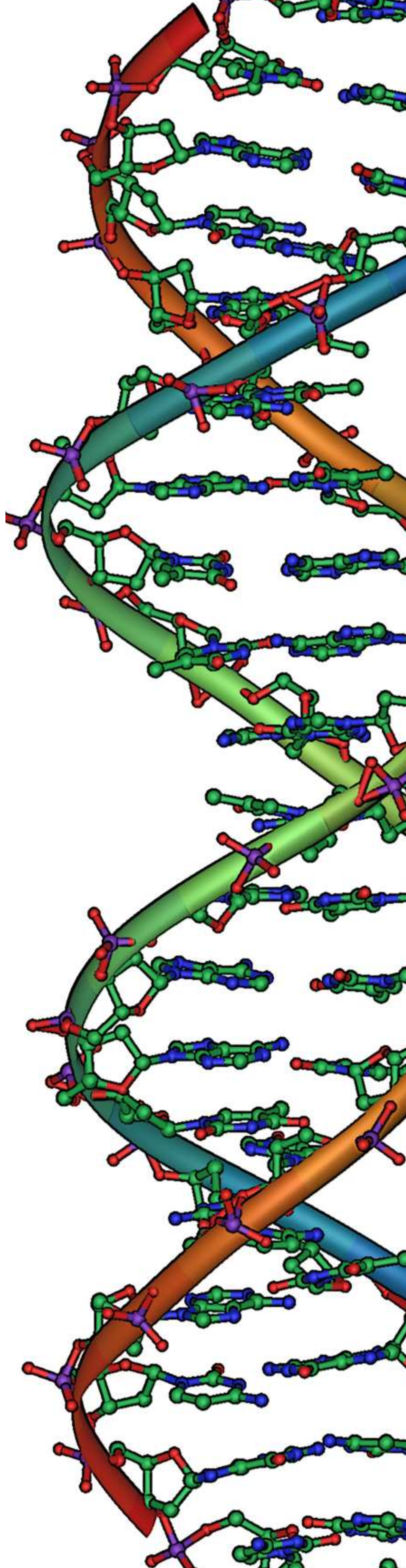
Evolutionary Reinforcement Learning

Jean-Baptiste Mouret

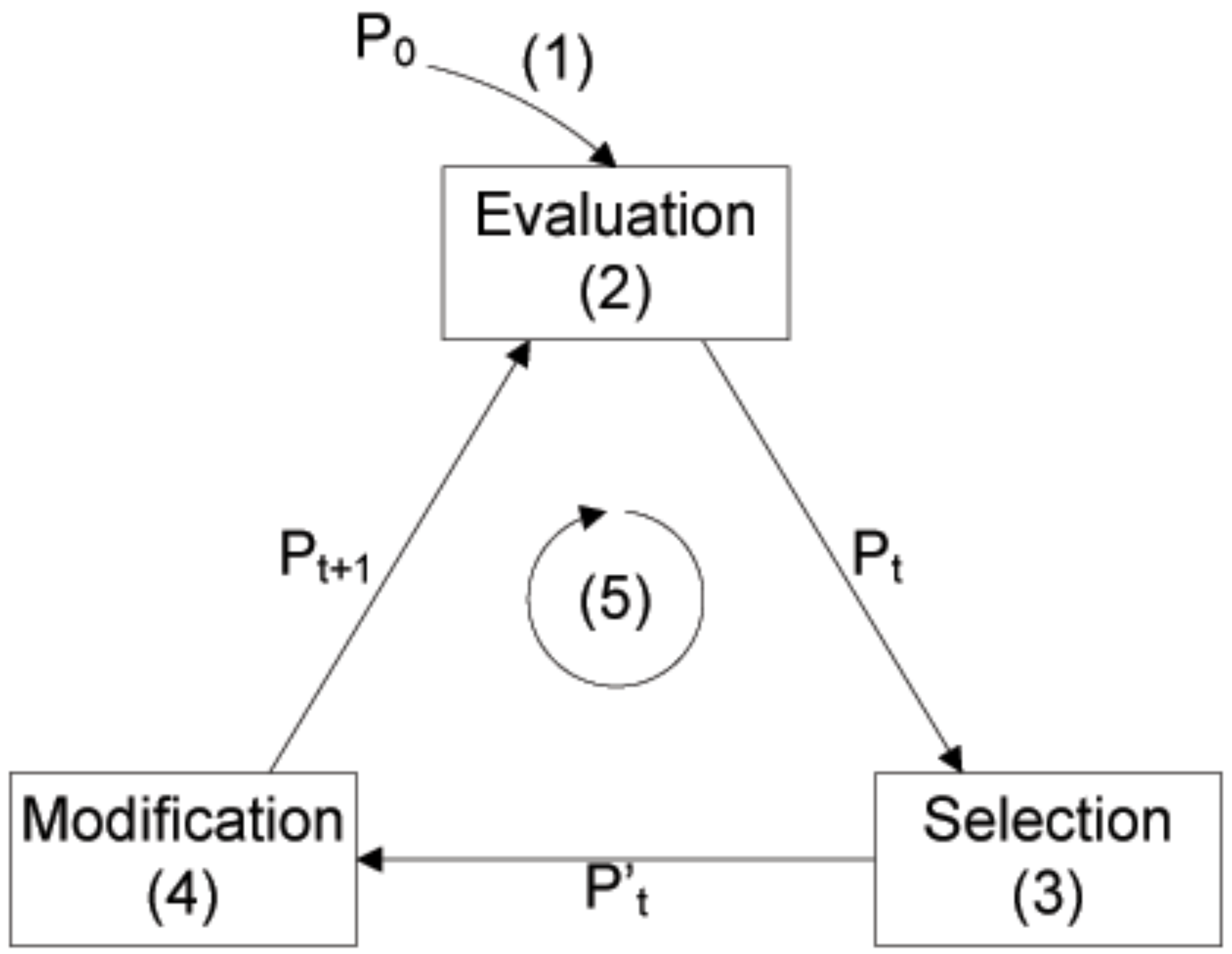
Dennis G. Wilson

@jbmouret @jb_mouret

@d9w @digiwilson



Evolution on computers



- Evolution is one of the most fundamental “algorithms” of life
- It is **iterative** (generation by generation)
- It works by **trial-and-error**
  - random mutation = exploration
  - fitness = reward

- Evolution solves a trial-and-error learning problem! like reinforcement learning
  - another metaphor (give other ideas / challenges)
  - a long history in computing (at least as long as learning)

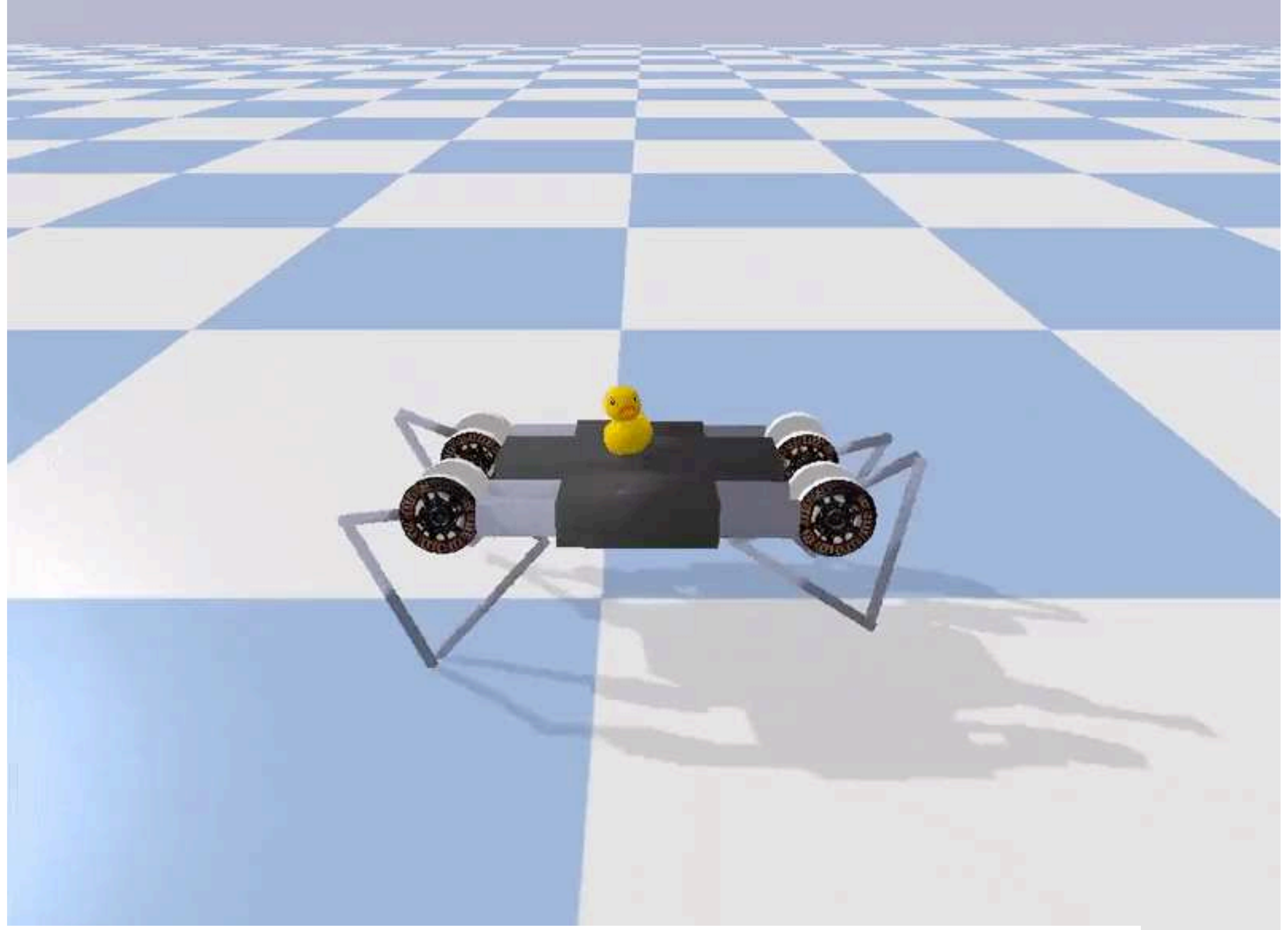
Evolution can optimize policies

like PPO/TRPO/etc.

Parameters of the policy

Optimize: $$J(\boldsymbol{\theta}) = \mathbb{E}\left[\sum_{t=1}^{T} r(\mathbf{x}_t) \,\middle|\, \boldsymbol{\theta}\right]$$

Reward for state $\mathbf{x}_t$

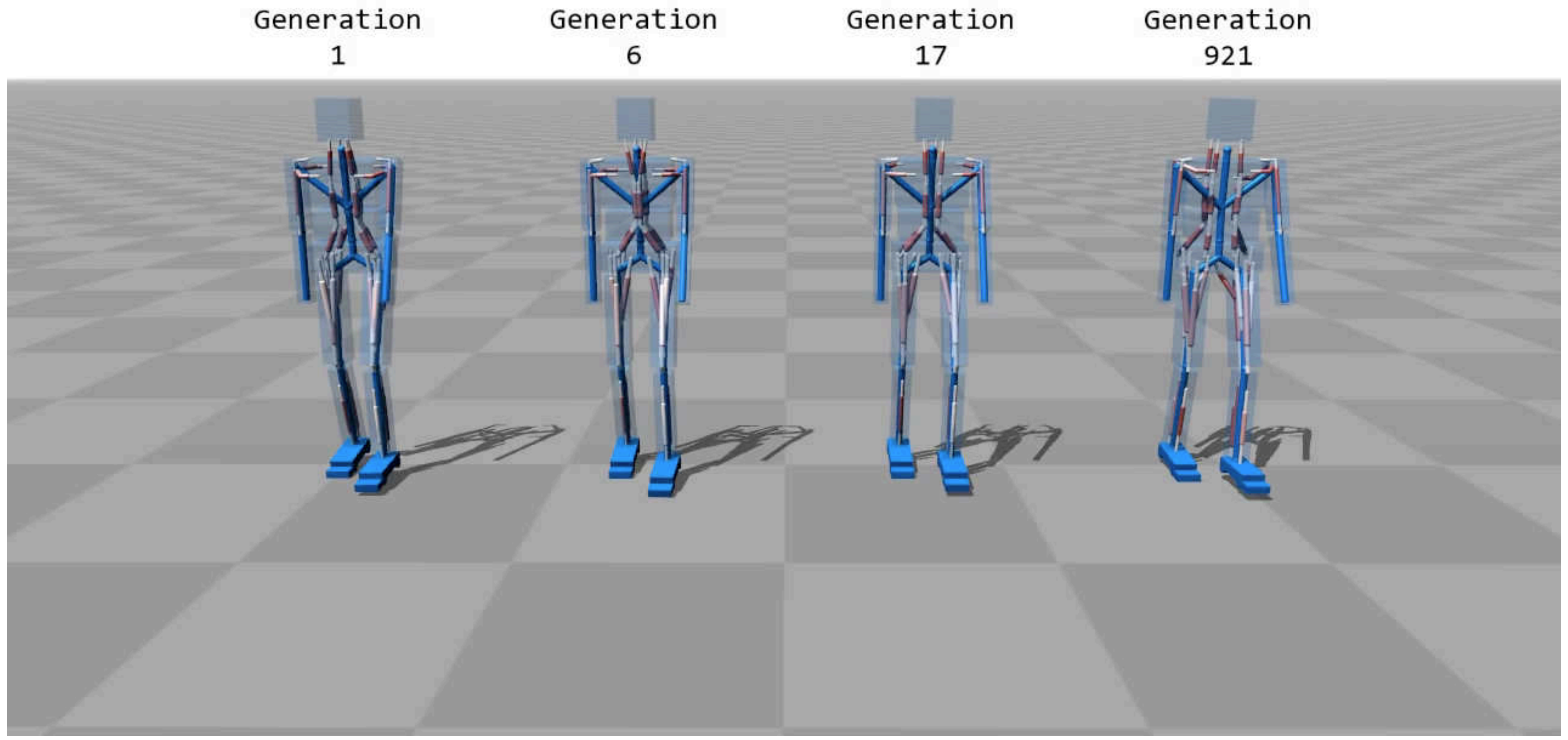


https://blog.otoro.net/2017/11/12/evolving-stable-strategies/

- Black-box optimizer (no gradient)
- Multi-objective (Pareto-based) optimization
- Scale to parallel computers easily

- Holistic view (discard intermediate steps)
  - no credit assignment problem
  - no problem with large (continuous) states

… but discard a lot of useful data



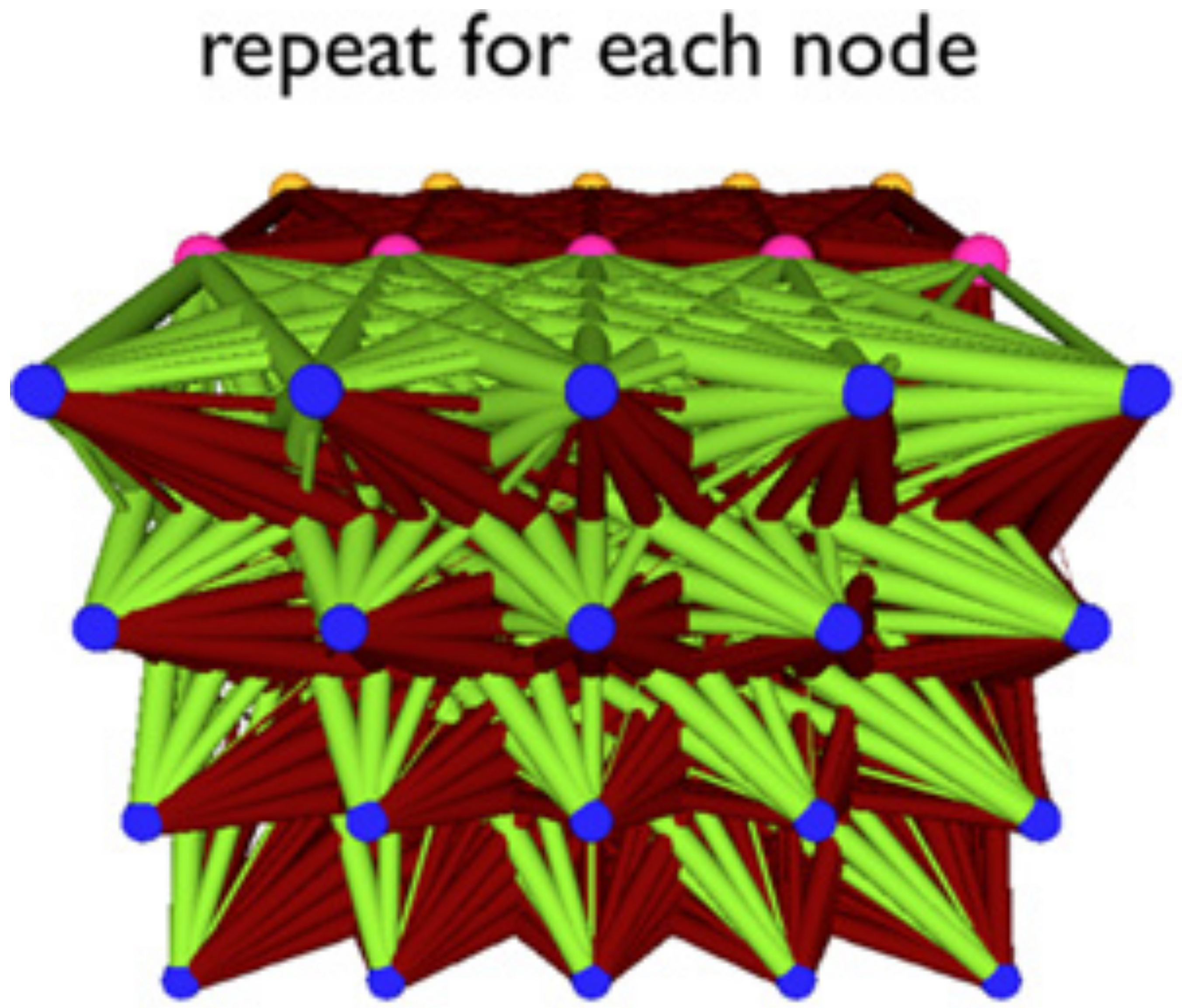
Geijtenbeek, T., van de Panne, M., & van der Stappen, A. F. (2013). "*Flexible muscle-based locomotion for bipedal creatures*". ACM Transactions on Graphics (TOG), 32(6), 206.

Evolution can optimize structures

neural architecture search, morphology

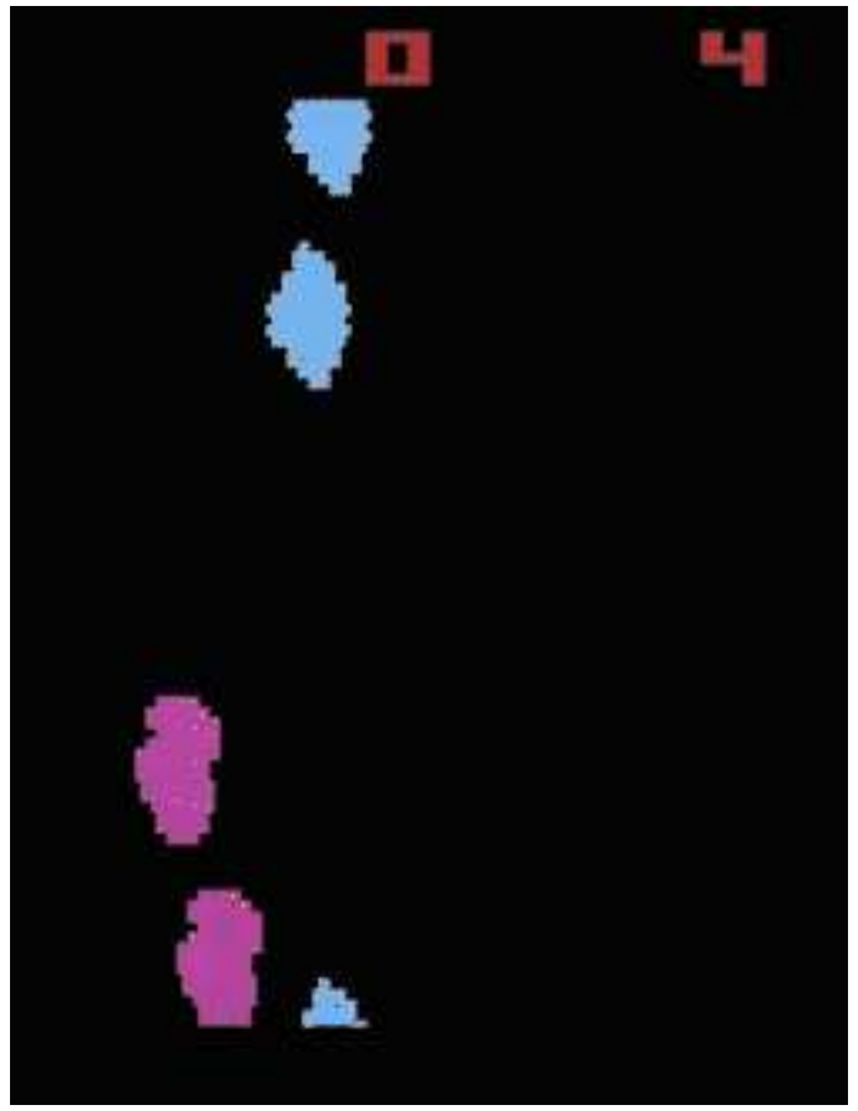
A question that was studied for a very long time in evolution!

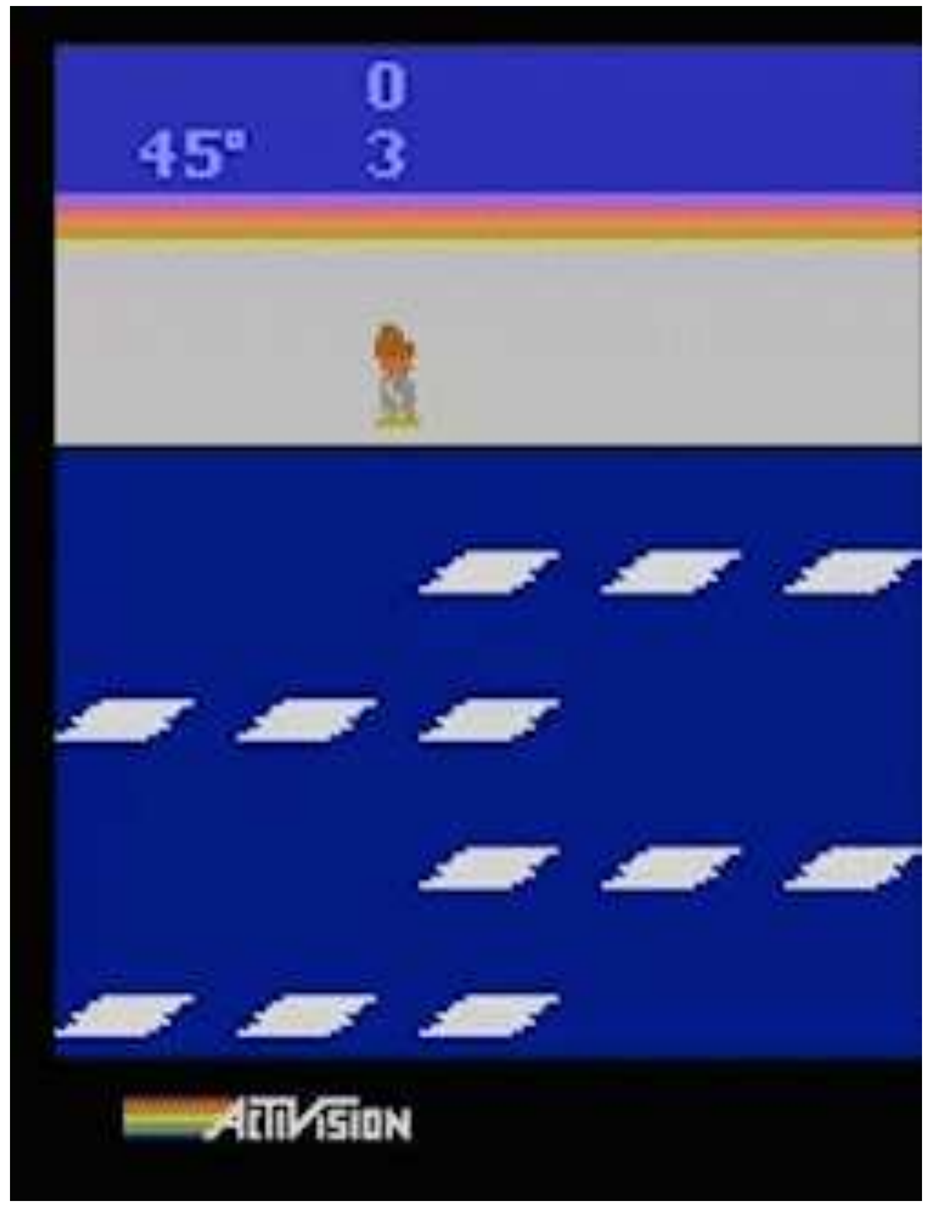
Sims, Karl. (1994) “Evolving 3D morphology and behavior by competition."
Artificial life 1.4 (1994): 353-372.



Clune J, Stanley KO, Pennock RT, Ofria C. (2011) On the performance of indirect encoding across the continuum of regularity. *IEEE Transactions on Evolutionary Computation.* 2011 Jun;15(3):346-67.

Evolution can be competitive for deep RL tasks





	DQN	ES	A3C	RS	GA	GA
Frames	200M	1B	1B	1B	1B	6B
Time	~7-10d	~ 1h	~ 4d	~ 1h or 4h	~ 1h or 4h	~ 6h or 24h
Forward Passes	450M	250M	250M	250M	250M	1.5B
Backward Passes	400M	0	250M	0	0	0
Operations	1.25B U	250M U	1B U	250M U	250M U	1.5B U
amidar	**978**	112	264	143	263	377
assault	4,280	1,674	**5,475**	649	714	814
asterix	4,359	1,440	**22,140**	1,197	1,850	2,255
asteroids	1,365	1,562	**4,475**	1,307	1,661	2,700
atlantis	279,987	**1,267,410**	911,091	26,371	76,273	129,167
enduro	**729**	95	-82	36	60	80
frostbite	797	370	191	1,164	**4,536**	**6,220**
gravitar	473	**805**	304	431	476	764
kangaroo	7,259	**11,200**	94	1,099	3,790	**11,254**
seaquest	**5,861**	1,390	2,355	503	798	850
skiing	-13,062	-15,443	-10,911	-7,679	†**-6,502**	†**-5,541**
venture	163	760	23	488	**969**	†**1,422**
zaxxon	5,363	6,380	**24,622**	2,538	6,180	7,864

Such FP, Madhavan V, Conti E, Lehman J, Stanley KO, Clune J. Deep neuroevolution: Genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning. arXiv preprint arXiv:1712.06567. **2017**.

Outline

1. **Evolutionary Strategies**
   from (μ,λ)-ES to CMA-ES
2. **Pareto-based Multi-objective evolutionary algorithms**
   NSGA-II
3. **Neuroevolution**
   from NEAT to HyperNEAT
4. **Beyond the fitness function**
   from novelty search to quality diversity

Keynote: Sebastian Risi (IT University of Copenhagen)
Evolving Agents that Learn More Like Animals

Choices:

- ➡ *introduce the “classic algorithms” of the field (not always the most competitive)*
- ➡ *present a few algorithms “in depth” (we will not be exhaustive at all)*
- ➡ *In many cases, a notebook makes it possible to reproduce the examples*