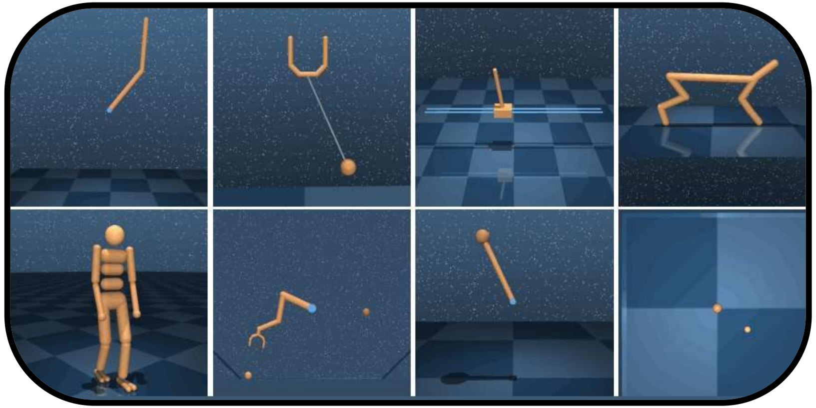
Scaling MAP-Elites to Deep Neuroevolution

Cédric Colas, Joost Huizinga, Vashisht Madhavan, Jeff Clune

UBER AI Labs

Solving games and control problems with deep neural networks



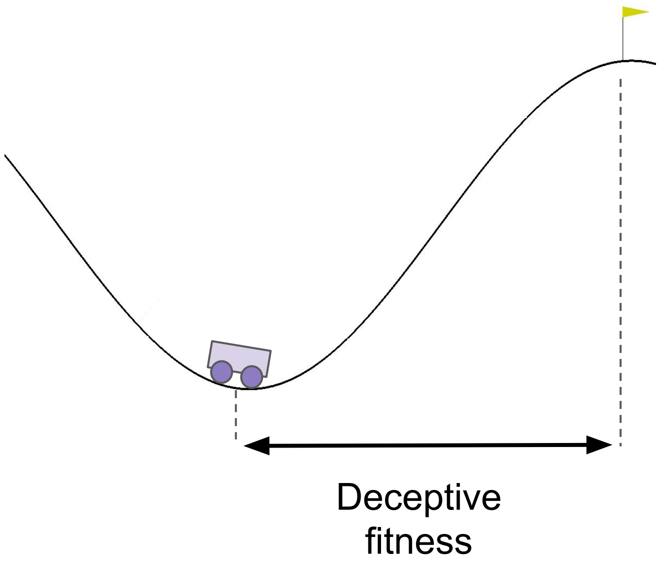
Tassa et al. (2018)

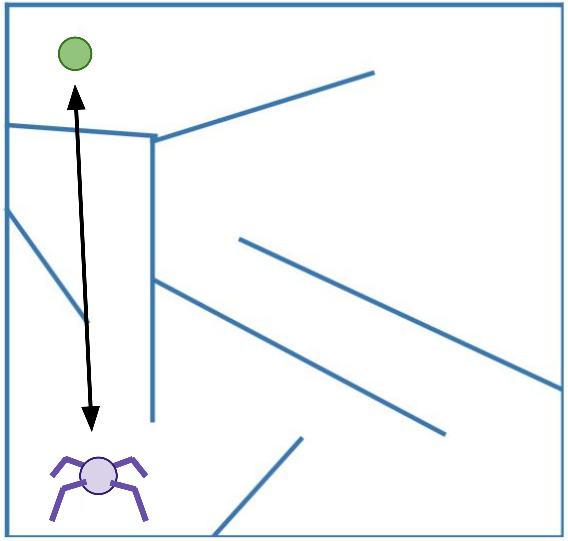
Bellemare et al. (2013)

… and many more (driving simulators, navigation in realistic houses etc)

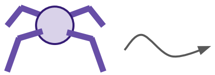
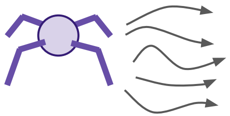
SGD-based (Deep RL) or black-box (Evolution Strategies) modern optimization methods leverage deep neural networks to solve high-dimensional problems.

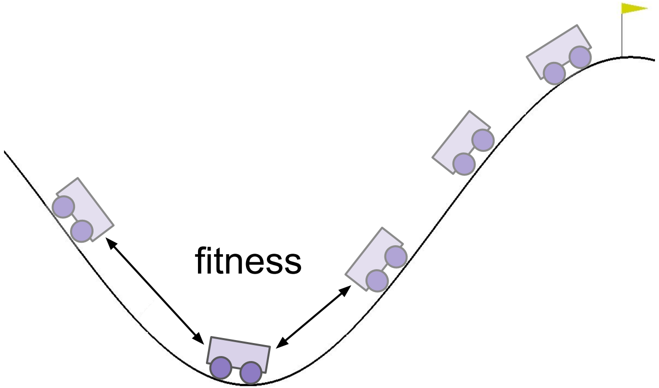
Deception: the path to success is not always a straight line

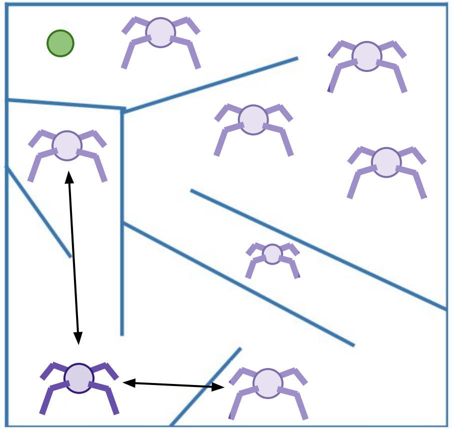




Searching for diversity instead

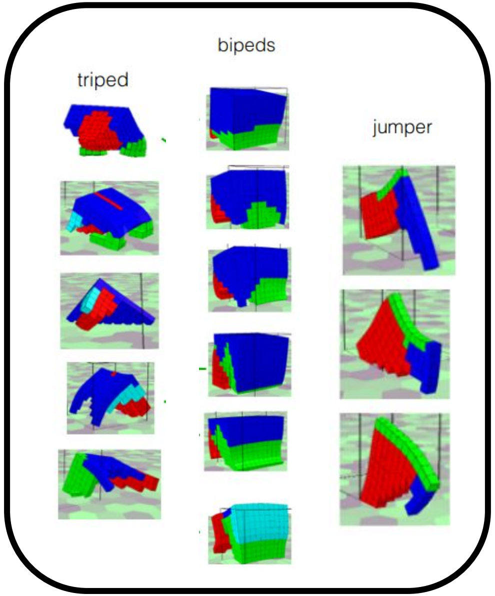
Instead of training one controller to solve the task  train many and maximize diversity 



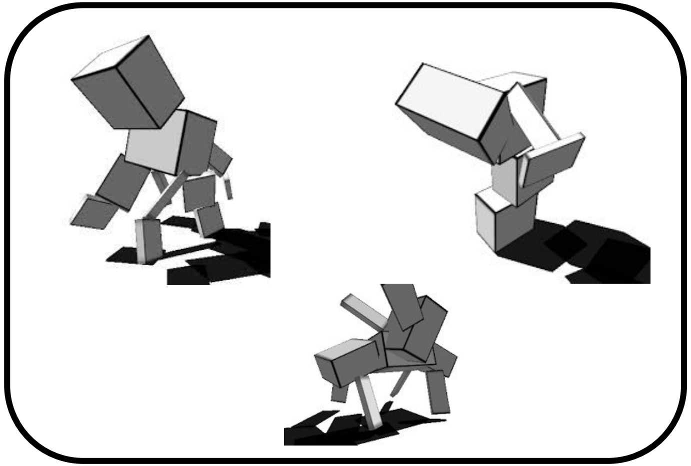


Goal Exploration Processes (GEP): Baranes & Oudeyer (2009, 2013)
Novelty-search (NS): Lehman & Stanley (2011)

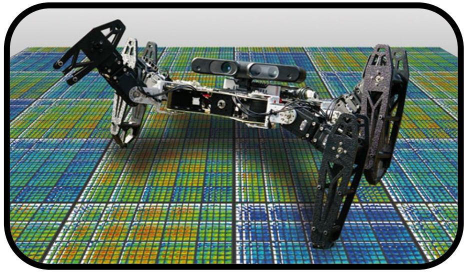
Searching for diversity and performance with Quality-Diversity



MAP-Elites: Mouret & Clune (2015)



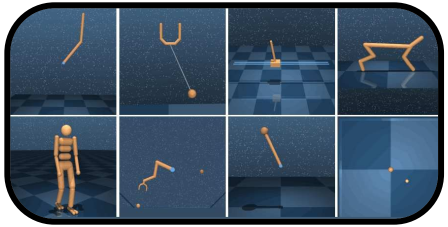
NS with local-competition:
Lehman & Stanley (2011)



Map-Elites + IT&E
Cully et al. (2015)

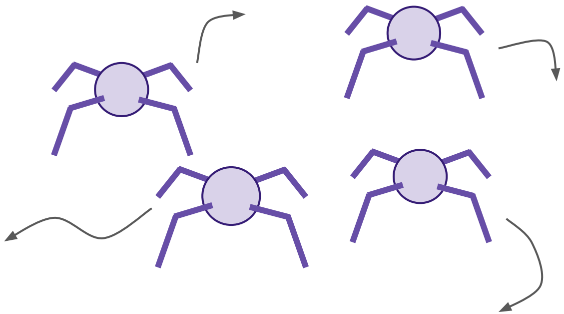
Quality & Diversity

Quality-Diversity currently limited to low-d controllers

		
Genetic Algorithms for deep nets Such et al. (2017)	✓	✗
Evolution Strategies for deep nets Salimans et al. (2017)	✓	✓

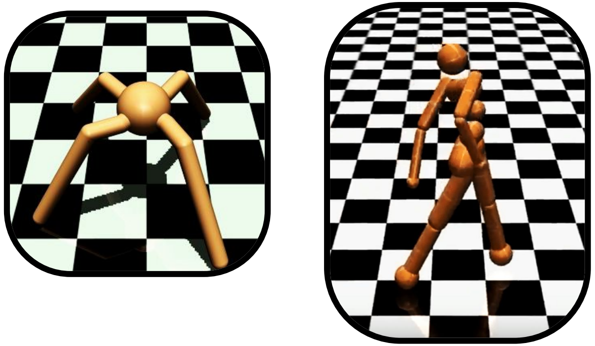
Quality-Diversity, powered by Evolution Strategies

A collection of
high-performing behaviors



MAP-Elites

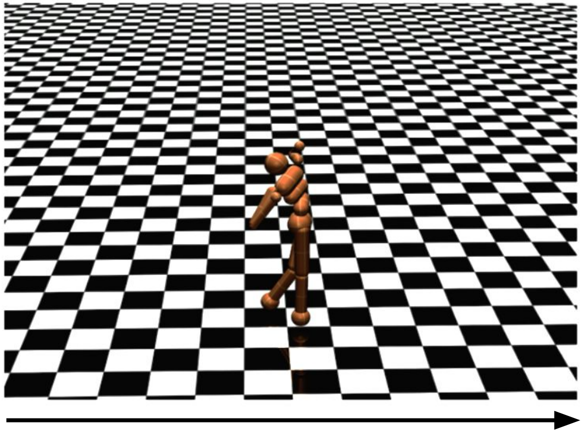
High-dimensional
controllers



Evolution Strategies

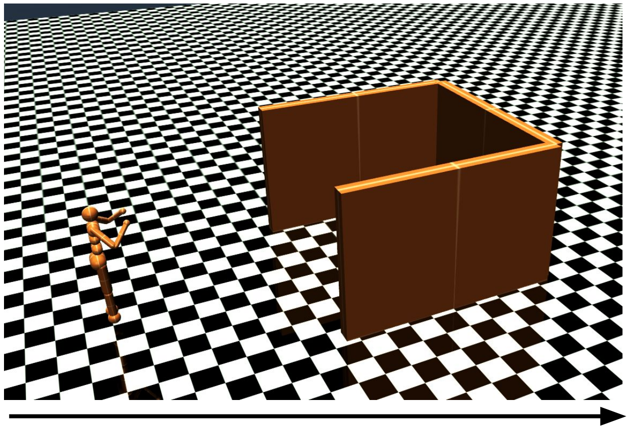
First steps: Novelty-Search, powered by ES

Humanoid

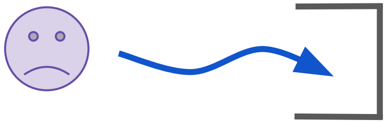


Gradient of performance

Deceptive Humanoid

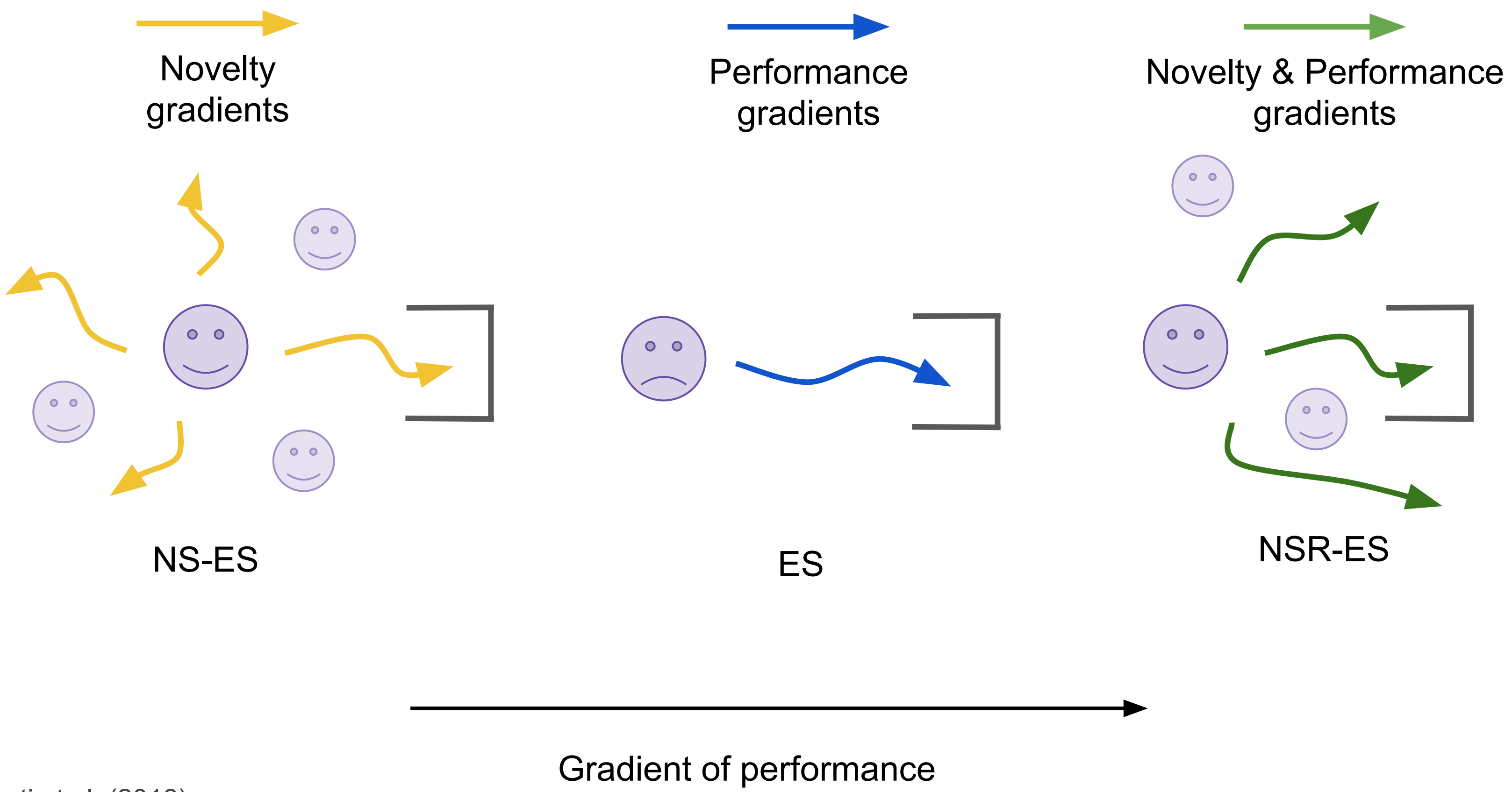


Gradient of performance (deceptive)



Salimans et al. (2017)
Conti et al. (2018)

First steps: Novelty-Search, powered by ES



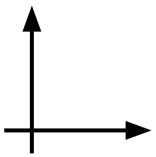
Conti et al. (2018)

Map-Elites with Evolution Strategies

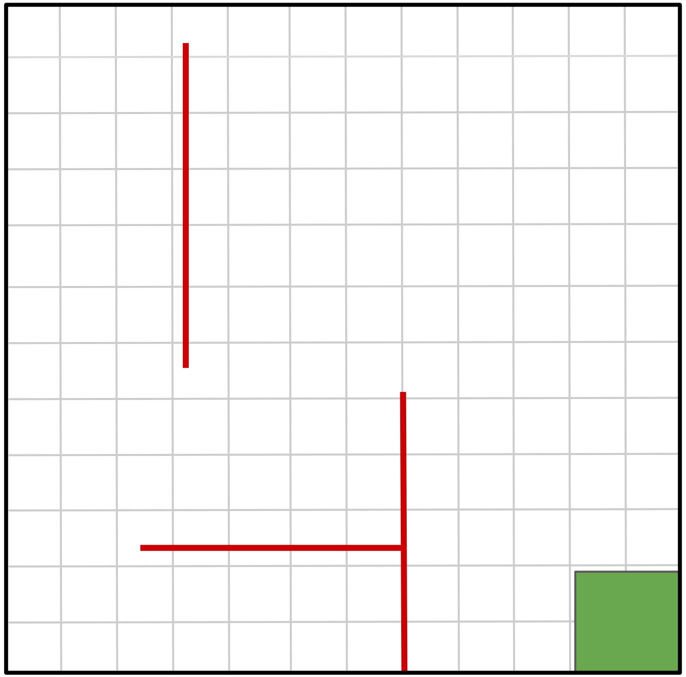
Objective: Build a behavioral repertoire of high-performing controllers.

Step 1: Define a behavioral characterization and a discretized behavioral map (BM)

Example: Final 2D position in a maze.



BC = 2D
spatial space



MAP-Elite based on Evolutionary Strategy (ME-ES)

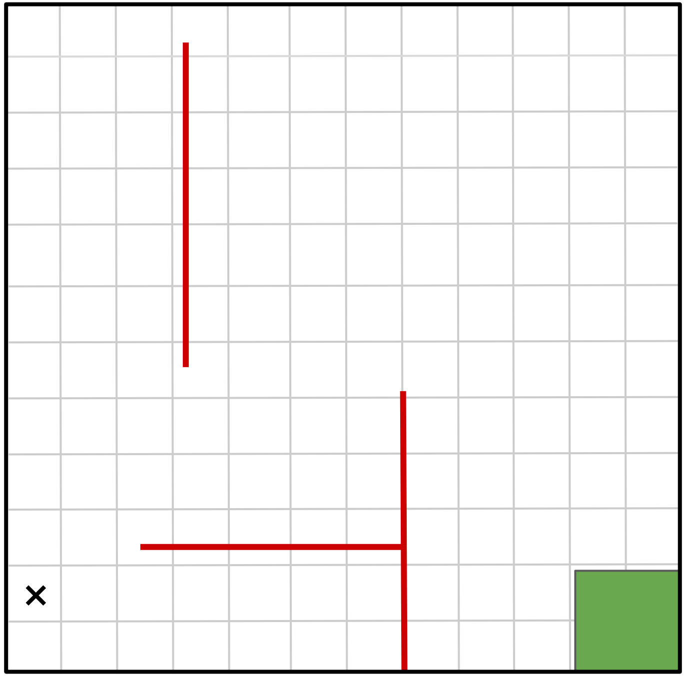
Objective: Build a behavioral repertoire of high-performing controllers.

Step 1: Define a behavioral characterization and a discretized behavioral map (BM)

Step 2: Fill the map!

Pick a cell and its controller

×



MAP-Elite based on Evolutionary Strategy (ME-ES)

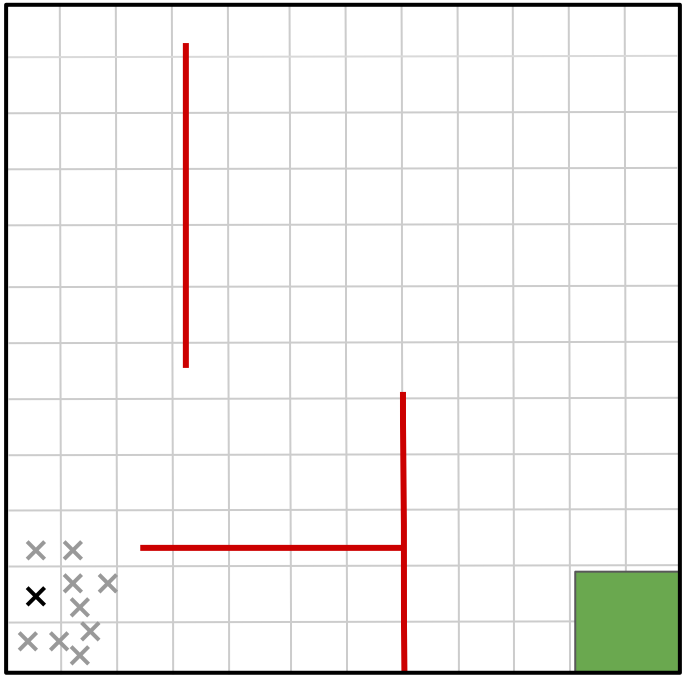
Objective: Build a behavioral repertoire of high-performing controllers.

Step 1: Define a behavioral characterization and a discretized behavioral map (BM)

Step 2: Fill the map!

Pick a cell and its controller 

Run mutated controllers 



MAP-Elite based on Evolutionary Strategy (ME-ES)

Objective: Build a behavioral repertoire of high-performing controllers.

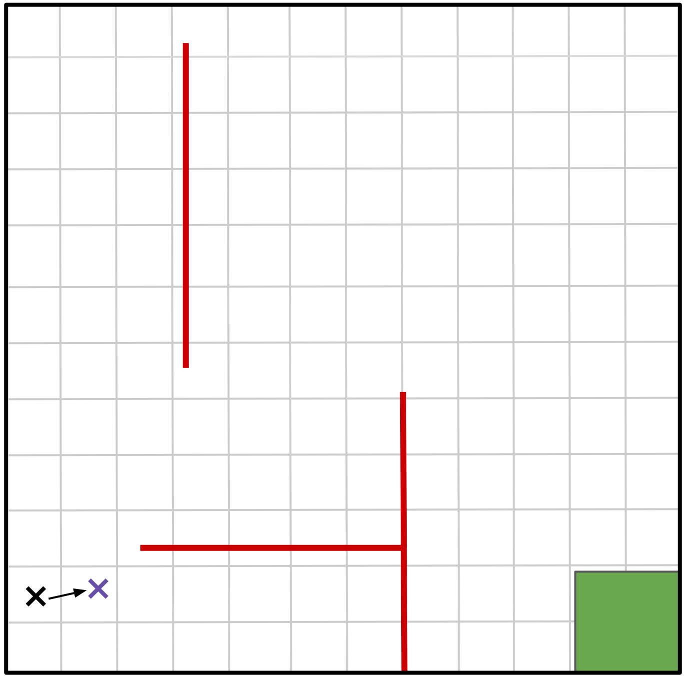
Step 1: Define a behavioral characterization and a discretized behavioral map (BM)

Step 2: Fill the map!

Pick a cell and its controller

Run mutated controllers

Compute ES update and evaluate controller



MAP-Elite based on Evolutionary Strategy (ME-ES)

Objective: Build a behavioral repertoire of high-performing controllers.

Step 1: Define a behavioral characterization and a discretized behavioral map (BM)

Step 2: Fill the map!

Pick a cell and its controller

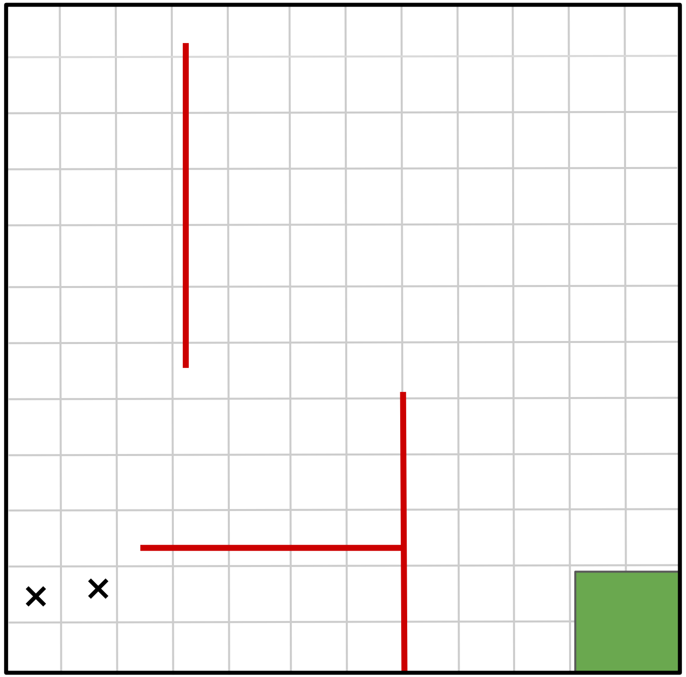
Run mutated controllers

Compute ES update and evaluate controller

Add it to the BM if:
- falls in a new cell

OR

- achieves high performance



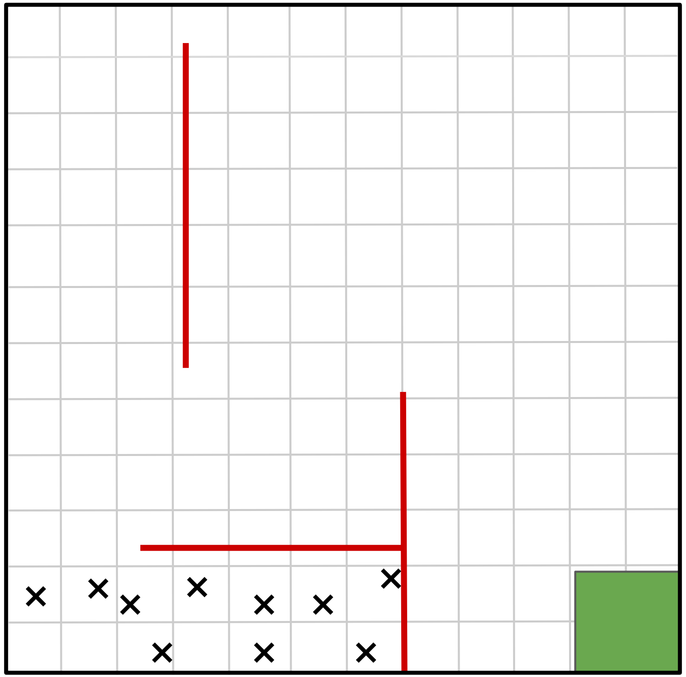
MAP-Elite based on Evolutionary Strategy (ME-ES)

Objective: Build a behavioral repertoire of high-performing controllers.

Step 1: Define a behavioral characterization and a discretized behavioral map (BM)

Step 2: Fill the map!

Repeat !



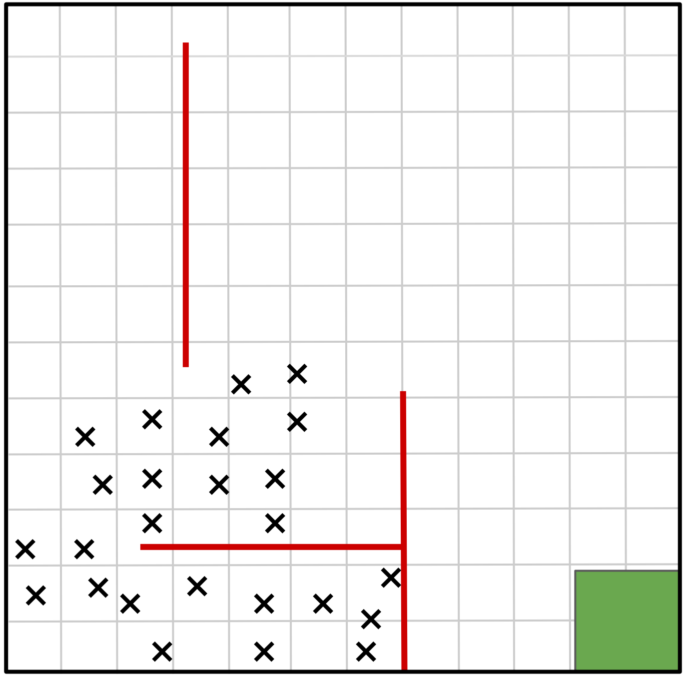
MAP-Elite based on Evolutionary Strategy (ME-ES)

Objective: Build a behavioral repertoire of high-performing controllers.

Step 1: Define a behavioral characterization and a discretized behavioral map (BM)

Step 2: Fill the map!

Repeat !



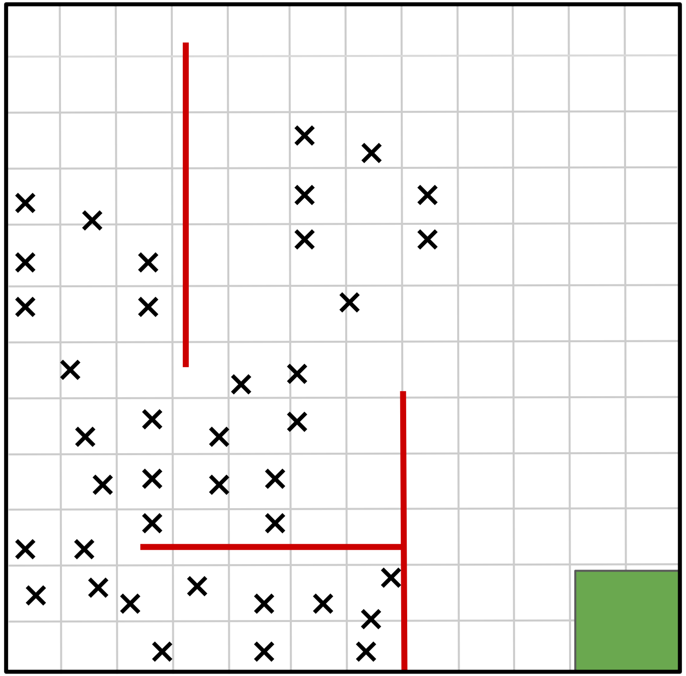
MAP-Elite based on Evolutionary Strategy (ME-ES)

Objective: Build a behavioral repertoire of high-performing controllers.

Step 1: Define a behavioral characterization and a discretized behavioral map (BM)

Step 2: Fill the map!

Repeat !



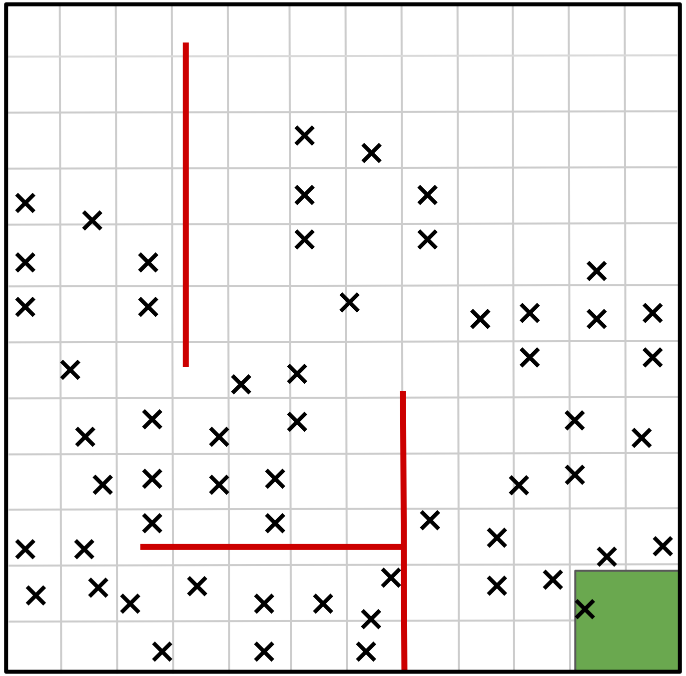
MAP-Elite based on Evolutionary Strategy (ME-ES)

Objective: Build a behavioral repertoire of high-performing controllers.

Step 1: Define a behavioral characterization and a discretized behavioral map (BM)

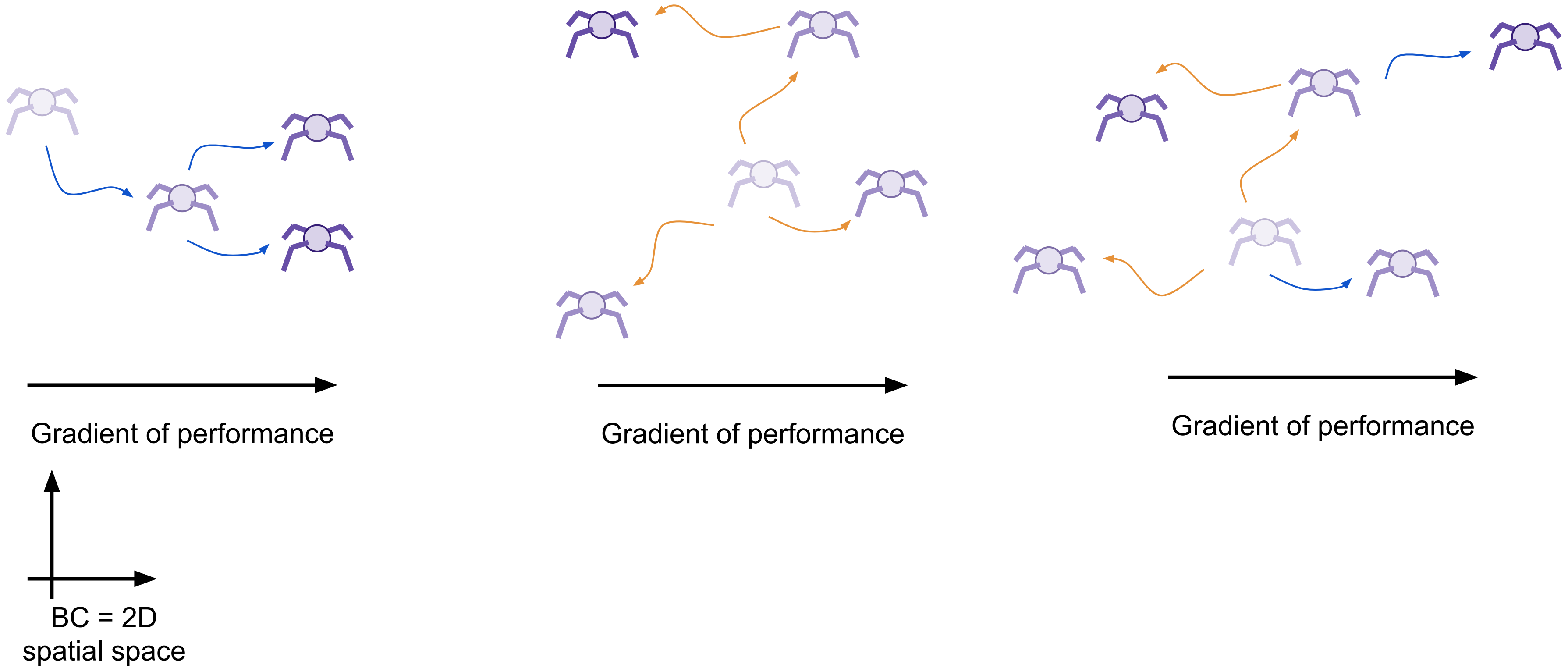
Step 2: Fill the map!

Repeat !

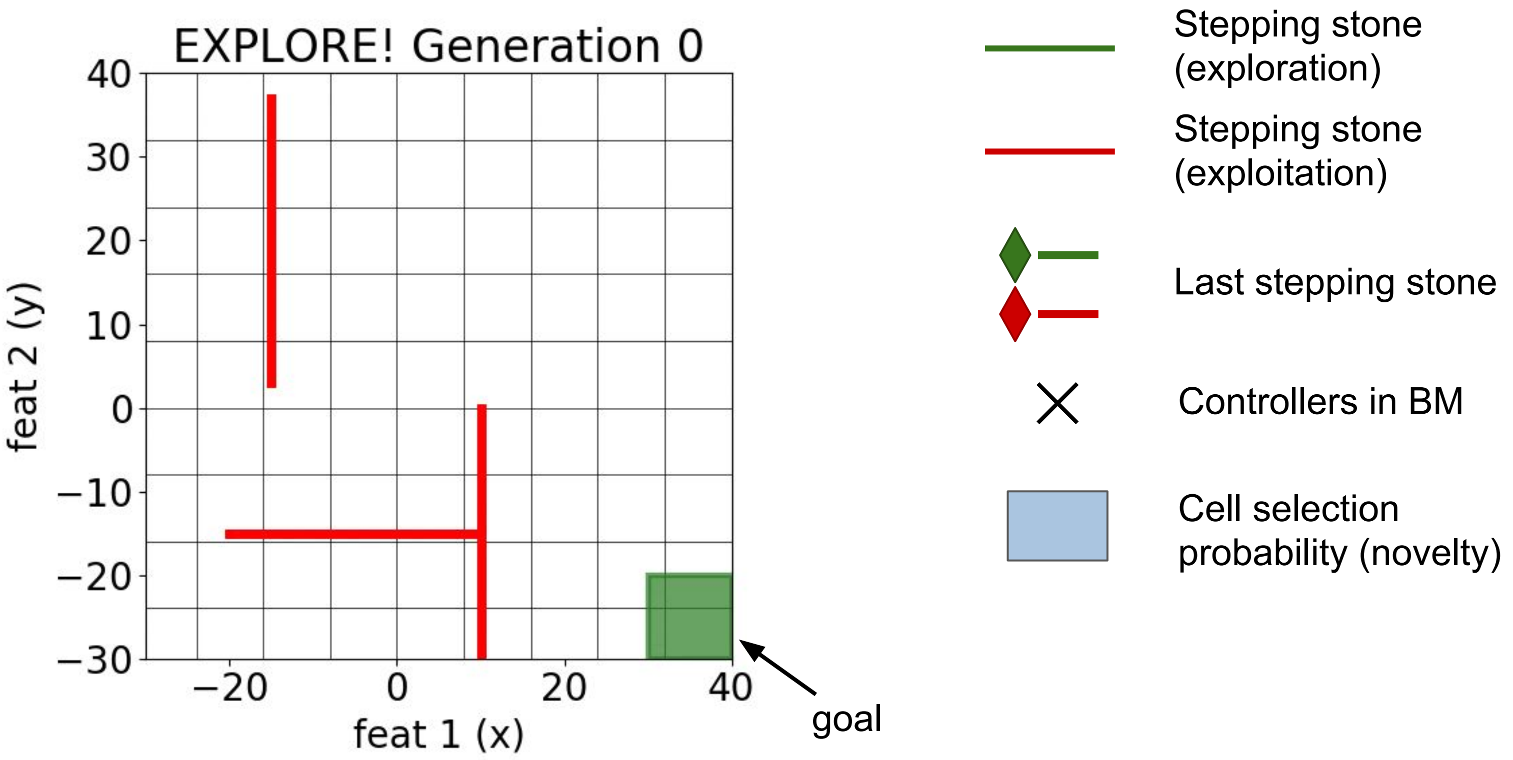


Three variants of ME-ES

ME-ES exploit   ME-ES explore   ME-ES explore - exploit



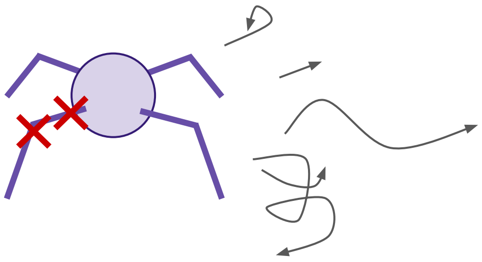
ME-ES at play


EXPLORE! Generation 0
feat 2 (y)
feat 1 (x)
40
30
20
10
0
−10
−20
−30
−20
0
20
40
goal
Stepping stone (exploration)
Stepping stone (exploitation)
Last stepping stone
Controllers in BM
Cell selection probability (novelty)

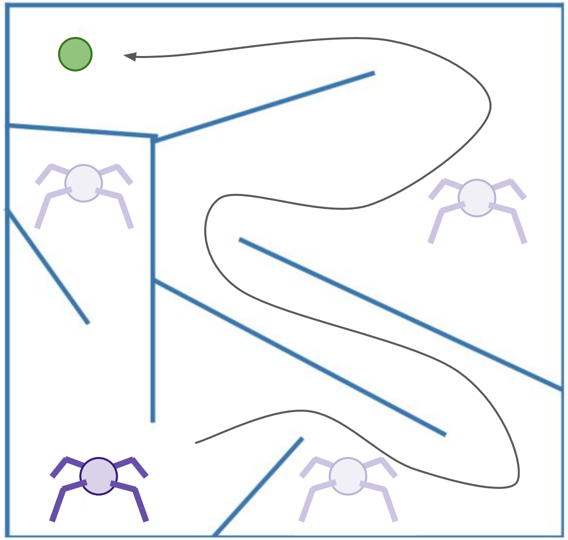
Experiments

Two applications: Damage Adaptation & Exploration

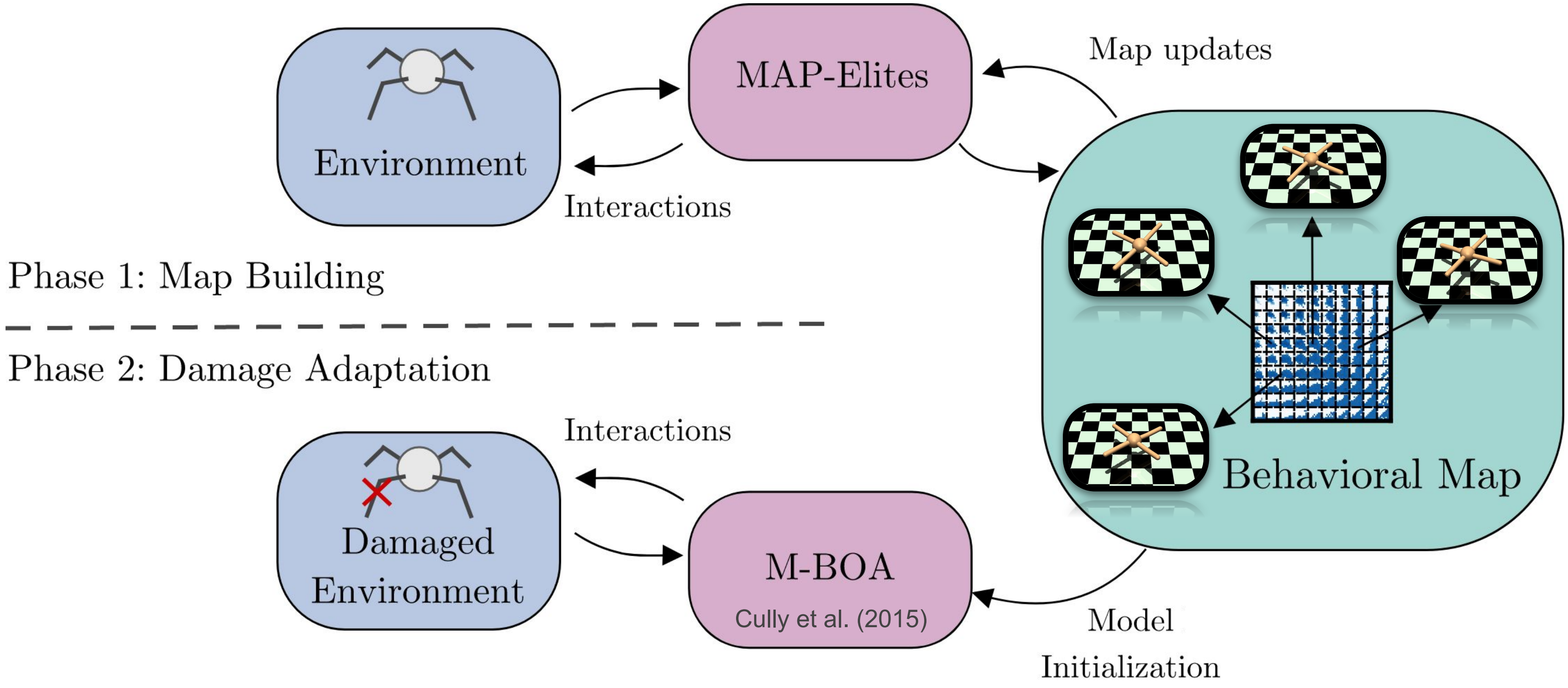
Damage Adaptation



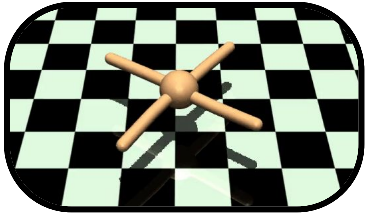
Exploration



Application #1 - Damage Adaptation



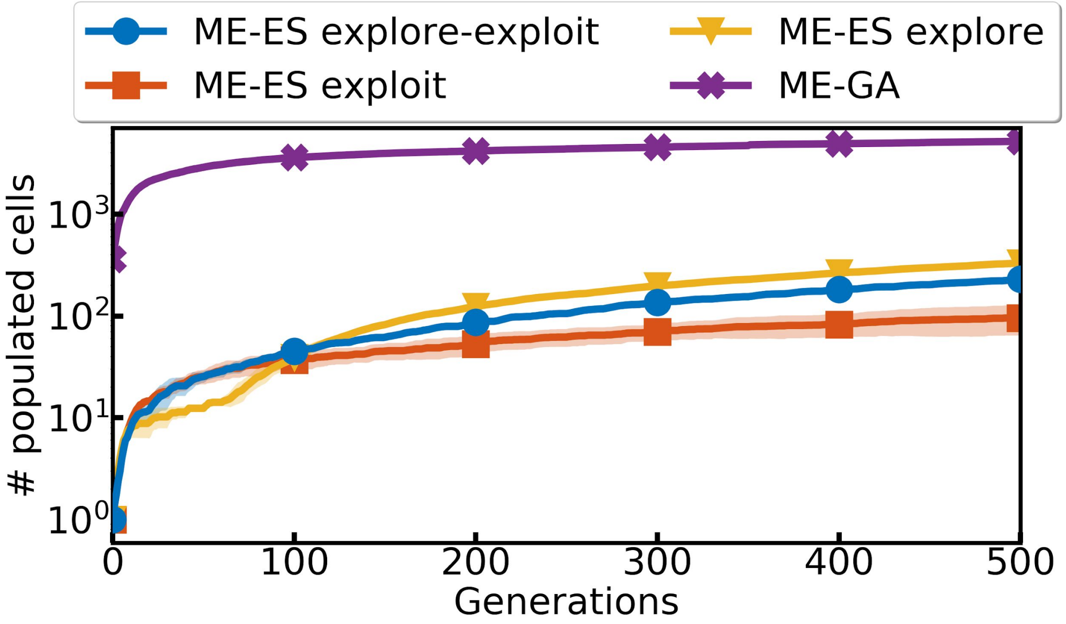
Cell Coverage



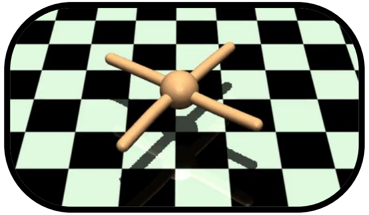
Ant-v2

Reward
default Gym

BC
[% leg contact]$_{L1:L4}$



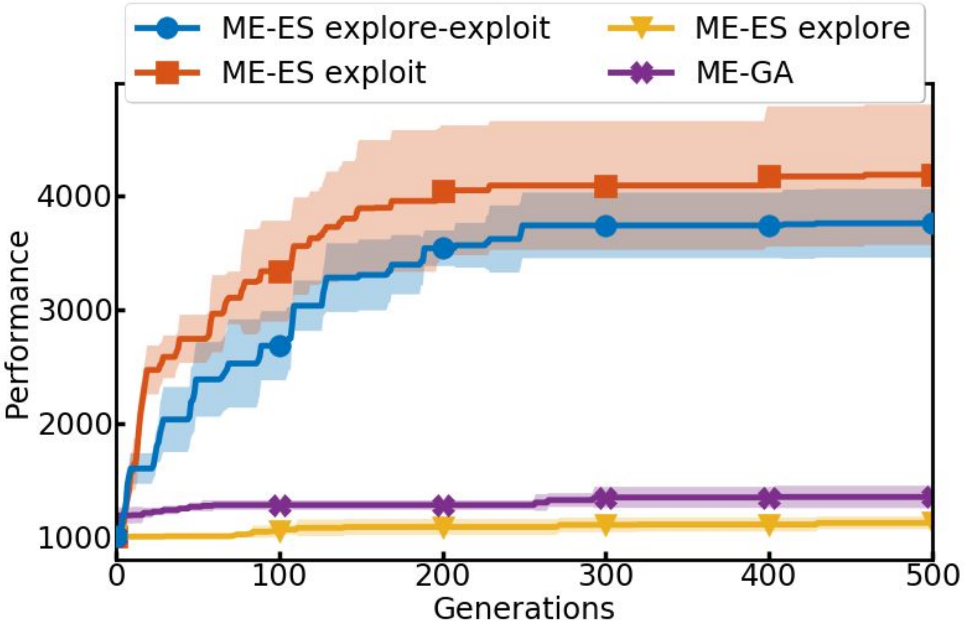
Maximum performance



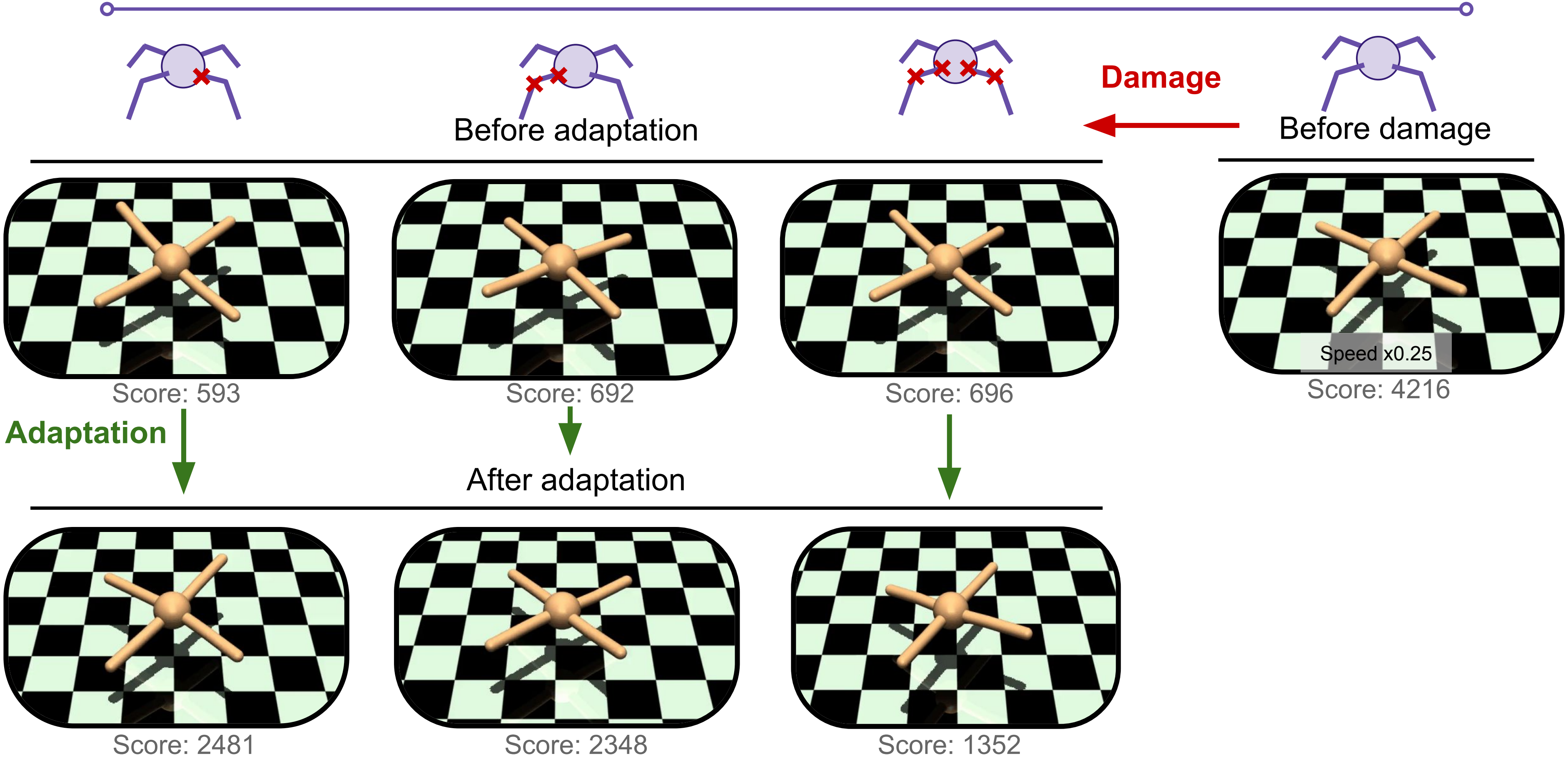
Ant-v2

Reward
default Gym

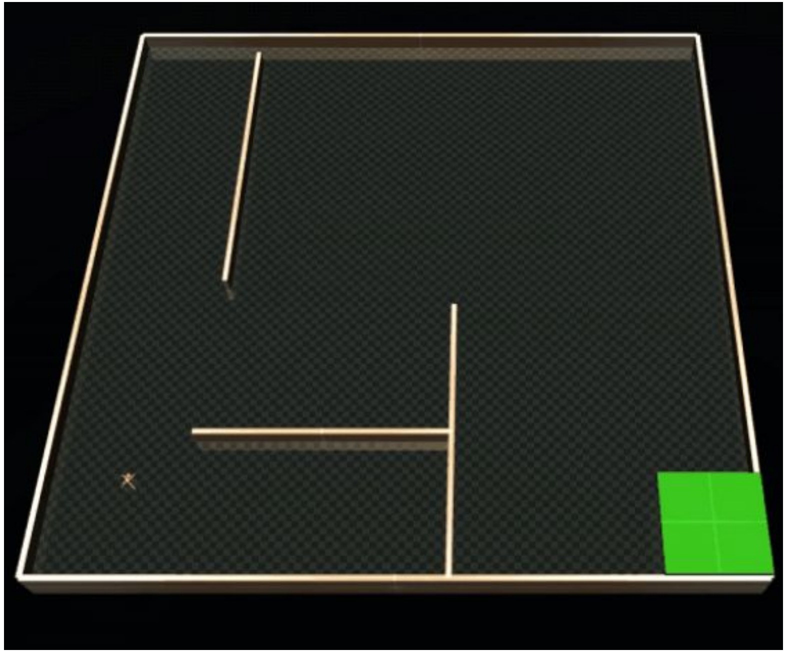
BC
[% leg contact]$_{L1:L4}$

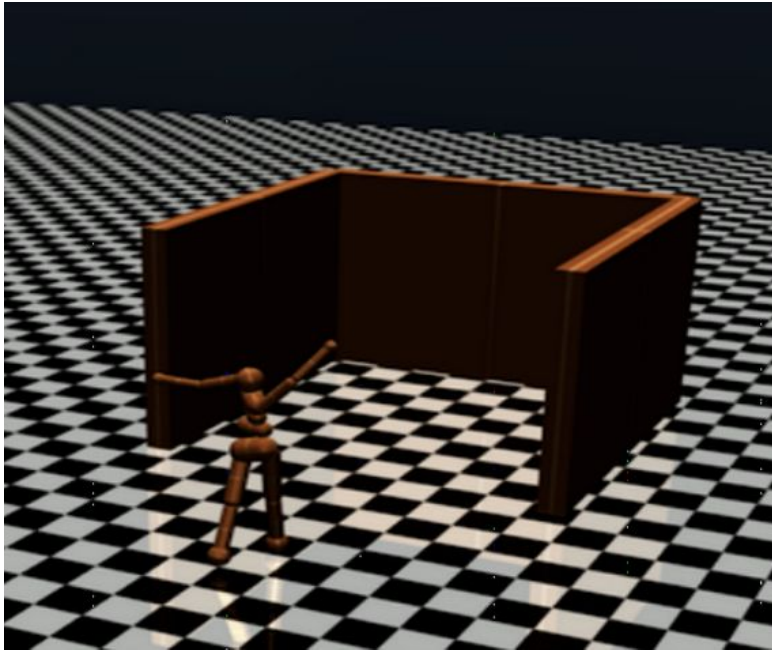


Phase 2: Damage Adaptation



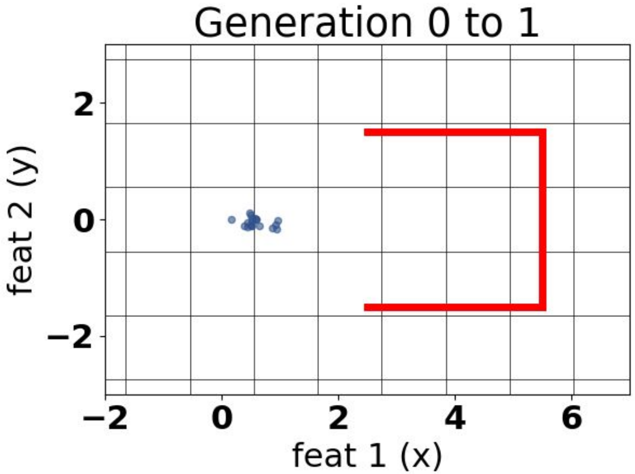
Application #2 - Exploration



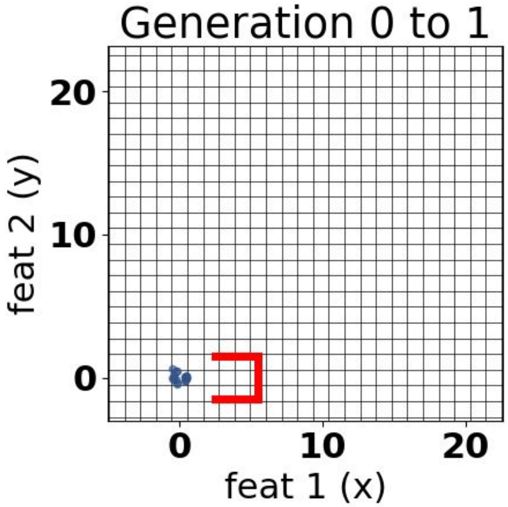


Exploration: Humanoid Deceptive

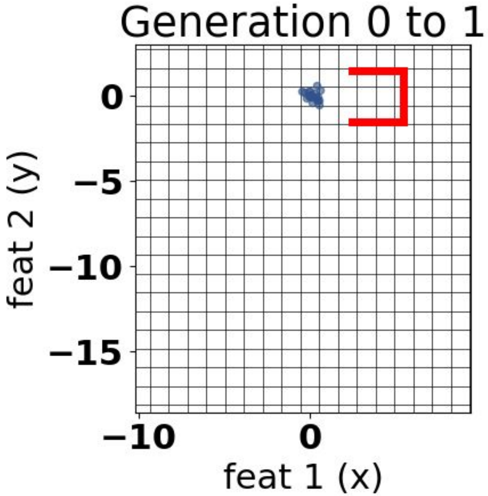
ME-ES exploit



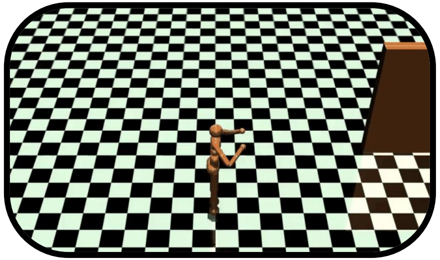
ME-ES
explore-exploit



ME-ES explore



Exploration: Humanoid Deceptive

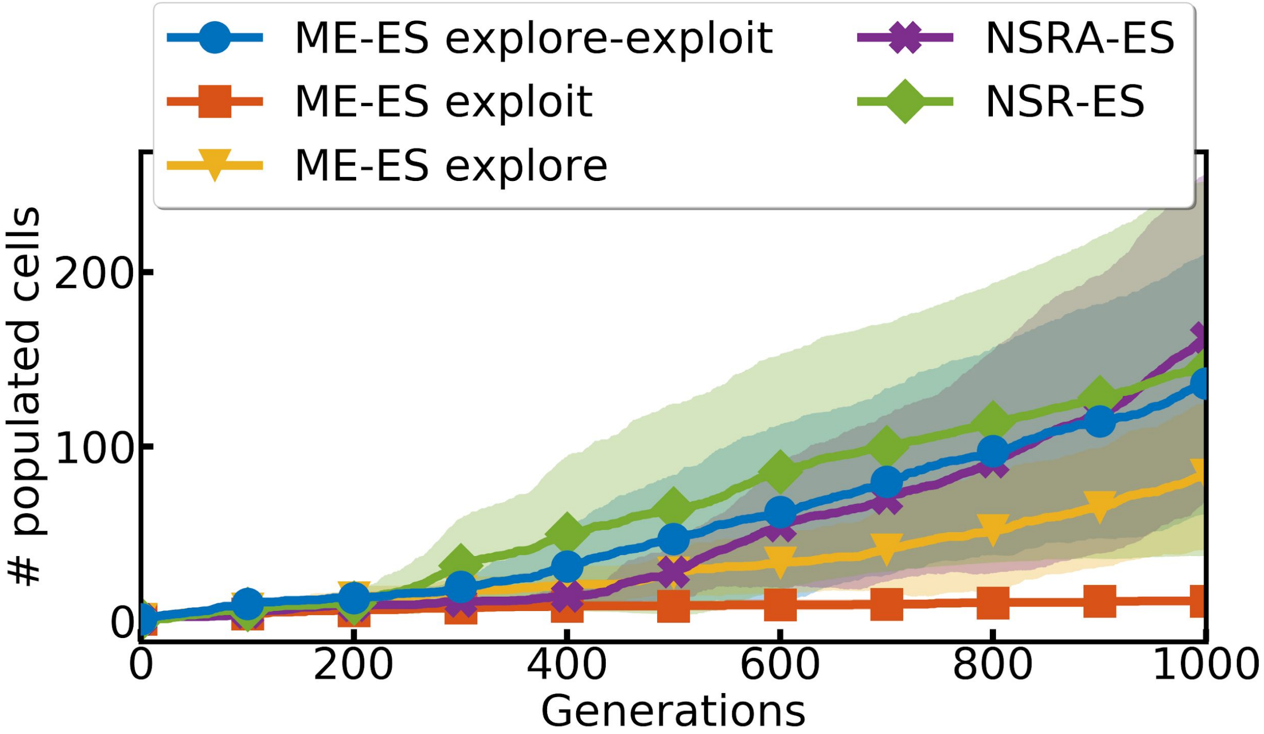


HumanoidDeceptive

Reward
default Gym

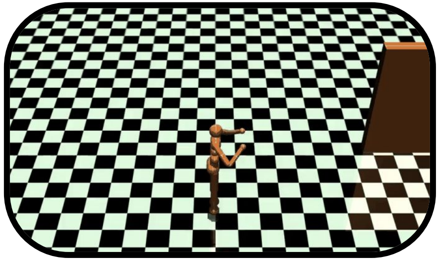
BC
Final (x, y)

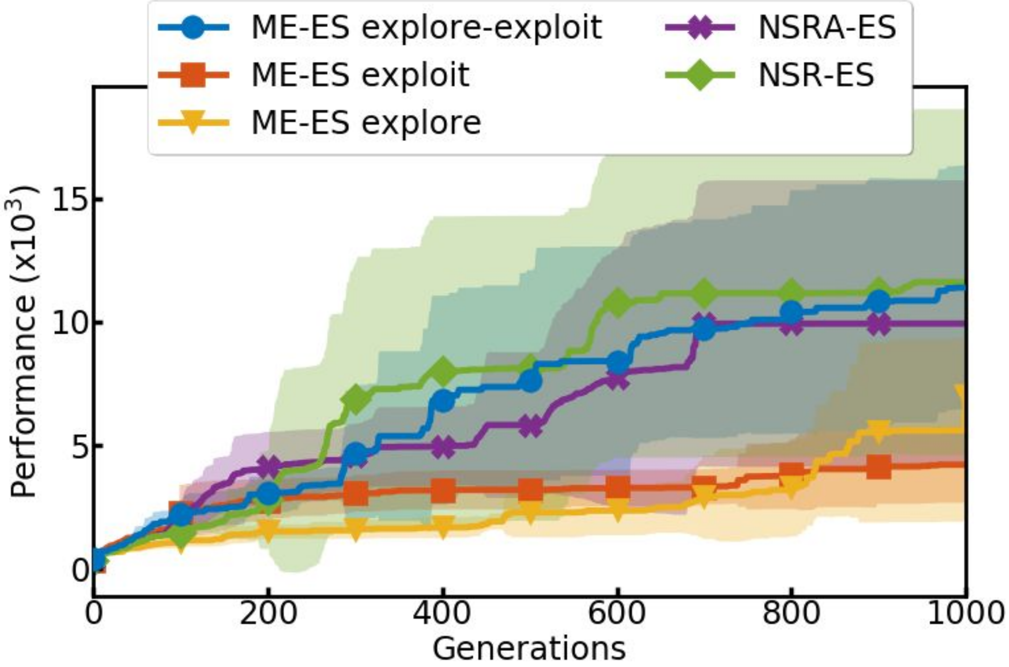
Cell Coverage



Exploration: Humanoid Deceptive

Best performance





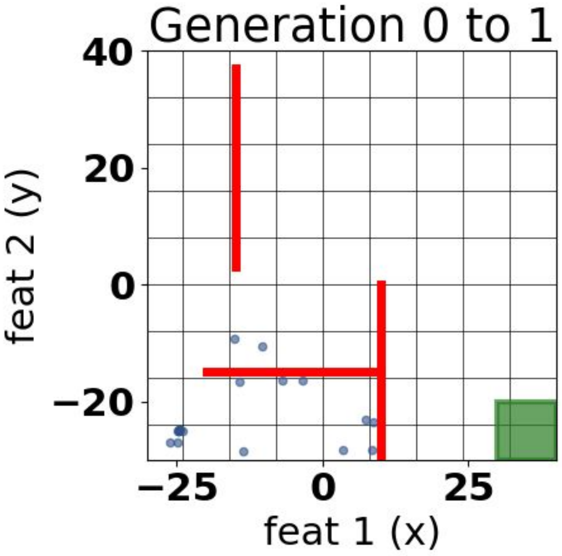
HumanoidDeceptive

Reward
default Gym

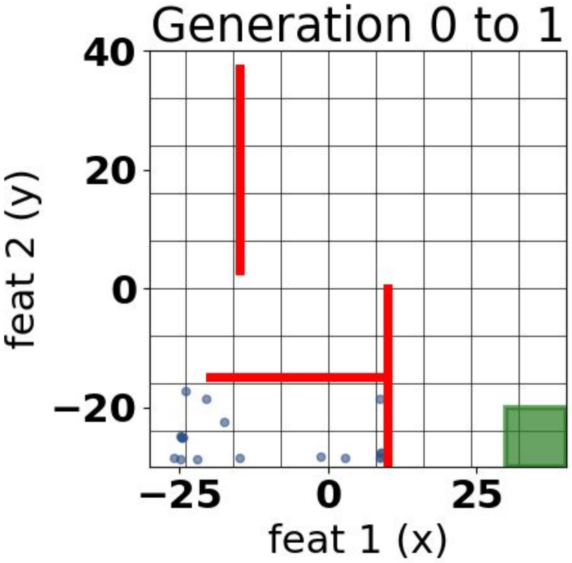
BC
Final (x, y)

Exploration: Ant Maze

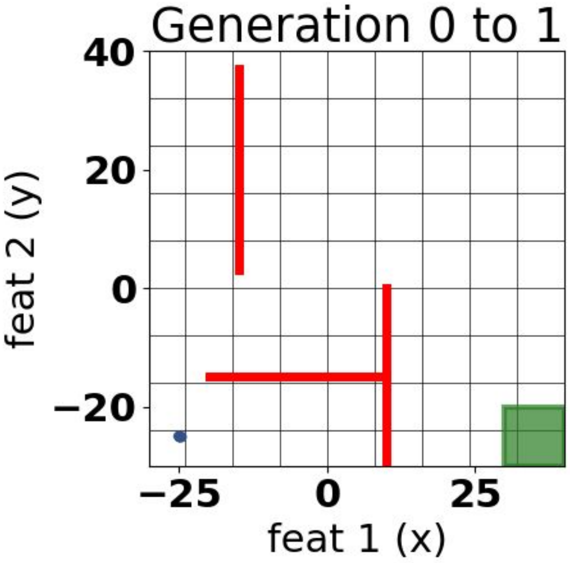
ME-ES exploit



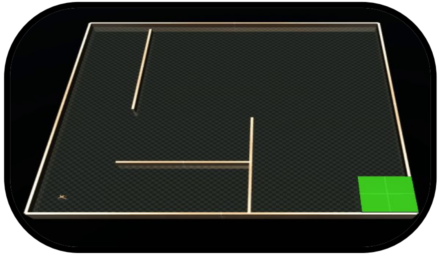
ME-ES explore-exploit



ME-ES explore



Exploration: Ant Maze

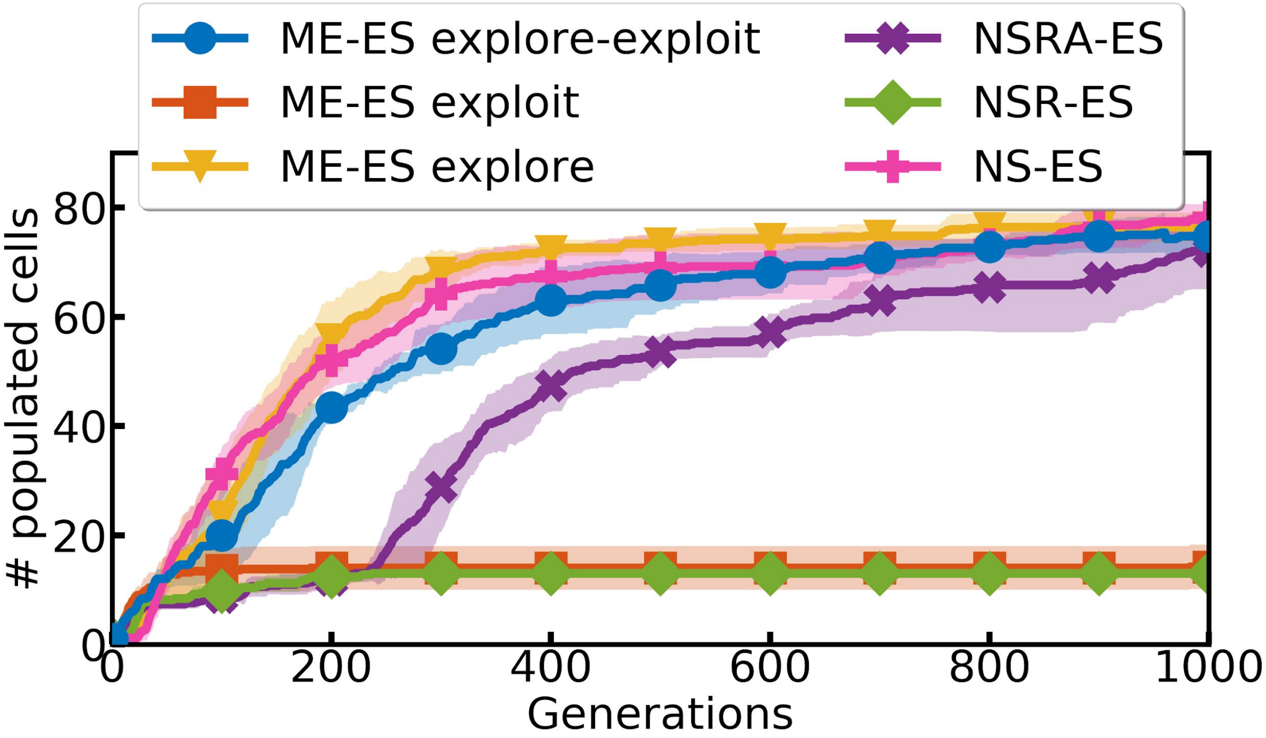


AntMaze

Reward
Euclidean distance to the goal

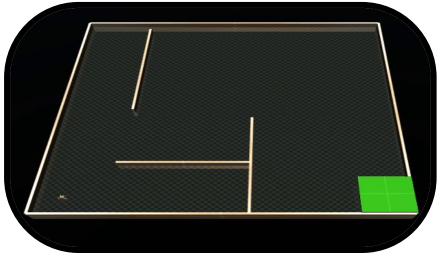
BC
Final (x, y)

Cell Coverage



Exploration: Ant Maze

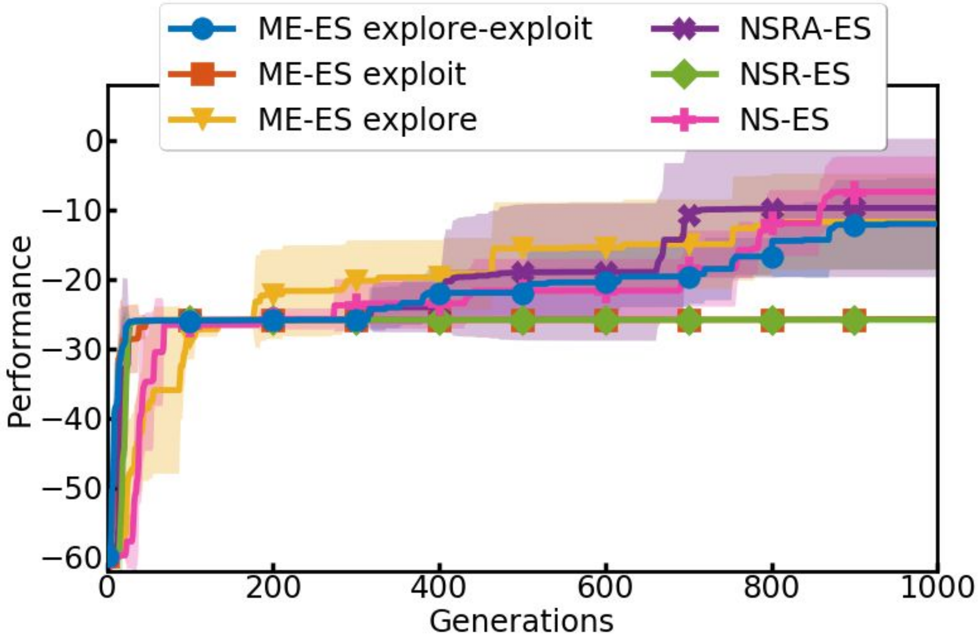
Best performance



AntMaze

Reward
Euclidean distance to the goal

BC
Final (x, y)



Discussion

Related work:

- Map-based Multi-Policy RL: an RL-based Map-Elites Kume et al. (2017)
- CMA-ME: parallel work using CMA-ES Fontaine et al. (GECCO 2020)

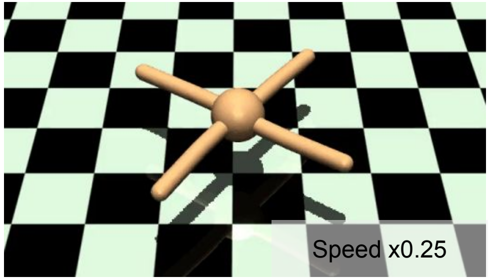
Future work:

- Automatize the exploration-exploitation tradeoff
- Sample reuse: reuse the offspring evaluations to compute many candidate child controllers with different objectives (e.g. novelty, performance, mixtures of these, evolvability etc).

MAP-Elites based on Evolution Strategies

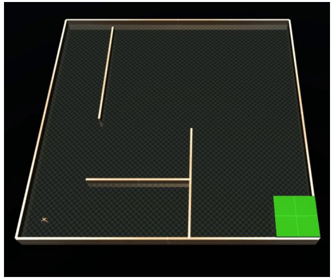
Build high-quality behavioral repertoires

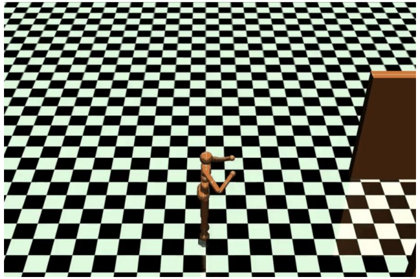
ES enables QD algorithms to be scaled to hard control tasks (Ant).
The archive can be used for damage adaptation.



Solve hard exploration problems

It decouples exploration and exploitation for efficient deep exploration and leverages ES to scale to hard control tasks (Humanoid, Ant)





GitHub uber-research/Map-Elites-Evolutionary

Exploration: Ant Maze

- Lehman, J., & Stanley, K. O. (2008). Exploiting open-endedness to solve problems through the search for novelty. In *ALIFE* (pp. 329-336).

- Lehman, J., & Stanley, K. O. (2011). Evolving a diversity of virtual creatures through novelty search and local competition. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation* (pp. 211-218). ACM.

- Cully, A., & Mouret, J. B. (2013). Behavioral repertoire learning in robotics. In *Proceedings of the 15th annual conference on Genetic and evolutionary computation* (pp. 175-182). ACM.

- Mouret, J. B., & Clune, J. (2015). Illuminating search spaces by mapping elites. *arXiv preprint arXiv:1504.04909*.

- Cully, A., Clune, J., Tarapore, D., & Mouret, J. B. (2015). Robots that can adapt like animals. *Nature*, *521*(7553), 503.

- Forestier, S., Mollard, Y., & Oudeyer, P. Y. (2017). Intrinsically motivated goal exploration processes with automatic curriculum learning. *arXiv preprint arXiv:1708.02190*.

- Kume, A., Matsumoto, E., Takahashi, K., Ko, W., & Tan, J. (2017). Map-based multi-policy reinforcement learning: enhancing adaptability of robots by deep reinforcement learning. *arXiv preprint arXiv:1710.06117*.

- Salimans, T., Ho, J., Chen, X., Sidor, S., & Sutskever, I. (2017). Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint arXiv:1703.03864*.

- Such, F. P., Madhavan, V., Conti, E., Lehman, J., Stanley, K. O., & Clune, J. (2017). Deep neuroevolution: Genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning. *arXiv preprint arXiv:1712.06567*.

- Conti, E., Madhavan, V., Such, F. P., Lehman, J., Stanley, K., & Clune, J. (2018). Improving exploration in evolution strategies for deep reinforcement learning via a population of novelty-seeking agents. In *Advances in Neural Information Processing Systems*(pp. 5027-5038).

- Bellemare, M. G., Naddaf, Y., Veness, J., & Bowling, M. (2013). The arcade learning environment: An evaluation platform for general agents. Journal of Artificial Intelligence Research, 47, 253-279.

- Tassa, Y., Doron, Y., Muldal, A., Erez, T., Li, Y., Casas, D. D. L., ... & Lillicrap, T. (2018). Deepmind control suite. arXiv preprint arXiv:1801.00690.

- Baranes, A., & Oudeyer, P. Y. (2009). R-iac: Robust intrinsically motivated exploration and active learning. IEEE Transactions on Autonomous Mental Development, 1(3), 155-169.

- Fontaine, M. C., Togelius, J., Nikolaidis, S., & Hoover, A. K. (2019). Covariance Matrix Adaptation for the Rapid Illumination of Behavior Space. arXiv preprint arXiv:1912.02400.

- Kume, A., Matsumoto, E., Takahashi, K., Ko, W., & Tan, J. (2017). Map-based multi-policy reinforcement learning: enhancing adaptability of robots by deep reinforcement learning. arXiv preprint arXiv:1710.06117.