On the Role of Planning in Model-Based Deep Reinforcement Learning

Jessica B. Hamrick
jhamrick@deepmind.com
DeepMind

CCN GAC: "What is the place of planning?"
August 28, 2022

Abraham (2020). The Cambridge Handbook of the Imagination.
Agostinelli et al. (2019). Solving the Rubik's Cube with Deep Reinforcement Learning and Search. NMI.
Allen et al. (2019). The tools challenge: Rapid trial-and-error learning in physical problem solving. CogSci 2019
Amos et al (2018). Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.
Amos et al (2019). The Differentiable Cross-Entropy Method. arXiv.
Anthony et al (2017). Thinking Fast and Slow with Deep Learning and Tree Search. NeurIPS.
Bellemare et al (2016). Unifying count-based exploration and intrinsic motivation. NeurIPS.
Buesing et al. (2018). Learning and Querying Fast Generative Models for Reinforcement Learning. ICML 2018.
Burgess et al. (2019). MONet: Unsupervised Scene Decomposition and Representation. arXiv.
Byravan et al (2019). Imagined Value Gradients. CoRL 2019.
Chiappa, Racaniere, Wierstra, Mohamed (2017). Recurrent environment simulators. ICLR 2017.
Choromanski et al (2019). Provably Robust Blackbox Optimization for Reinforcement Learning. CoRL 2019.
Chua, Calandra, McAllister, & Levine (2018). Deep reinforcement learning in a handful of trials using probabilistic dynamics models. NeurIPS 2018.
Corneil et al. (2018). Efficient Model-Based Deep Reinforcement Learning with Variational State Tabulation. ICML.
Depeweg et al. (2017). Learning and policy search in stochastic dynamical systems with bayesian NNs. ICLR 2017.
Du et al (2019). Model-Based Planning with Energy Based Models. CoRL 2019
Dubey, Agrawal, Pathak, Griffiths, & Efros (2018). Investigating human priors for playing video games. ICML 2018.
Ebert, Finn, et al. (2018). Visual foresight: Model-based deep RL for vision-based robotic control. arXiv.
Ecoffet et al. (2019). Go-explore: a new approach for hard-exploration problems. arXiv.
Edwards, Downs, & Davidson (2018). Forward-Backward Reinforcement Learning. arXiv.
Ellis et al. (2019). Write, Execute, Assess: Program Synthesis with a REPL. NeurIPS.
Eysenbach, Salakhutdinov, & Levine (2019). Search on the replay buffer: Bridging planning and RL. NeurIPS.
Farquhar et al (2017). TreeQN and ATreeC: Differentiable Tree-Structured Models for Deep RL. ICLR 2018.
Fazeli et al. (2019). See, feel, act: Hierarchical learning for complex manipulation skills with multisensory fusion. Science Robotics, 4(26).
Finn & Levine (2017). Deep visual foresight for planning robot motion. ICRA.
Finn, Goodfellow, & Levine (2016). Unsupervised learning for physical interaction through video prediction. NeurIPS.
Fisac et al. (2019). A General Safety Framework for Learning-Based Control in Uncertain Robotic Systems. IEEE Transactions on Automatic Control.
Gal et al. (2016). Improving PILCO with Bayesian neural network dynamics models. In Data-Efficient Machine Learning workshop, ICML.
Grill et al. (2020). Monte-Carlo tree search as regularized policy optimization. ICML.
Gu, Lillicrap, Sutskever, & Levine (2016). Continuous Deep Q-Learning with Model-based Acceleration. ICML 2016.
Guez et al (2019). An Investigation of Model-Free Planning. ICML 2019.
Ha & Schmidhuber (2018). World Models. NeurIPS 2018.
Hafner et al (2019). Dream to Control: Learning Behaviors by Latent Imagination. ICLR 2020.
Hamrick (2019). Analogues of mental simulation and imagination in deep learning. Current Opinion in Behavioral Sciences, 29, 8-16.
Hamrick et al (2020). Combining Q-Learning and Search with Amortized Value Estimates. ICLR 2020.
Hamrick et al. (2017). Metacontrol for adaptive imagination-based optimization. ICLR 2017.
Heess et al (2015). Learning Continuous Control Policies by Stochastic Value Gradients. NeurIPS 2015.
Houthooft et al (2016). VIME: Variational Information Maximizing Exploration. NeurIPS 2016.
Jaderberg et al. (2017). Reinforcement learning with unsupervised auxiliary tasks. ICLR 2017.
Jang, Gu, & Poole (2017). Categorical Reparameterization with Gumbel-Softmax. ICLR 2017.
Janner et al (2019). When to Trust Your Model: Model-Based Policy Optimization. NeurIPS 2019.
Jurgenson et al. (2019). Sub-Goal Trees -- A Framework for Goal-Directed Trajectory Prediction and Optimization. arXiv.
Kidambi et al (2020). MOReL: Model-Based Offline Reinforcement Learning. arXiv.
Konidaris, Kaelbling, & Lozano-Pérez (2018). From Skills to Symbols: Learning Symbolic Representations for Abstract High-Level Planning. JAIR.
Kurutach et al. (2018). Learning Plannable Representations with Causal InfoGAN. NeurIPS.
Laskin, Emmons, Jain, Kurutach, Abbeel, & Pathak (2020). Sparse Graphical Memory for Robust Planning. arXiv.
Levine, Wagener, & Abbeel (2015). Learning Contact-Rich Manipulation Skills with Guided Policy Search. ICRA 2015.
Levine et al (2020). Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems. arXiv.
Lin et al (2020). Model-based Adversarial Meta-Reinforcement Learning. arXiv.
Lowrey et al. (2019). Plan Online, Learn Offline: Efficient Learning and Exploration via Model-Based Control. ICLR 2019.
Lu, Mordatch, & Abbeel (2019). Adaptive Online Planning for Continual Lifelong Learning. NeurIPS Deep RL Workshop.
Maddison, Mnih, & Teh (2017). The Concrete Distribution. ICLR 2017.
Mordatch et al (2015). Ensemble-CIO: Full-body dynamic motion planning that transfers to physical humanoids. IROS 2015.
Mordatch et al (2015). Interactive Control of Diverse Complex Characters with Neural Networks. NeurIPS 2015.
Nagabandi et al (2019). Deep Dynamics Models for Learning Dexterous Manipulation. CoRL 2019.
Nagabandi et al. (2019). Learning to Adapt in Dynamic, Real-World Environments through Meta-Reinforcement Learning. ICLR.
Nair, Babaeizadeh, Finn, Levine & Kumar (2020). Time Reversal as Self-Supervision. ICRA 2020.
Nair, Pong, et al. (2018). Visual Reinforcement Learning with Imagined Goals. NeurIPS.
Nasiriany et al. (2019). Planning with Goal-Conditioned Policies. NeurIPS.
Oh, Guo, Lee, Lewis, & Singh (2015). Action-Conditional Video Prediction using Deep Networks in Atari Games. NIPS 2015.
Oh et al. (2017). Value Prediction Network. NeurIPS.
OpenAI et al. (2020). Learning Dexterous In-Hand Manipulation. International Journal of Robotics Research, 39(1), 3-20.
OpenAI et al. (2019). Solving Rubik's Cube with a Robot Hand. arXiv.
Osband et al (2018). Randomized Prior Functions for Deep Reinforcement Learning. NeurIPS 2018.
Parascandolo, Buesing, et al. (2020). Divide-and-Conquer Monte Carlo Tree Search For Goal-Directed Planning. arXiv.
Pascanu, Li, et al. (2017). Learning model-based planning from scratch. arXiv.
Pathak et al. (2017). Curiosity-driven exploration by self-supervised prediction. ICML.
Peters, Mulling, & Altun (2010). Relative Entropy Policy Search. AAAI 2010.
Rajeswaran et al. (2017). EPOpt: Learning Robust Neural Network Policies Using Model Ensembles. ICLR 2017.
Rajeswaran et al. (2020). A Game Theoretic Framework for Model Based Reinforcement Learning. arXiv.
Sadigh et al. (2016). Planning for autonomous cars that leverage effects on human actions. RSS 2016.
Sanchez-Gonzalez et al. (2018). Graph Networks as Learnable Physics Engines for Inference and Control. ICML 2018.
Savinov, Dosovitskiy, & Koltun (2018). Semi-parametric topological memory for navigation. ICLR 2018.
Schrittwieser et al. (2019). Mastering Atari, Go, Chess and Shogi by planning with a learned model. arXiv.
Segler, Preuss, & Waller (2018). Planning chemical syntheses with deep neural networks and symbolic AI. Nature, 555(7698).
Sharma et al. (2020). Dynamics-Aware Unsupervised Discovery of Skills. ICLR.
Shen et al. (2019). M-Walk: Learning to Walk over Graphs using Monte Carlo Tree Search. NeurIPS.
Silver, van Hasselt, Hessel, Schaul, Guez, Harley, Dulac-Arnold, Reichert, Rabinowitz, Barreto, Degris (2017). The Predictron: End-To-End Learning and Planning. ICML 2017.
Silver et al. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484.
Silver et al. (2017). Mastering the game of Go without human knowledge. Nature, 550, 354-359.
Silver et al. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science.
Sutton and Barto (2018). Reinforcement Learning: An Introduction.
Tamar et al. (2016). Value iteration networks. NeurIPS 2016.
Tamar et al. (2017). Learning from the Hindsight Plan – Episodic MPC Improvement. ICRA 2017.
Tzeng et al. (2017). Adapting Deep Visuomotor Representations with Weak Pairwise Constraints.
van den Oord, Li, & Vinyals (2019). Representation Learning with Contrastive Predictive Coding. arXiv.
van Hasselt, Hessel, & Aslanides (2019). When to use parametric models in reinforcement learning? NeurIPS 2019.
Veerapaneni, Co-Reyes, Chang, et al. (2019). Entity Abstraction in Visual Model-Based Reinforcement Learning. CoRL 2019.
Watter, Springenberg, Boedecker, & Riedmiller (2015). Embed to Control: A Locally Linear Latent Dynamics Model for Control from Raw Images. NeurIPS 2015.
Weber et al. (2017). Imagination-augmented agents for deep reinforcement learning. NeurIPS 2017.
Williams et al. (2017). Information Theoretic MPC for Model-Based Reinforcement Learning. ICRA 2017.
Wu et al. (2015). Galileo: Perceiving Physical Object Properties by Integrating a Physics Engine with Deep Learning. NeurIPS 2015.
Yu et al (2020). MOPO: Model-based Offline Policy Optimization. arXiv.
Zhang, Lerer, et al. (2018). Composable Planning with Attributes. ICML 2018.

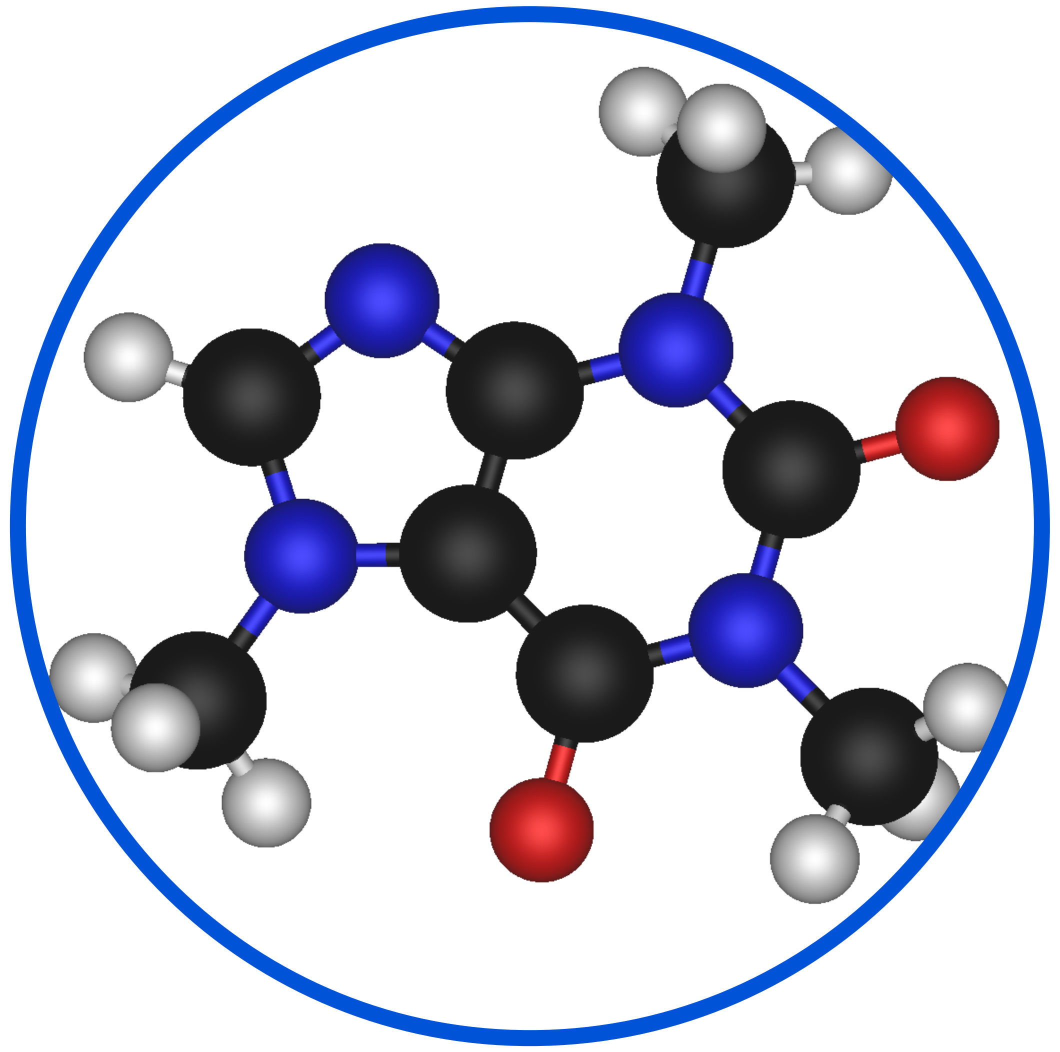
Silver et al. (2016)

Silver et al. (2016)

OpenAI et al. (2019)

Silver et al. (2016)

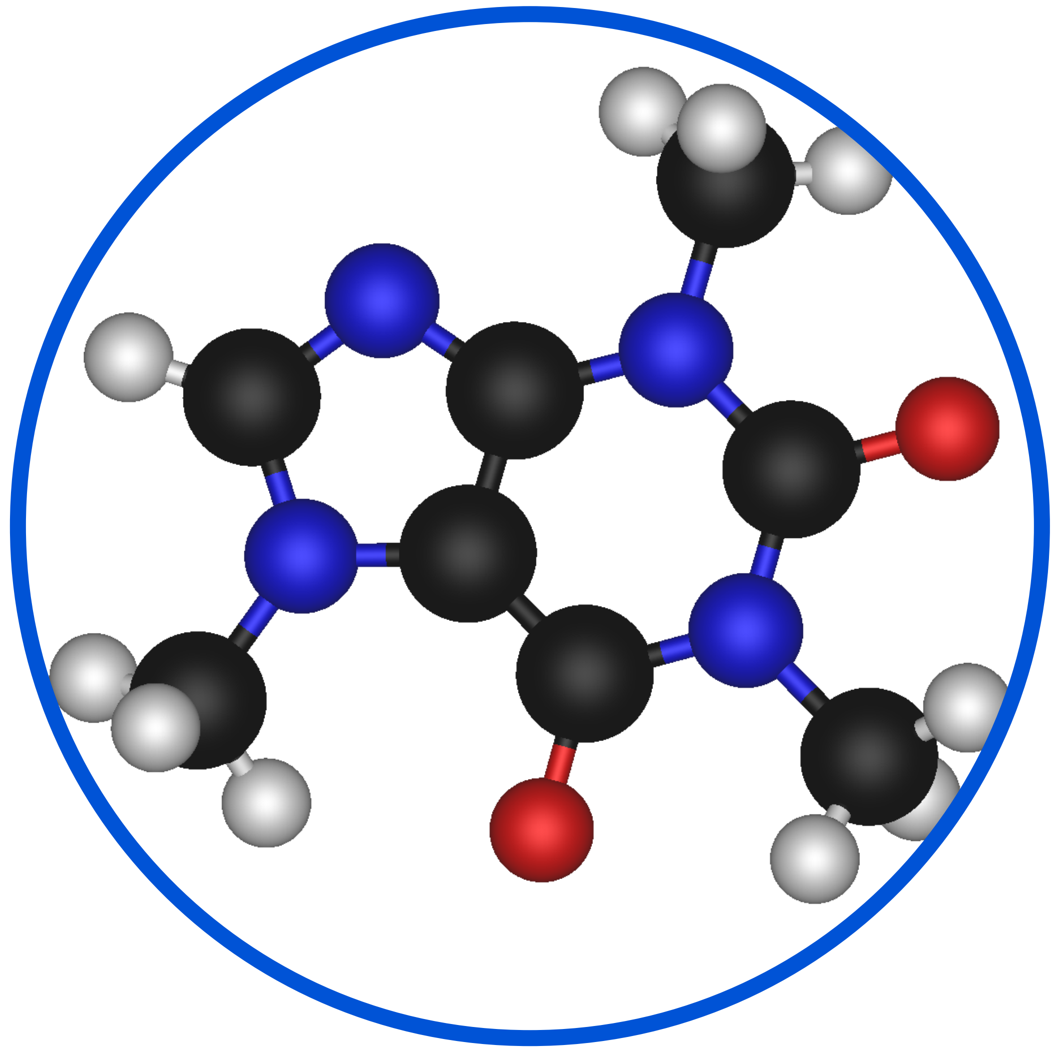
OpenAI et al. (2019)



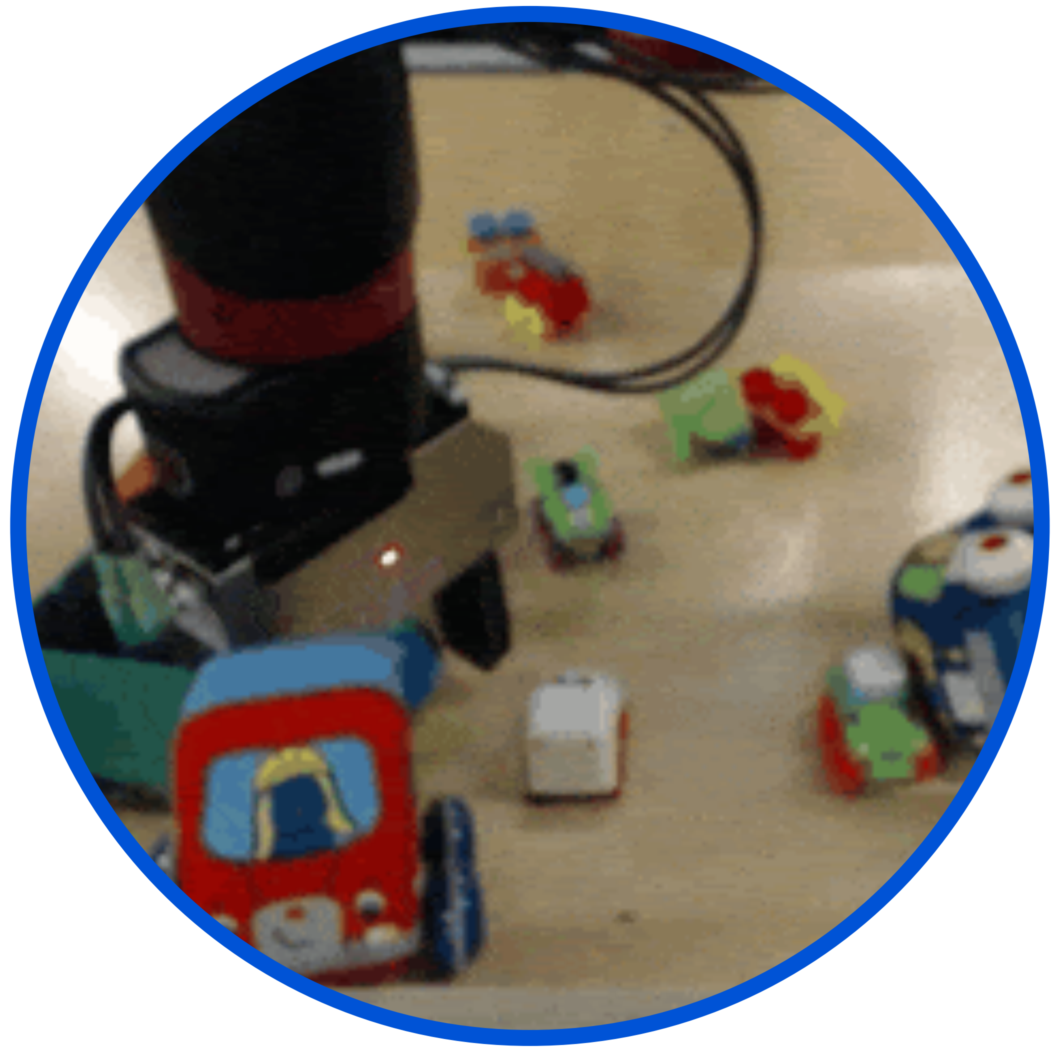
Segler et al. (2018)

Silver et al. (2016)

OpenAI et al. (2019)



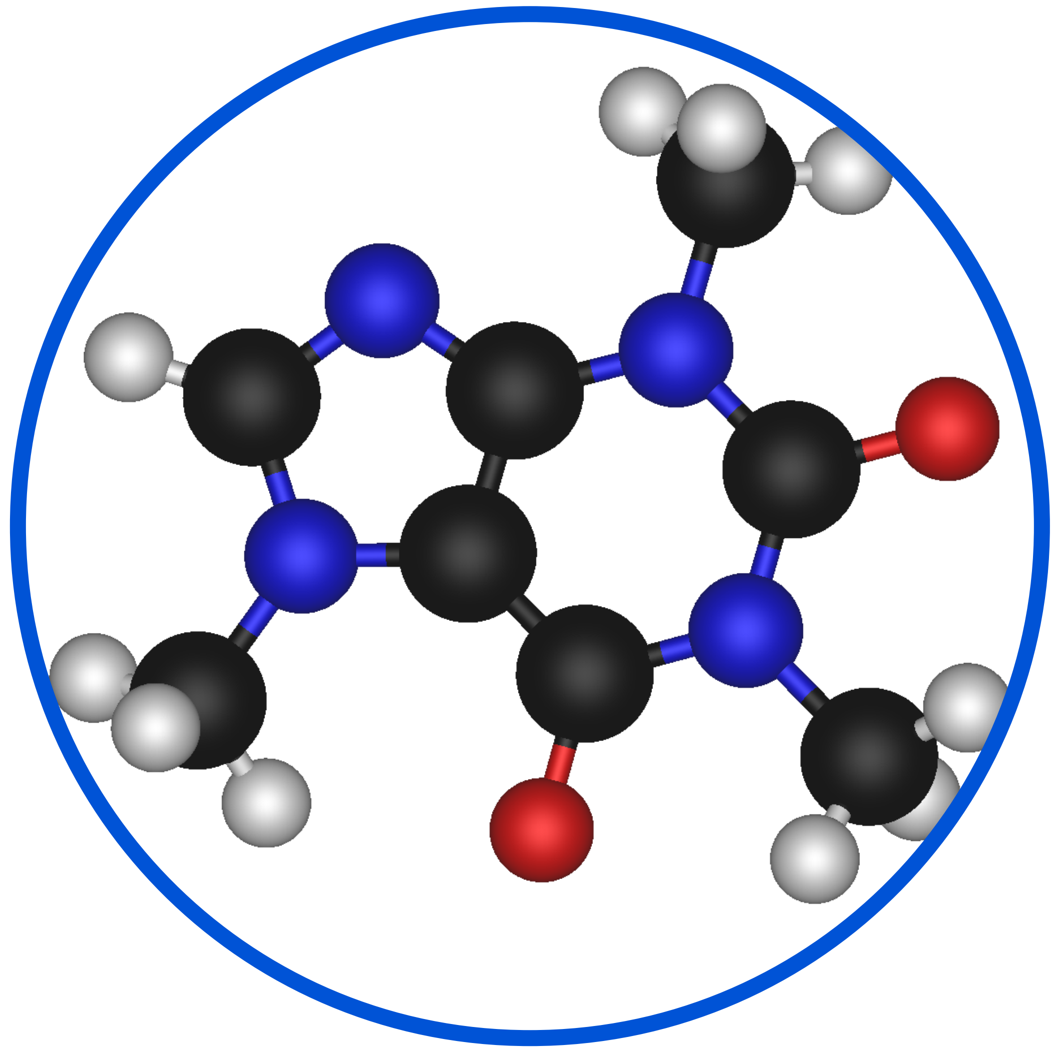
Segler et al. (2018)



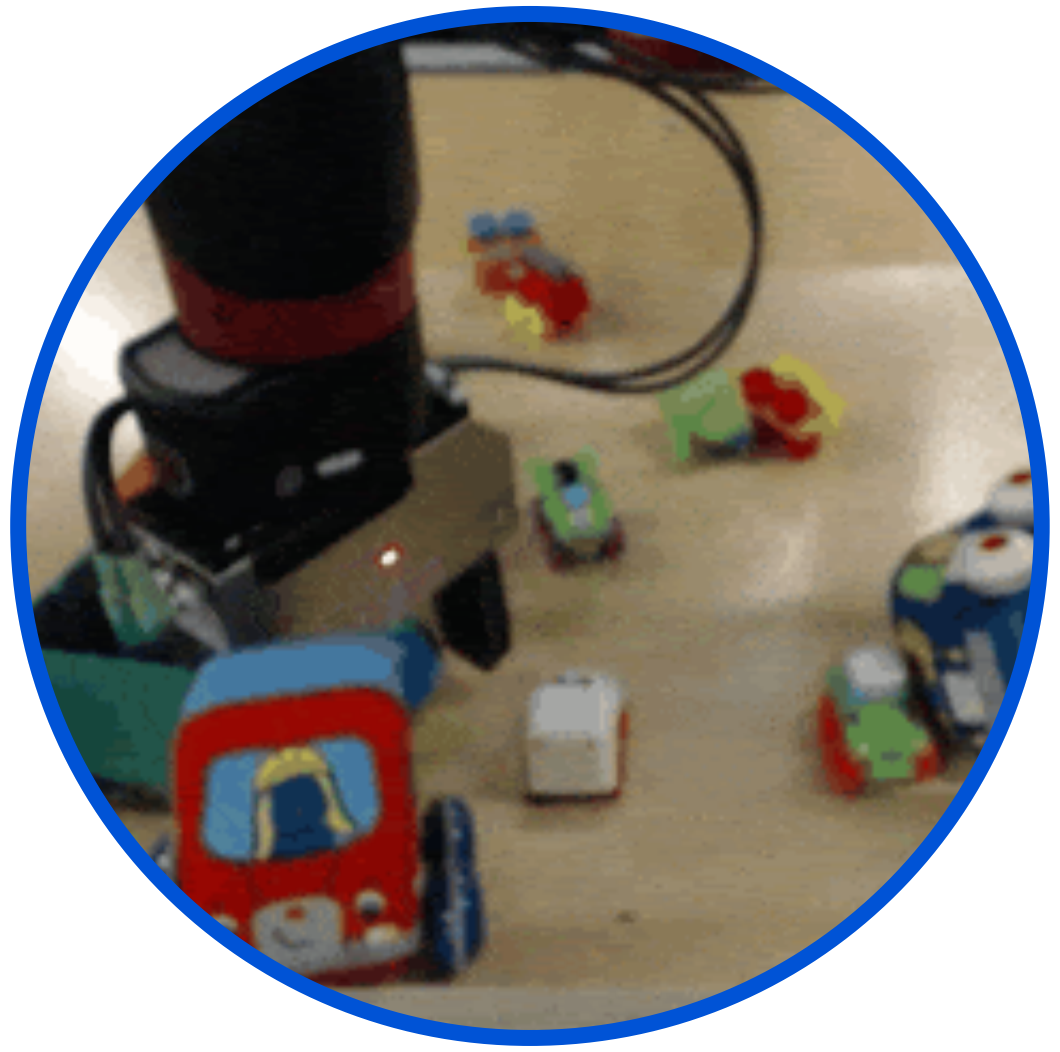
Finn et al. (2018)

Silver et al. (2016)

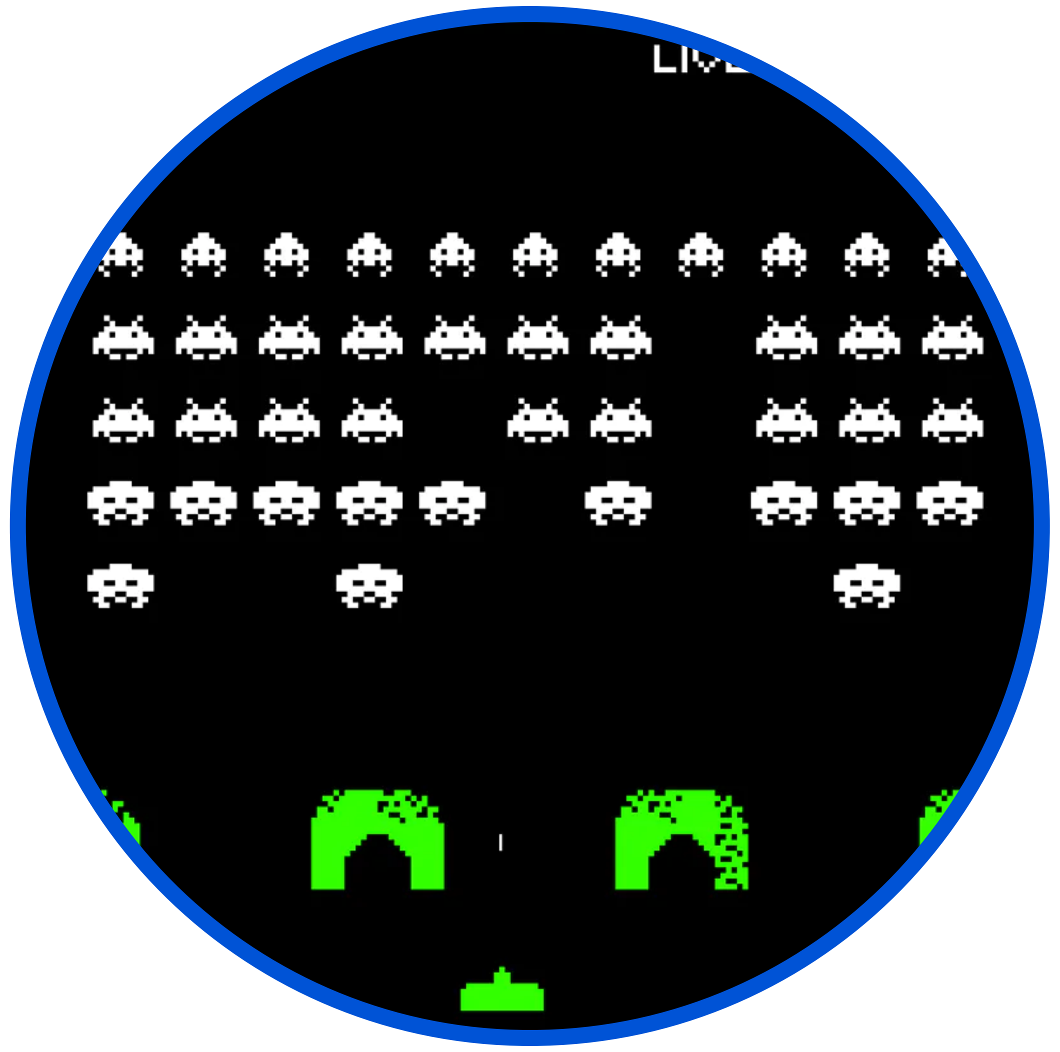
OpenAI et al. (2019)



Segler et al. (2018)



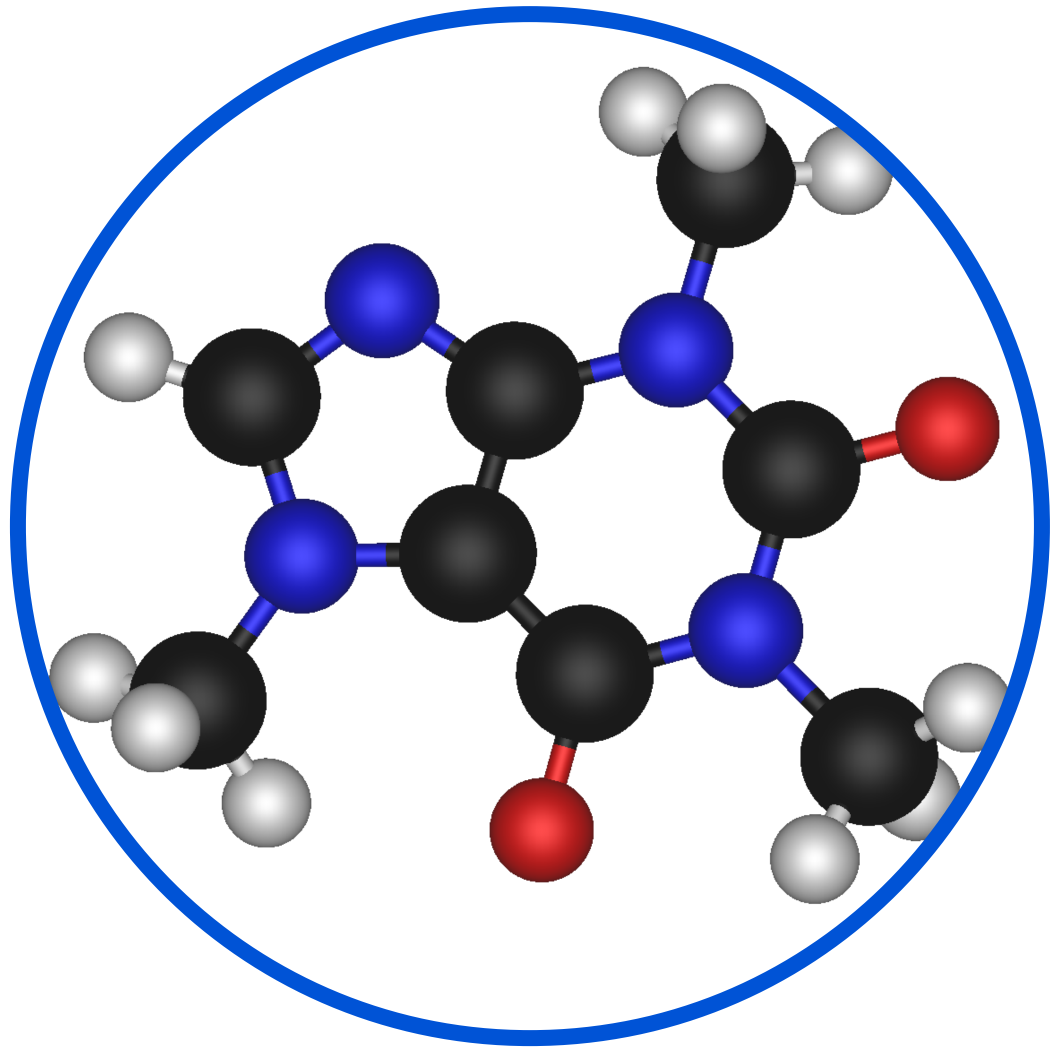
Finn et al. (2018)



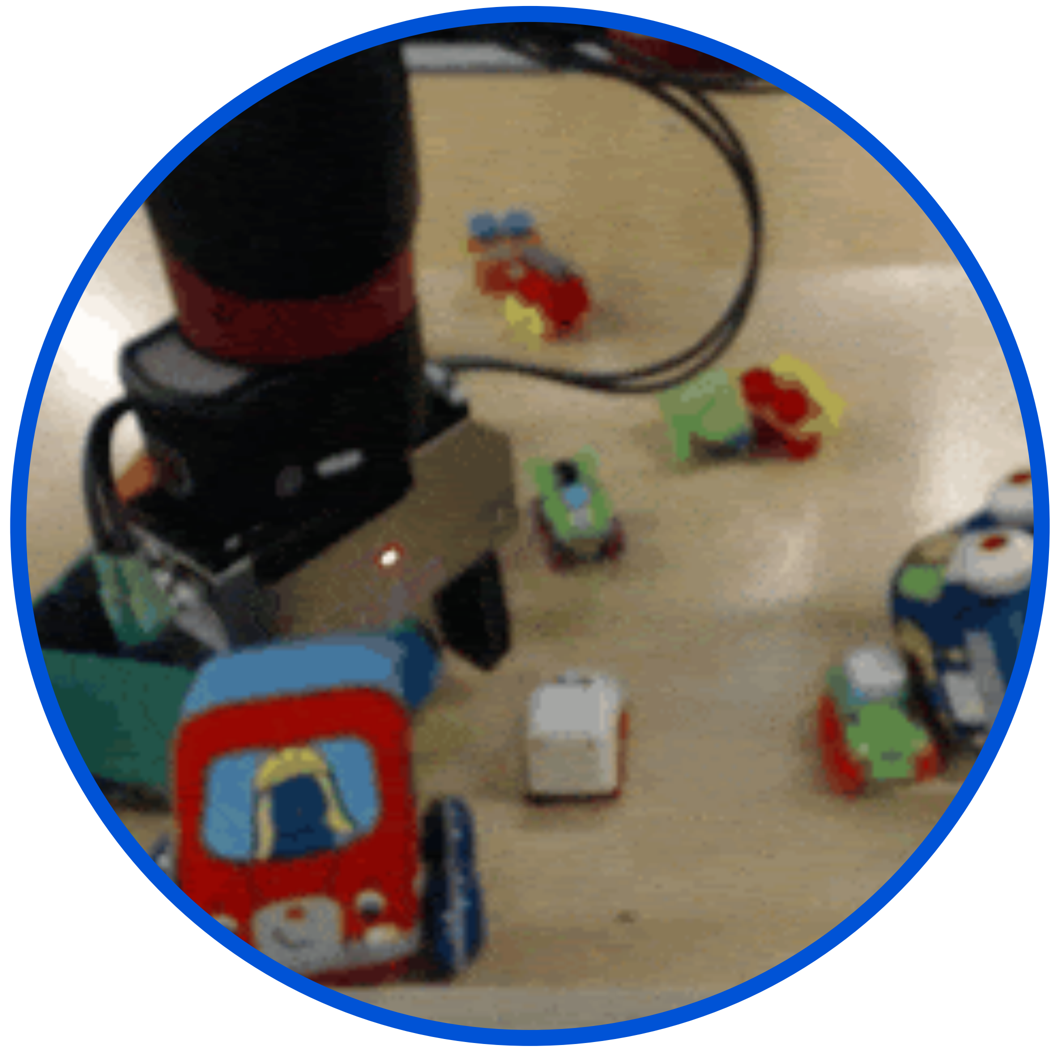
Schrittwieser et al. (2020)

Silver et al. (2016)

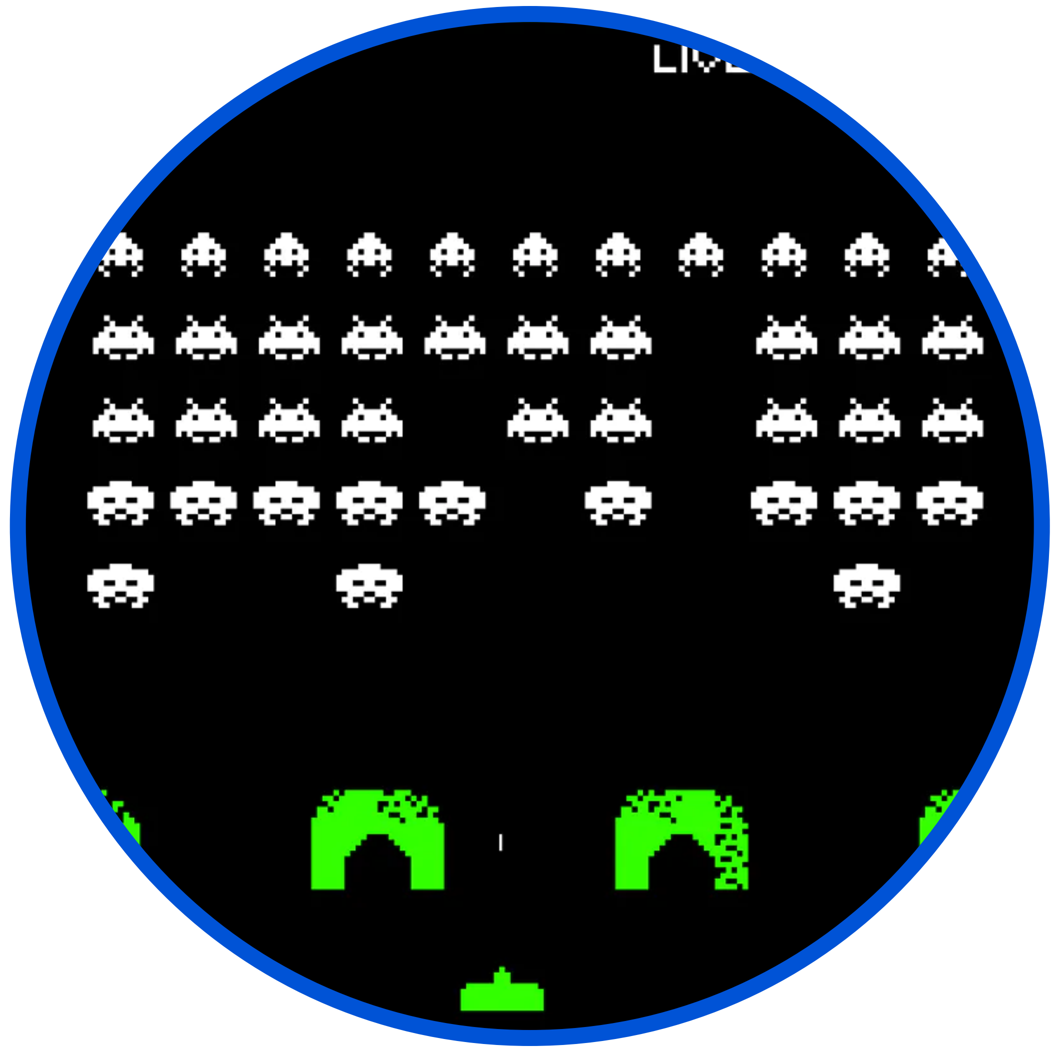
OpenAI et al. (2019)



Segler et al. (2018)



Finn et al. (2018)

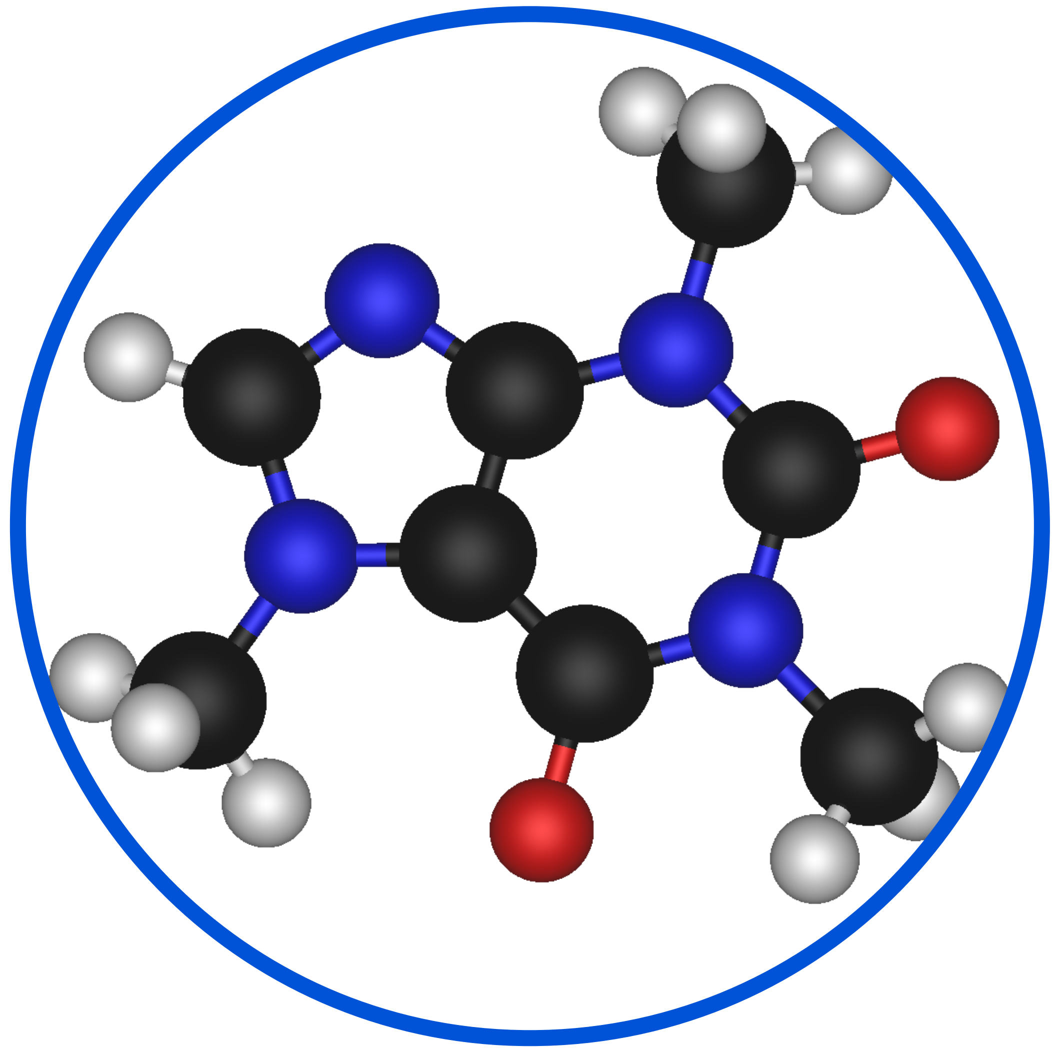


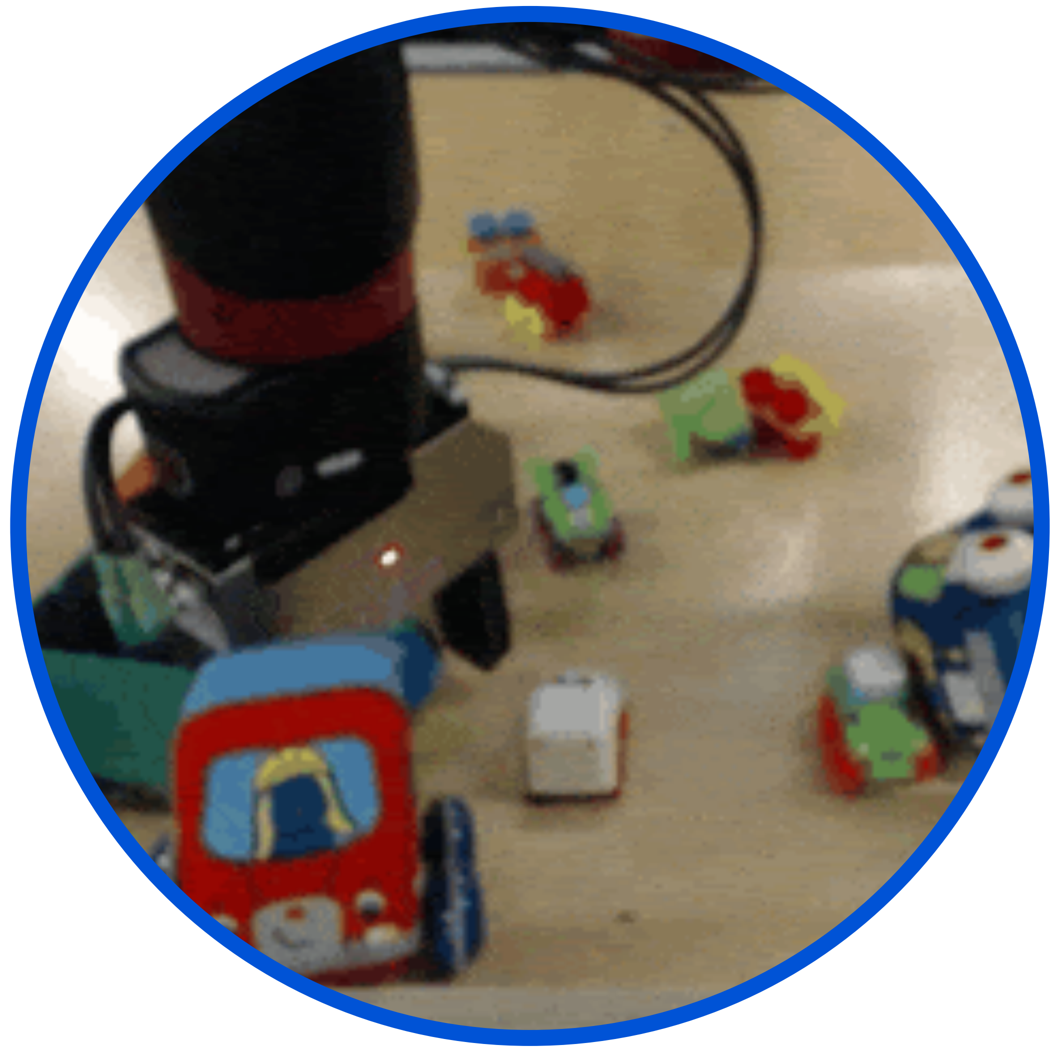
Schrittwieser et al. (2020)

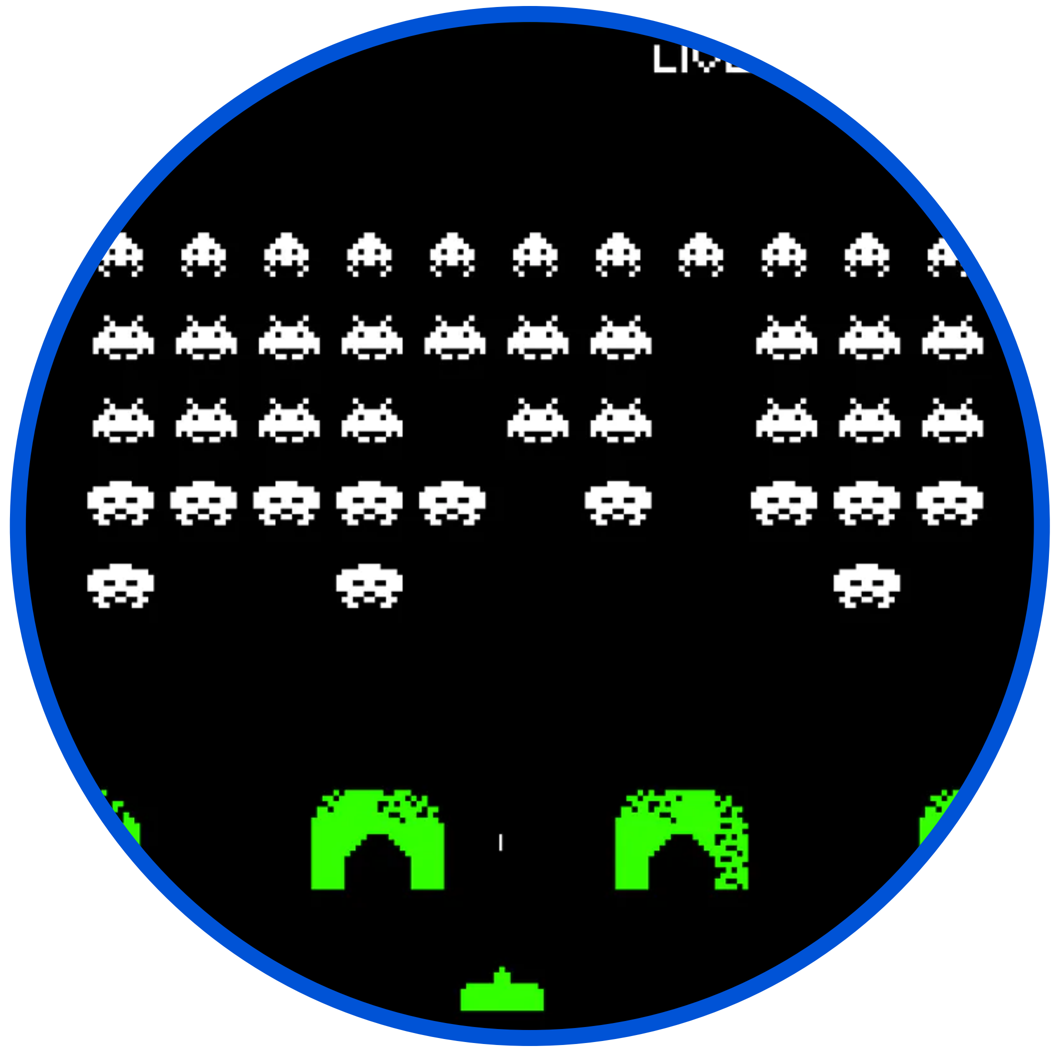
Luo et al. (2019)


Silver et al. (2016)

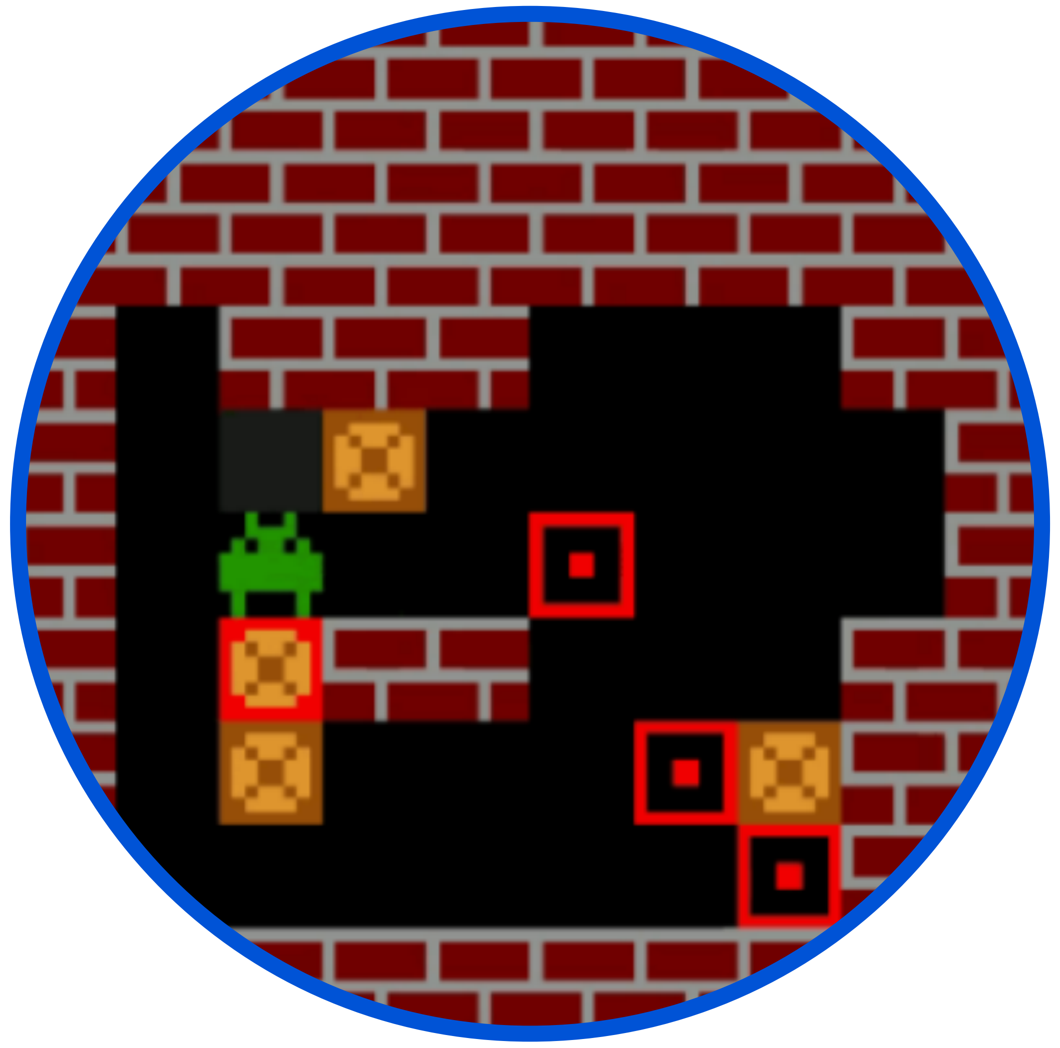

OpenAI et al. (2019)


Segler et al. (2018)


Finn et al. (2018)

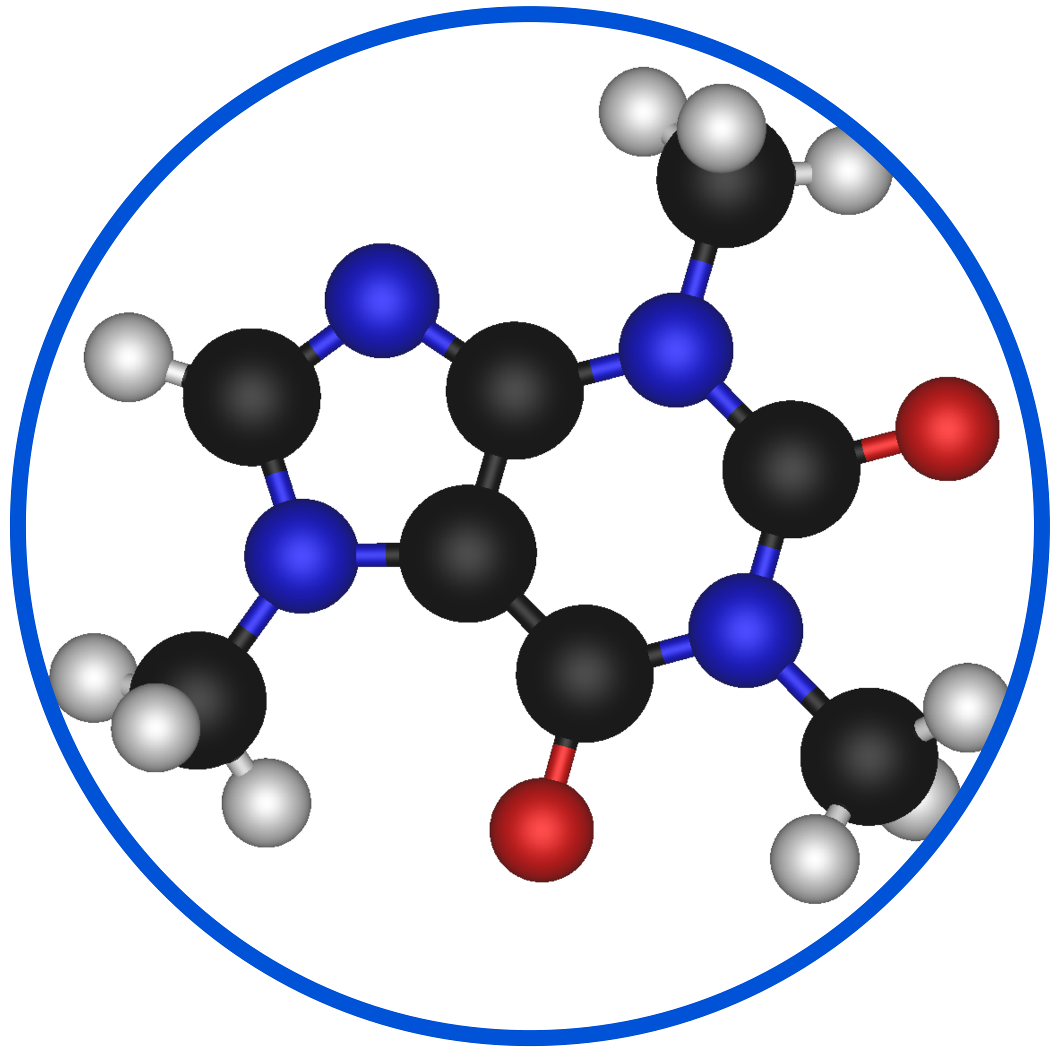

Schrittwieser et al. (2020)

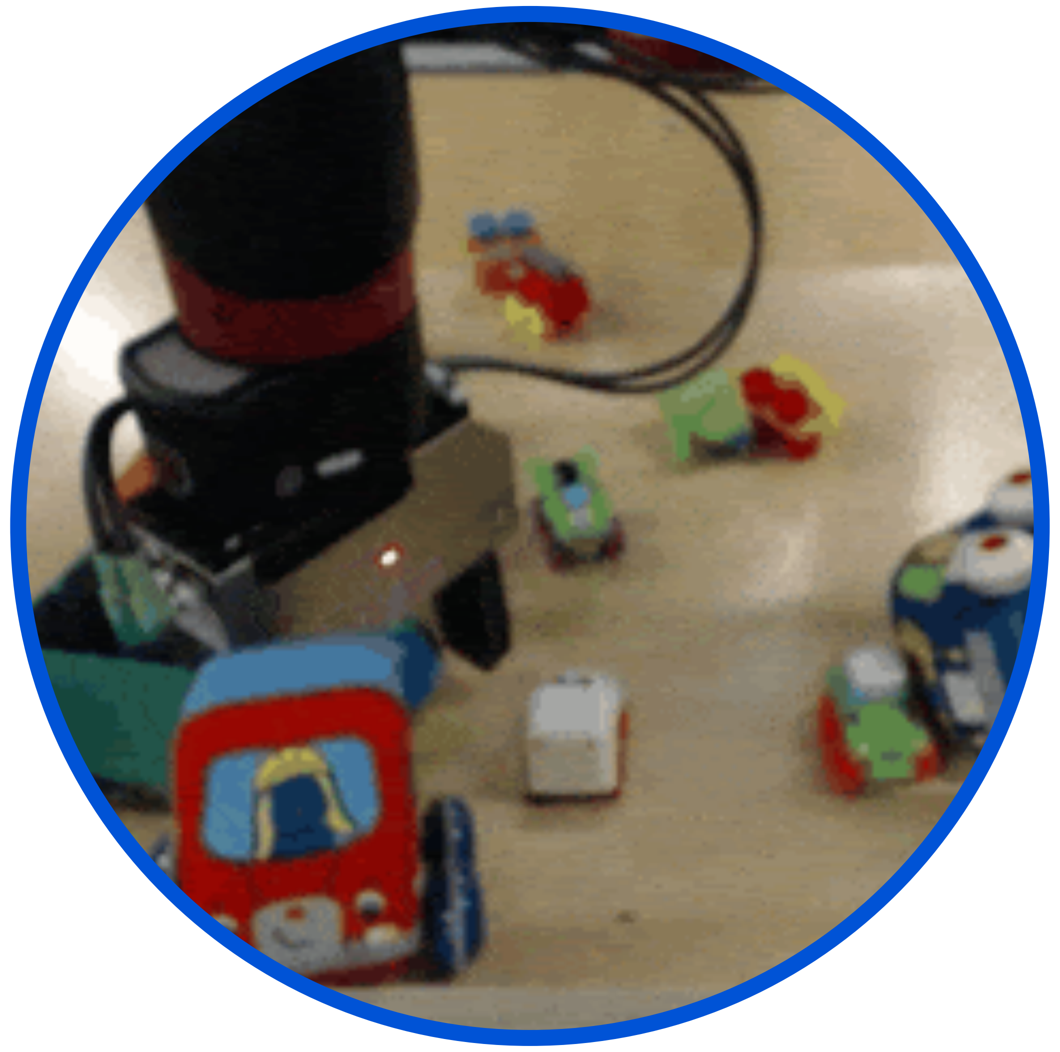

Luo et al. (2019)

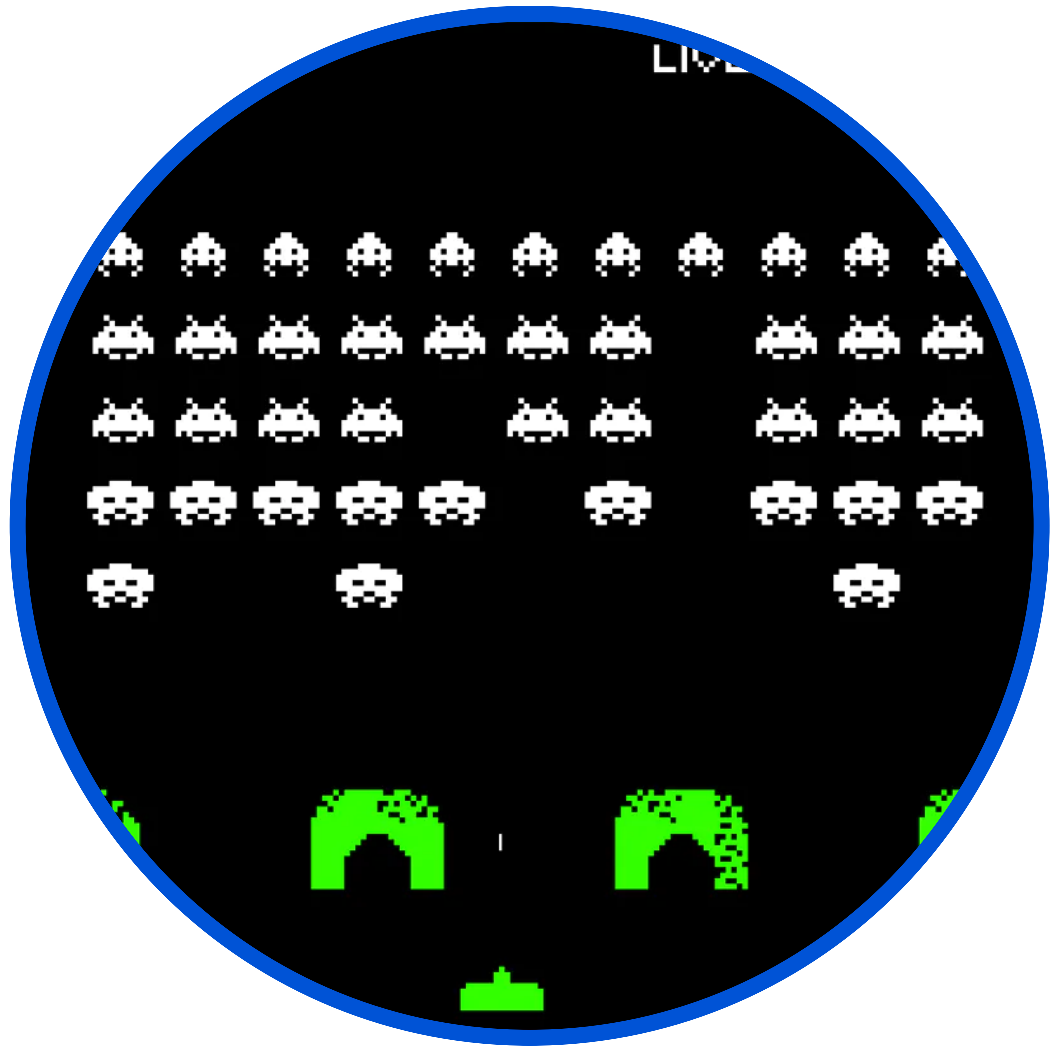

Weber et al. (2017)


Silver et al. (2016)

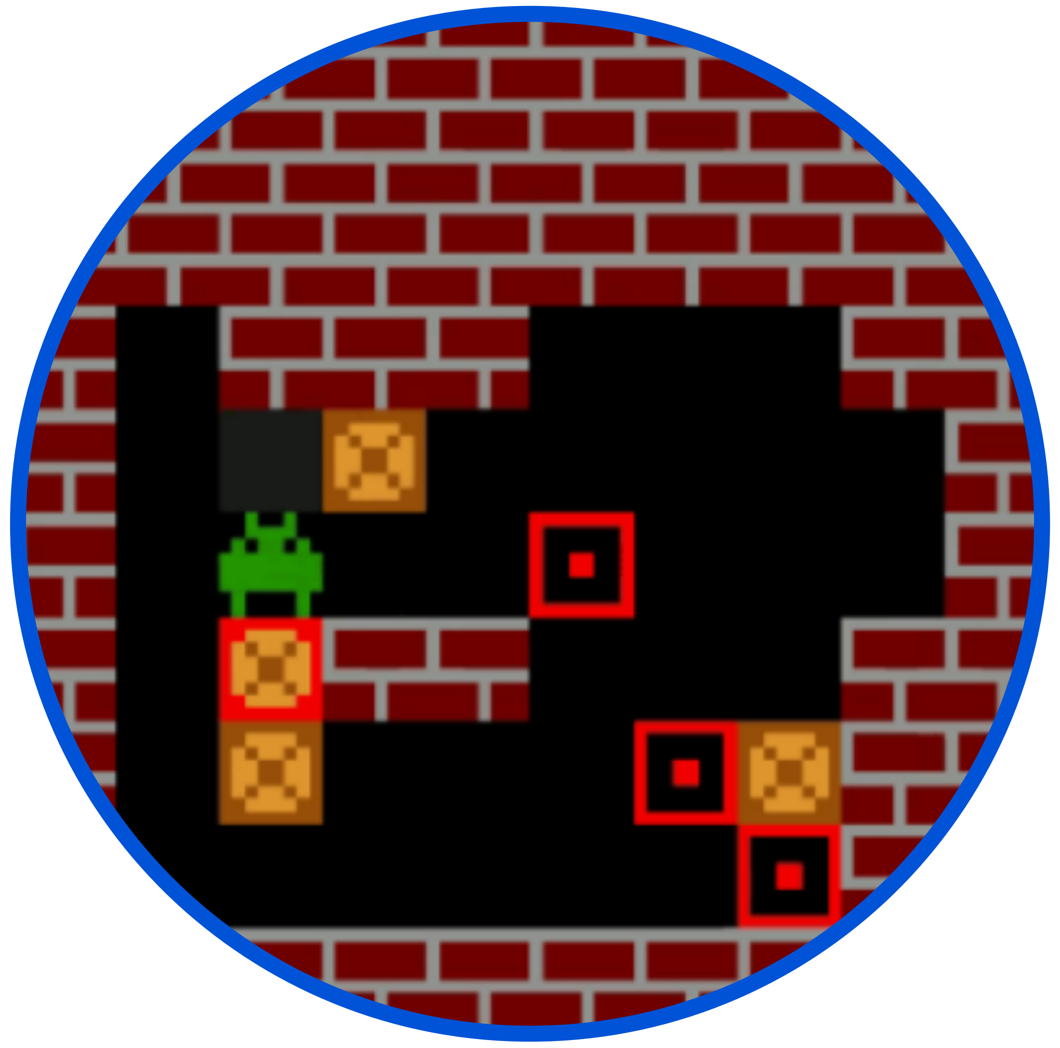

OpenAI et al. (2019)

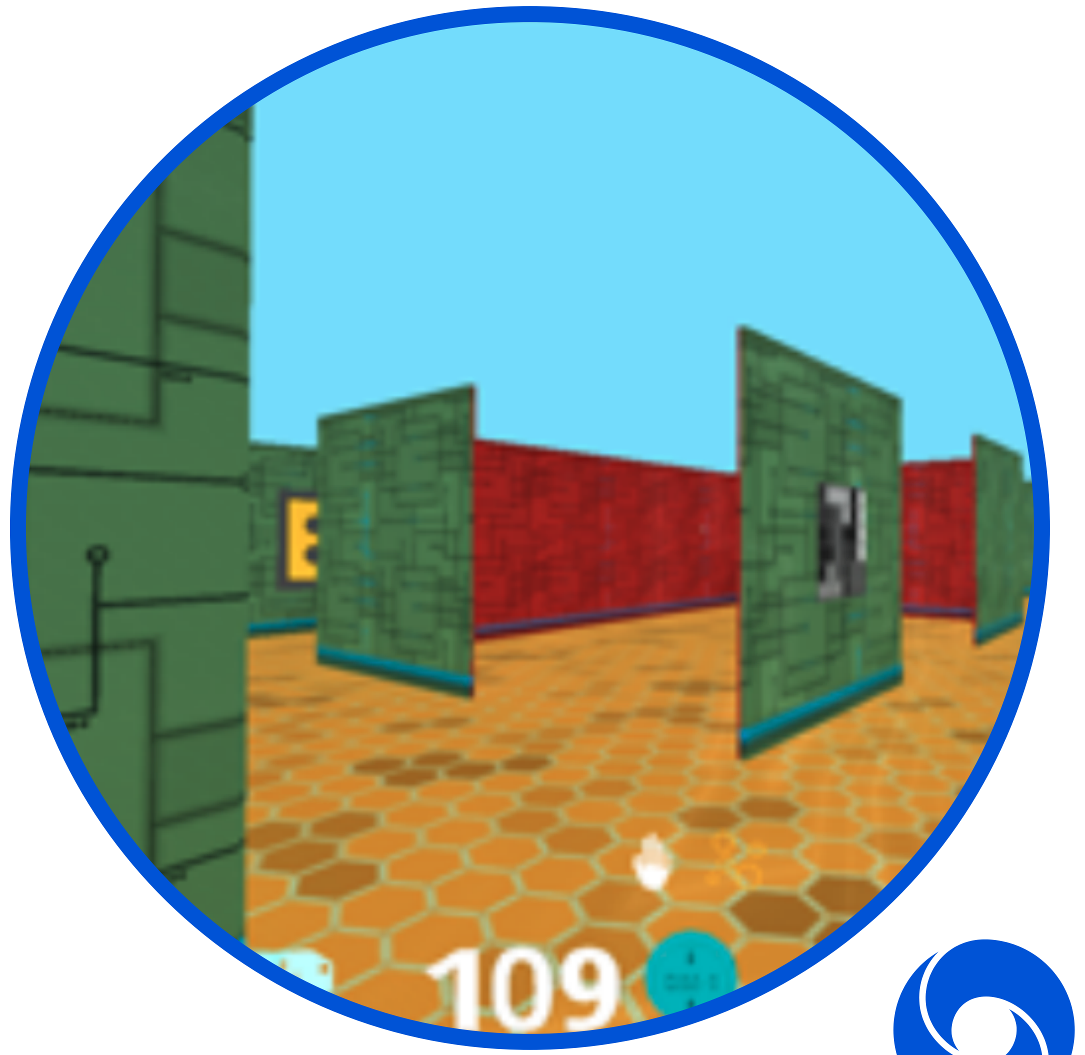

Segler et al. (2018)


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)

"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)

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

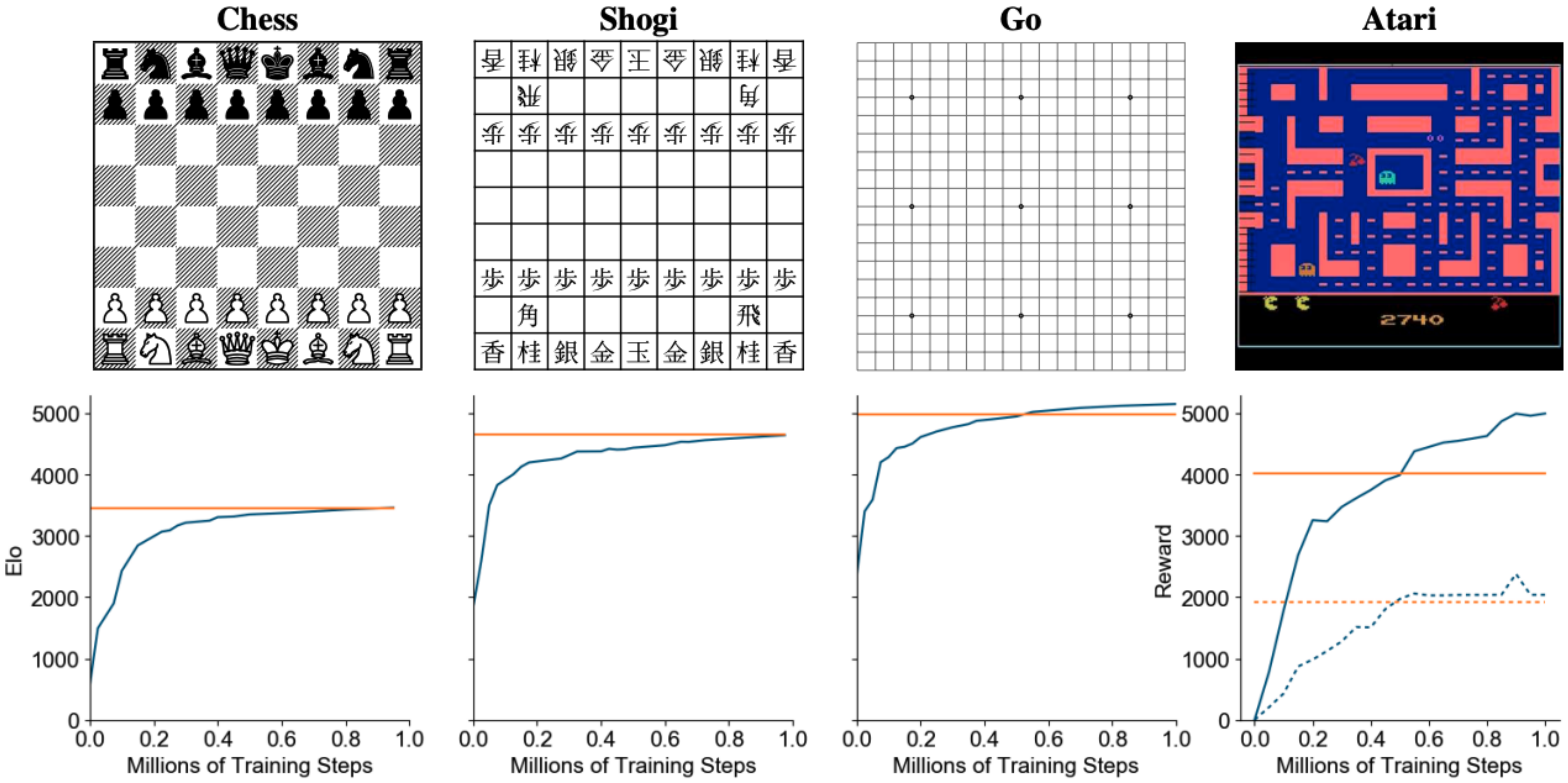
"Model-based planning is an essential ingredient of human intelligence, enabling ***flexible adaptation*** to new tasks and goals"
-Lake et al. (2016)

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

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

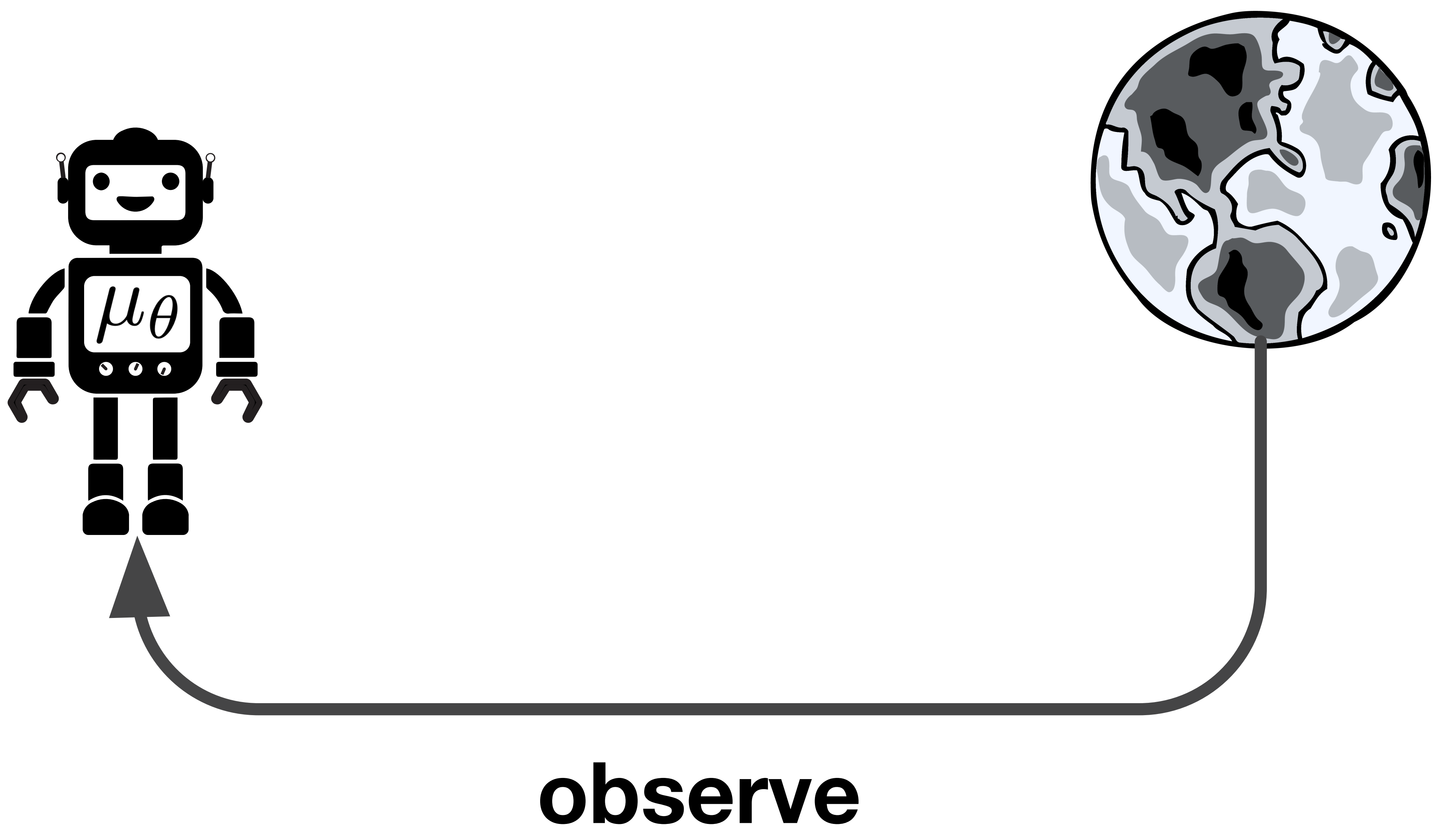
MuZero

Schrittwieser et al. (2019)



MuZero

Schrittwieser et al. (2019)

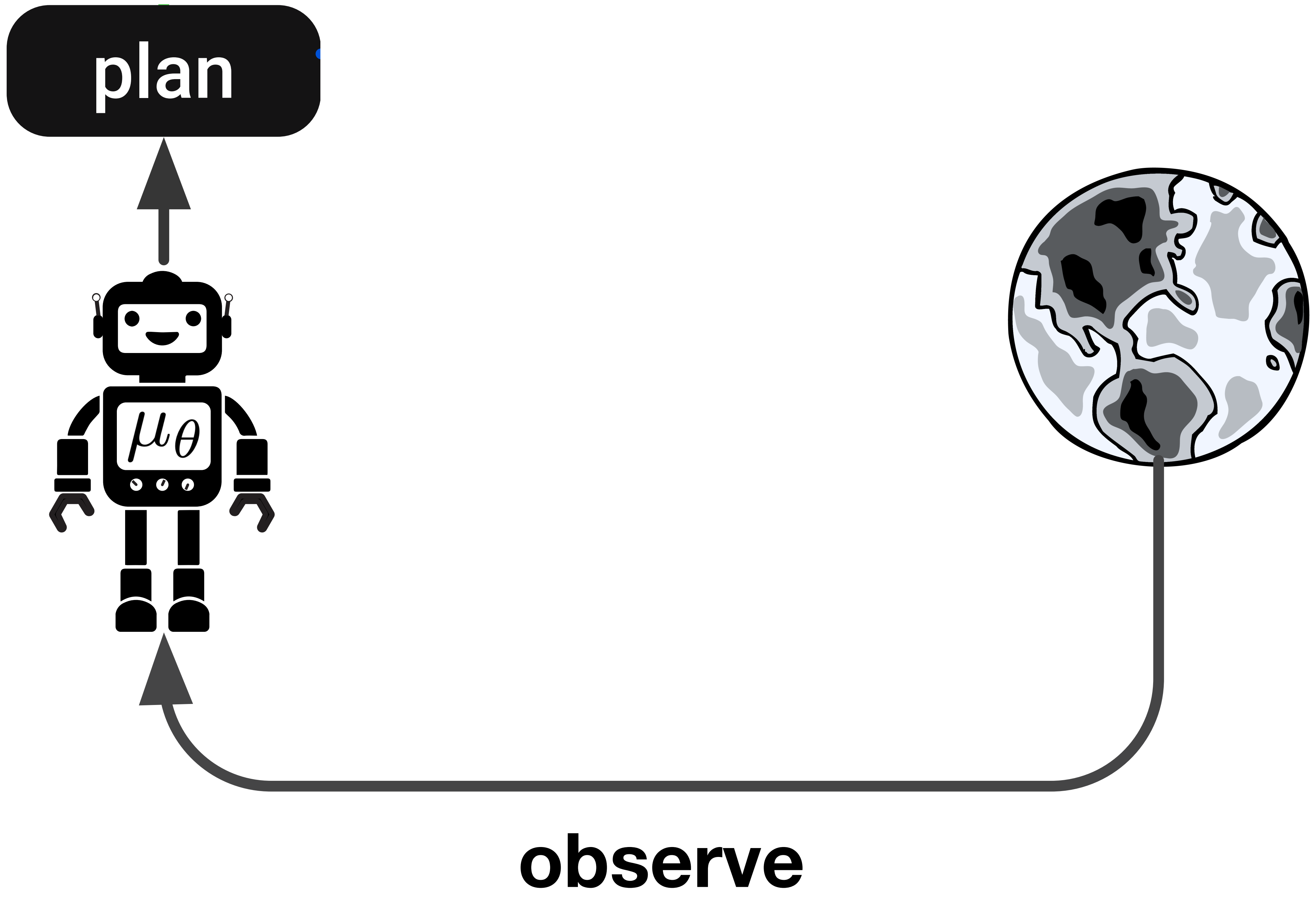


MuZero

Schrittwieser et al. (2019)

*Guide MCTS using learned **policy and value functions***

policy: where to search?
model: what will happen?
value: is what will happen good?



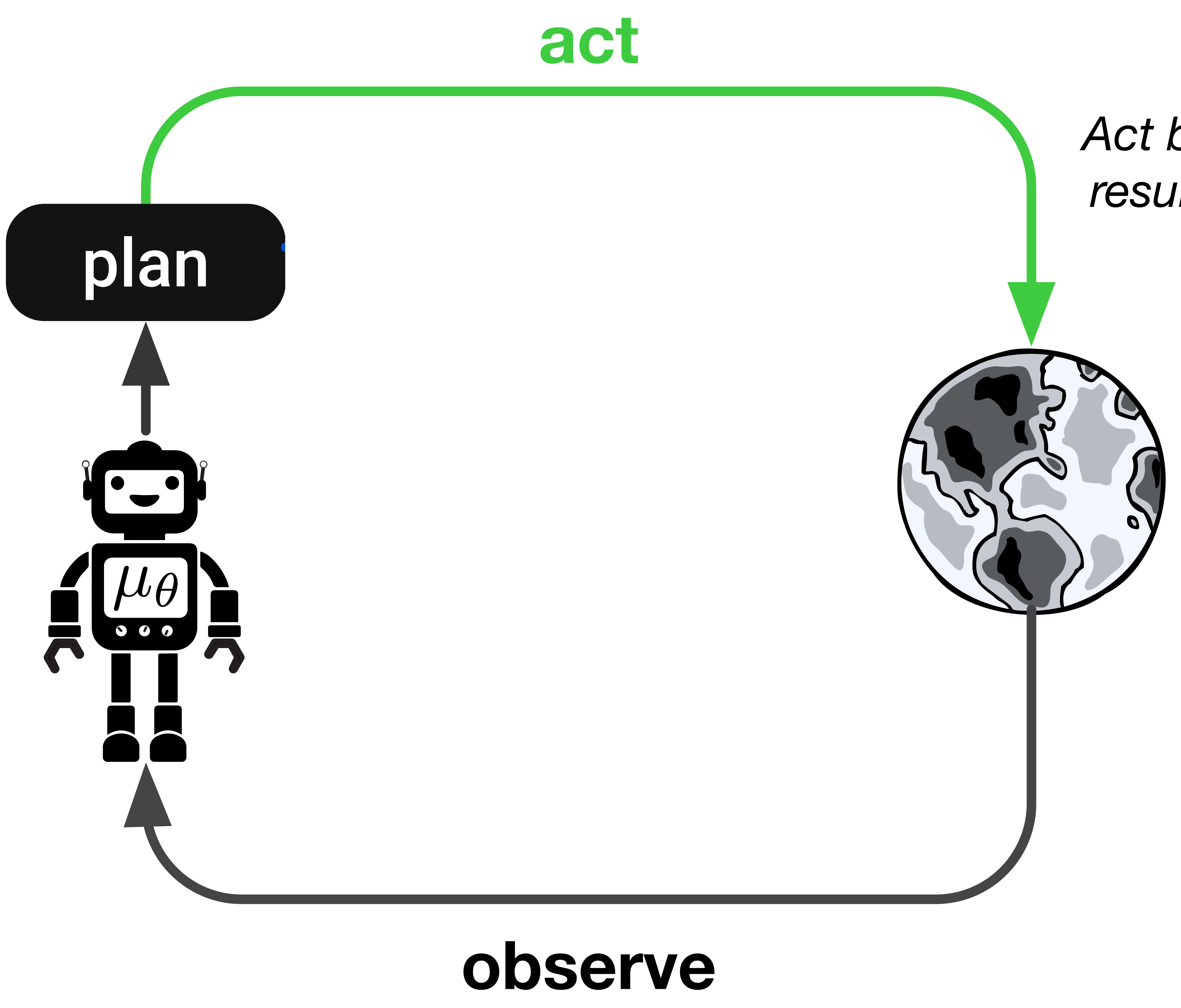
(MCTS = Monte Carlo Tree Search)

MuZero

Schrittwieser et al. (2019)

*Guide MCTS using learned **policy and value functions***

policy: where to search?
model: what will happen?
value: is what will happen good?



Act based on the results of search

(MCTS = Monte Carlo Tree Search)

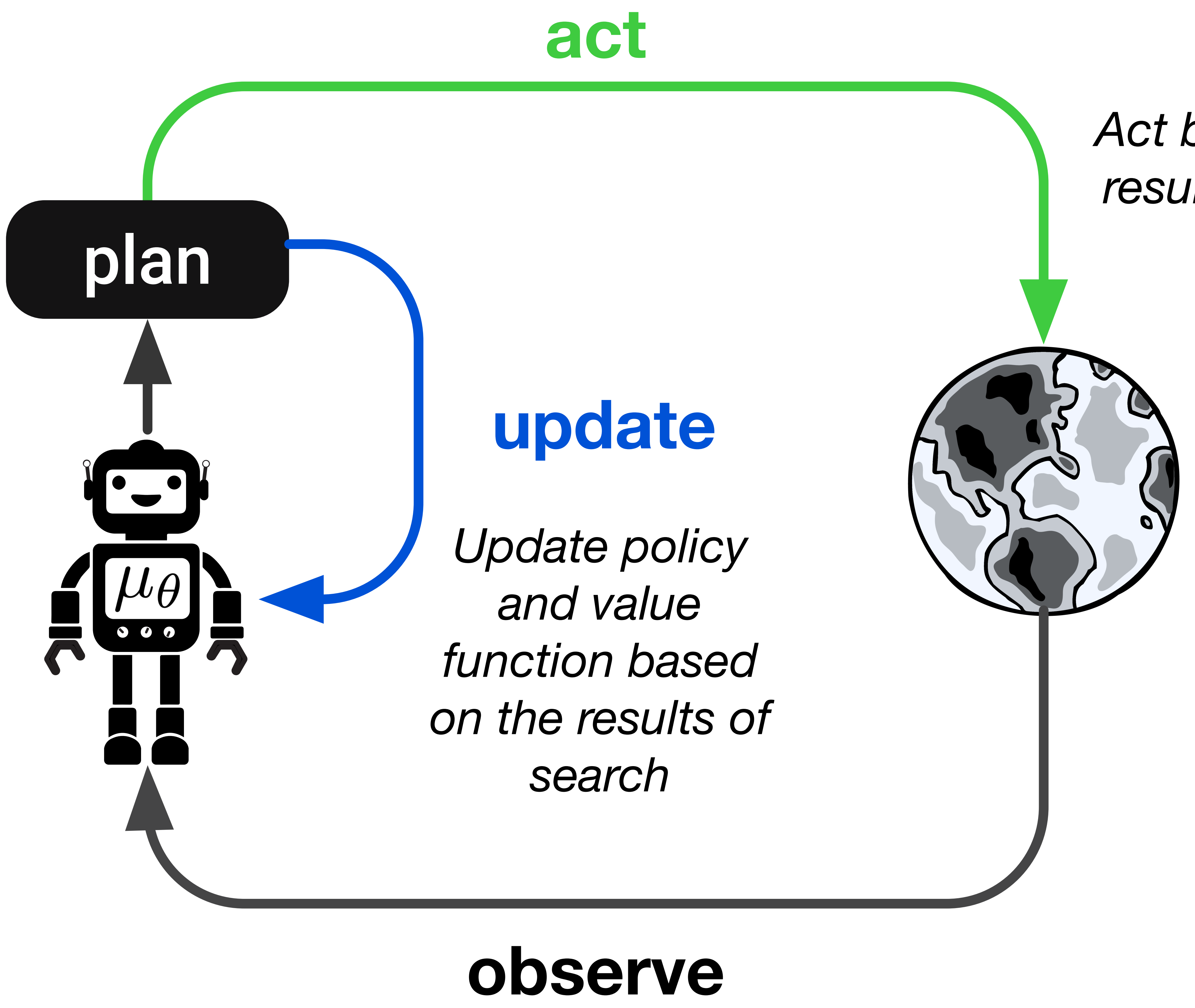
Jessica Hamrick - jhamrick@deepmind.com

MuZero

Schrittwieser et al. (2019)

*Guide MCTS using learned **policy and value functions***

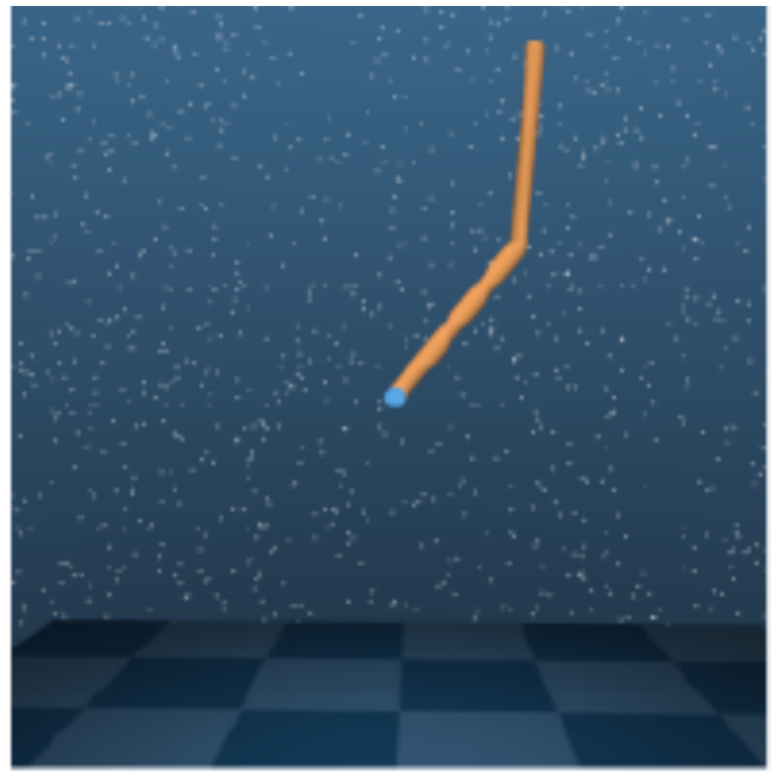
policy: where to search?
model: what will happen?
value: is what will happen good?



Act based on the results of search

(MCTS = Monte Carlo Tree Search)

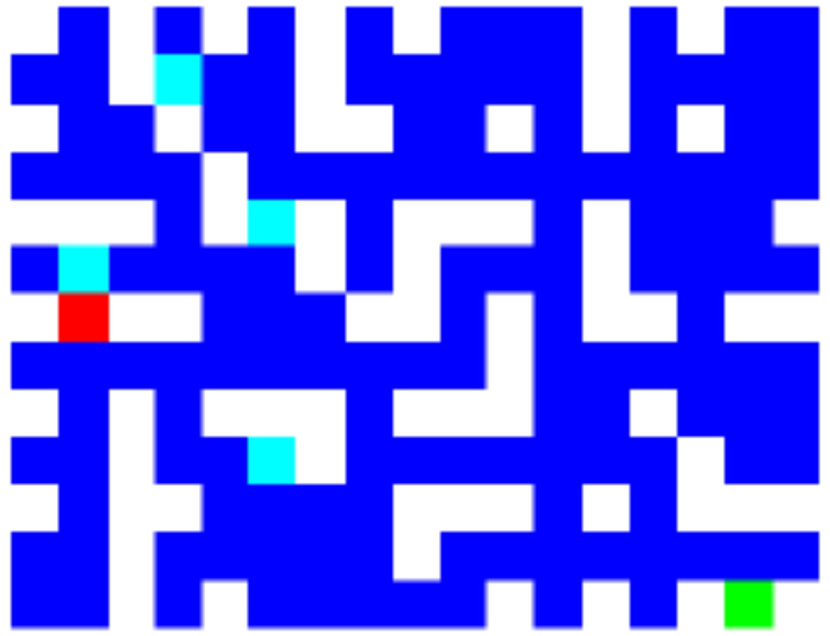
Environments



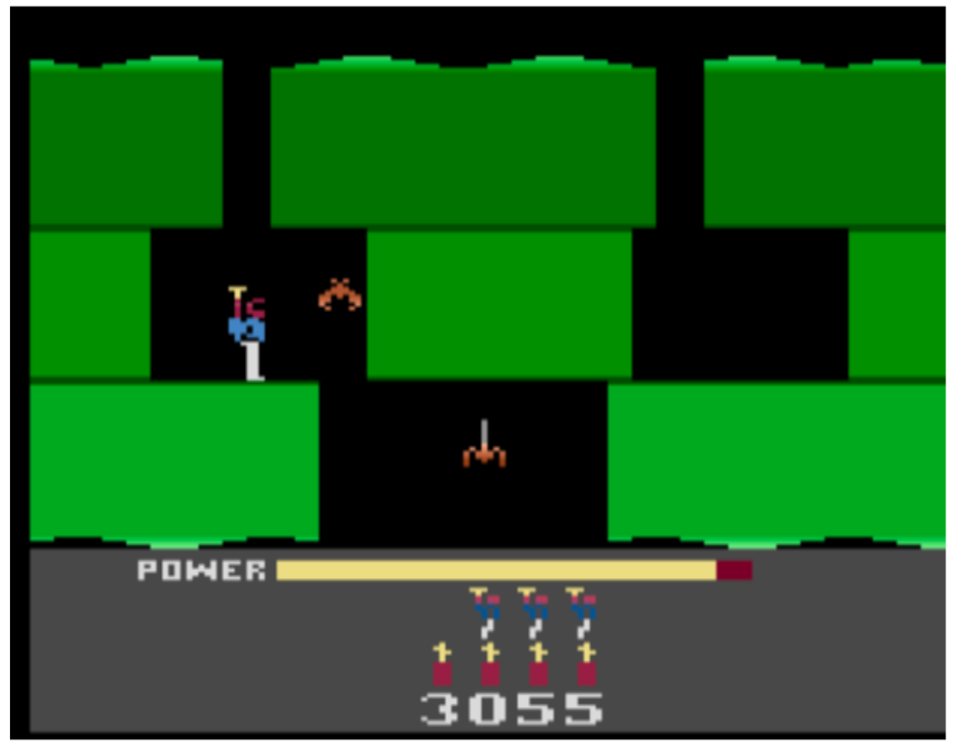
Acrobot
(Swingup Sparse)

Cheetah
(Run)

Humanoid
(Stand)

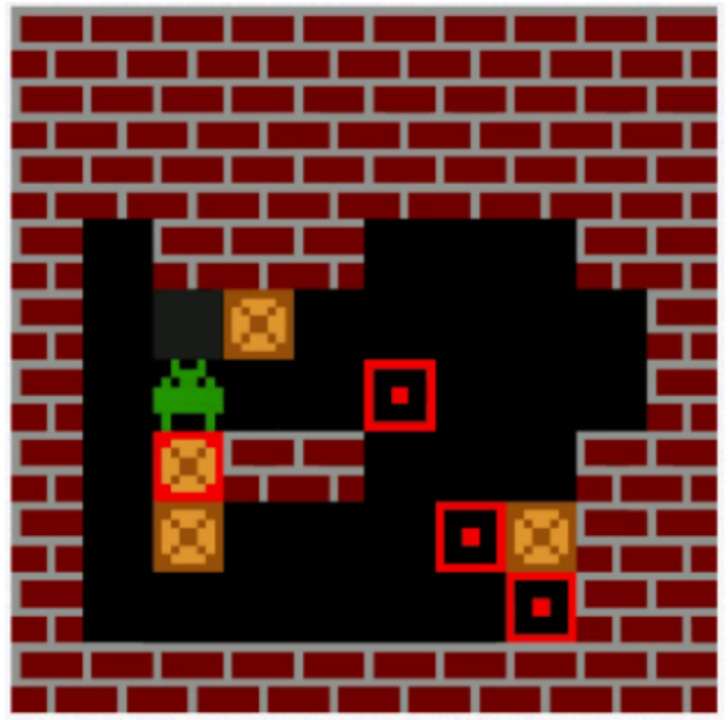


Minipacman
(Procedural)

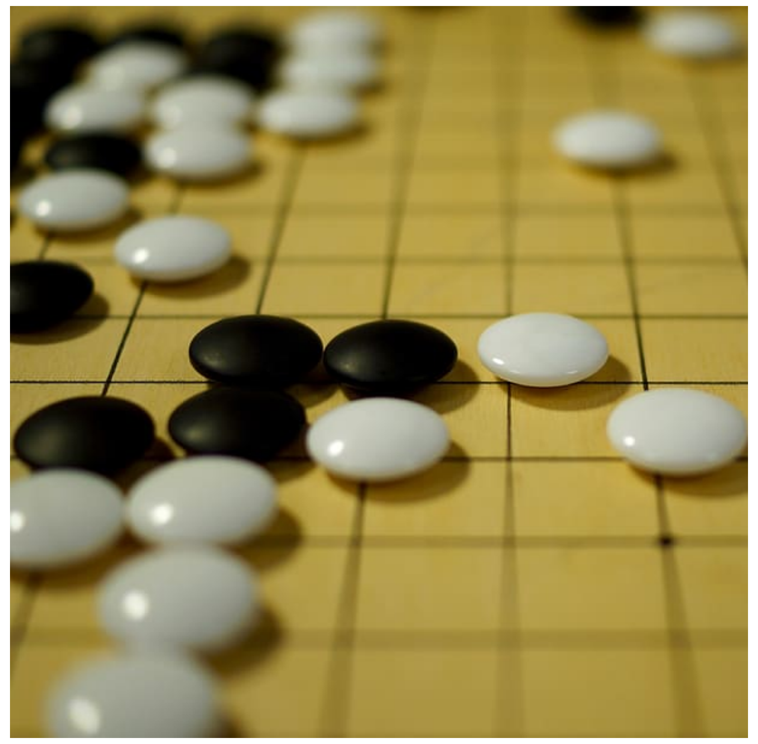


Hero

Ms. Pacman

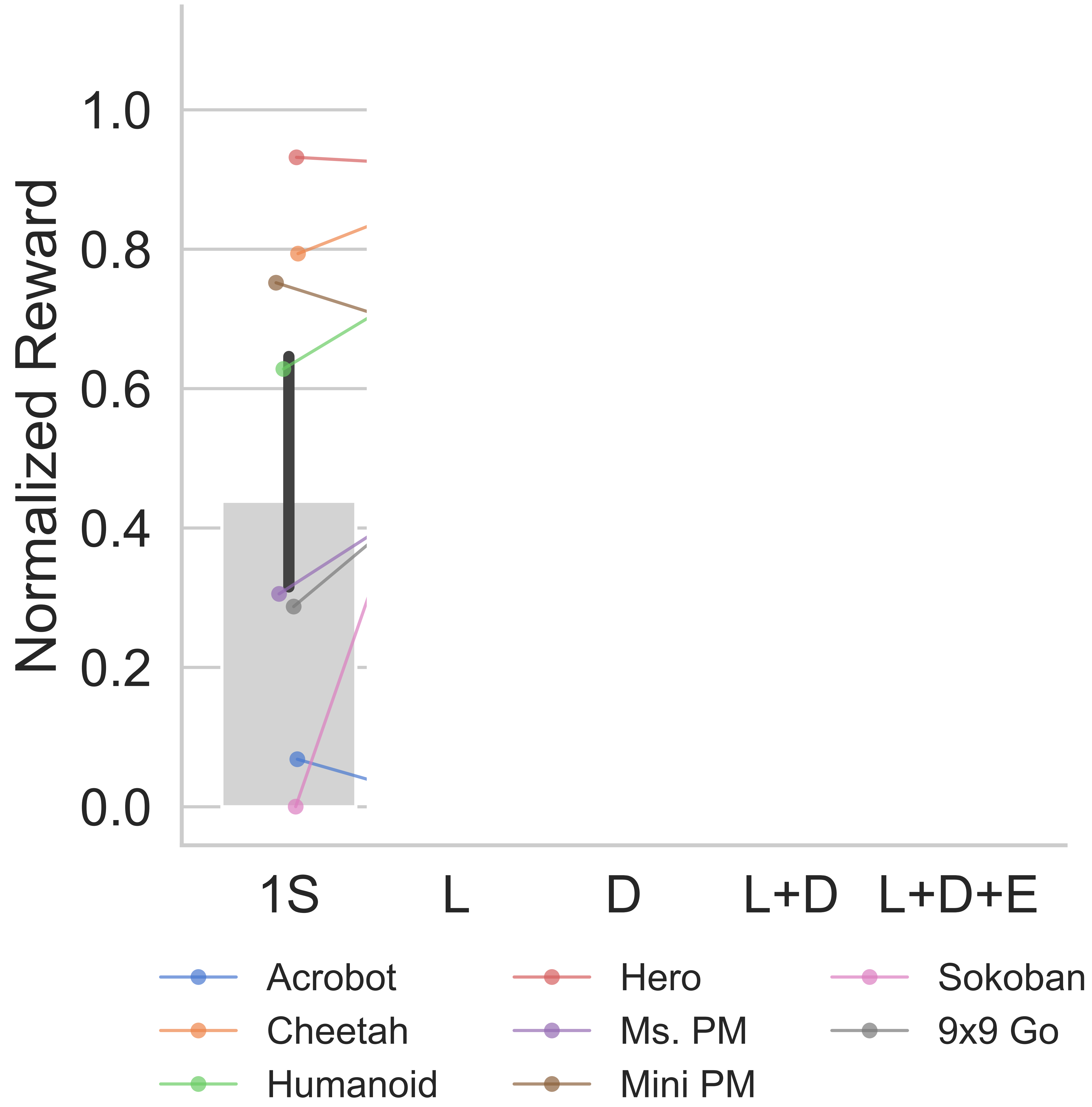


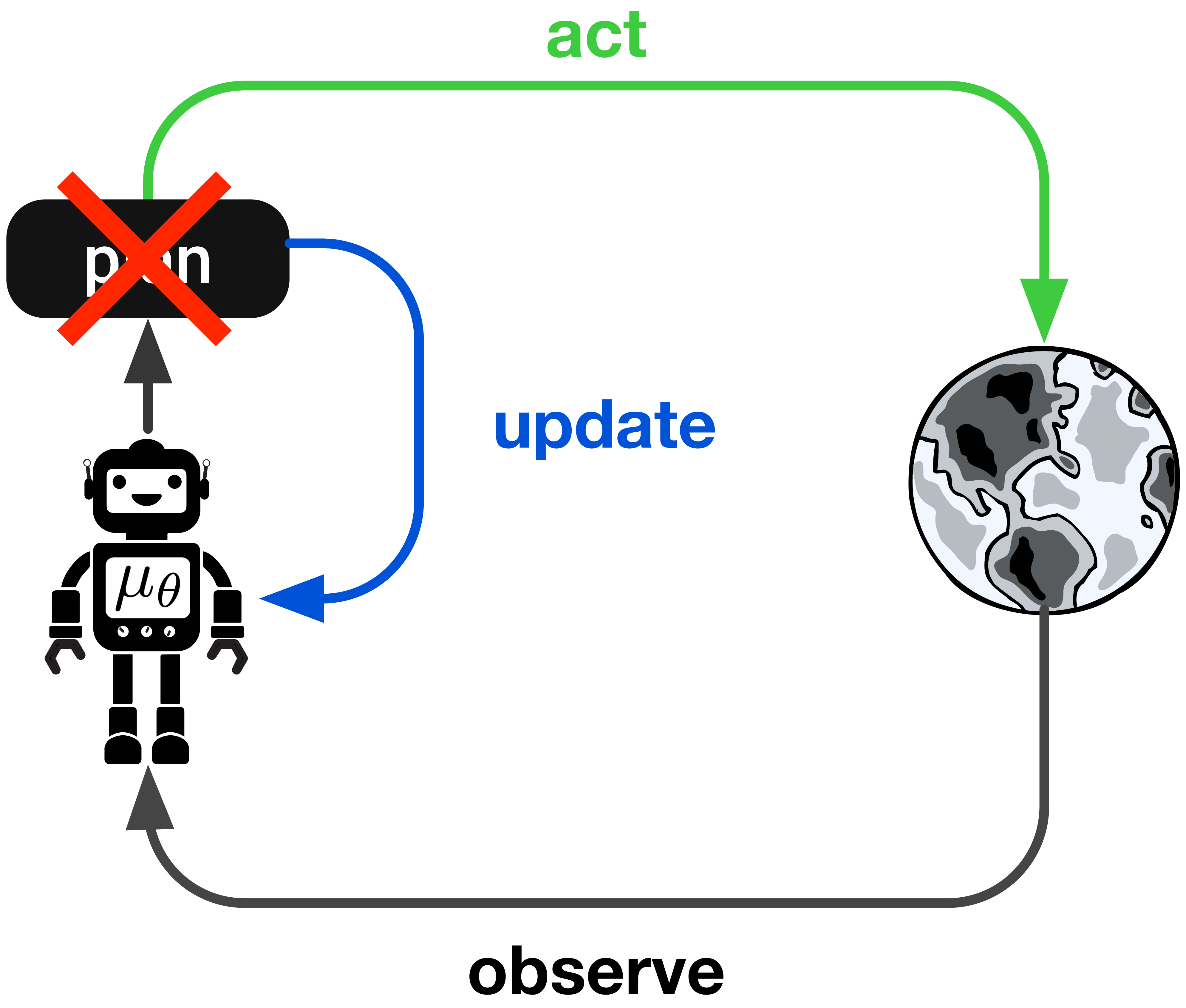
Sokoban



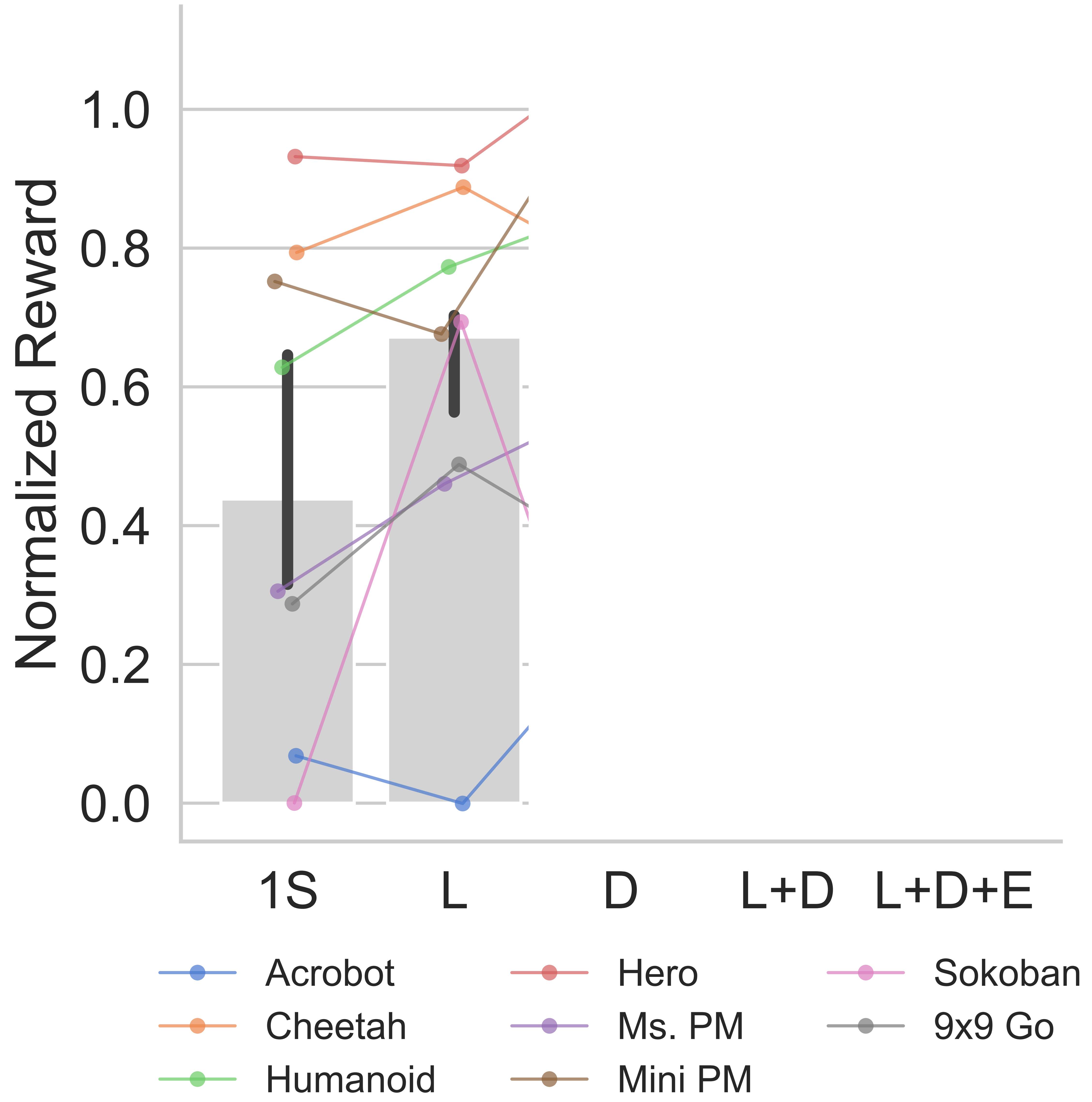
9x9 Go

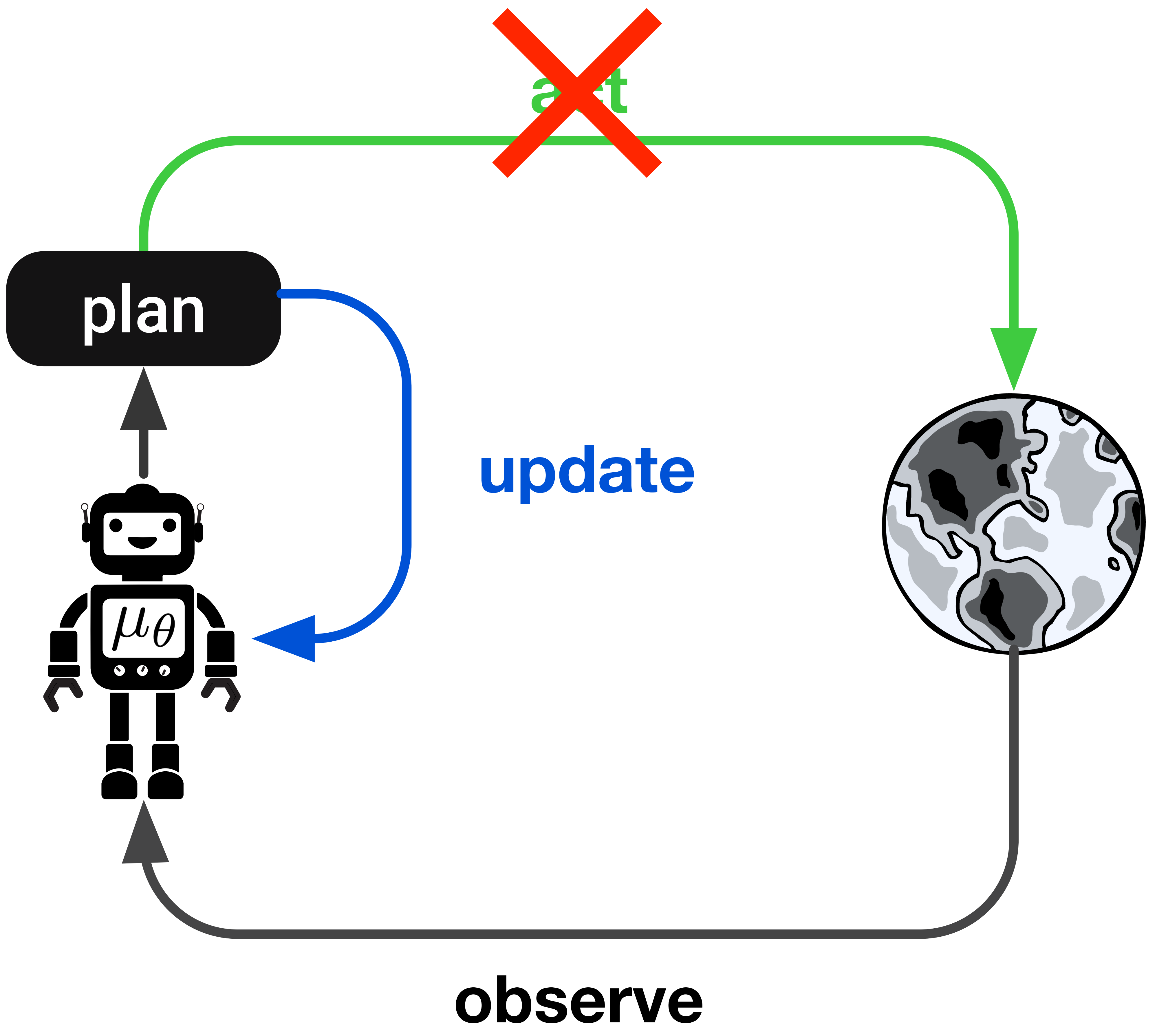
Jessica Hamrick - jhamrick@deepmind.com


Normalized Reward
1.0
0.8
0.6
0.4
0.2
0.0
1S
L
D
L+D
L+D+E
Acrobot
Cheetah
Humanoid
Hero
Ms. PM
Mini PM
Sokoban
9x9 Go

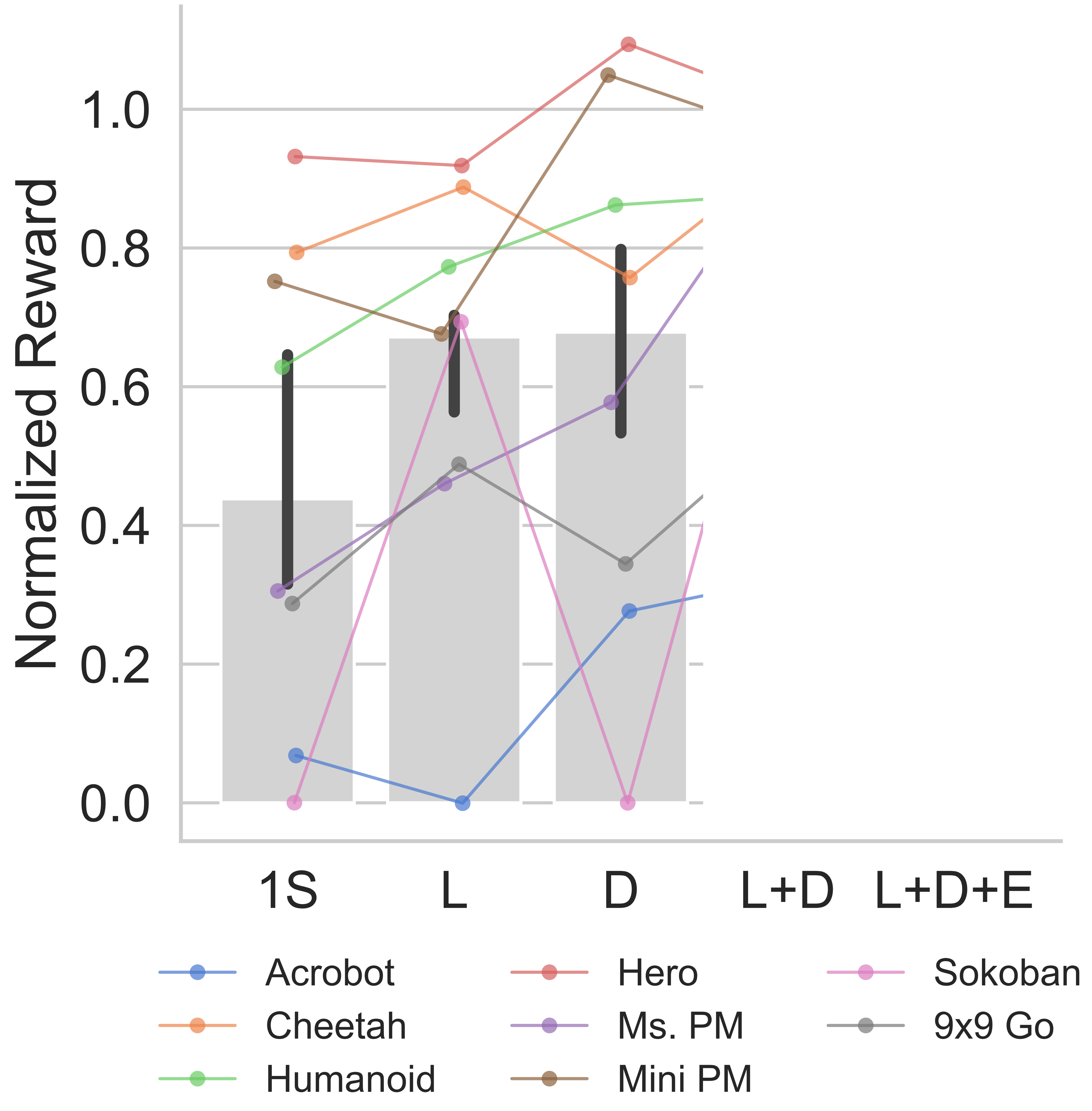

act
update
μ_θ
observe

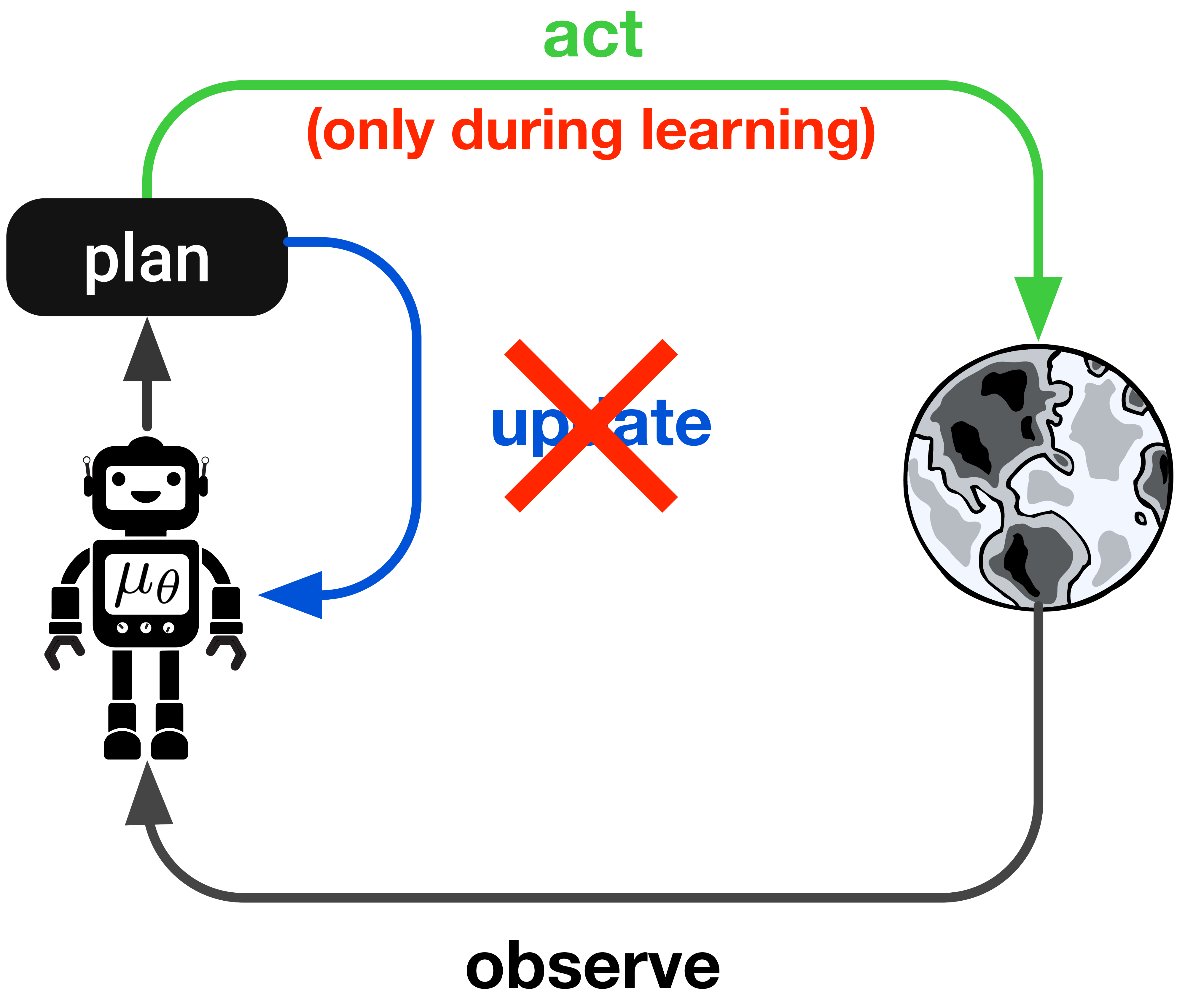
Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.



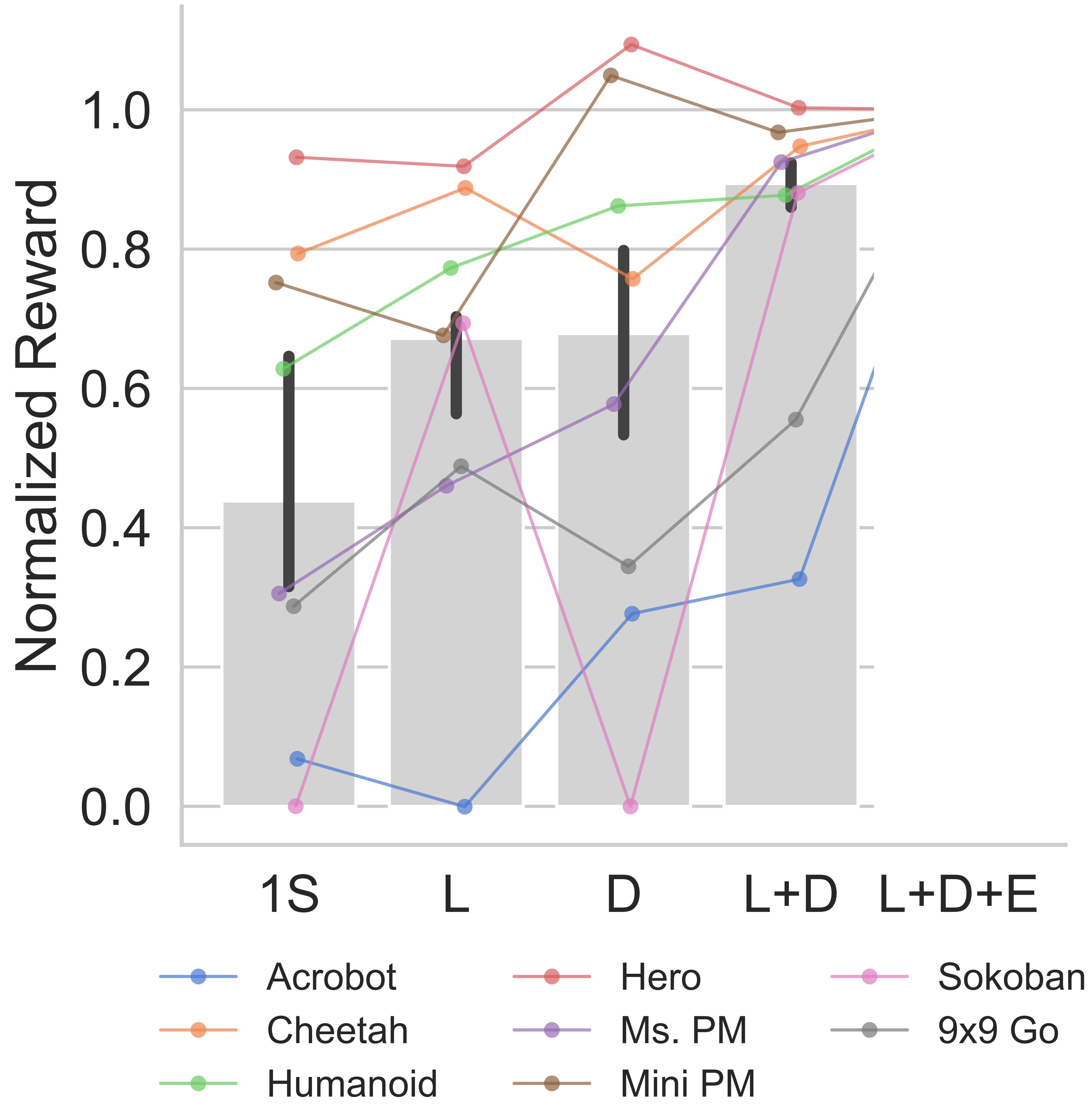


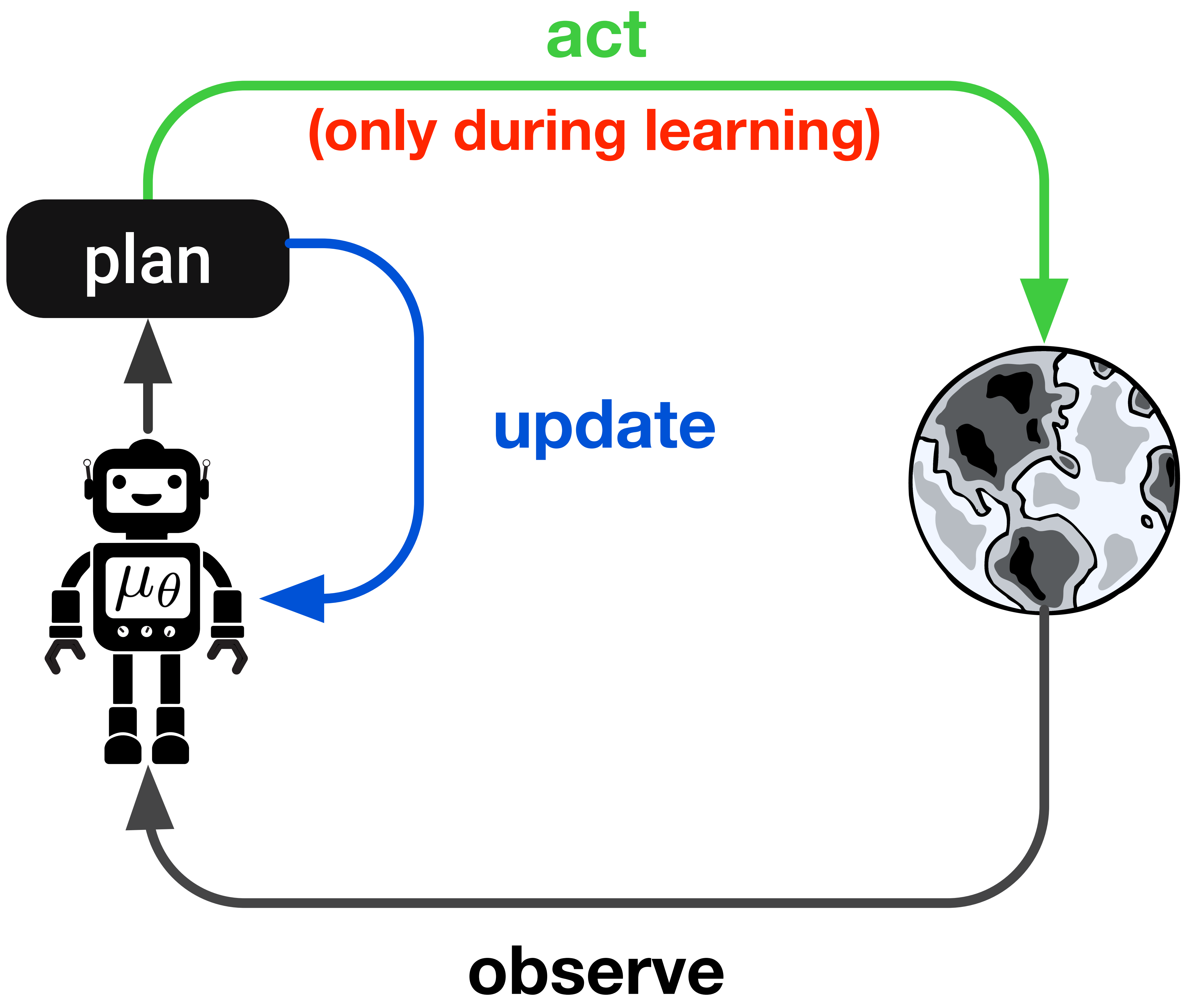
Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.



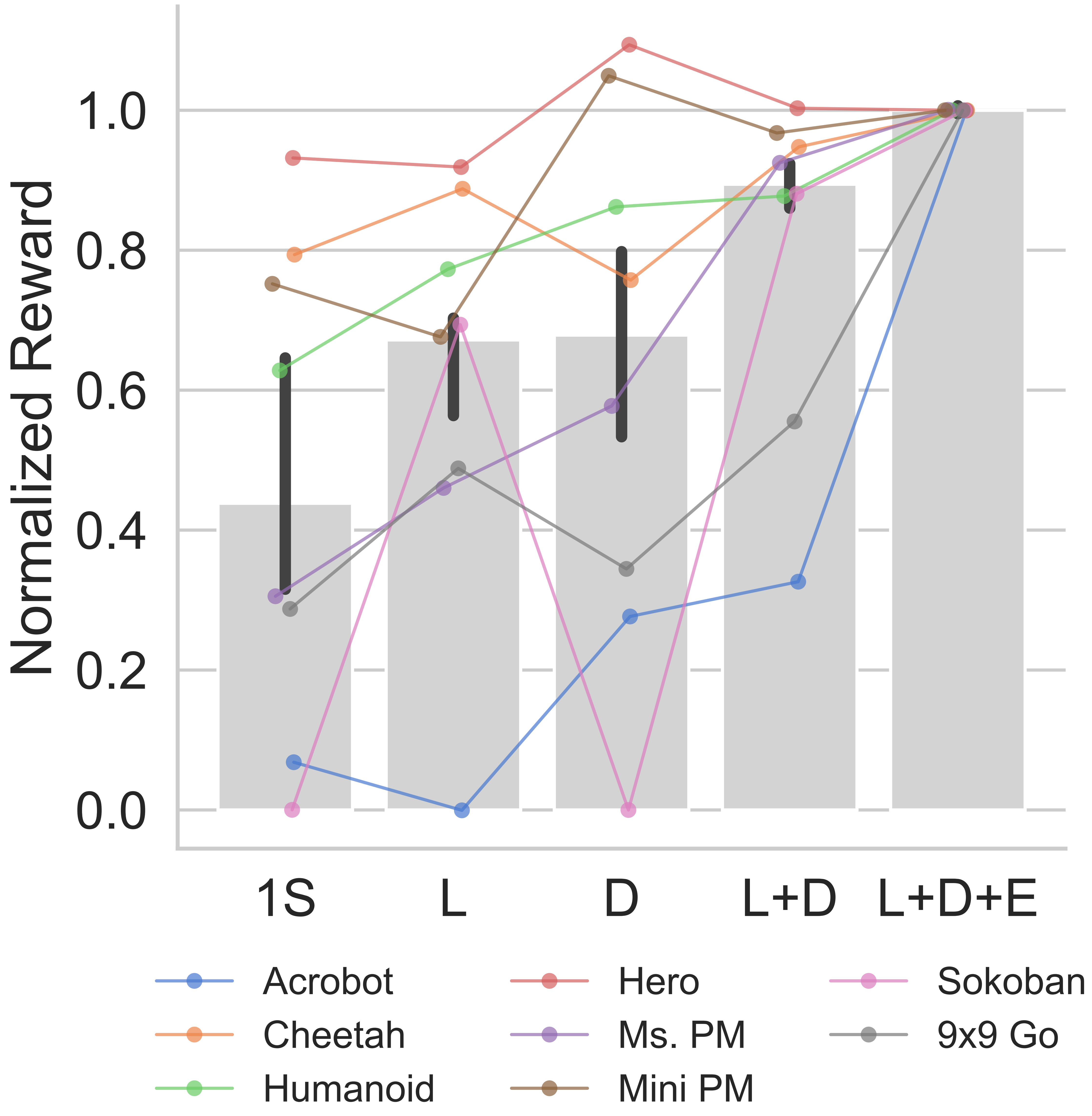


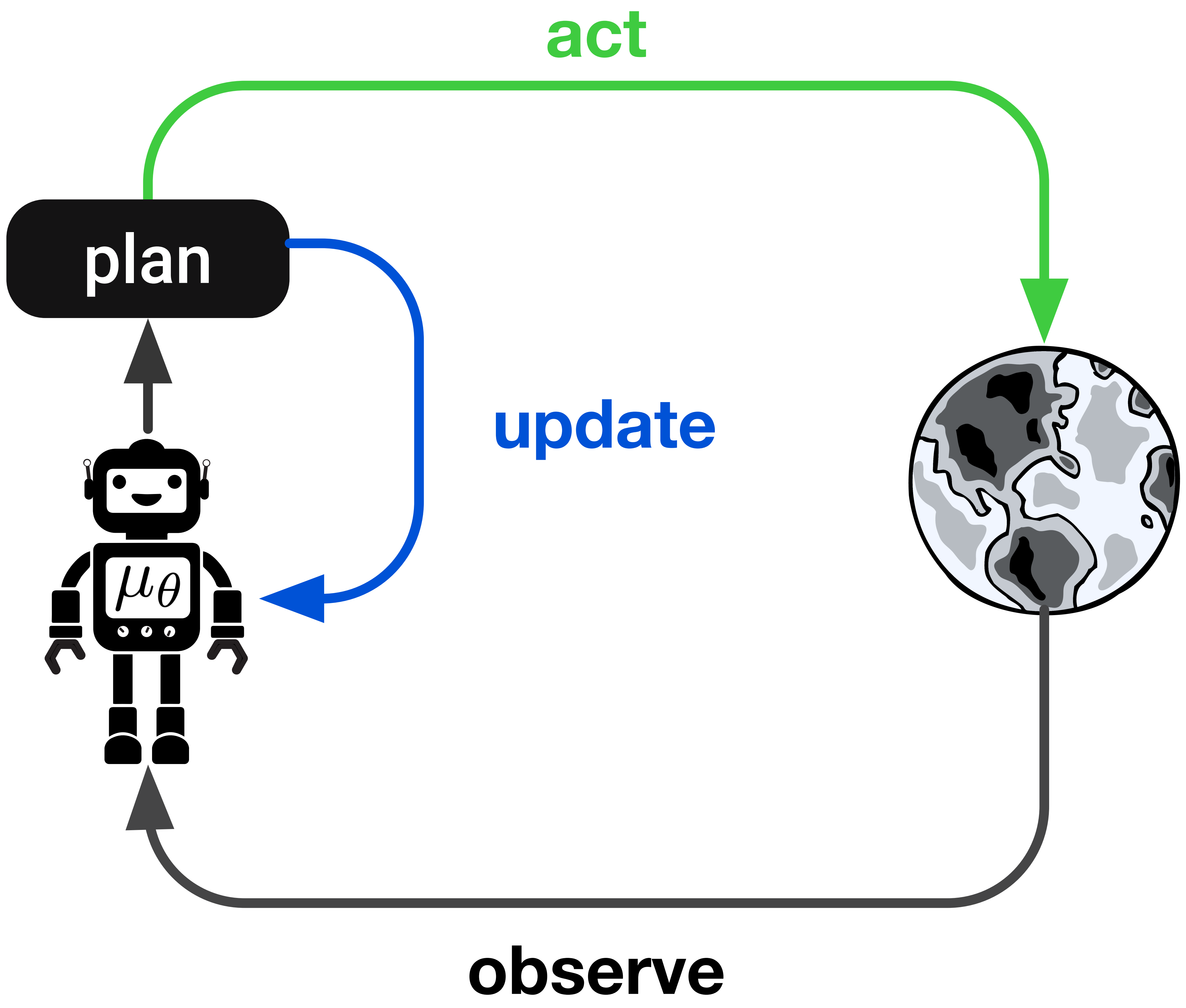
Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.





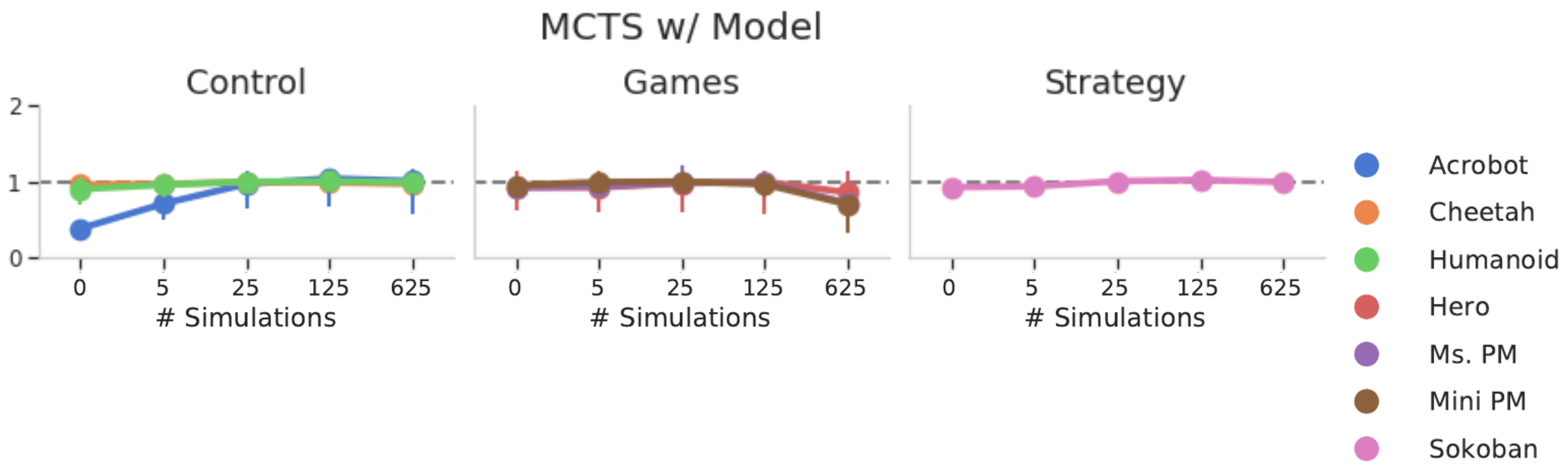
Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.





Varying the amount of search at test time

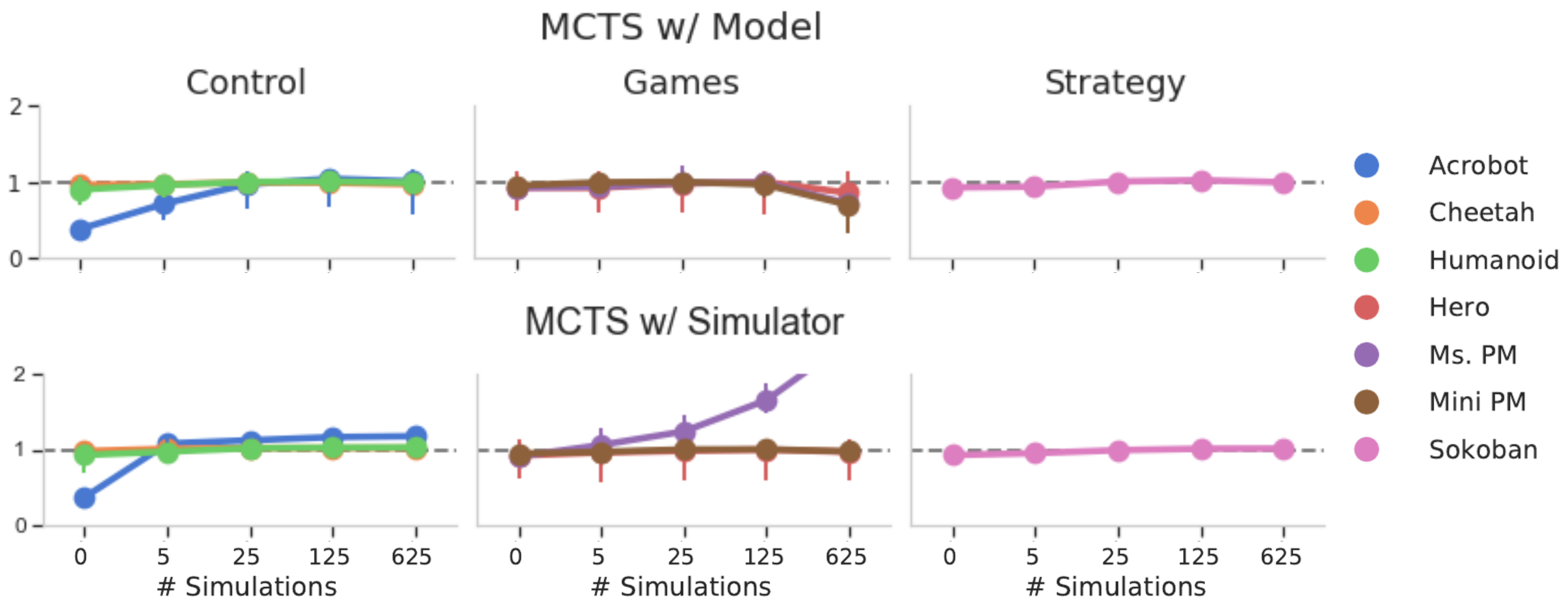
Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.



Hamrick et al. (2021)

Varying the amount of search at test time

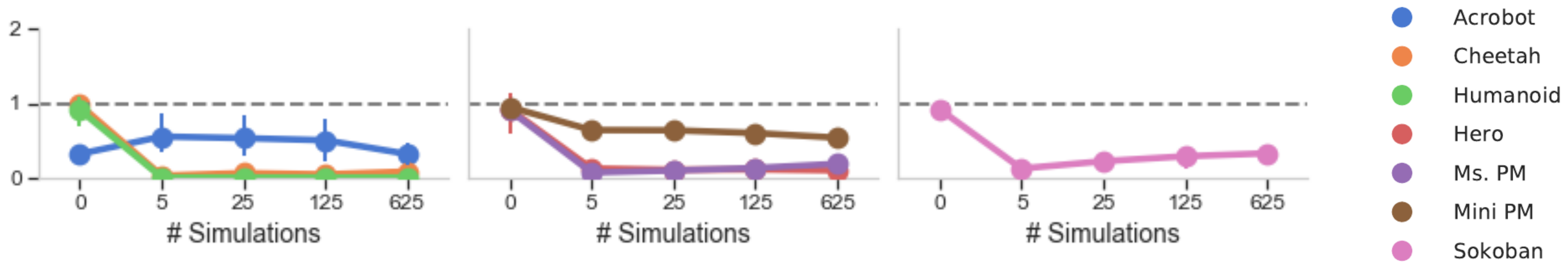
Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.



Hamrick et al. (2021)

Stress-testing the value function

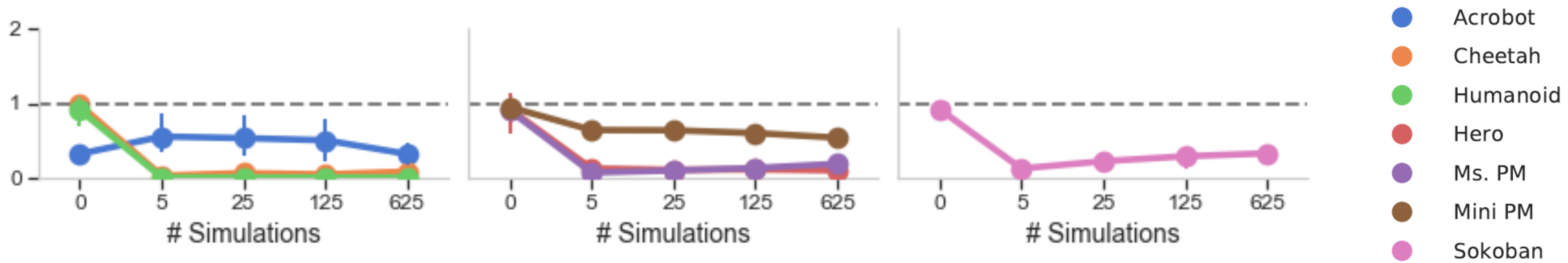
Hamrick et al. (2021). On the role of planning in model-based deep reinforcement learning. ICLR.



Hamrick et al. (2021)

Stress-testing the value function

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

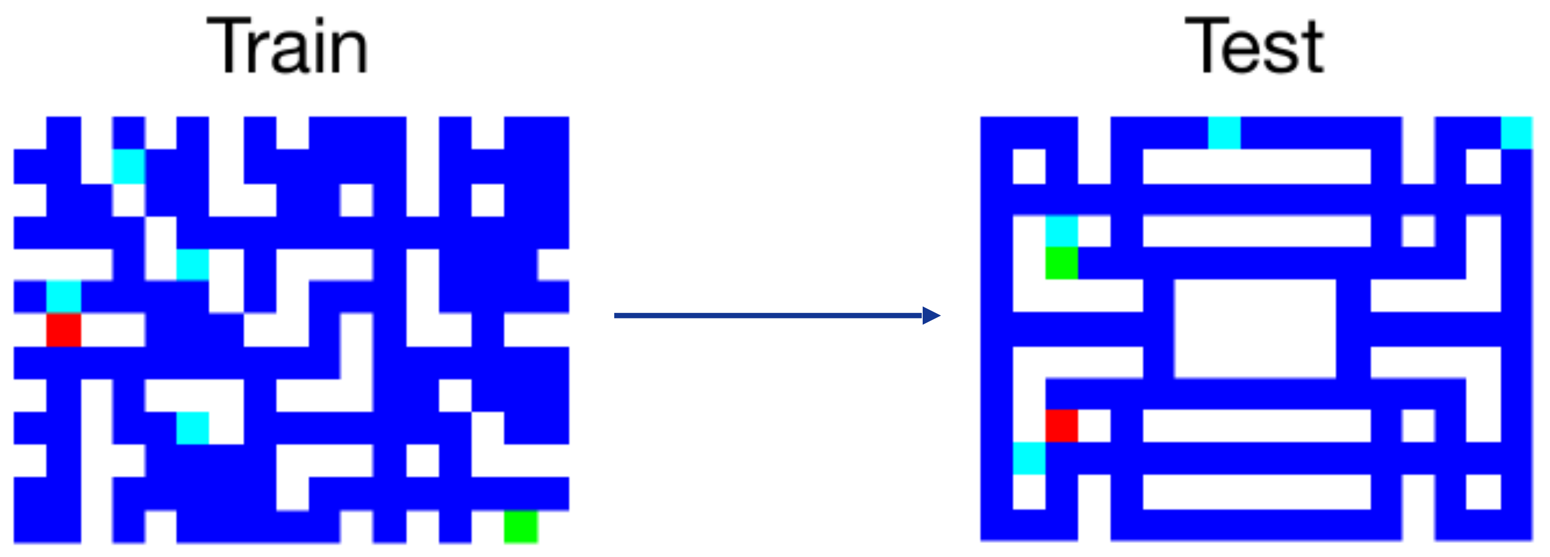


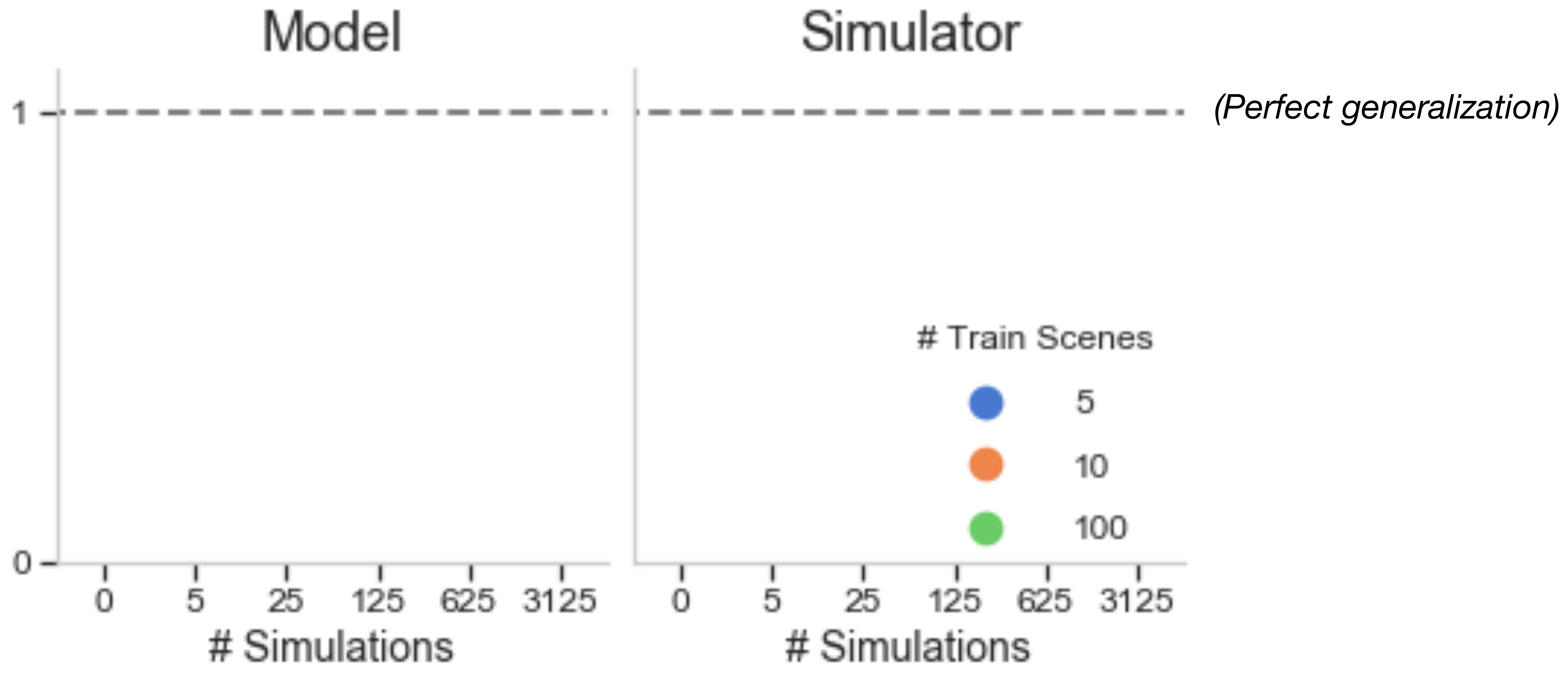
Errors in the model of the world (i.e. transition function) are not the only types of error to be concerned about.

Hamrick et al. (2021)

Generalizing to new mazes

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



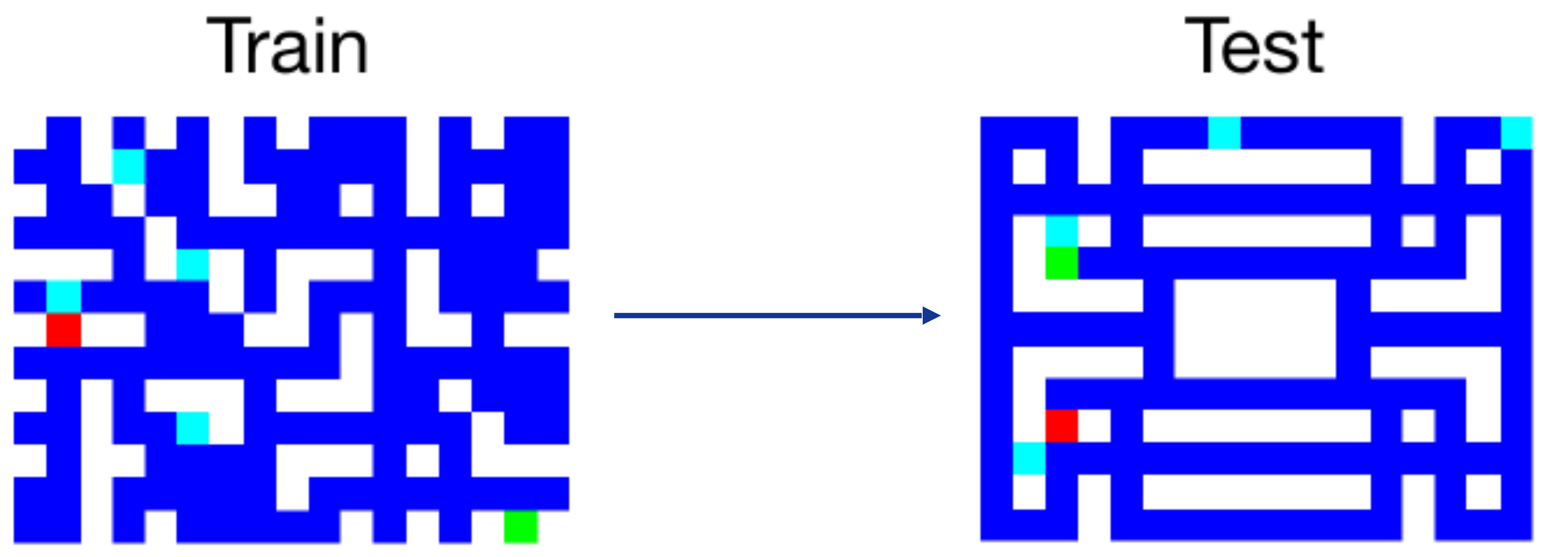


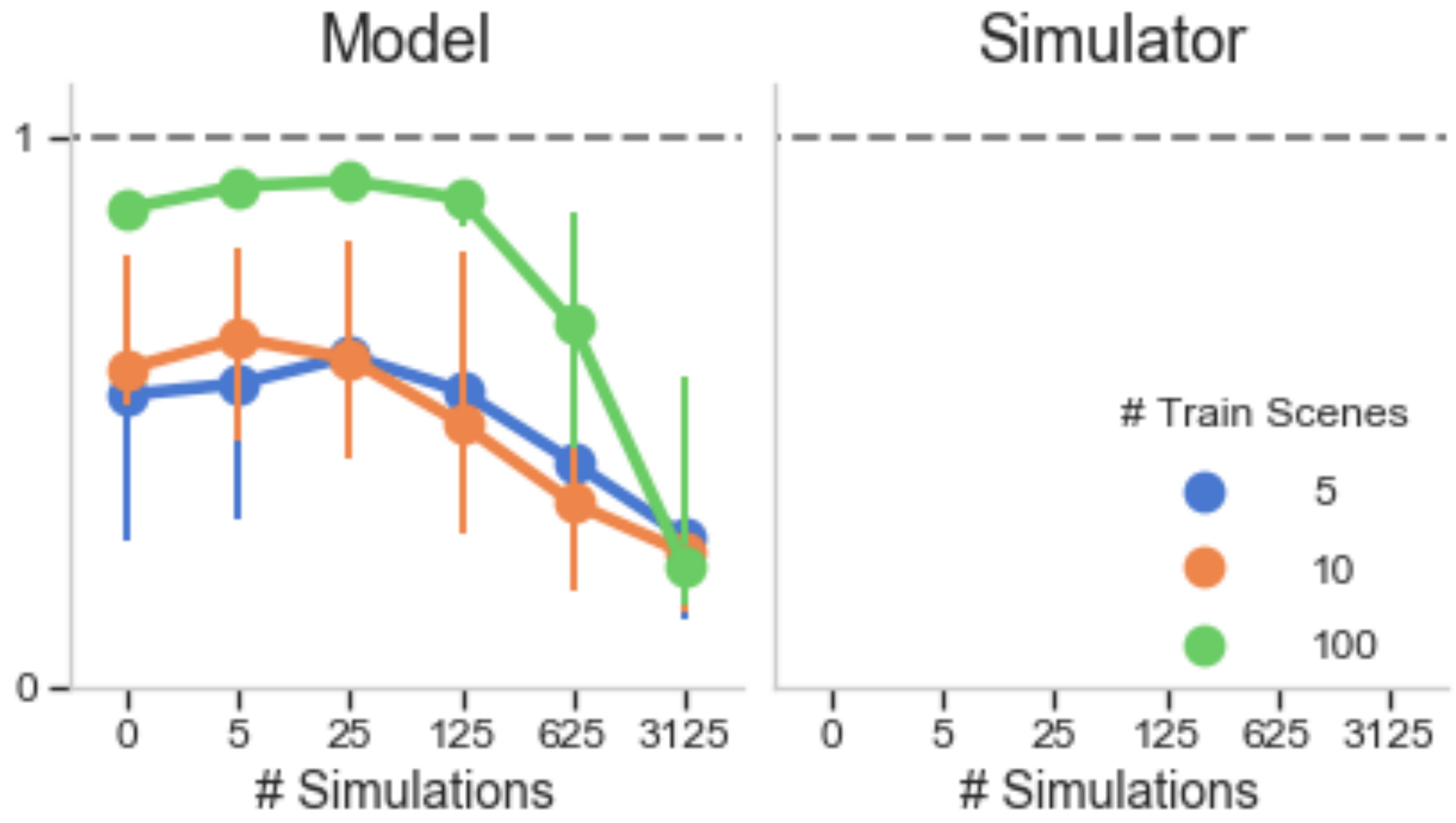
Hamrick et al. (2021)

Jessica Hamrick - jhamrick@deepmind.com

Generalizing to new mazes

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



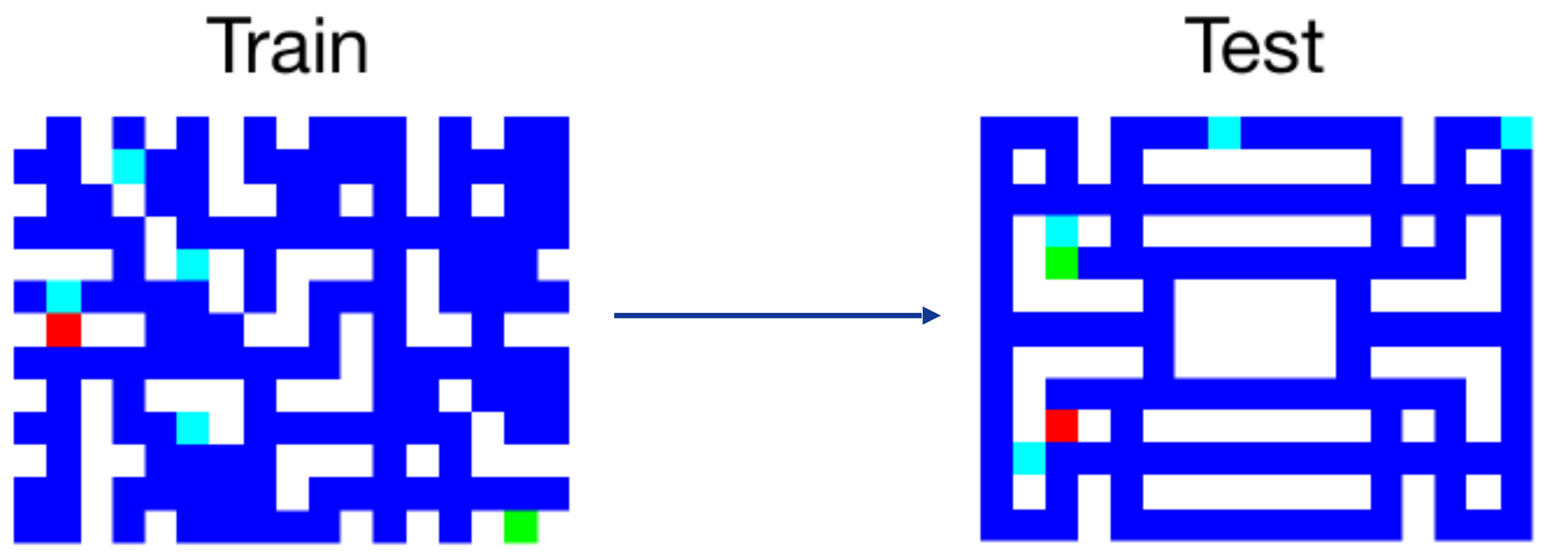


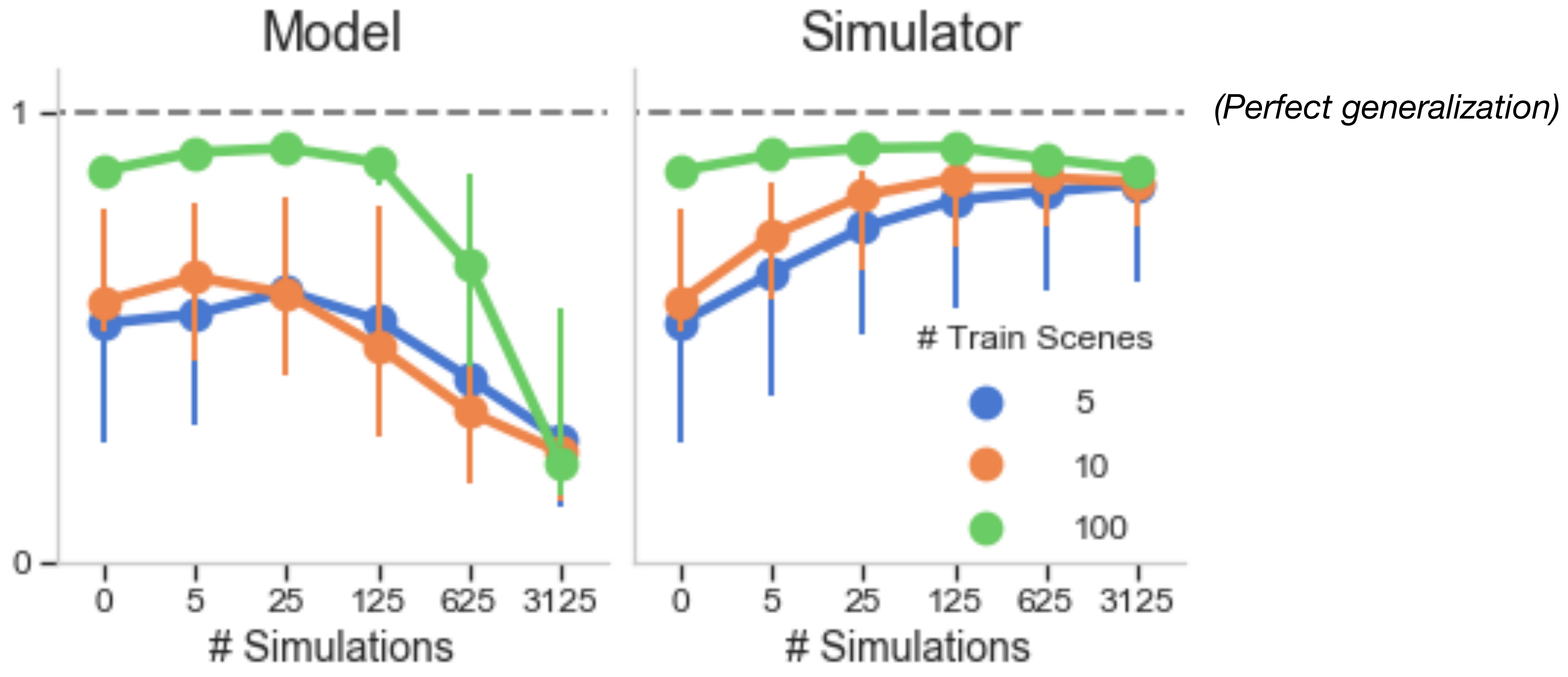
(Perfect generalization)

Hamrick et al. (2021)

Generalizing to new mazes

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

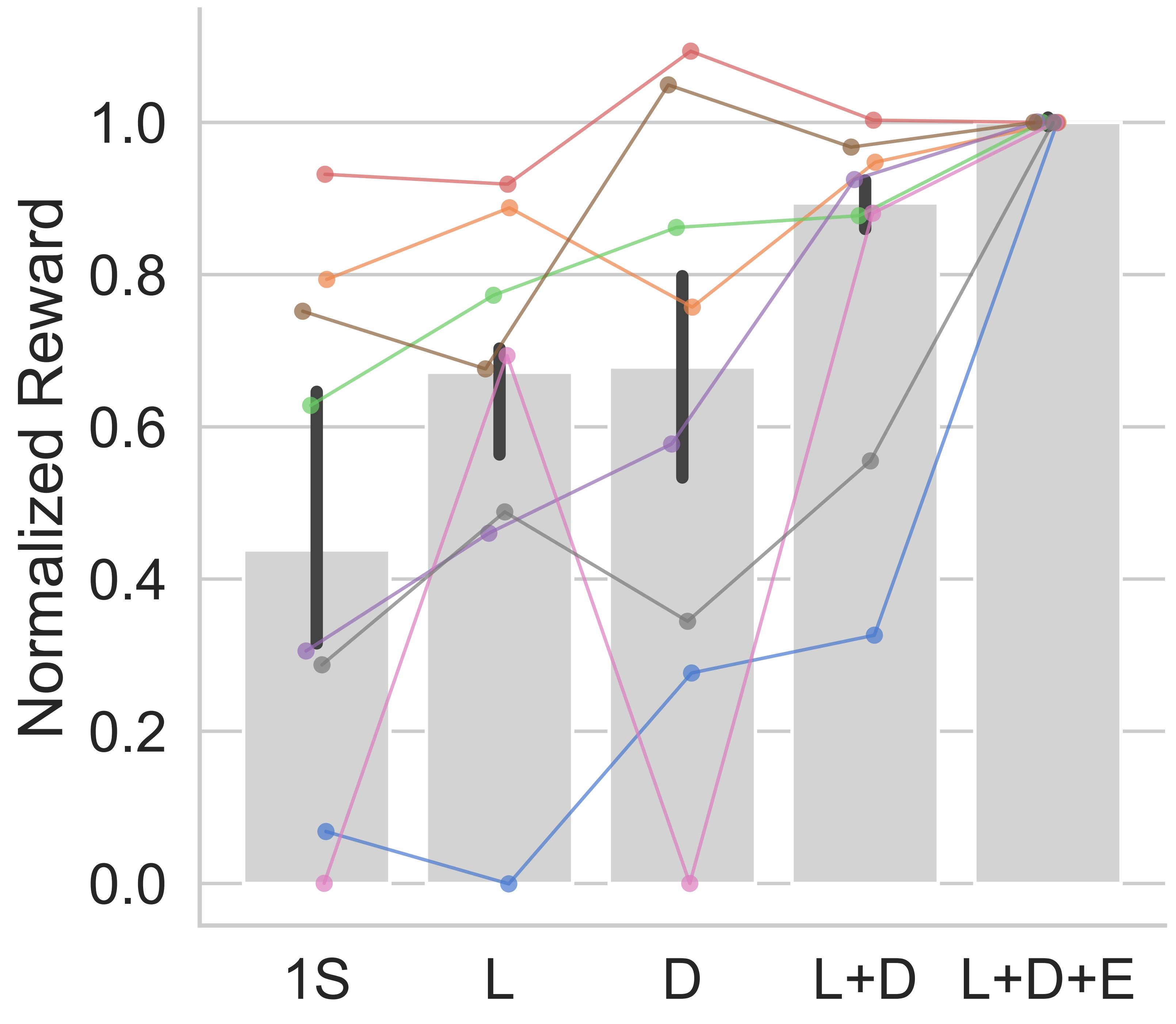




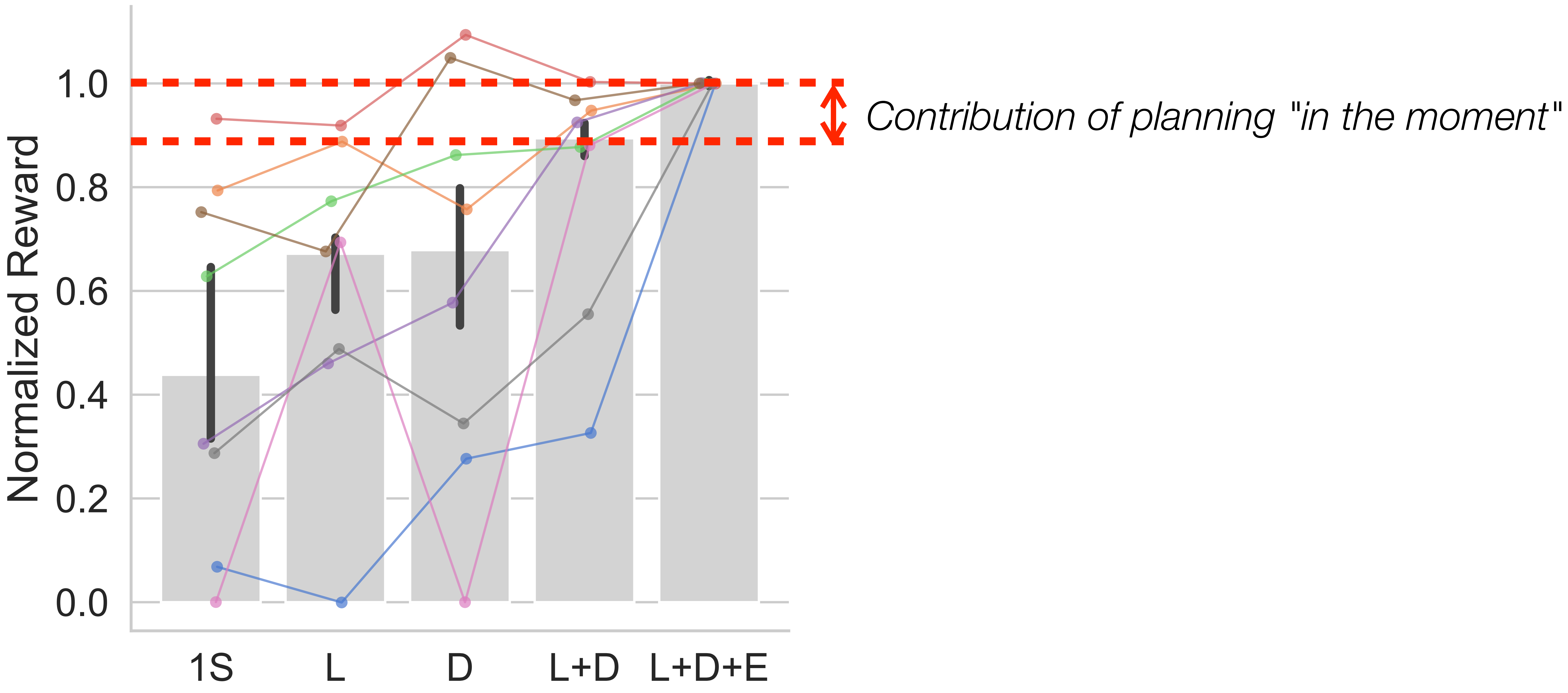
Planning—even with a perfect model—does not guarantee good generalization performance.

Hamrick et al. (2021)

Takeaway #1: Planning seems to be most useful during learning and less so at test time (in most environments).

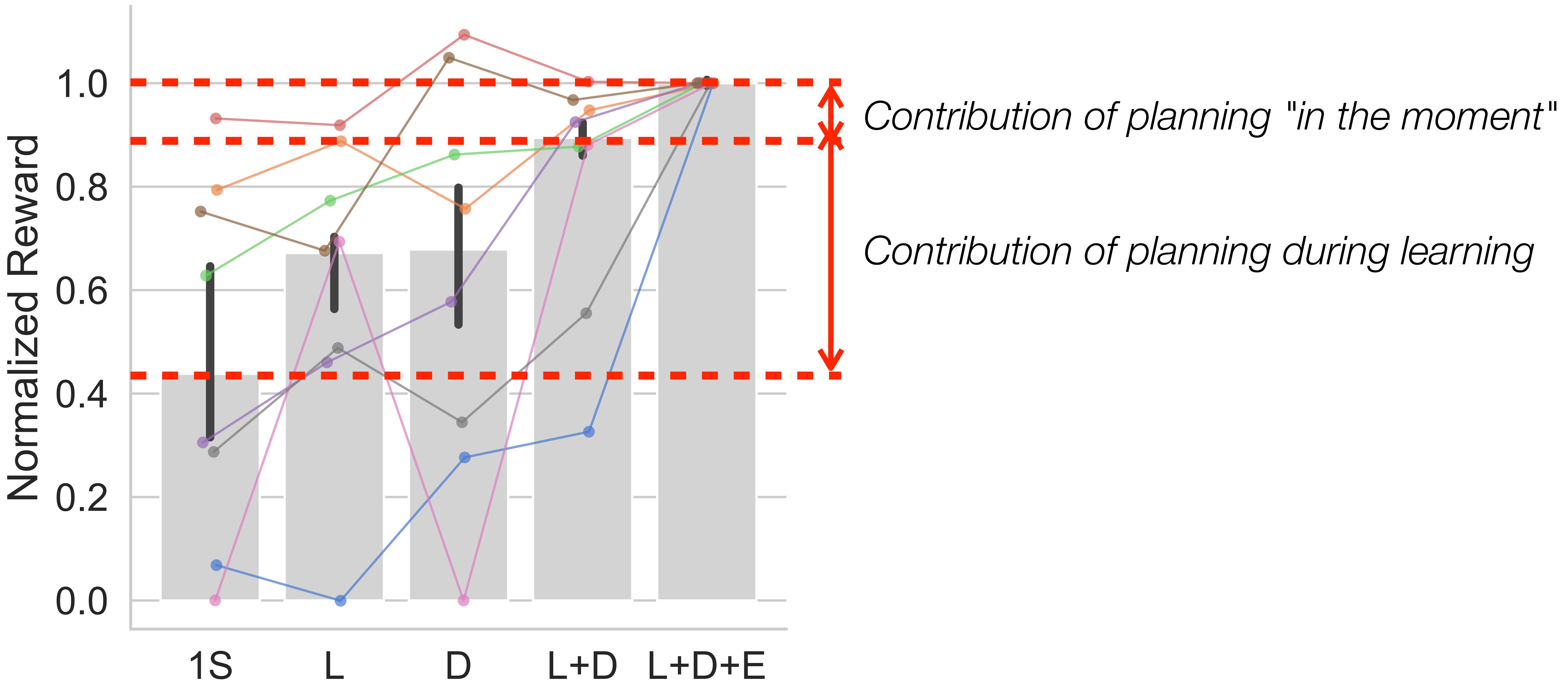


Takeaway #1: Planning seems to be most useful during learning and less so at test time (in most environments).



Jessica Hamrick - jhamrick@deepmind.com

Takeaway #1: Planning seems to be most useful during learning and less so at test time (in most environments).



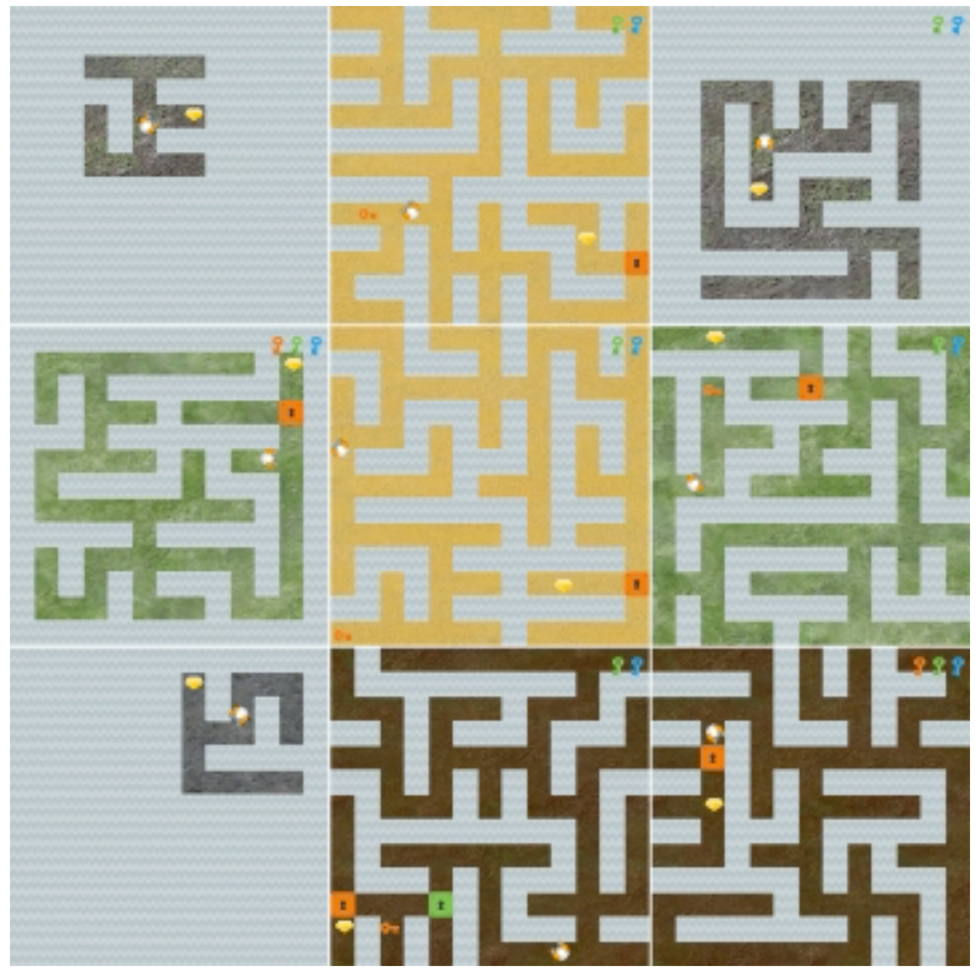
Jessica Hamrick - jhamrick@deepmind.com

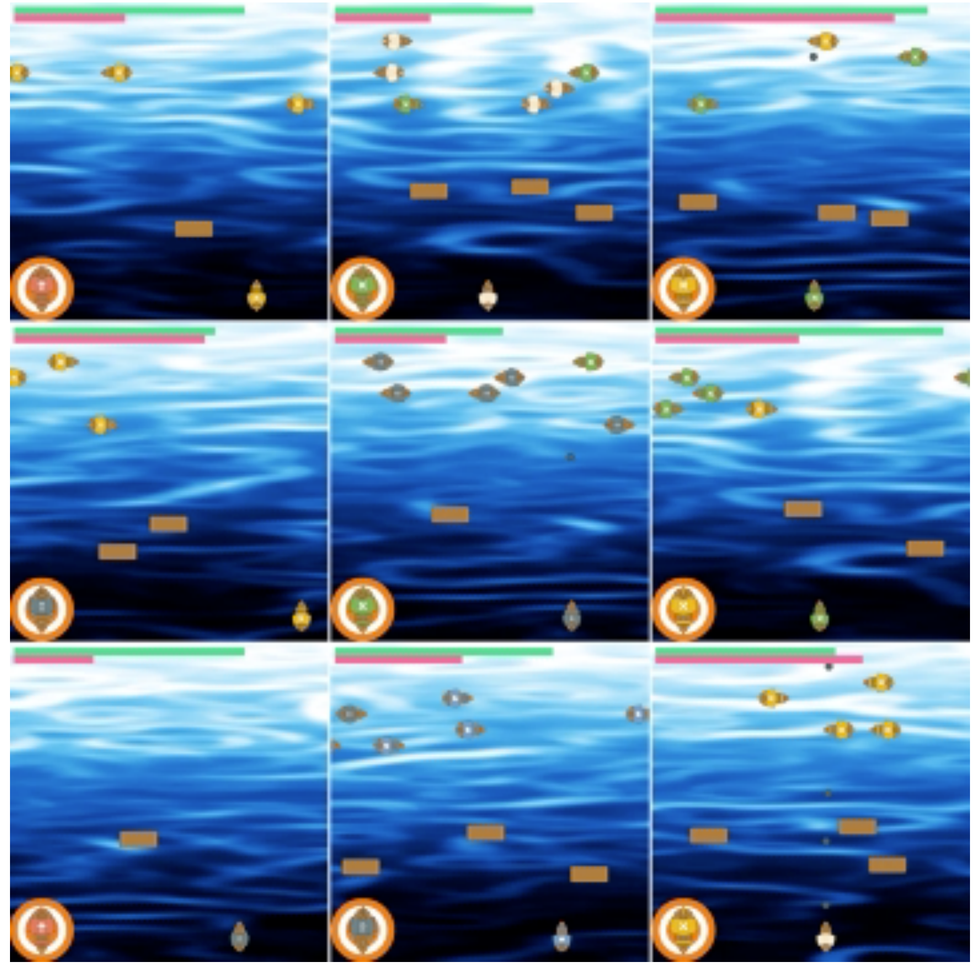
Takeaway #2: Effective planning requires having good representations for multiple components (policy/value/model).

Takeaway #2: Effective planning requires having good representations for multiple components (policy/value/model).

Anand, Walker et al. (2022). *ICLR*

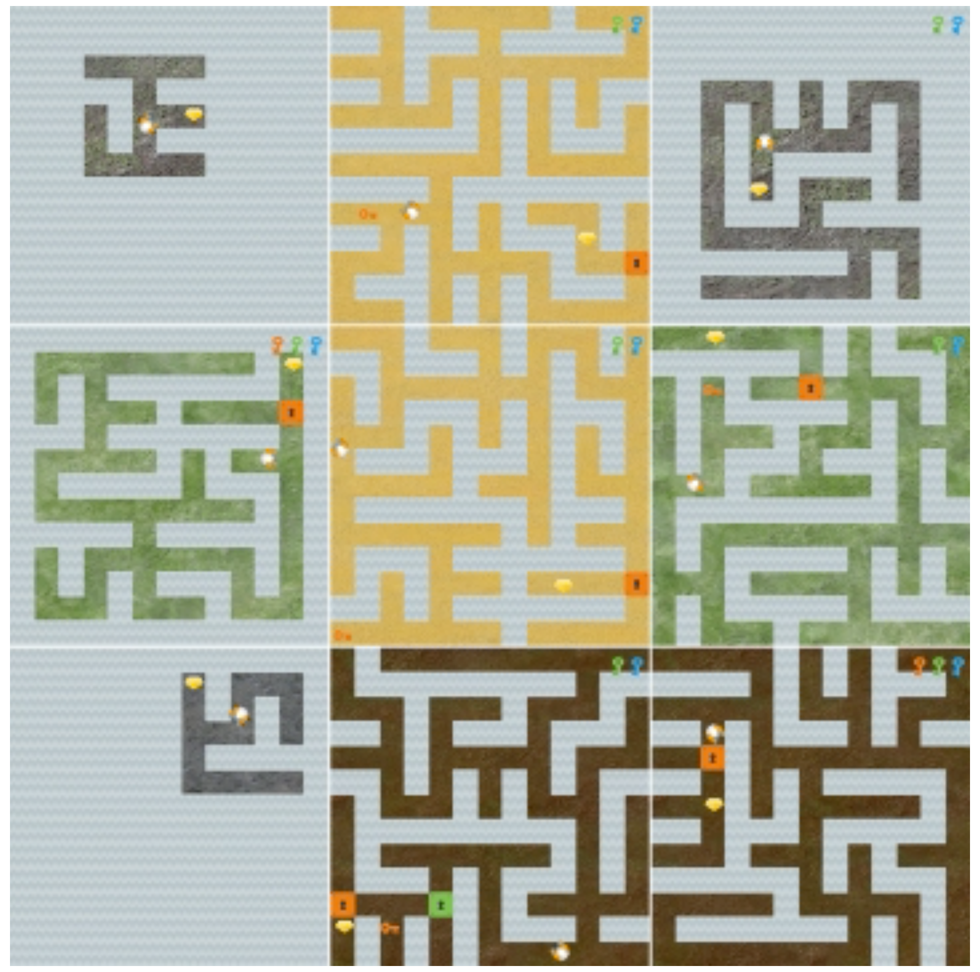
Takeaway #2: Effective planning requires having good representations for multiple components (policy/value/model).

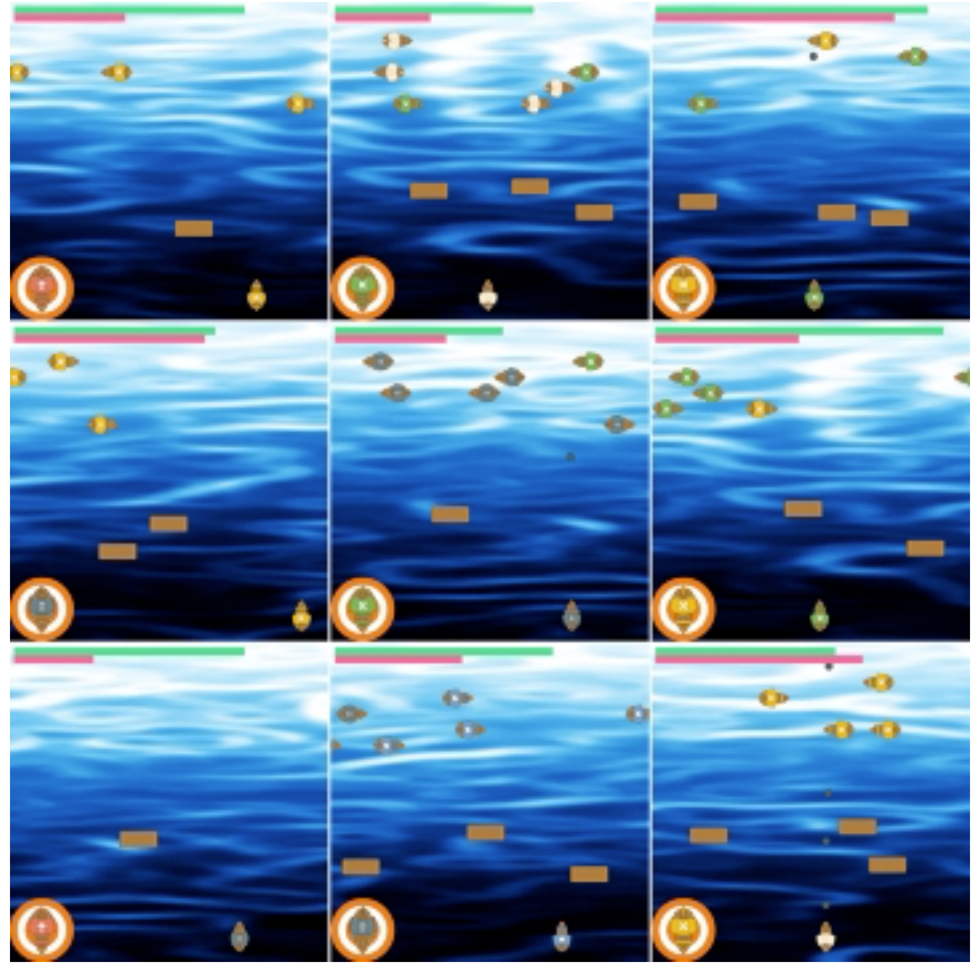




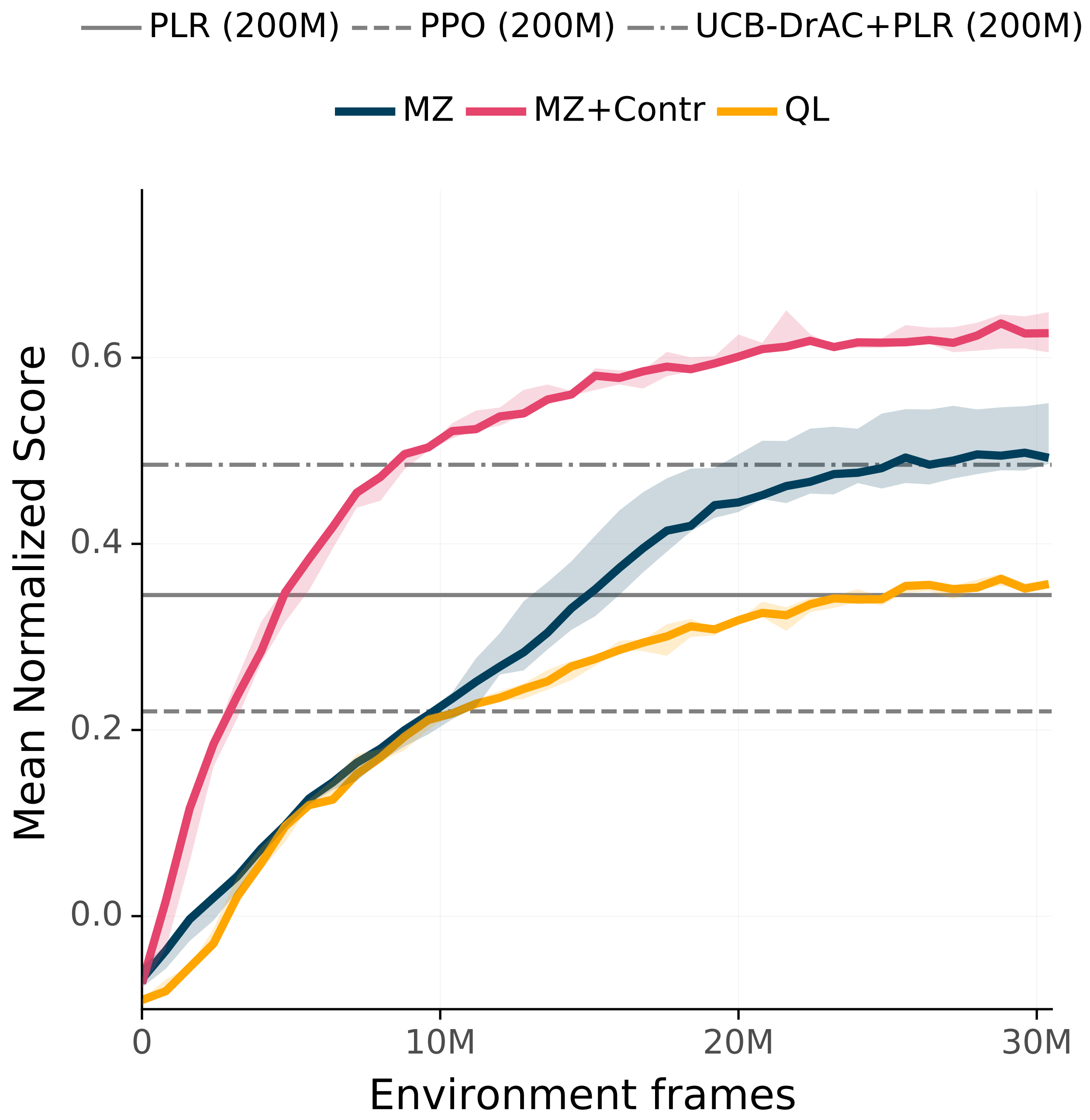

Anand, Walker et al. (2022). *ICLR*

Takeaway #2: Effective planning requires having good representations for multiple components (policy/value/model).





Anand, Walker et al. (2022). *ICLR*



Conundrum: If we have good enough value functions and policies, do we even need planning at all? 🤯

Conundrum: If we have good enough value functions and policies, do we even need planning at all? 🤯

For learning? **Yes**, planning helps!
At test time? **...maybe?**

Thanks to:
Ankesh Anand
Thomas Anthony
Feryal Behbahani
Lars Buesing
Abe Friesen
Arthur Guez
Yazhe Li
Sherjil Ozair
Julian Schrittwieser
Petar Veličković
Eszter Vértes
Fabio Viola
Jacob Walker
Sims Witherspoon
Theo Weber

Conundrum: If we have good enough value functions and policies, do we even need planning at all? 🤯

For learning? **Yes**, planning helps!
At test time? **...maybe?**

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.*