

Agent-centric learning: from external reward maximization to internal knowledge curation

Hanqi Zhou^{1,2,3,4}, Fryderyk Mantiuk¹, David G. Nagy^{1,2,4}, Charley M. Wu^{1,2,4,5,6}

¹ Human and Machine Cognition Lab, University of Tübingen, Tübingen, Germany ² Max Planck Institute for Biological Cybernetics ³ IMPRS-IS ⁴ Tübingen AI Center ⁵ TU Darmstadt ⁶ Hessian.AI

The challenges of task- and environment-centric learning

We propose

- representational empowerment
- Away from external objectives: control over its environments or mastery of specific tasks
- Towards the capacity to shape one's own knowledge, achieving better **preparedness** distinct from direct environmental influence.

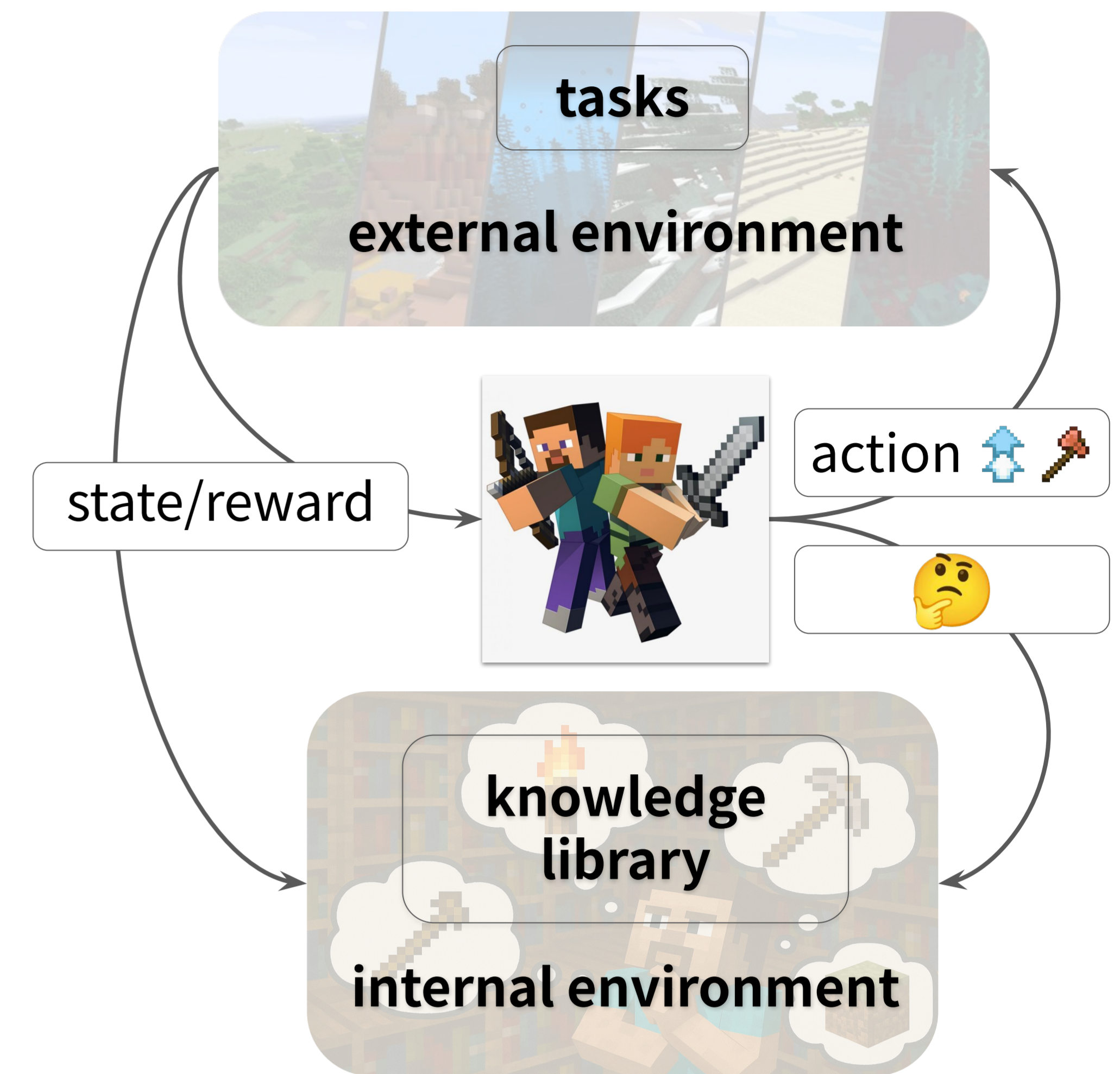
Task-centric extrinsic reward

Narrow tasks + fixed rewards → brittle agents that know what to do only when we tell them.

Environment-centric Intrinsic motivation

Curiosity (novelty, information gain, etc.) still ties learning to this environment, so agents over-fit the world they were born in.

To meet truly unforeseen challenges we must shift the locus of learning inward, from task-centric to agent-centric.



Controllability as preparedness for future tasks

Two perspectives: endless adaptation & instrumental convergence

Challenge: Since we cannot simulate all possible worlds, the current learning paradigm makes it hard to achieve broadly applicable intelligence -- a high level of preparedness for unforeseen challenges.

Empowerment

Where in this environment should I go to maximize my future influence?

$$\text{EnvEmp}(s) = \max_{\pi(a_{1:T})} I(s'; a_{1:T} | s)$$

$H(s' | s)$ diversity

$-H(s' | s, a_{1:T})$ controllability

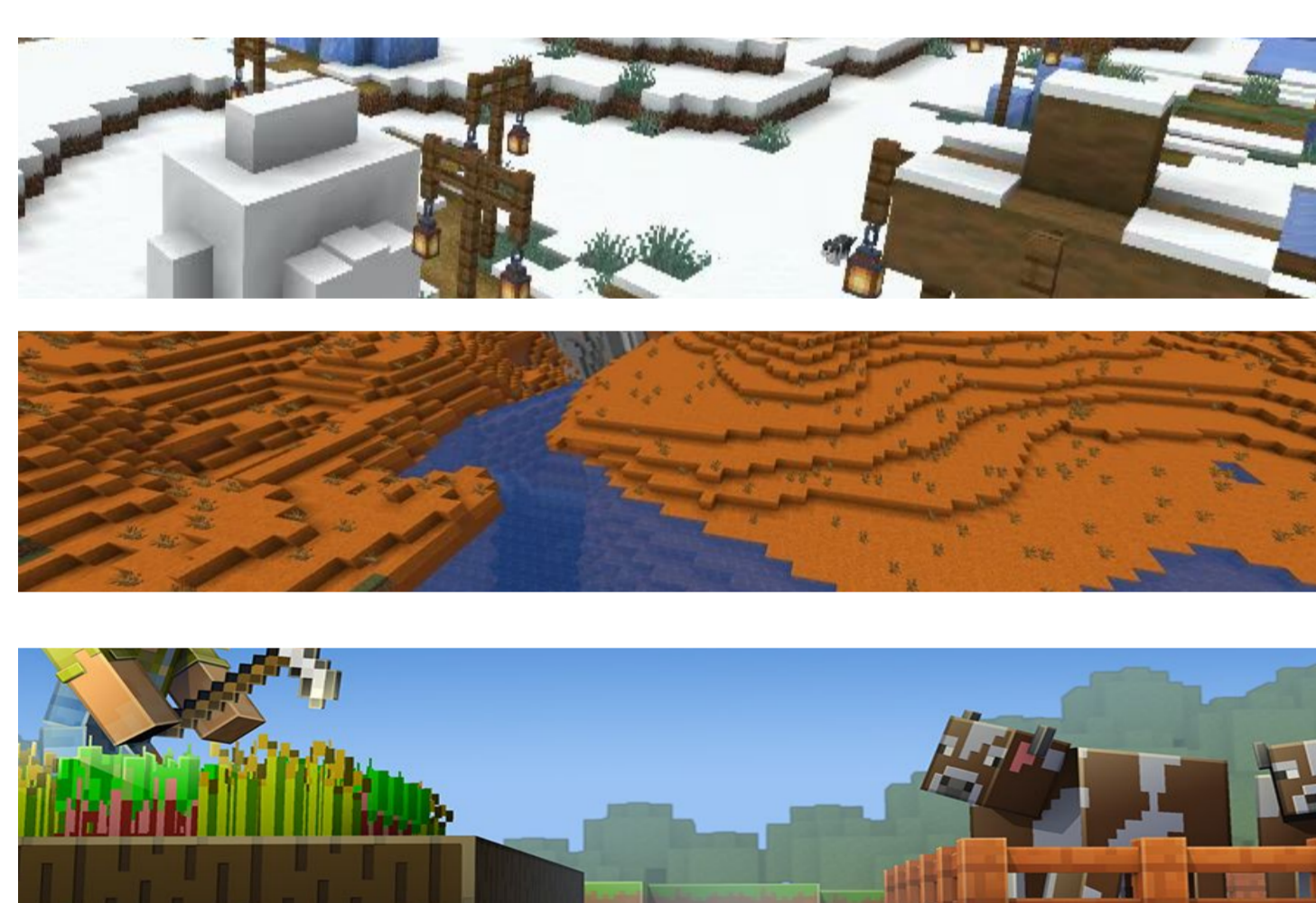
Representational empowerment

How should I structure my knowledge to be best prepared for any future challenge?

$$\text{RepEmp}(Z_k) = \max_{\omega_k \in \Omega^T} I(Z'_k; \omega_k | Z_k)$$

$H(Z'_k | Z_k)$ diverse candidates as starting point

$-H(Z'_k | Z_k, \omega_k)$ regularization on overly flexible knowledge



Seasonal goals

Tundra ❄️: Harvest fish with stone axe before blizzard

Badlands 🏜️: Mine terracotta with stone pickaxe

Farms 🌾: Till soil with wood hoe, plant wheat

Action a

- Step(n), Turn(dir), Sleep($time$), Place($tool$), Pickup($tool$)

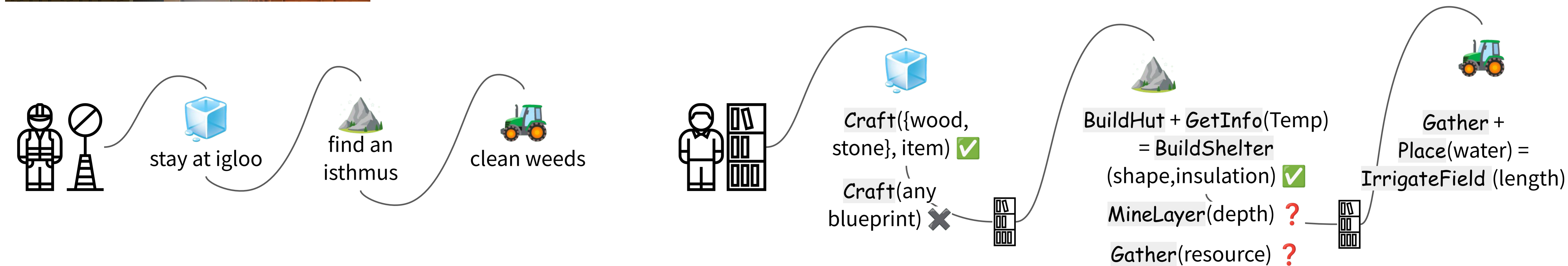
Knowledge library $\mathcal{L} = \{Z_1, Z_2, \dots\}$

- GetInfo(weather / animal)

- Gather(wood), Craft(wood, item), BuildHut(wall + roof)

Operations ω

- Specialize, Compose, Abstract, Delete

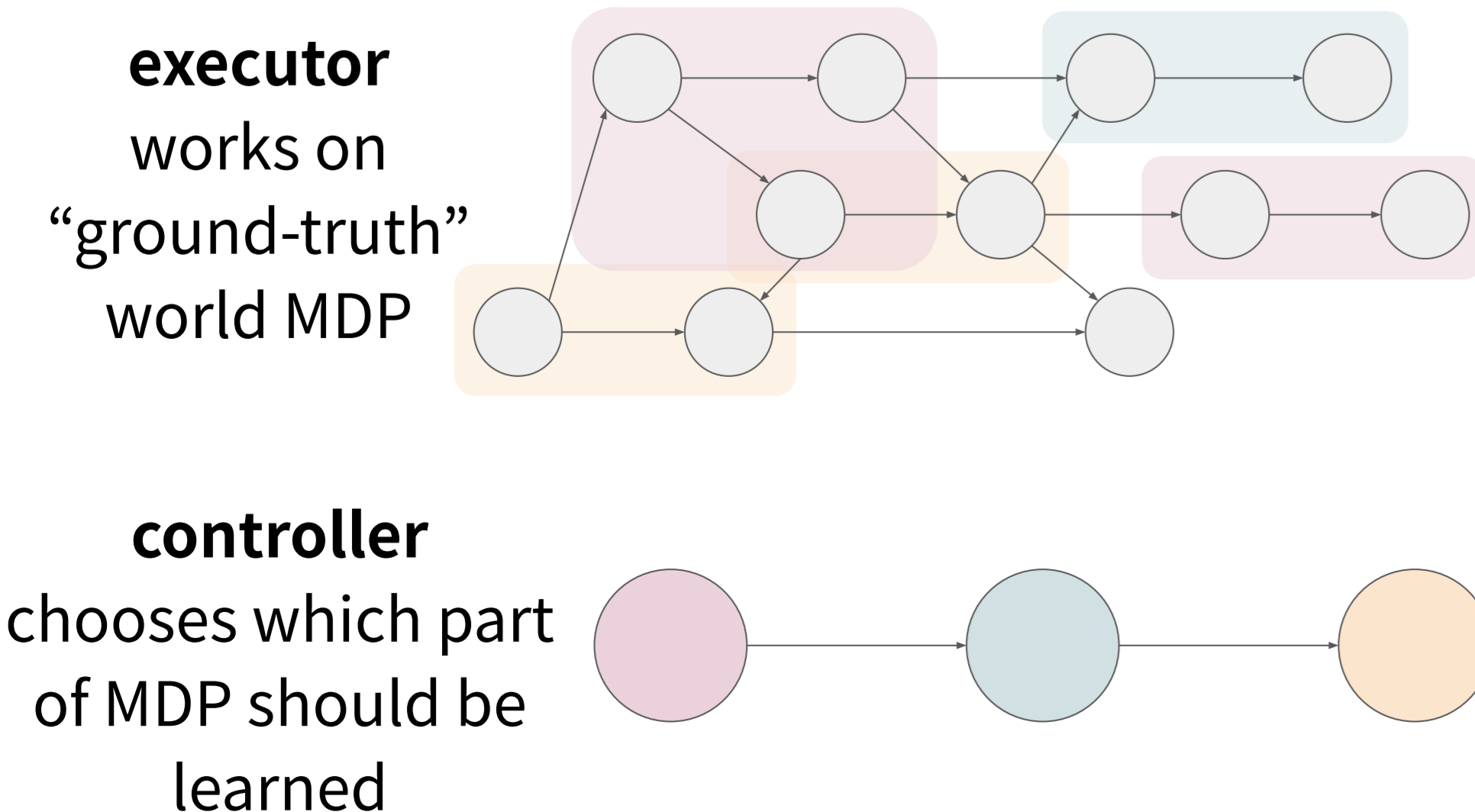


Limitation: When the environment changes, previous high-empowerment states might become useless.

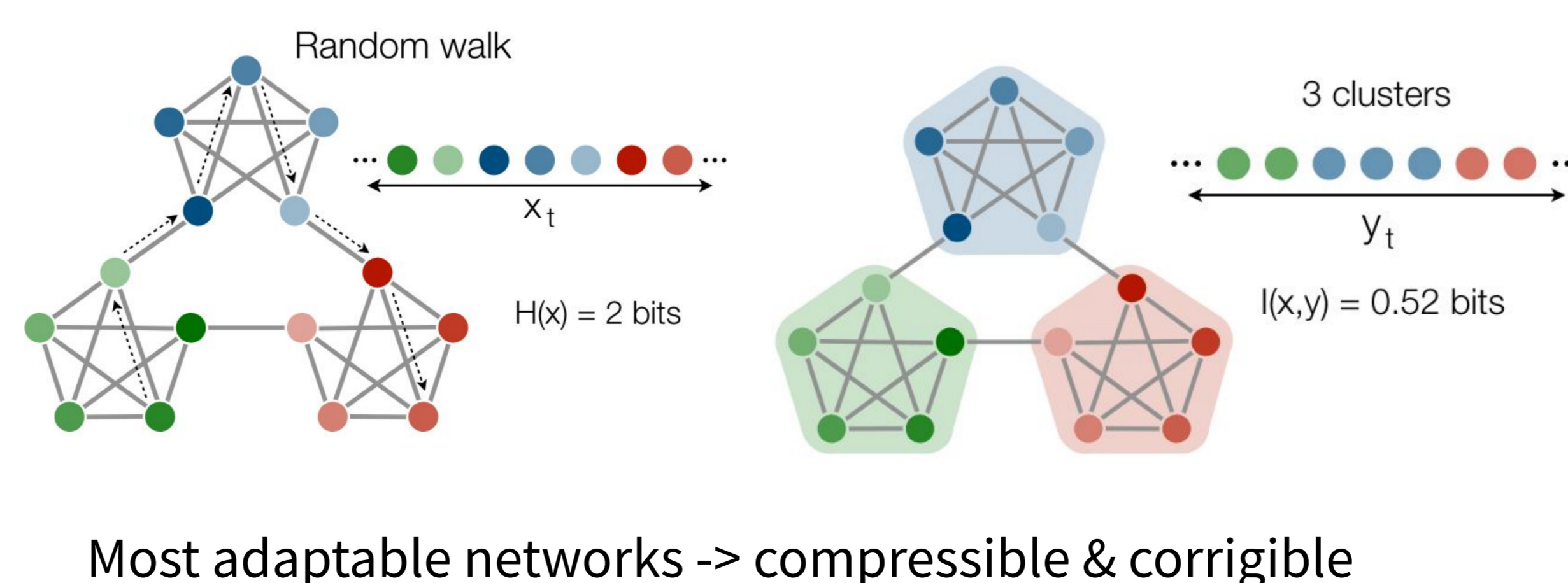
Limitation: Dependent on the choice of primitives, the kernel for the representation space, and the approximation of emp calculation.

Different representation forms

Hierarchical / Bayesian RL



Graph / Network



Most adaptable networks → compressible & corrigible

Future directions

- Multi-agent communication
- Curriculum design in pedagogy

[1] Singh, S., Lewis, R. L., & Barto, A. G. (2009, August). Where do rewards come from. In Proceedings of the annual conference of the cognitive science society (pp. 2601-2606). Cognitive Science Society.
 [2] Lynn, C. W., & Bassett, D. S. (2021). Quantifying the compressibility of complex networks. Proceedings of the National Academy of Sciences, 118(32), e2023473118.