



Ingredients for generalization & transfer

Model-based learning

High-quality world model

High-quality exploration prior

Missing ingredient: (Pre-)training
on lots of high-quality, diverse data

Foundation models as the missing ingredient?

SayCan (Ahn et al., 2022)

… and yet …

Hallucinate / make stuff up

Get distracted by irrelevant context

Struggle with symbolic/abstract reasoning

Make simple calculation errors

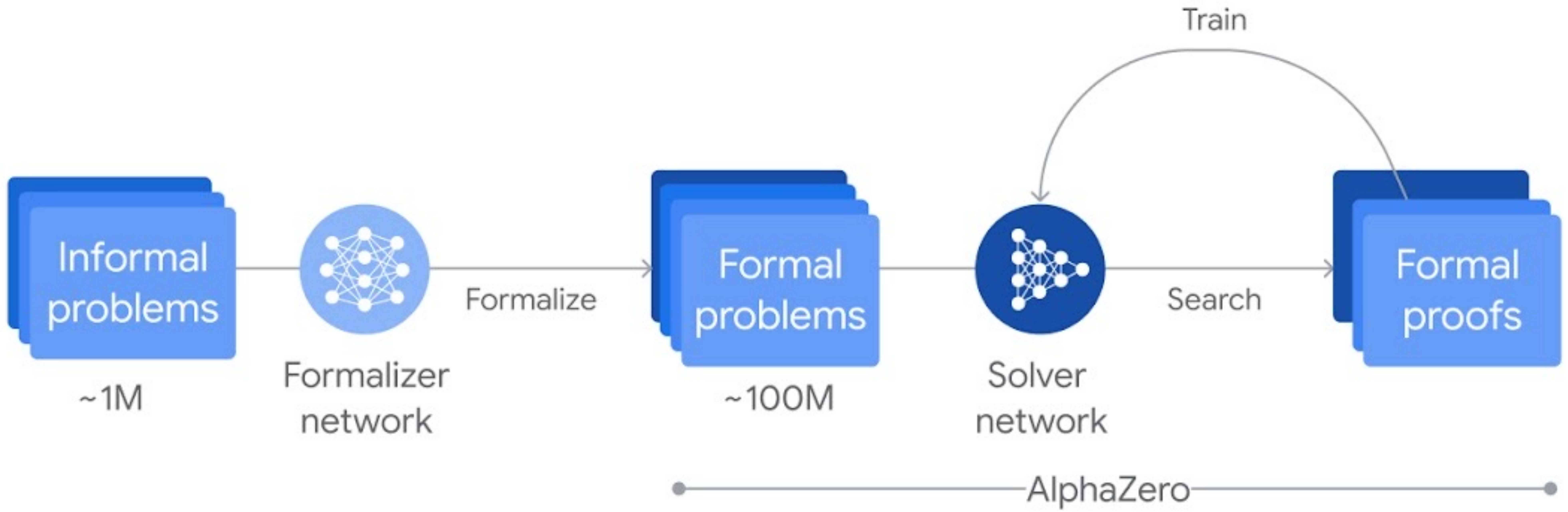
Get stuck in loops

Model-based learning

+

Foundation models

Model-based learning + foundation models



AlphaProof & AlphaGeometry 2 (2024)
Together achieved silver medal standard at the IMO!

Model-based learning + foundation models

+

A **golden age** for
neurosymbolic AI?

Thanks!

Kelsey Allen
Ankesh Anand
Victor Bapst
Peter Battaglia
Lars Buesing
Thomas Anthony
Feryal Behbahani
Lars Buesing
Carl Doersch
Gabriel Dulac-Arnold
Abe Friesen
Arthur Guez
Pushmeet Kohli

Yazhe Li
Sherjil Ozair
Tobias Pfaff
Alvaro Sanchez-Gonzalez
Julian Schrittwieser
Kim Stachenfeld
Petar Veličković
Eszter Vértes
Fabio Viola
Jacob Walker
Sims Witherspoon
Theo Weber

Bapst*, Sanchez-Gonzalez*, Doersch, Stachenfeld, Kohli, Battaglia, Hamrick (2019). Structured agents for physical construction. *ICML.*

Hamrick, Friesen, Behbahani, Guez, Viola, Witherspoon, Anthony, Buesing, Veličković, & Weber (2021). On the role of planning in model-based deep reinforcement learning. *ICLR.*

Anand*, Walker*, Li, Vértes, Schrittwieser, Ozair, Weber, & Hamrick (2022). Procedural generalization by planning with self-supervised world models. *ICLR.*

Walker*, Vértes*, Li*, Dulac-Arnold, Anand, Weber, & Hamrick (2023). Investigating the role of model-based learning in exploration and transfer. *ICML.*