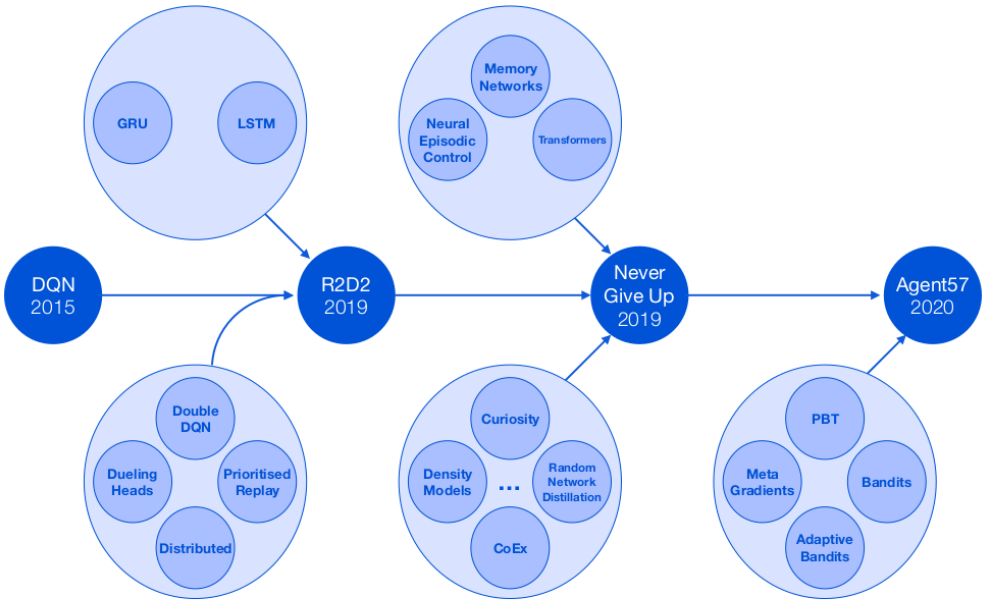
Beyond (the) RAINBOW

Deep RL research after DQN

Olivier Sigaud
(building on a talk from Stéphane Doncieux)

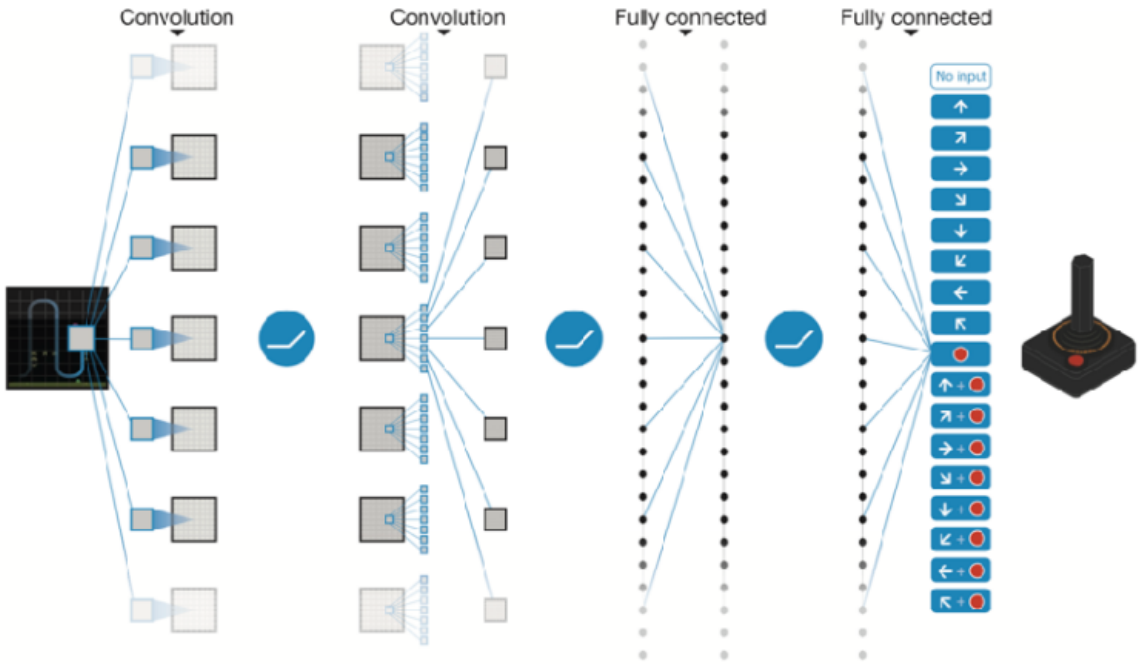
Sorbonne Université
http://people.isir.upmc.fr/sigaud

Introduction



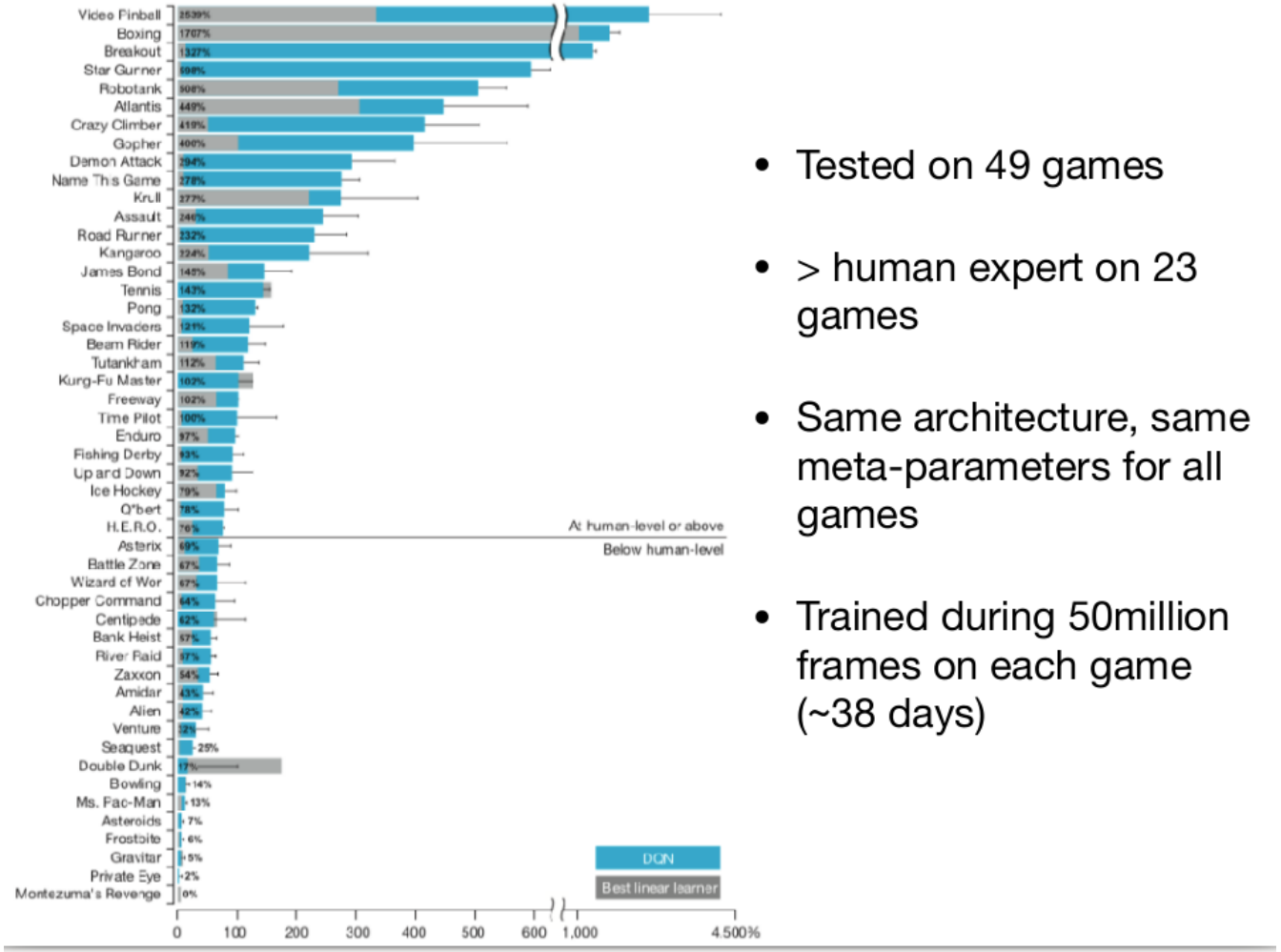
- DQN solved 23/57 ATARI games better than humans
- AGENT57 solved them all better
- The Deepmind story: from DQN to AGENT57

DQN in ATARI



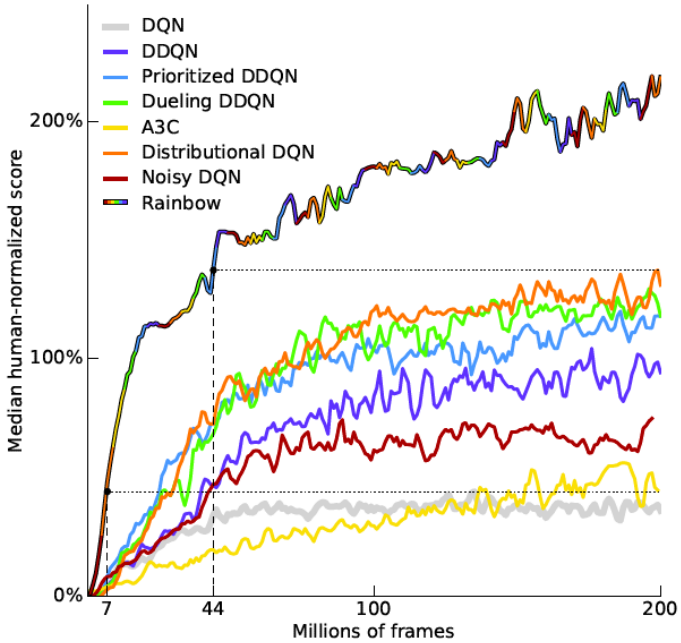
Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015) Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533.

DQN in ATARI: results



- Tested on 49 games
- > human expert on 23 games
- Same architecture, same meta-parameters for all games
- Trained during 50million frames on each game (~38 days)

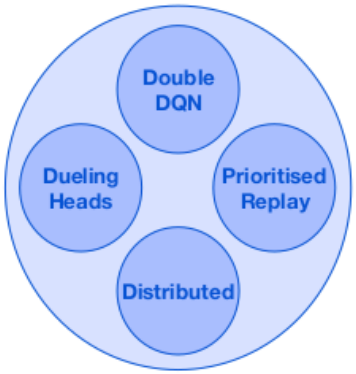
RAINBOW

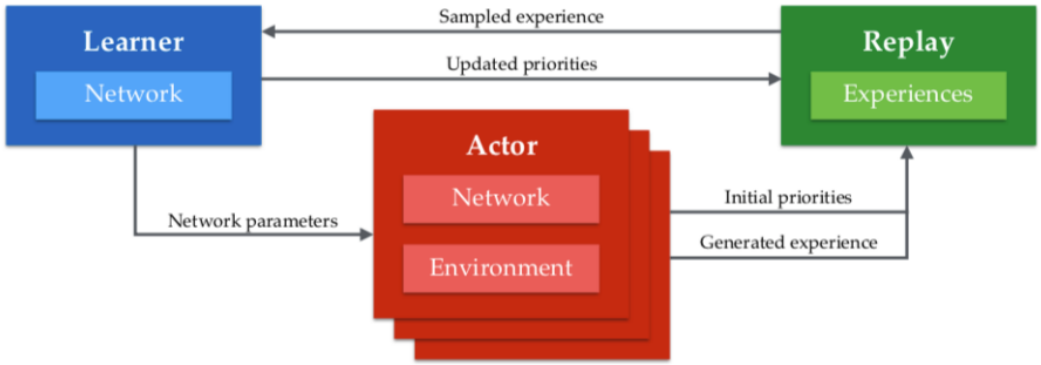


- A combination of several improvements

Hessel, M., Modayil, J., Van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., Horgan, D., Piot, B., Azar, M., & Silver, D. (2017) Rainbow: Combining improvements in deep reinforcement learning. *arXiv preprint arXiv:1710.02298*

From RAINBOW to APE-X

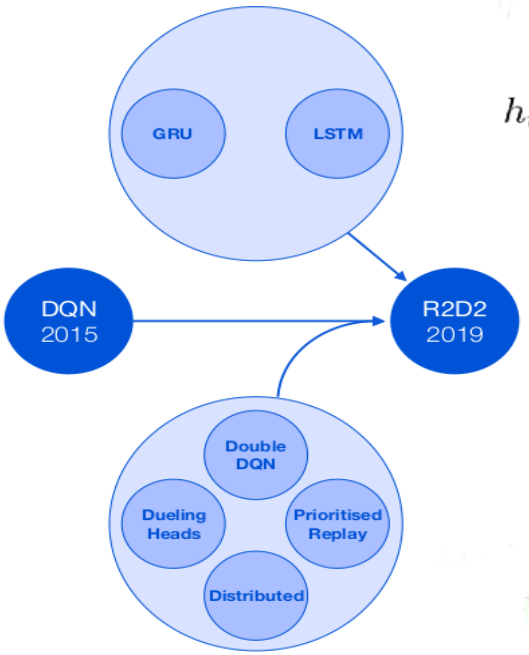


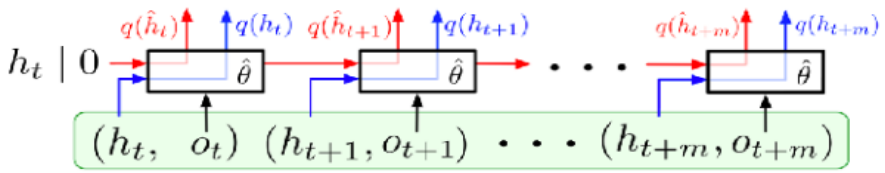


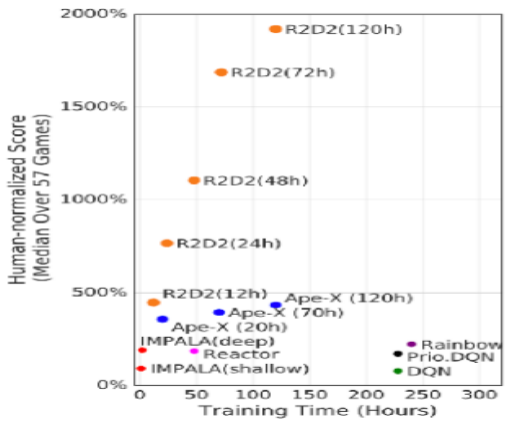
- Drops Distributional RL
- Adds Distributed Prioritized Experience Replay
- Time of focus on distributed architectures: Gorila, Impala, ...

Horgan, D., Quan, J., Budden, D., Barth-Maron, G., Hessel, M., van Hasselt, H., and Silver, D. (2018) Distributed prioritized experience replay. In *6th International Conference on Learning Representations*, ICLR 2018. OpenReview.net, URL https://openreview.net/forum?id=H1Dy---0Z

From APE-X to R2D2



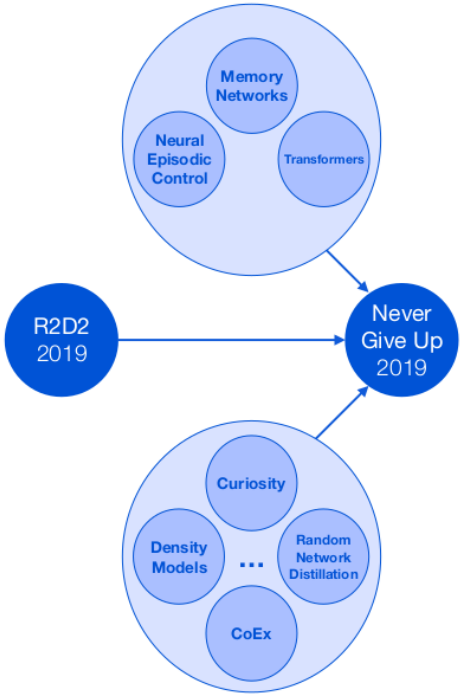




- Adds some short term memory (LSTM or GRU)

Kapturowski, S., Ostrovski, G., Quan, J., Munos, R., and Dabney, W. (2019 Recurrent experience replay in distributed reinforcement learning. In *International conference on learning representations*

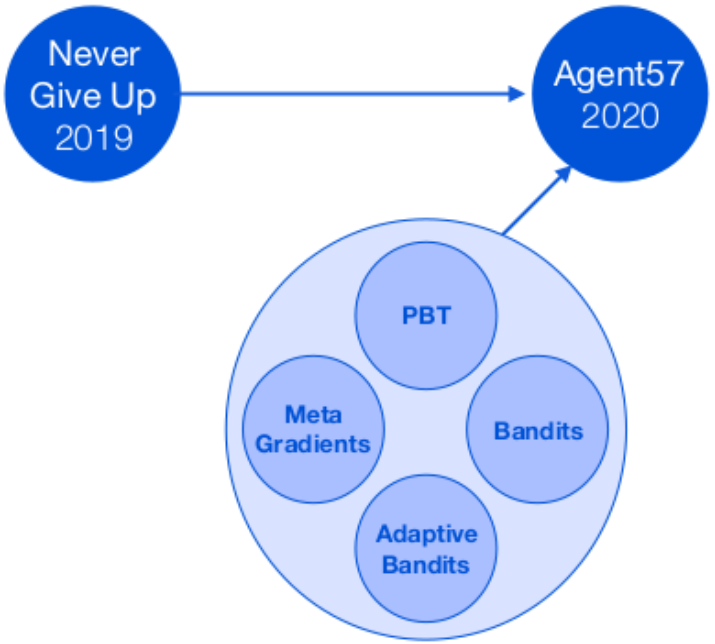
From R2D2 to NGU



- Never Give Up, two components:
  - More memory, using transformers, memory networks, neural episodic control
  - More exploration, using Random Network Distillation, and curiosity as an intrinsic motivation

Badia, A. P., Sprechmann, P., Vitvitskyi, A., Guo, D., Piot, B., Kapturowski, S., Tieleman, O., Arjovsky, M., Pritzel, A., Bolt, A., et al. (2020) Never give up: Learning directed exploration strategies. *arXiv preprint arXiv:2002.06038*

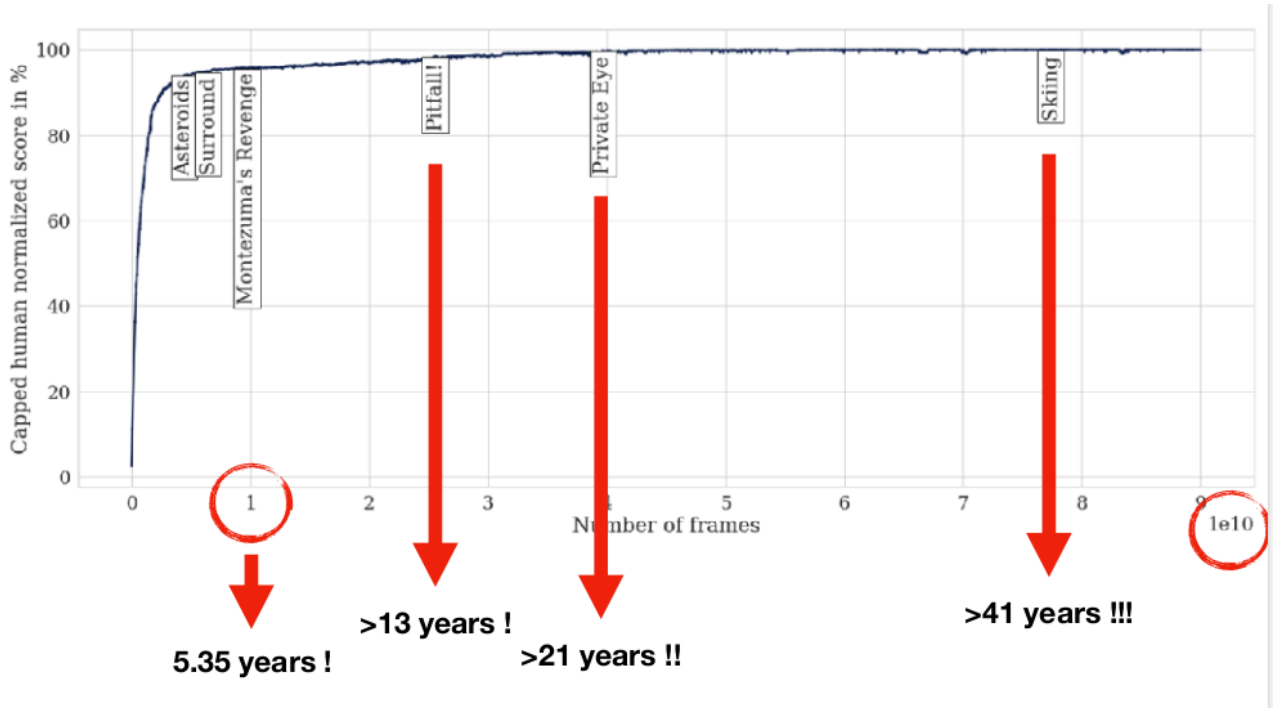
Finally AGENT57



- Additional components:
  - Population-based training (as AlphaStar)
  - Dynamic adjustement of exploration and discount factors (β and γ)
  - Larger window for BPTT (80 steps $\rightarrow$ 160)

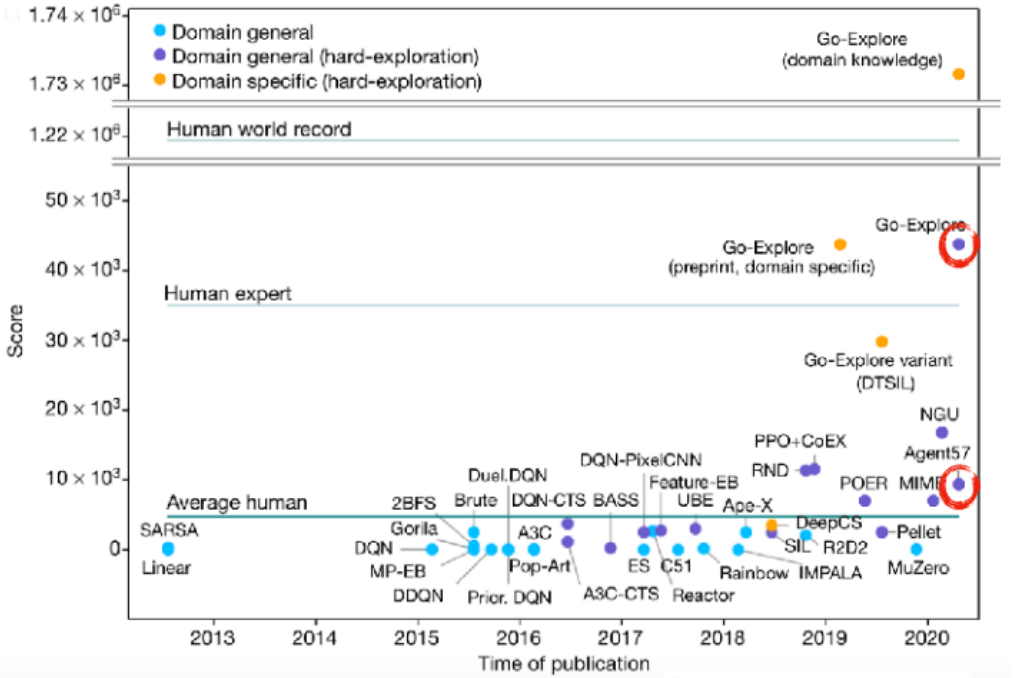
Badia, A. P., Piot, B., Kapturowski, S., Sprechmann, P., Vitvitskyi, A., Guo, D., and Blundell, C. (2020) Agent57: Outperforming the atari human benchmark. In *International Conference on Machine Learning*, pp. 507–517. PMLR

Main results



- Pitfall!, PrivateEye and Skiing were unsolved games for RL
- But humans solve them much, much faster

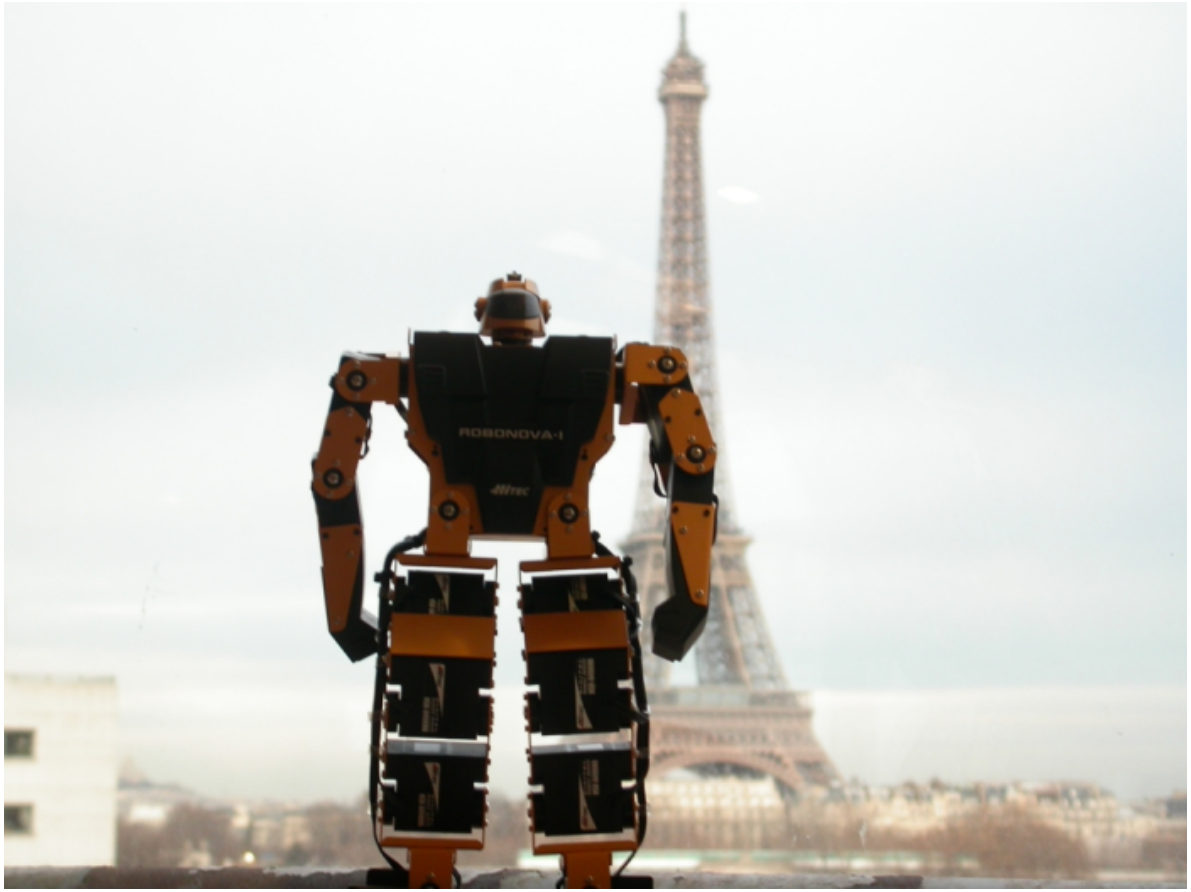
Comparison to Go-Explore



- Focus on Montezuma's Revenge (infamous Hard Exploration Problem)
- Need for very long horizon trajectories, stepping stones, skill chaining...

Ecoffet, A., Huizinga, J., Lehman, J., Stanley, K. O., and Clune, J. (2021) First return, then explore. *Nature*, 590(7847):580–586, 2021

Any question?



Send mail to: Olivier.Sigaud@upmc.fr

Badia, A. P., Piot, B., Kapturowski, S., Sprechmann, P., Vitvitskyi, A., Guo, Z. D., and Blundell, C.

Agent57: Outperforming the atari human benchmark.
In *International Conference on Machine Learning*, pp. 507–517. PMLR, 2020a.

Badia, A. P., Sprechmann, P., Vitvitskyi, A., Guo, D., Piot, B., Kapturowski, S., Tieleman, O., Arjovsky, M., Pritzel, A., Bolt, A., et al.

Never give up: Learning directed exploration strategies.
arXiv preprint arXiv:2002.06038, 2020b.

Ecoffet, A., Huizinga, J., Lehman, J., Stanley, K. O., and Clune, J.

First return, then explore.
Nature, 590(7847):580–586, 2021.

Hessel, M., Modayil, J., van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., Horgan, D., Piot, B., Azar, M. G., and Silver, D.

Rainbow: Combining improvements in deep reinforcement learning.
In McIlraith, S. A. and Weinberger, K. Q. (eds.), *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 3215–3222. AAAI Press, 2018.
URL https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17204.

Horgan, D., Quan, J., Budden, D., Barth-Maron, G., Hessel, M., van Hasselt, H., and Silver, D.

Distributed prioritized experience replay.
In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018.
URL https://openreview.net/forum?id=H1Dy---0Z.

Kapturowski, S., Ostrovski, G., Quan, J., Munos, R., and Dabney, W.

Recurrent experience replay in distributed reinforcement learning.
In *International conference on learning representations*, 2019.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al.

Human-level control through deep reinforcement learning.

Nature, 518(7540):529–533, 2015.