Artifacts as Memory Beyond the Agent Boundary

John D. Martin

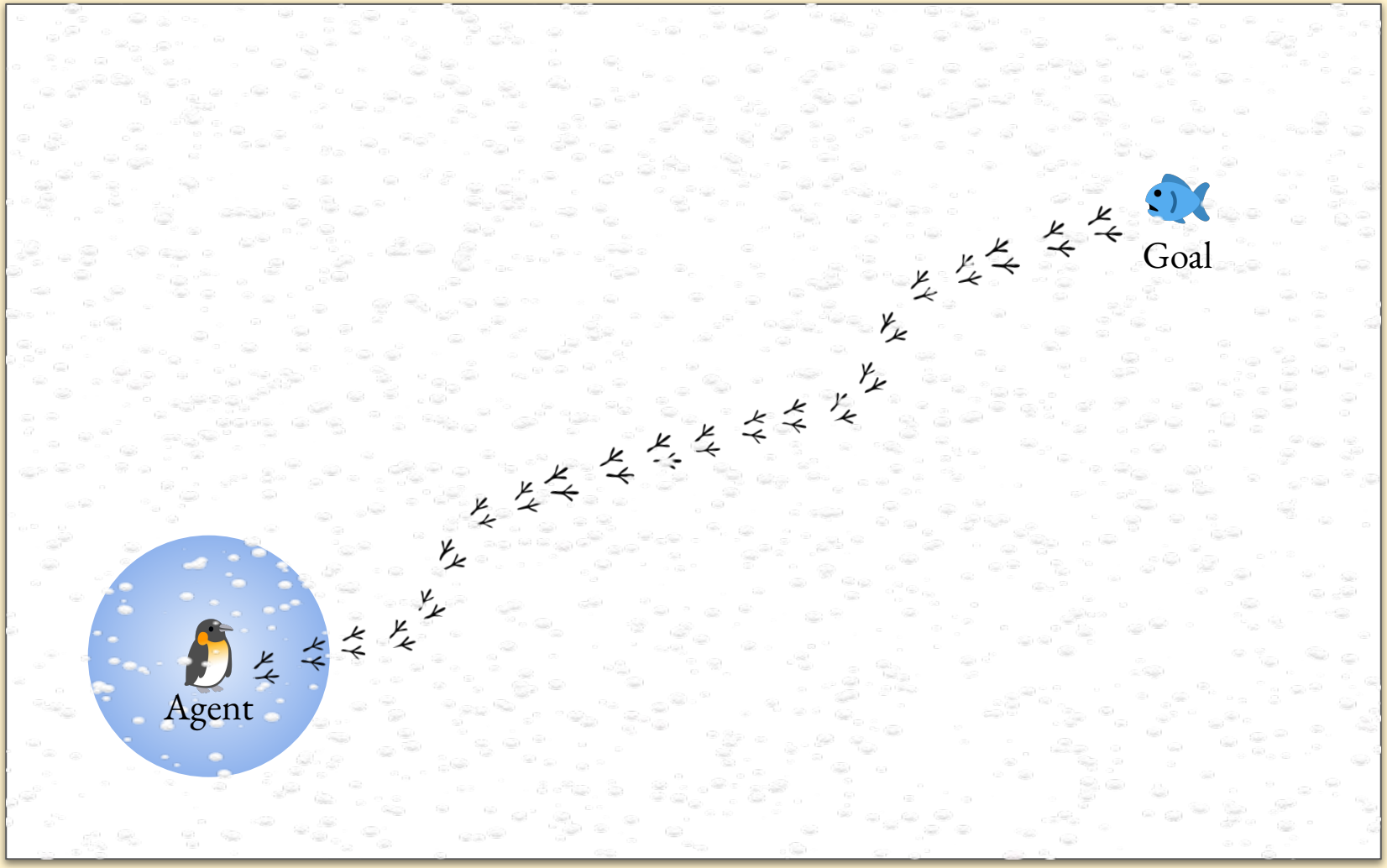
Research Fellow & Adjunct Professor

Openmind RESEARCH INSTITUTE

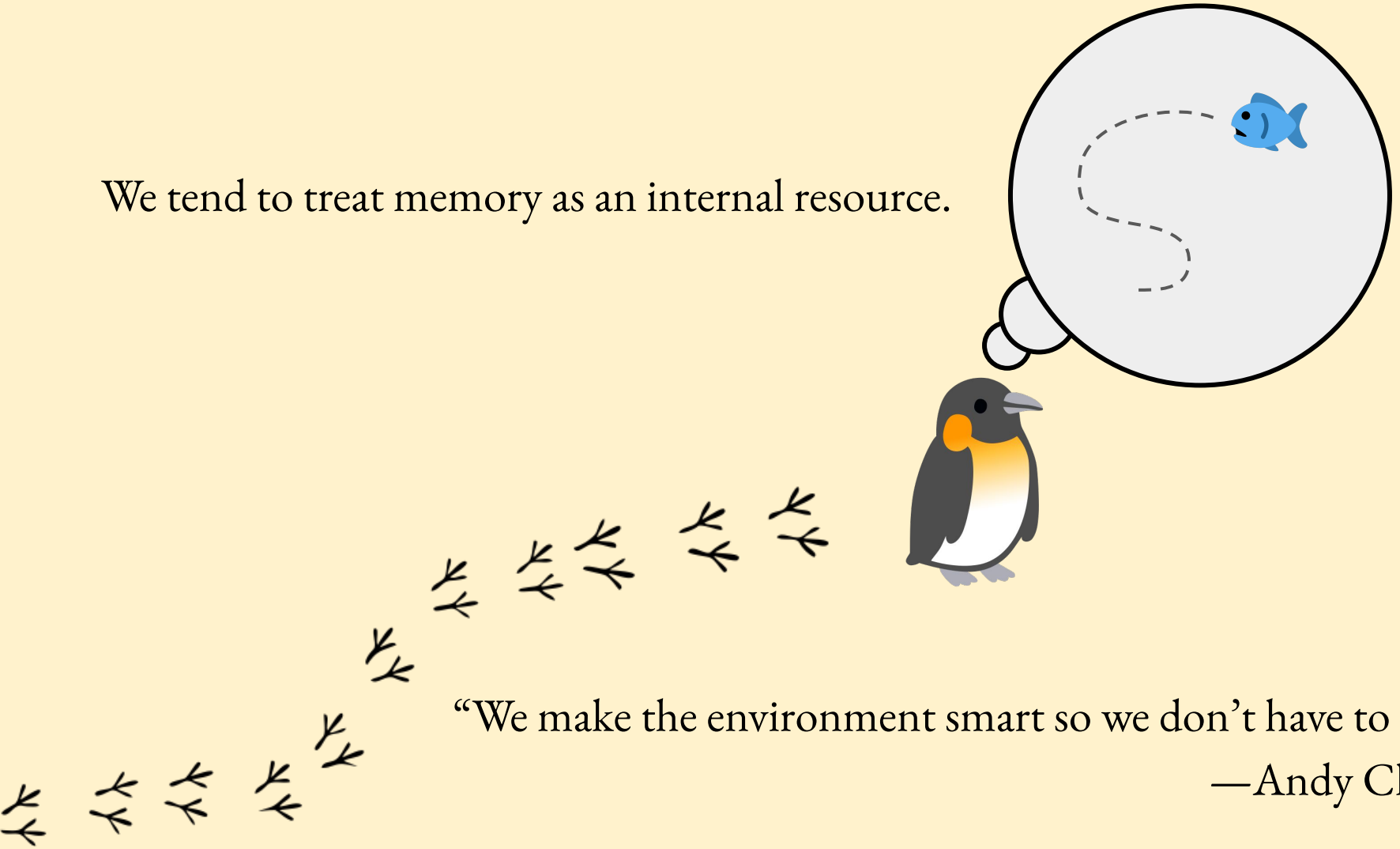
UNIVERSITY OF ALBERTA
QUAECUMQUE VERA

RL-Philosophy Symposium – May, 2026

A Cartoon Example



We tend to treat memory as an internal resource.



"We make the environment smart so we don't have to be."

—Andy Clark

Main Claim. The environment can function as an RL agent's computational memory.

Can we formalize this?

Key Results

- We formalize what it means for an environment to function as an agent's memory. Central to this is the concept of *artifacts*.
- Evidence that RL agents can *externalize* memory in spatial settings.
- Externalization need not be intentional to be experienced.

Why this matters

- Could reveal principled ways to exploit the environment as a substitute for explicit internal memory.
- Highlights the centrality of environmental resources to an agent's success.

My wonderful collaborators and colleagues.

Fraser Mince

Amy Pajak

Esra'a Saleh

Will Dabney

Joseph Modayil

David Abel

1. **Formalism**
2. **Experiments**

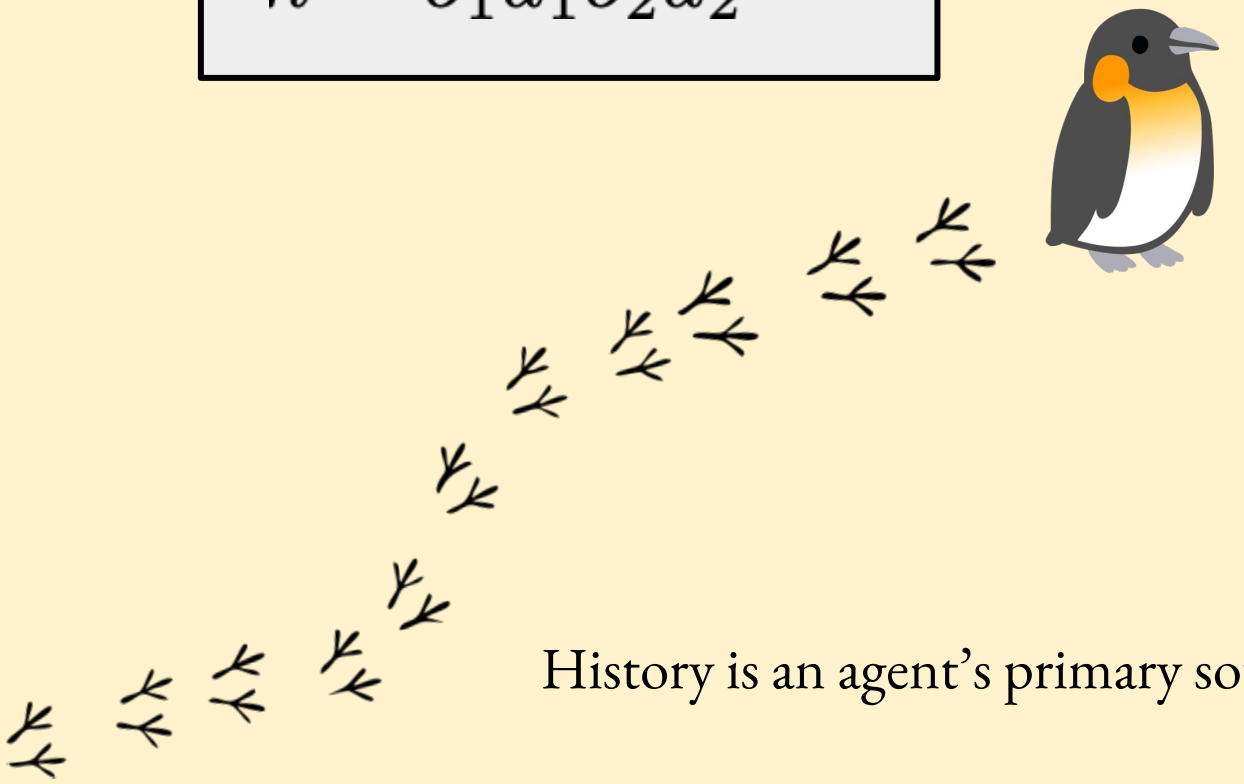
History

Histo

$$h \in \mathcal{H} = (\mathcal{O} \times \mathcal{A})^*$$

$$h = o_1 a_1 o_2 a_2 \cdots$$

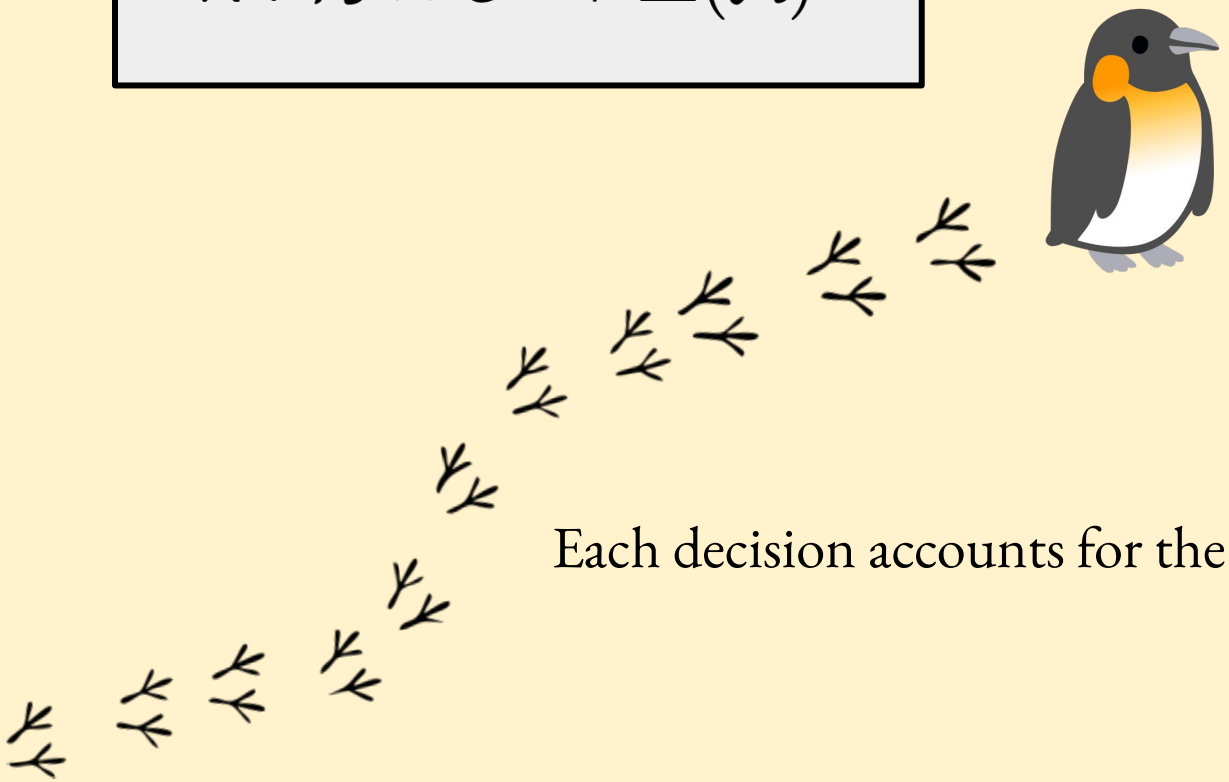
interaction.



History is an agent's primary source of knowledge.

Agent

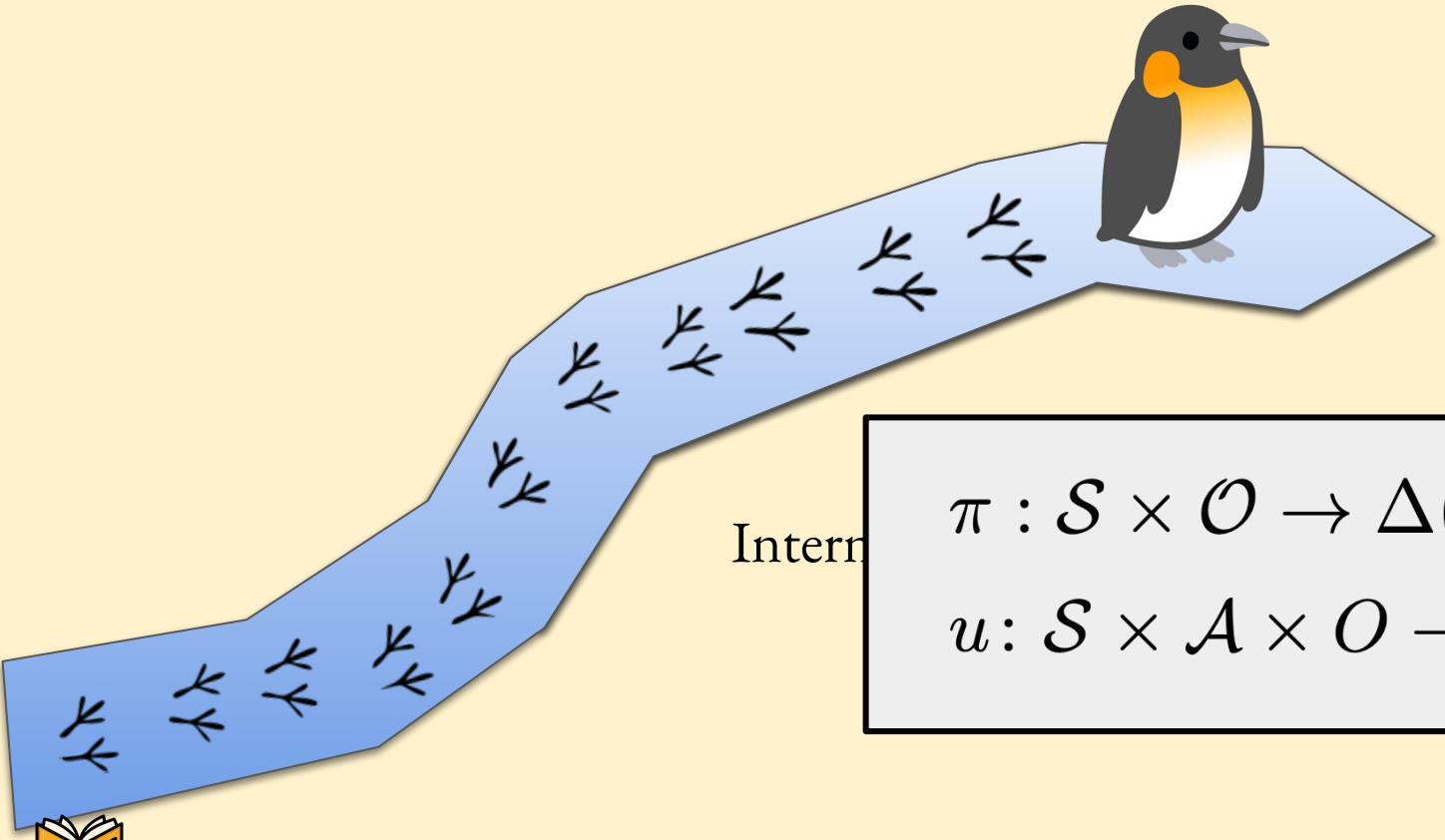
$$\lambda : \mathcal{H} \times \mathcal{O} \to \Delta(\mathcal{A})$$



Each decision accounts for the agent's entire lifetime.

Bounded Agent

Bounded agents encode history with some finite *internal state*.



Intern

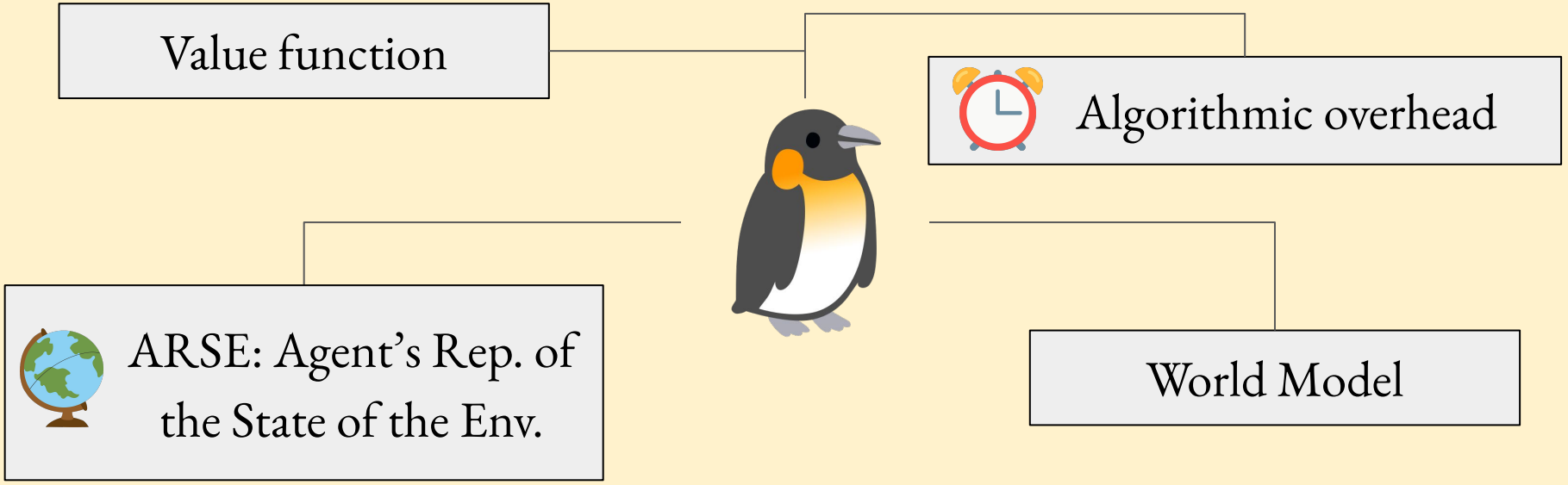
$$\pi : \mathcal{S} \times \mathcal{O} \to \Delta(\mathcal{A})$$
$$u : \mathcal{S} \times \mathcal{A} \times \mathcal{O} \to \mathcal{S}$$

tation.

On the Convergence of Bounded Agents, David Abel et al., 2023

On the Agent's Internal State

The internal state encodes multiple distinct kinds of knowledge.

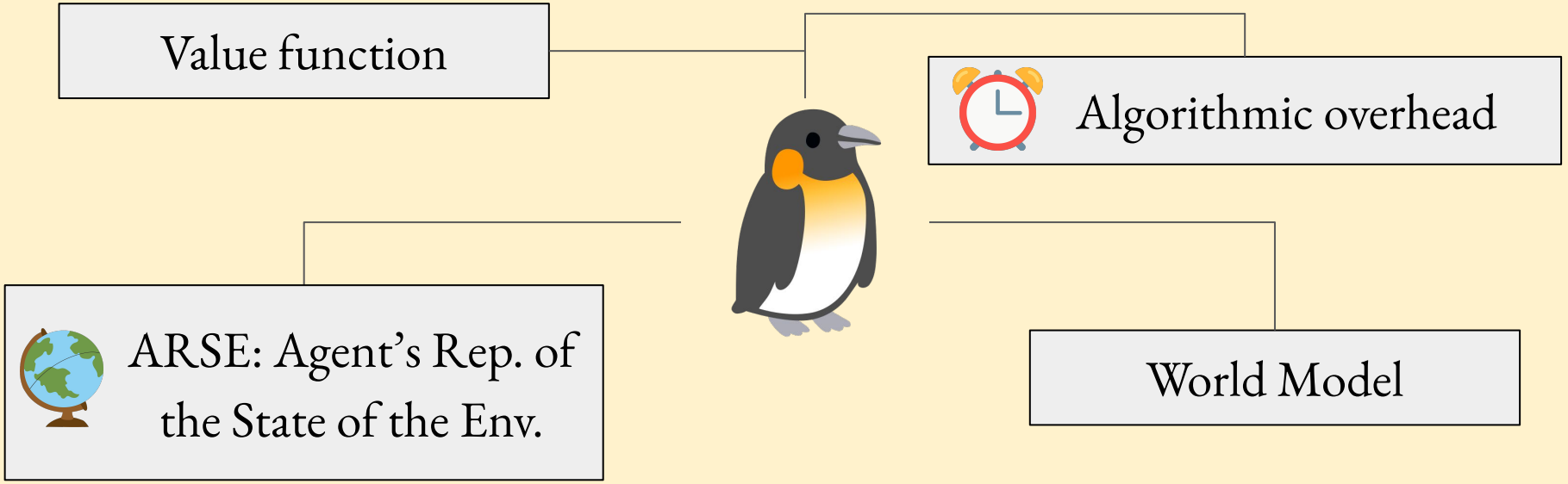


The Quest for a Common Model of the Intelligent Decision Maker, Richard Sutton, 2022

Simple Agent, Complex Environment: Efficient Reinforcement Learning with Agent States, Shi Dong, Benjamin Van Roy, Zhengyuan Zhou, 2022

On the Agent's Internal State

The internal state encodes multiple distinct kinds of knowledge.



👉 We group all kinds of knowledge into a single *computational state*.

📖 **On the Convergence of Bounded Agents**, David Abel et al., 2023

Computational State and Internal Memory

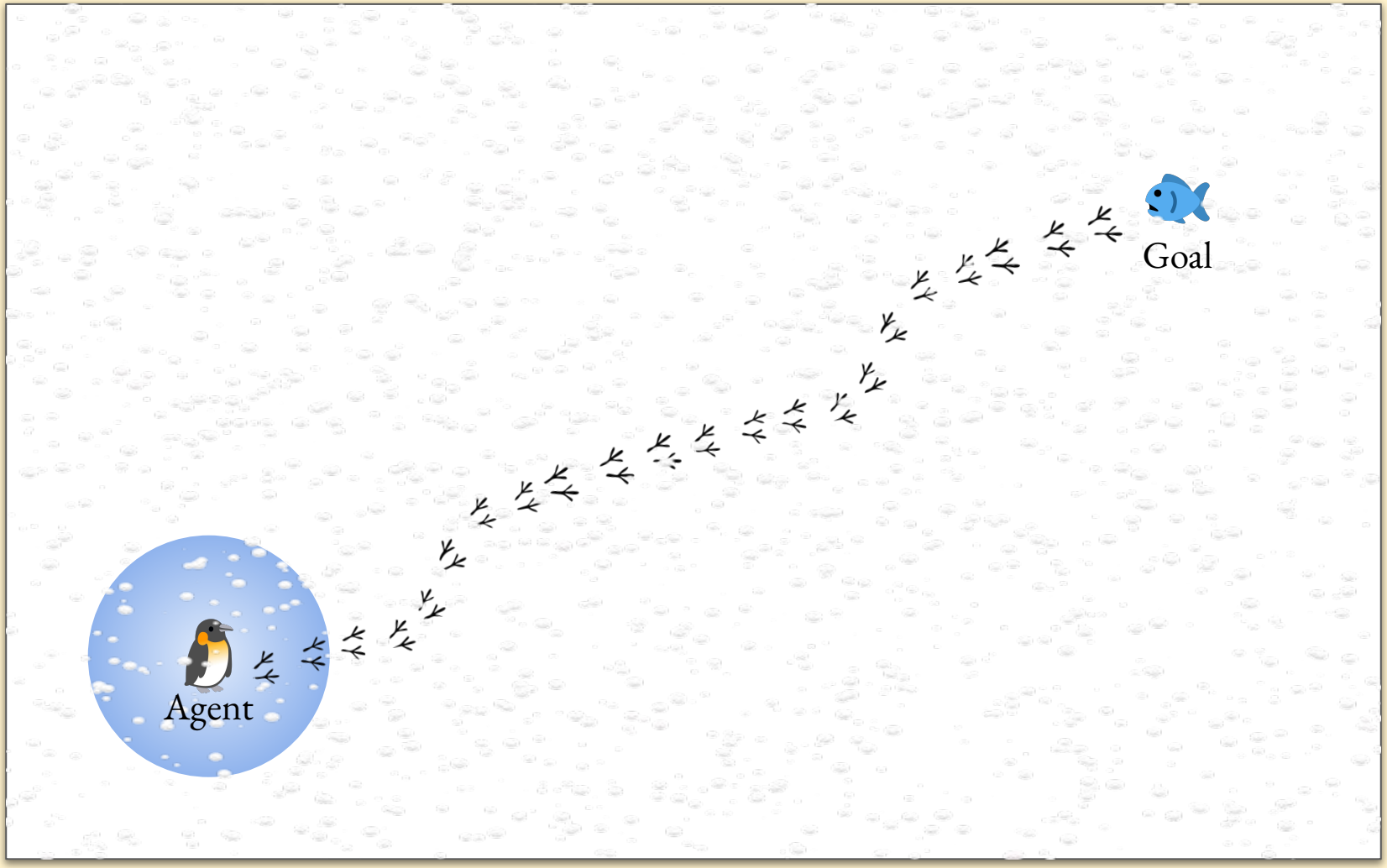
Action-value weights are a kind of internal memory

- Memory represents an agent's experience.
- Action-value weights are a component of the internal computational state.
- Action-value weights encode history.

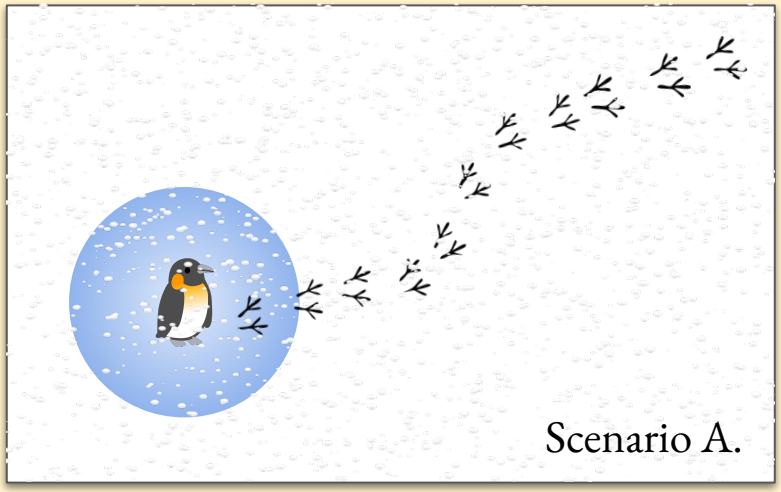
Capacity measures the amount of internal memory.

- Computer Science often equates memory with storage.
- Memory and storage are different concepts that sometimes come apart.
- *Capacity* is the total amount of system storage, e.g. bits of RAM, number of network weights.
- When we say memory, we mean capacity.

A Cartoon Example



Argument for Existence External Memory



Assumption: Achieving a goal requires a fixed amount of capacity.

Assumption: Capacity only comes from the agent or the environment.

Conclusion: If in A the agent achieves the same goal with less capacity, then the residual capacity must have come from the environment.

Formalizing artifacts

"objects for the purpose of aiding, enhancing, or improving cognition"

—Edwin Hutchins

Definition (Artifact): An observation o is an artifact of o' if and only if,

- $O_t = o$ exists for some $t > 0$,
- $o' \neq o$,
- For all $t > 0$, $O_t = o \implies O_{t'} = o'$, for some $t' < t$, and for all $t'' \in (t', t), O_{t''} \neq o'$.

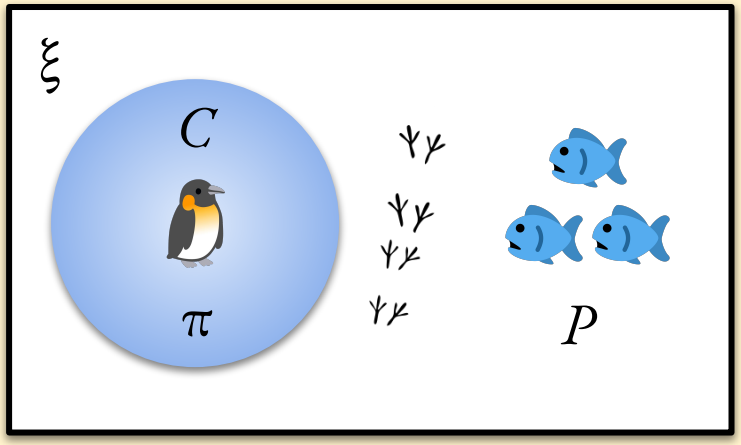
Artifact Reduction Theorem

Theorem. Let ξ be an artifactual environment, and let H be a history from ξ containing $m > 1$ observations and at least one artifact. There exists a reduced history H' with $m - 1$ observations, such that,

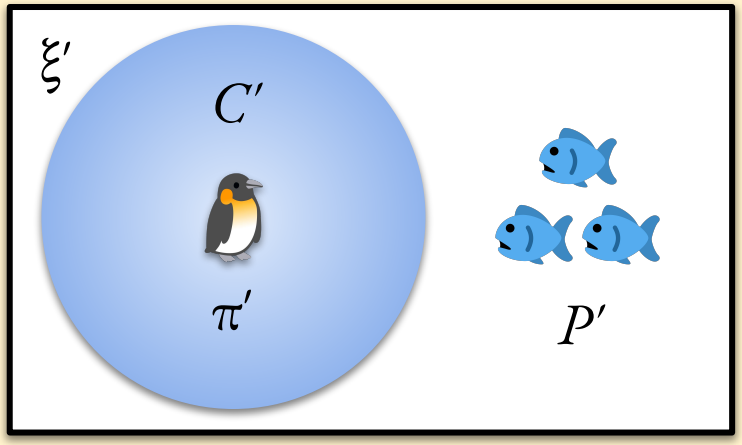
$$\mathbb{I}(O_{t+1}; H) = \mathbb{I}(O_{t+1}; H')$$

Key Idea: Artifacts reduce the information needed to represent the past.

Tightening the setup



Scenario A.

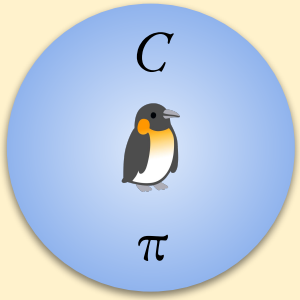


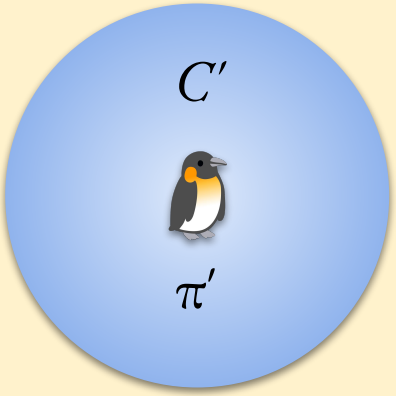
Scenario B.

Scenario A: Suppose agent π with capacity $C > 0$ achieves performance P in artifactual environment ξ.

Scenario B: Let ξ' be an *artifactless copy* of ξ. Denote π', C', P' as above.

Similar Designs, Different Scales

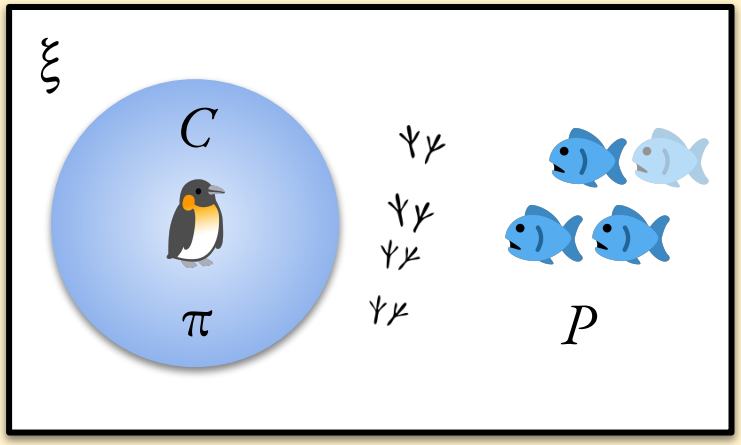




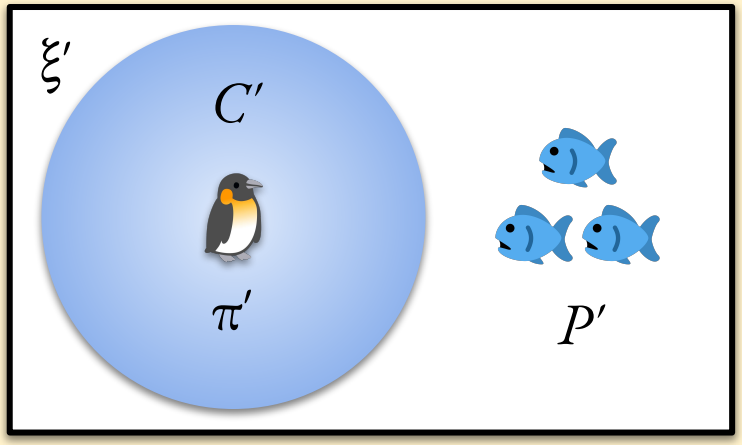
Assumption: π and π' only differ in the amount of capacity.

- Agents share the same general blueprint: interface, learning algorithm, representation.
- Differ in scale of capacity.

Actualizing Externalized Memory



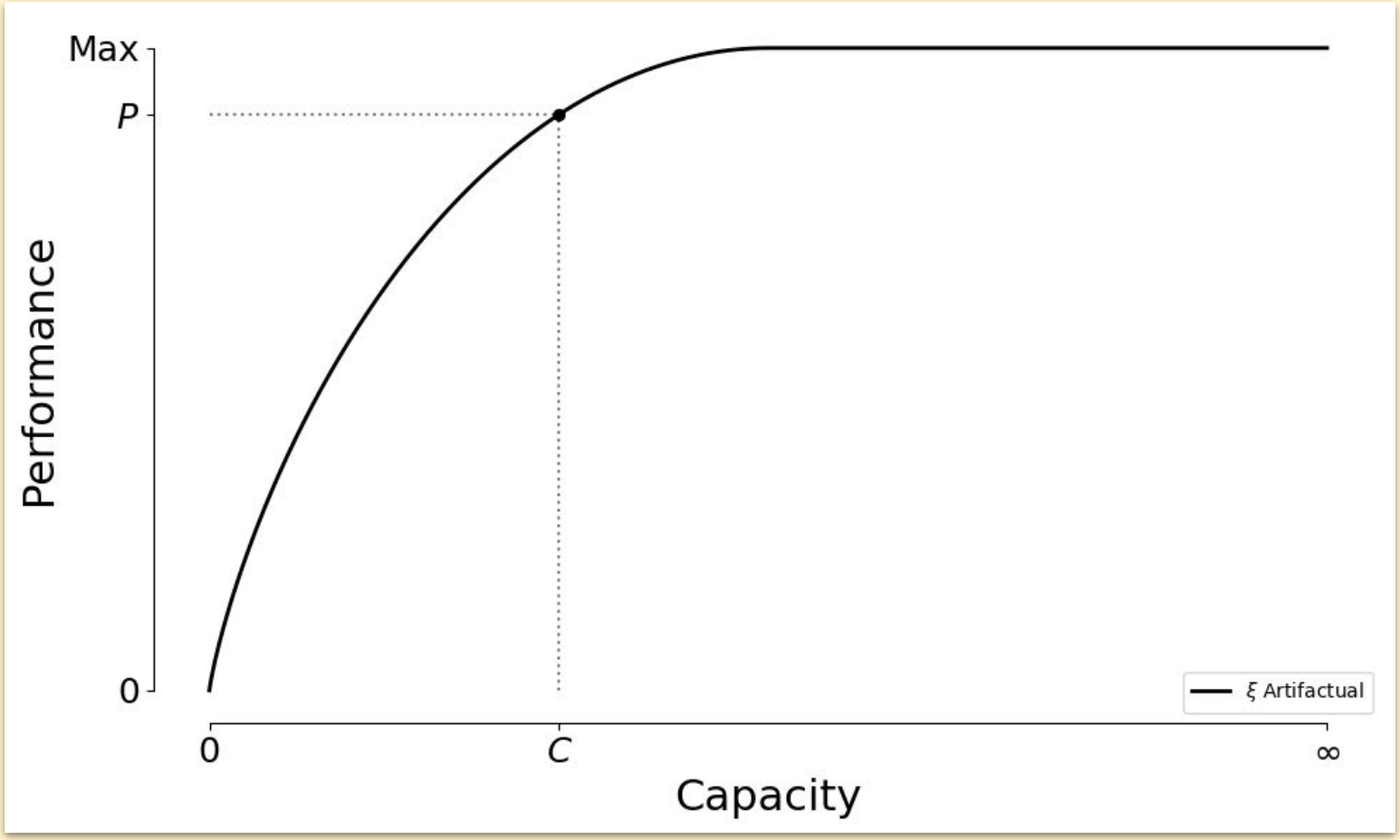
Scenario A.



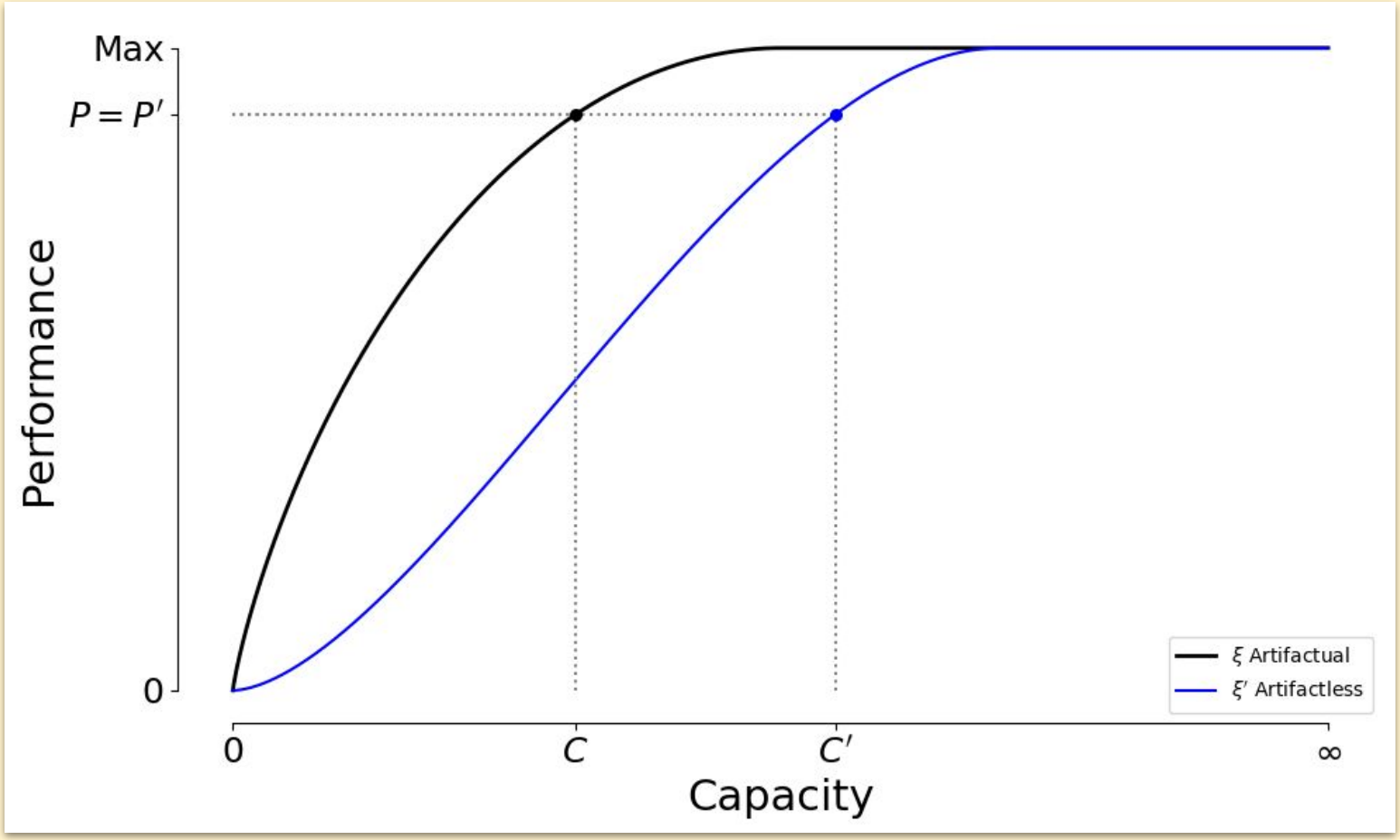
Scenario B.

Condition. $C < C'$ and $P \geq P'$, for some π'.

- Externalization occurs when the agent must use artifacts to match or exceed performance with less capacity.

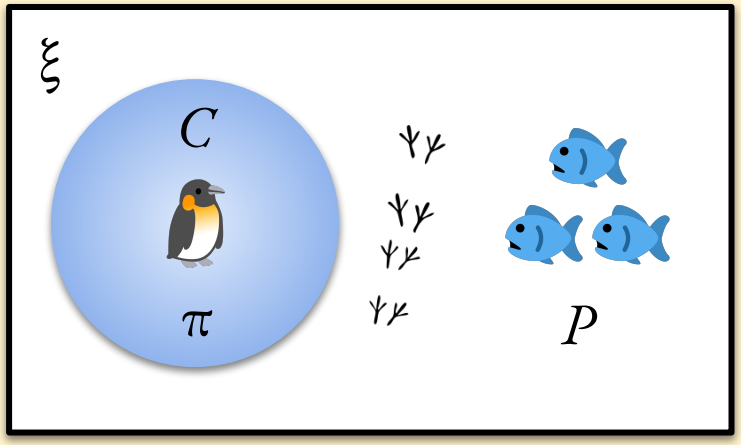


- Consider the hypothetical curve of all possible agents π learning in ξ.
- Performance is measured for some fixed amount of experience.
- Performance improves monotonically with increasing capacity.

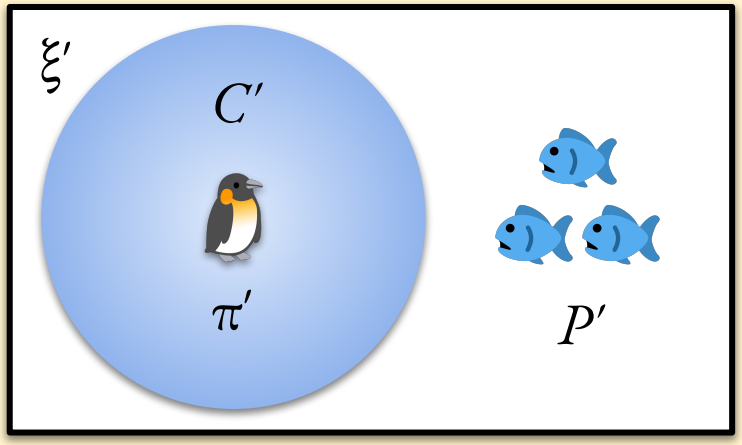


Condition. $C < C'$ and $P \geq P'$, for some π'.

A better definition



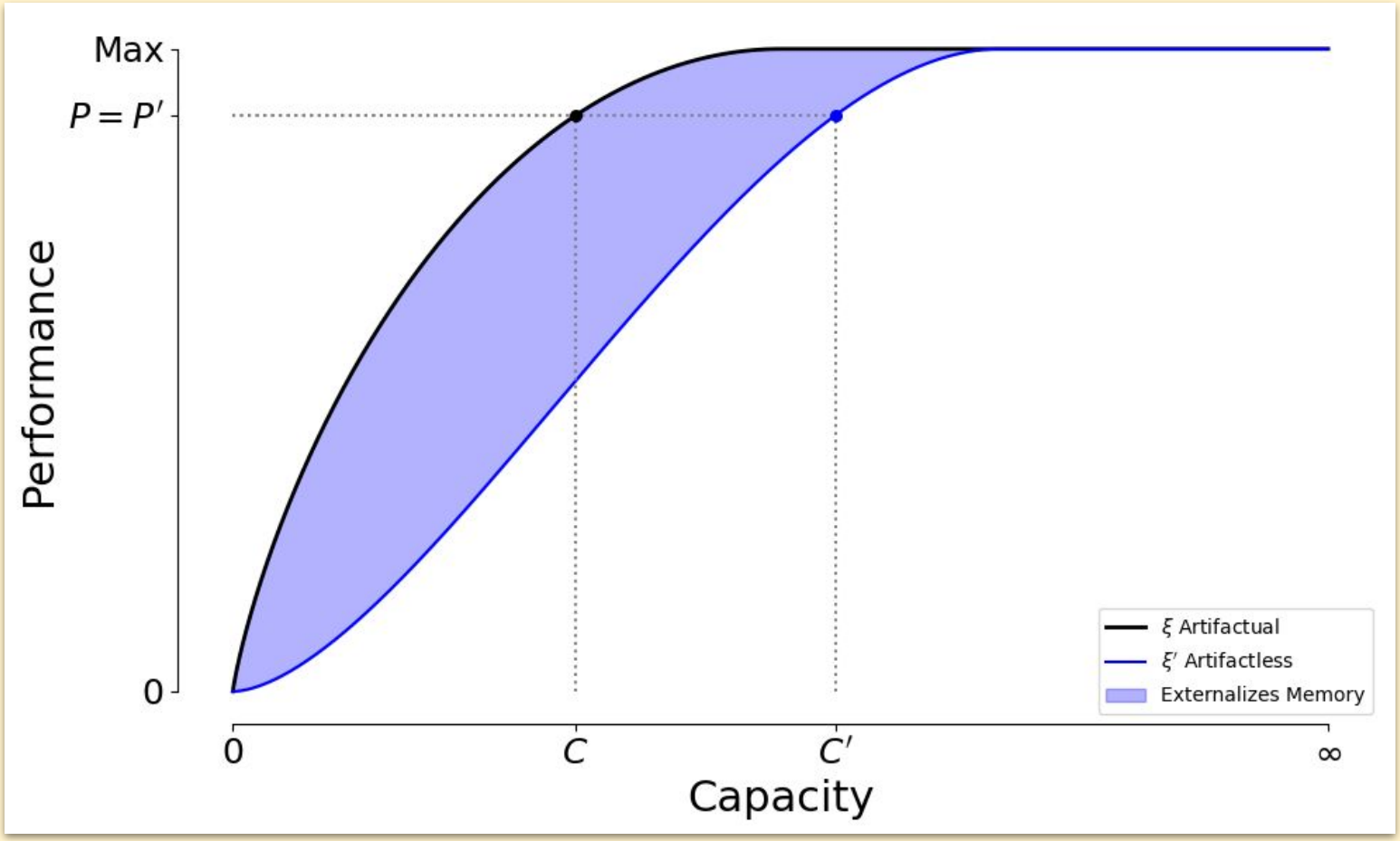
Scenario A.



Scenario B.

Externalizes Memory: Agent π **externalizes memory** to ξ if, and only if

1. $C < C'$ and $P \geq P'$, for some π'.
2. For all π'' learning in ξ', if $C'' \leq C$ then $P'' < P$.

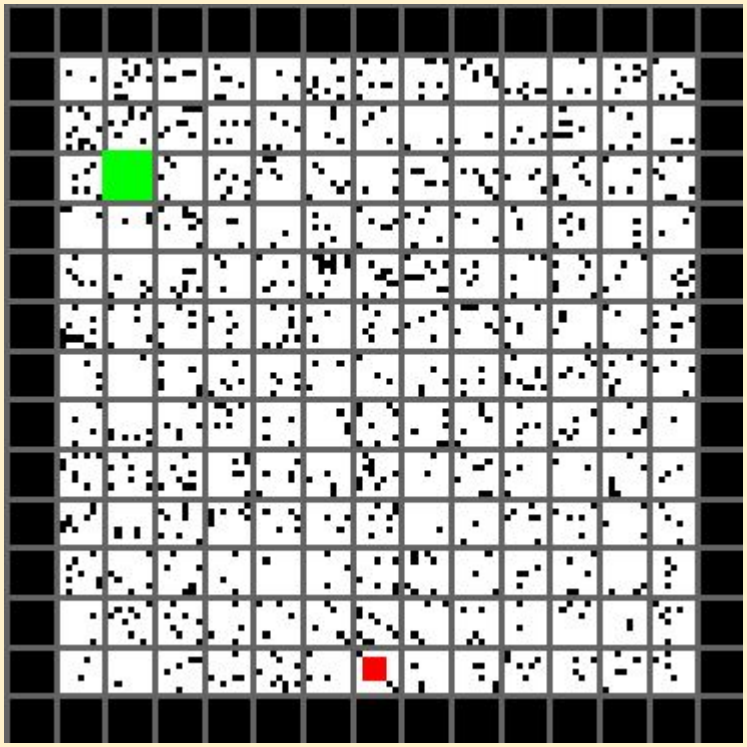


Externalizes Memory: Agent π **externalizes memory** to ξ if, and only if

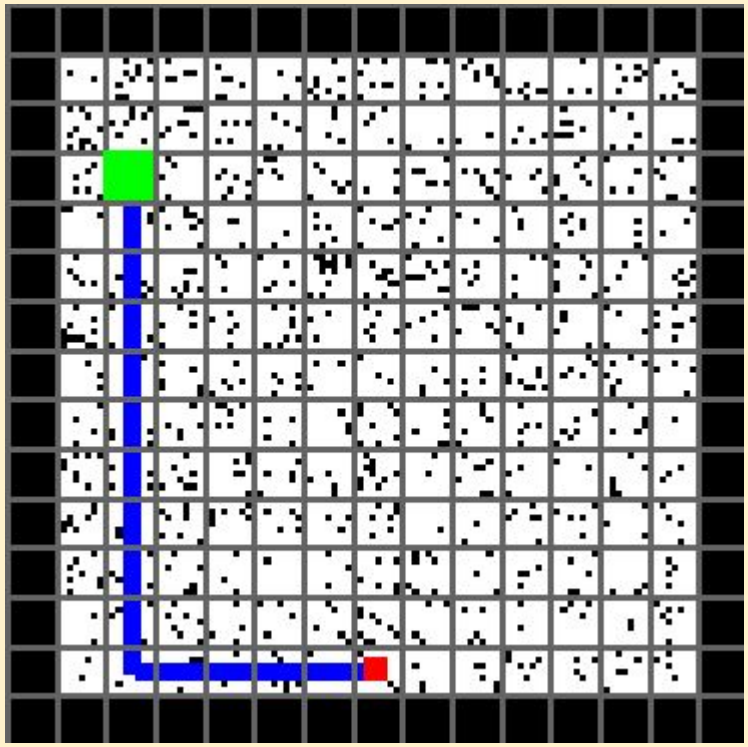
1. $C < C'$ and $P \geq P'$, for some π'.
2. For all π'' learning in ξ', $C'' \leq C \Rightarrow P'' < P$.

1. Formalism
2. **Experiments**

A Concrete Demonstration



Scenario A.



Scenario B.

- Each location contains an 8x8 binary image of salt and pepper noise.
- The agent observes a composite image from the surrounding 3x3 cells.
- A bonus of +1 is given for reaching the goal. Otherwise, reward is zero.

Our study considers two agent designs

1. **Linear Q-learning:** $\hat{q}(o, a; w) = w_a^\top o$

- Action-values are represented as *mixtures of the input-observations*.
- Weights are updated incrementally according to the Q-learning rule.
- System capacity scales with the number of weights w_a.

2. **DQN:** `q_hat = nn.Sequential([nn.Dense(32), nn.relu, ...])`

- Action-values are outputs of a multi-layer neural network.
- Weights are updated with backprop using mini-batches.
- System capacity is proportional to the number of network units.

👉 See paper for the full set of results with DQN.

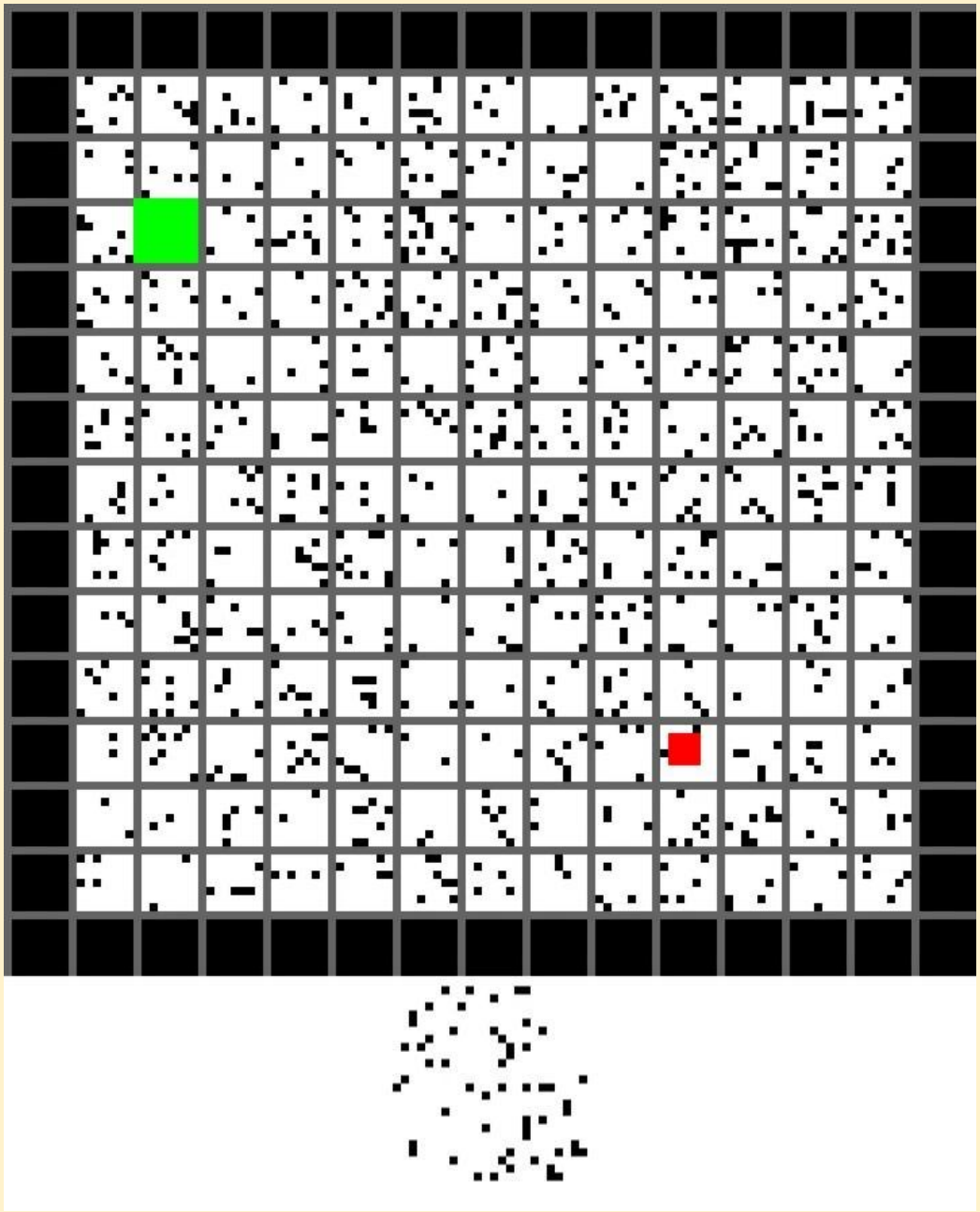
Operational Measures

Performance: Average reward up to time t.

$$P = \frac{1}{T}\sum_{t=1}^{T} r_t$$

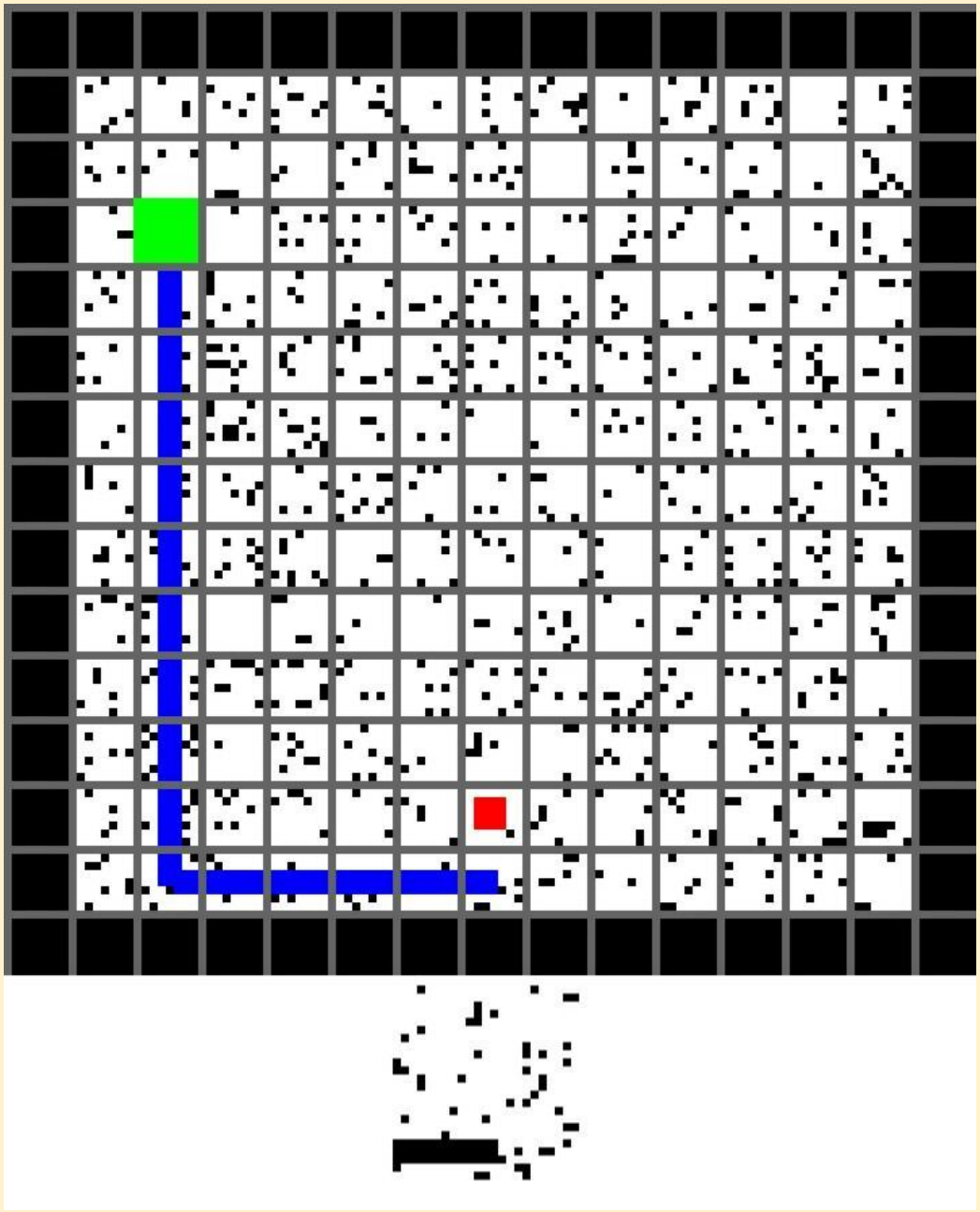
Capacity: a scalar C proportional to the number learnable parameters.

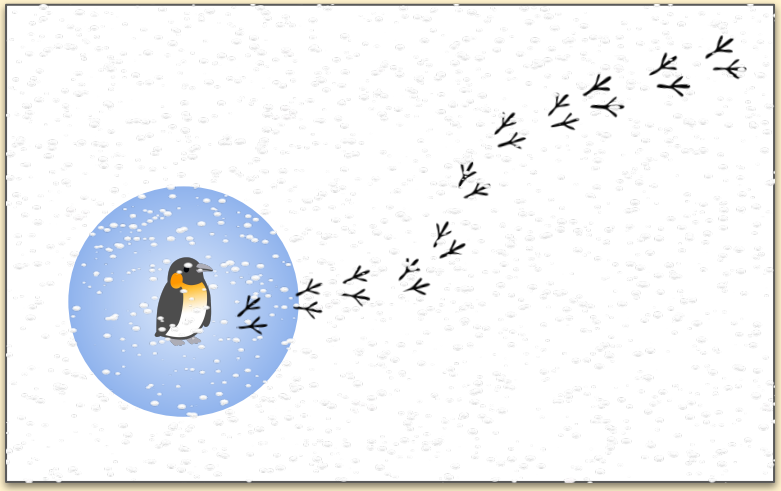
1. Linear-Q: number of linear weights per action-value.
2. DQN: number of hidden units.



Artifactless baseline

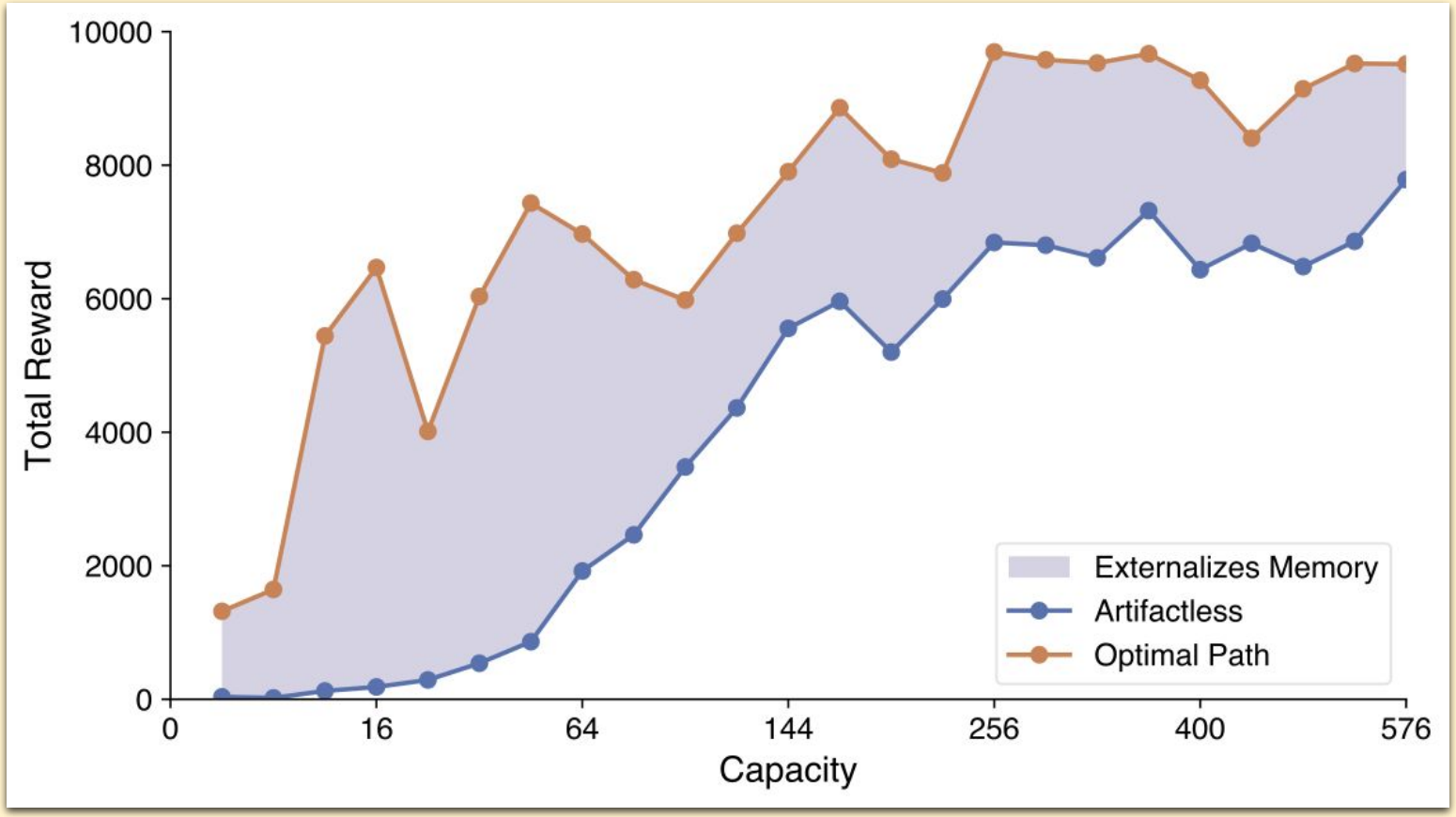
- Policy at 150k / 200k steps
- Behavior is mostly random





Optimal Path

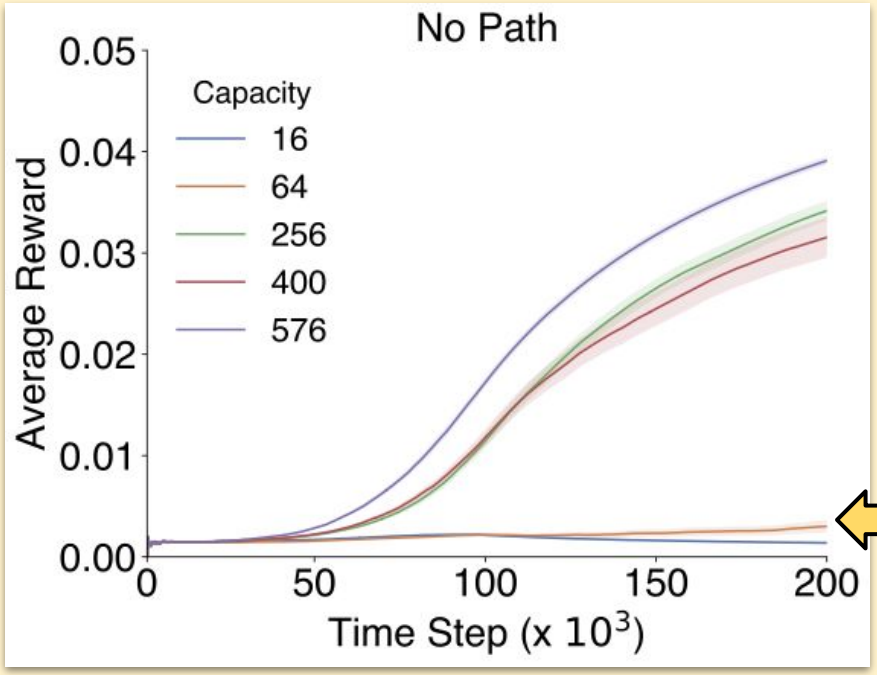
- Policy at 150k / 200k steps
- Behavior is goal-directed

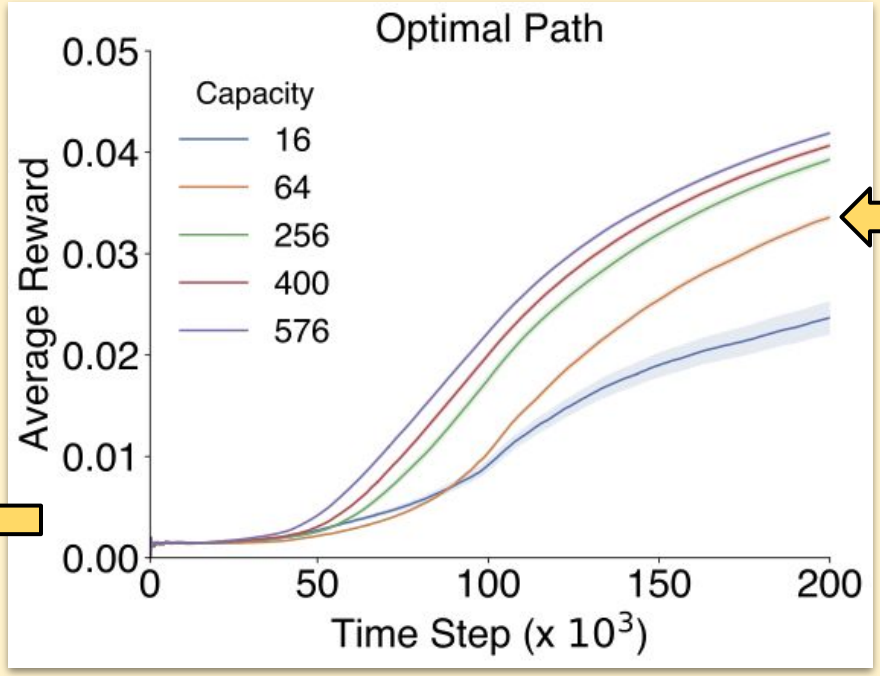


Linear Q-learning observing an optimal path

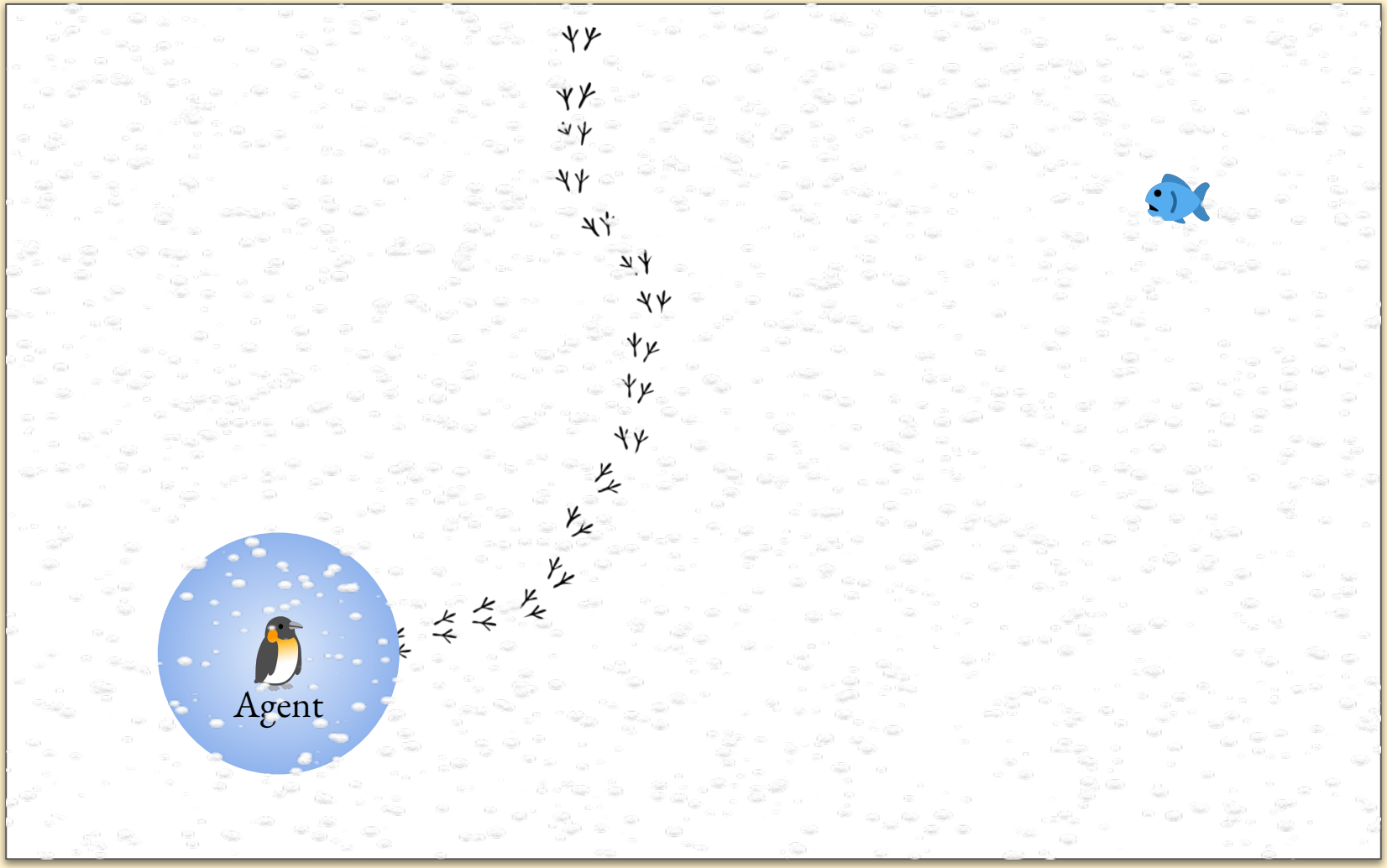
- For a given capacity, performance uniformly improves.
- Externalization is found across all capacities, excluding the endpoints.
- Effect is starkest for the low capacity regime. Theoretical minimum is 169.

Linear Q-learning observing an optimal path





What about other fixed artifacts?

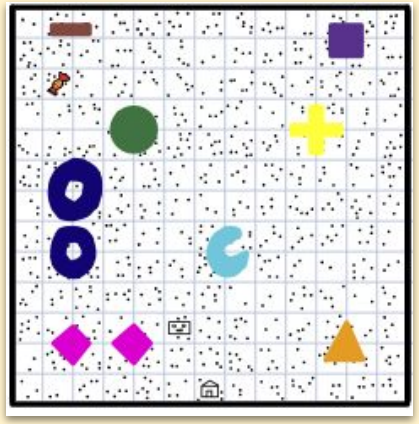


Experiment 2: Externalization from Other Artifacts?

Suboptimal

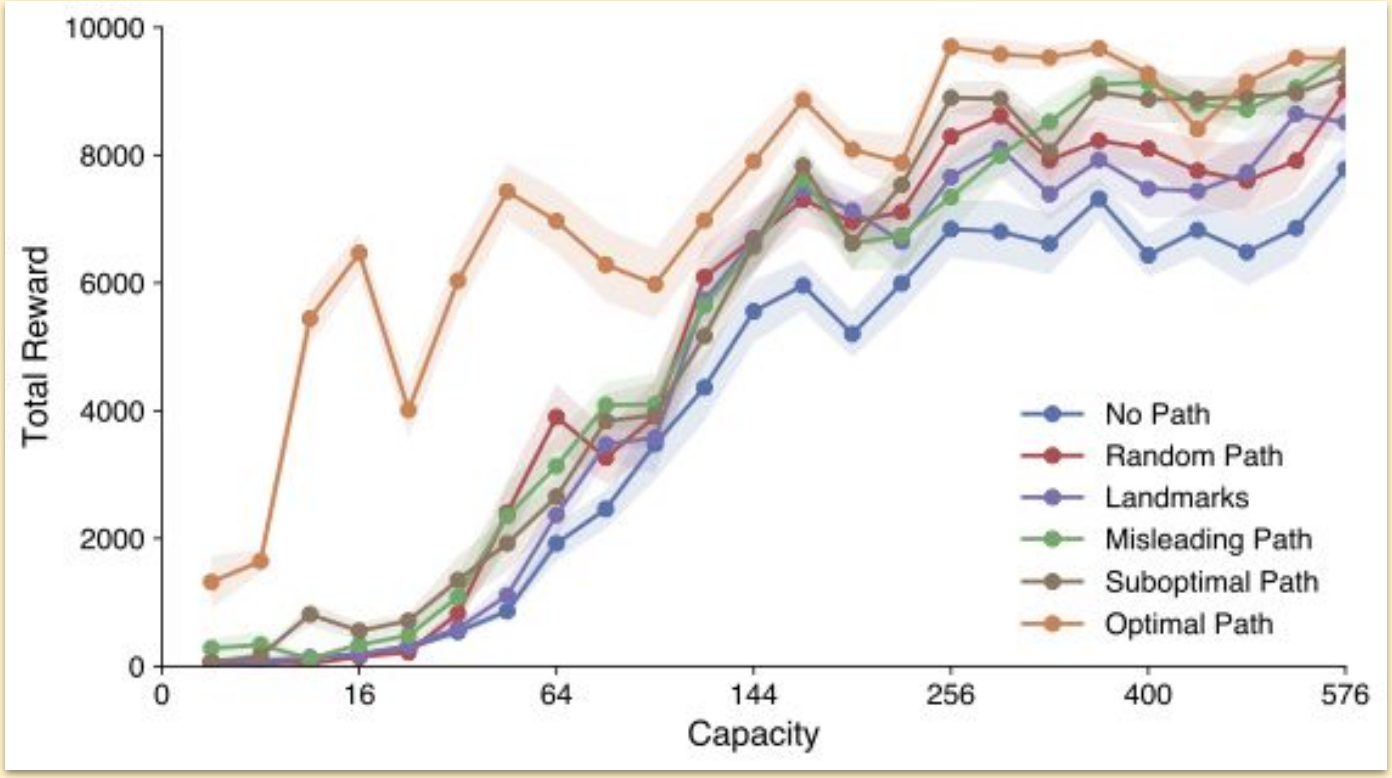
Random

Misleading



Landmarks

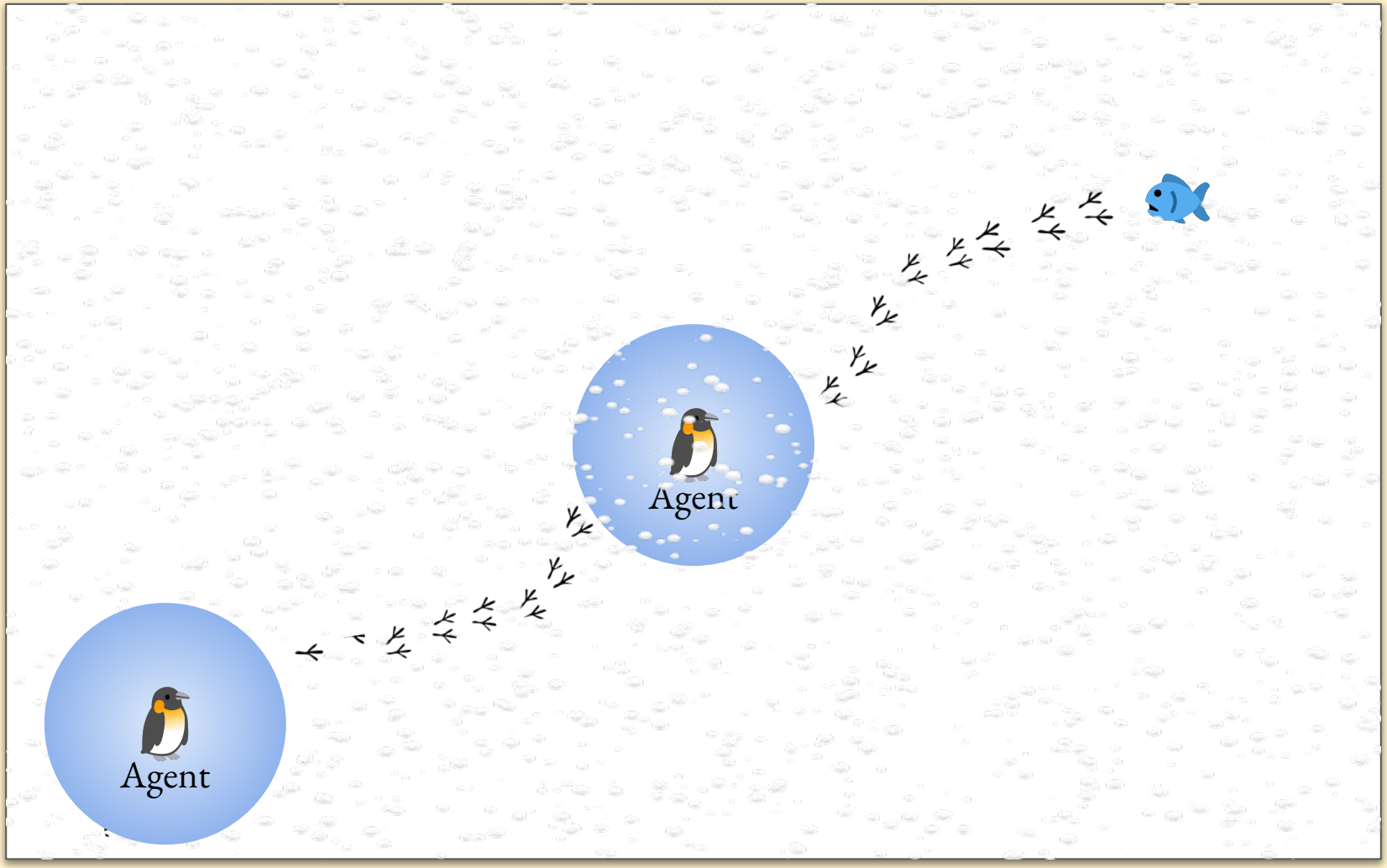
- Positive evidence of externalization suggests artifacts are selectively accessed as they relate to the goal.
- This experiment rules out the hypothesis that agents merely follow paths.
- We consider both behavioral and non-behavioral artifacts to understand how the strength of the effect covaries without an overt behavioral signal.

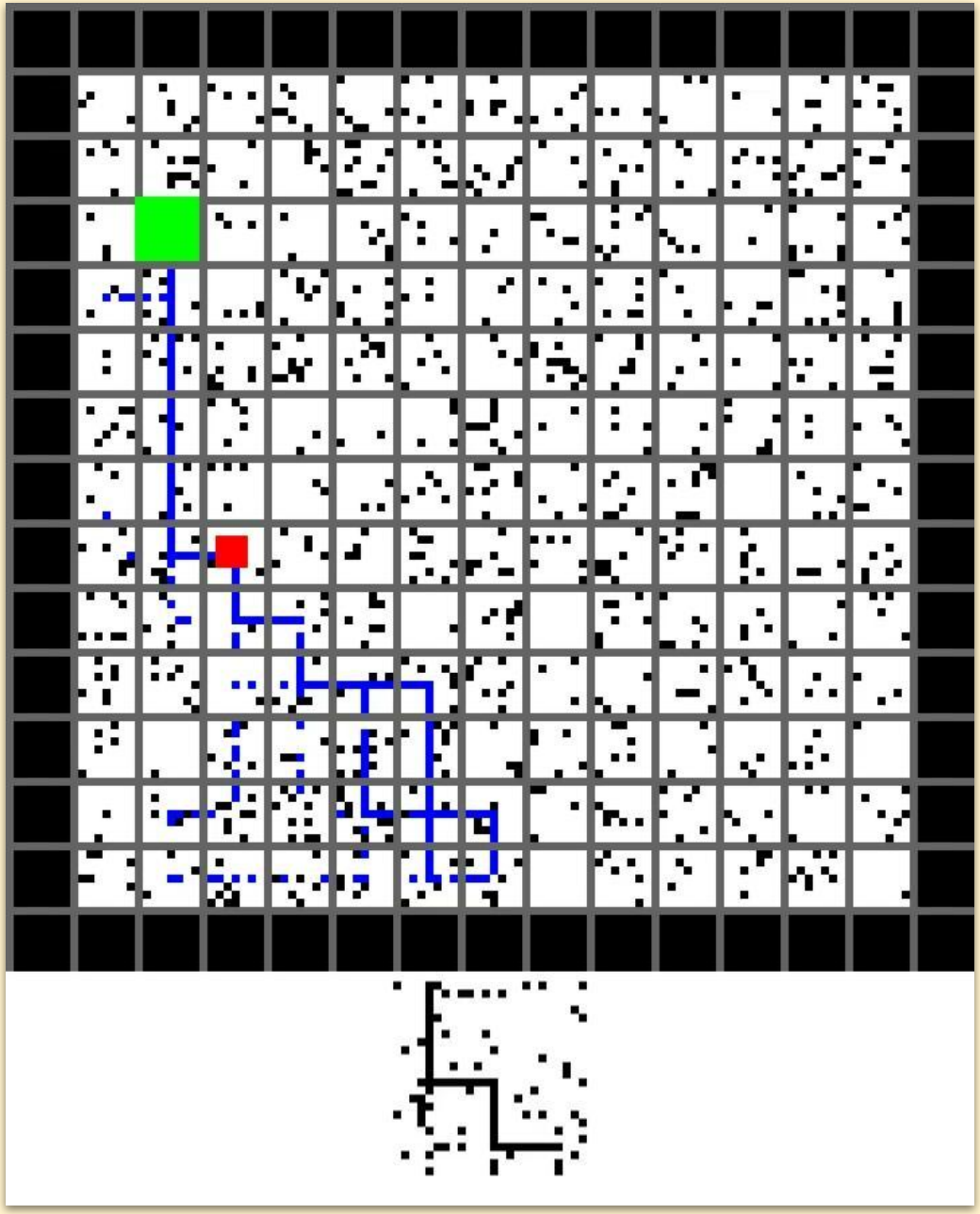


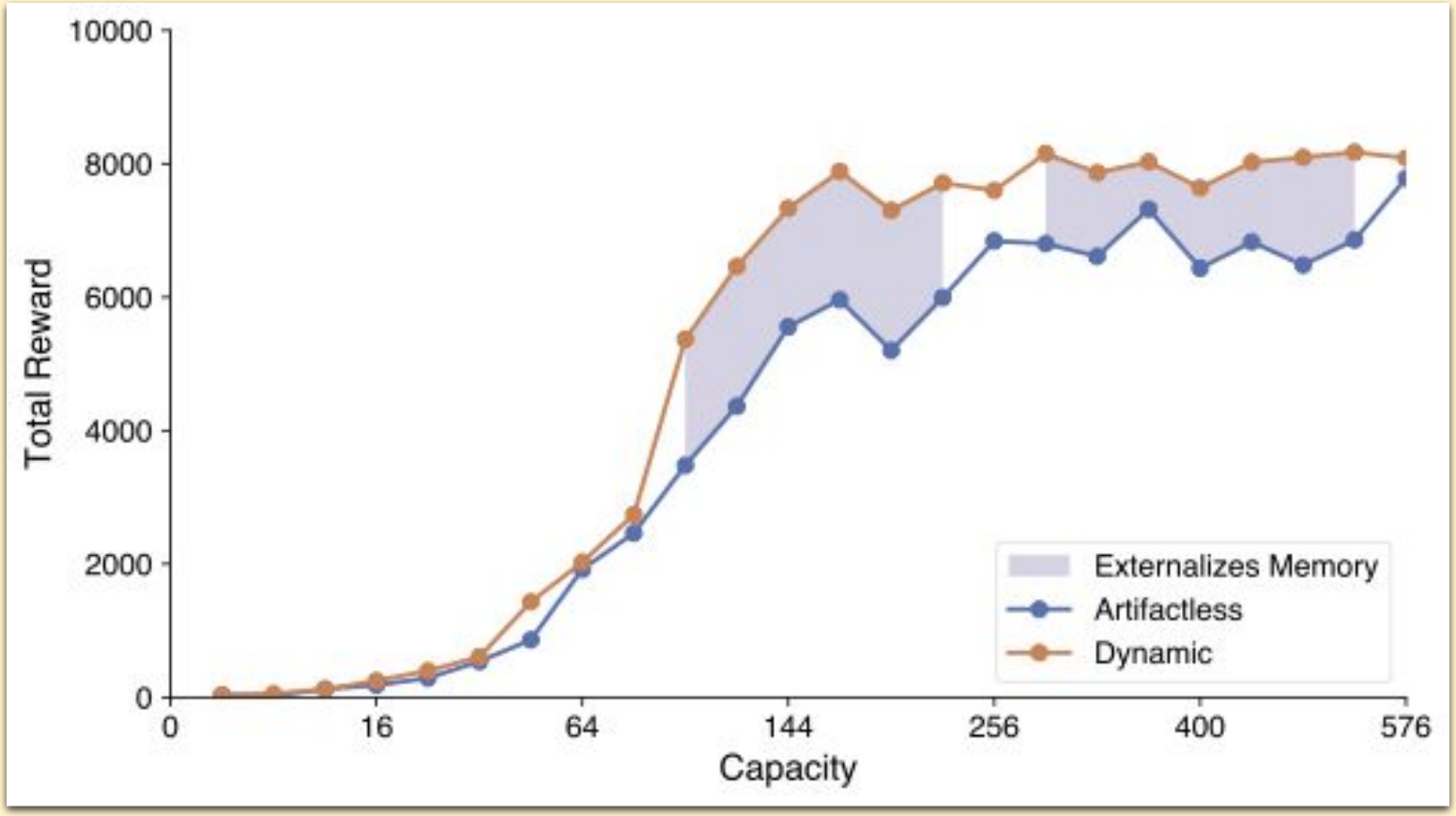
Linear Q-learning with other fixed artifacts

- Strength of the effect covaries with the type of artifact.
- Random artifacts and geometric landmarks satisfy empirical condition.
- Hierarchy of support for Linear:
  - Optimal > Suboptimal > Misleading > Random ~ Landmarks

Natural settings involve dynamic paths







Linear Q-learning observing a dynamic path

- Steady-state path is the mode of the occupancy distribution.
- As time goes on, the path becomes optimal.
- Definition is approximately satisfied in a few cases.

Takeaways

RL agents use the environment as an effective source of memory

- Evidence that RL agents can externalize memory in spatial settings.
- The amount of externalized memory can be quantified.
- Externalization need not be intentional to be experienced.

Open Questions (seeking collaborators)

- Can an RL agent adapt its capacity in response to artifacts?
- Can RL agents intentionally generate supportive artifacts?
- Corroborate with physical experimentation.
- Study other relationships between agent computation and the environment.

Thank you!