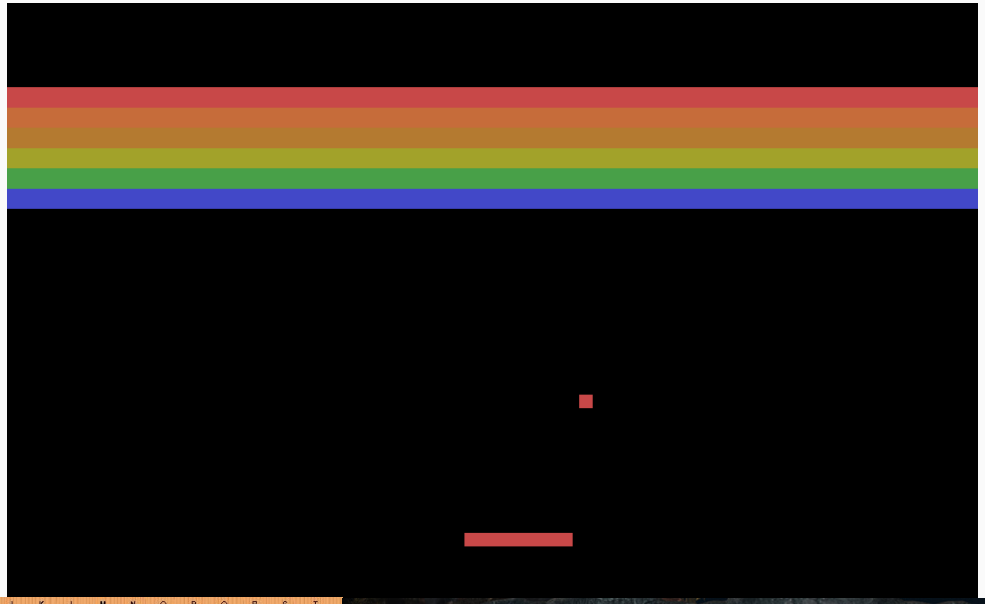
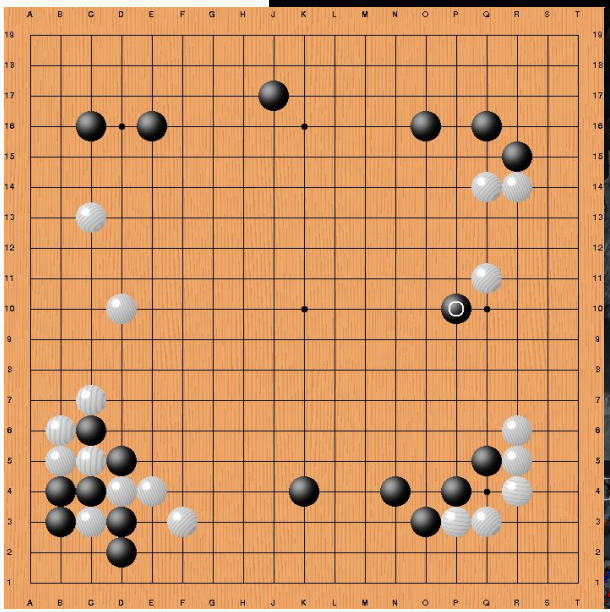
Embodiment and environmental realism lead to more systematic generalization

Felix Hill, Andrew Lampinen, Rosalia Schneider, Stephen Clark, Matthew Botvinick, James L. McClelland, Adam Santoro

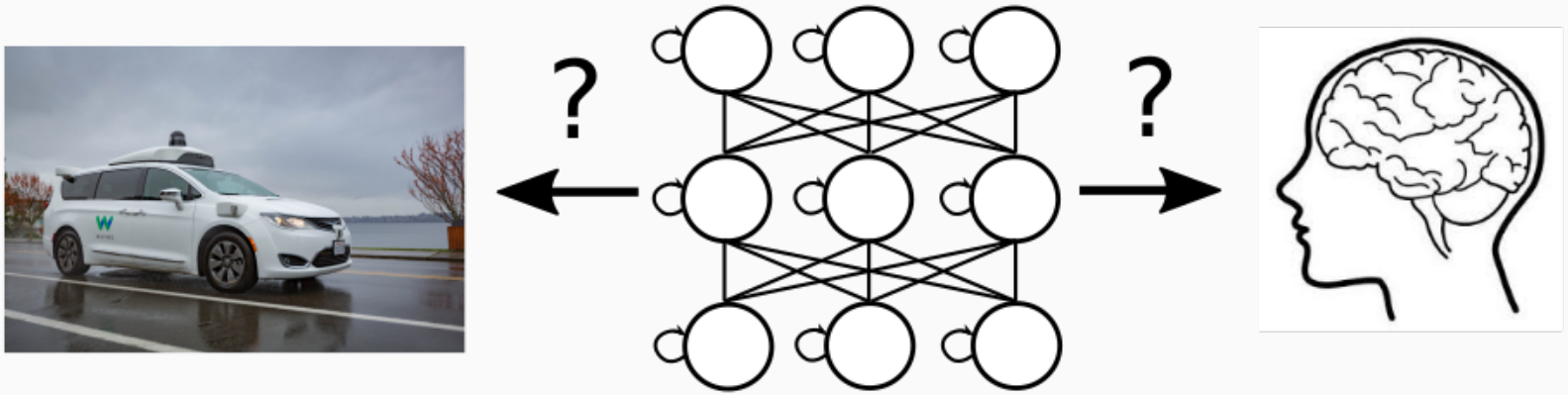
Deep RL is cool!



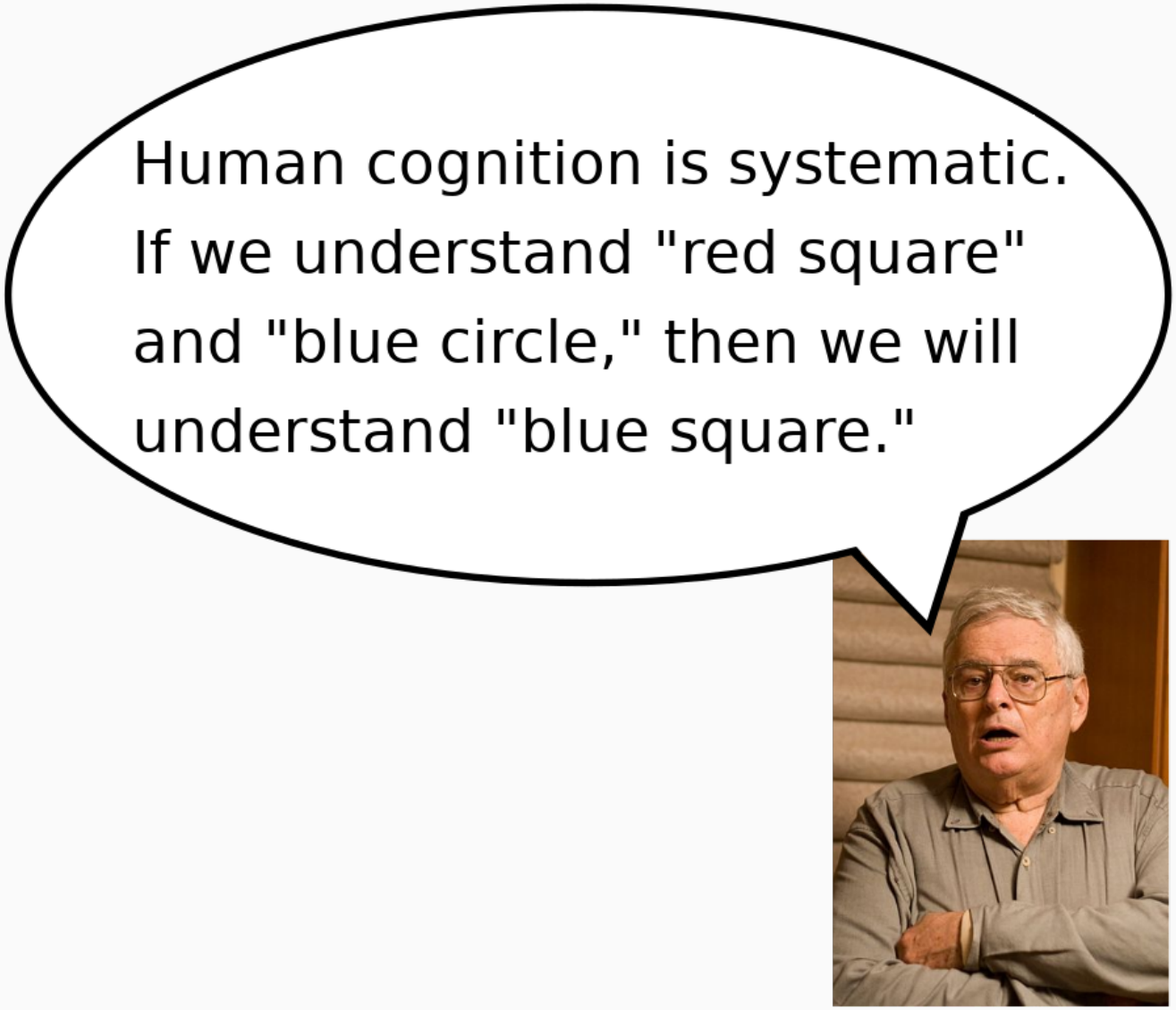


But when do they generalize what they learn?

An important question

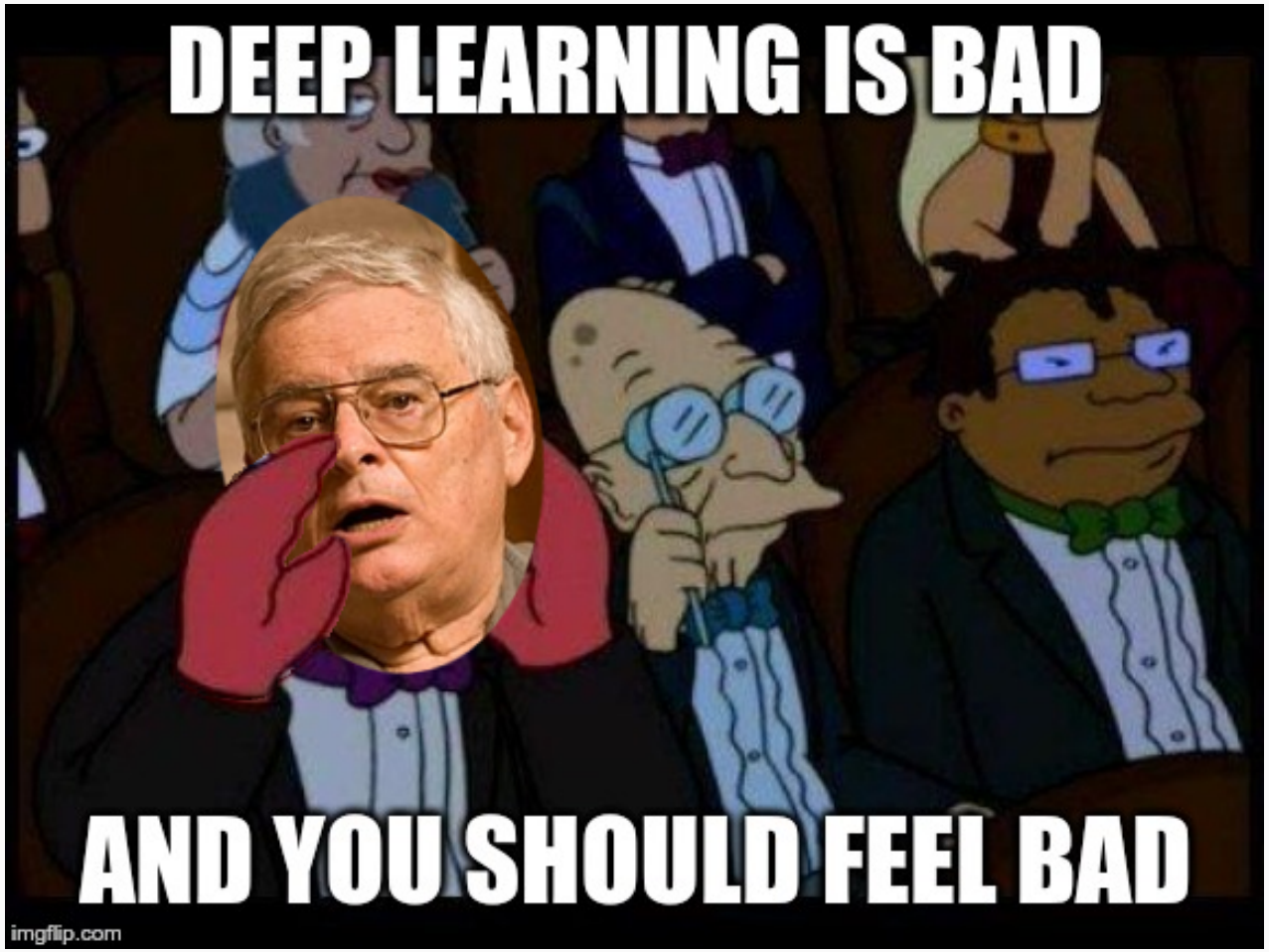


Systematic compositional generalization

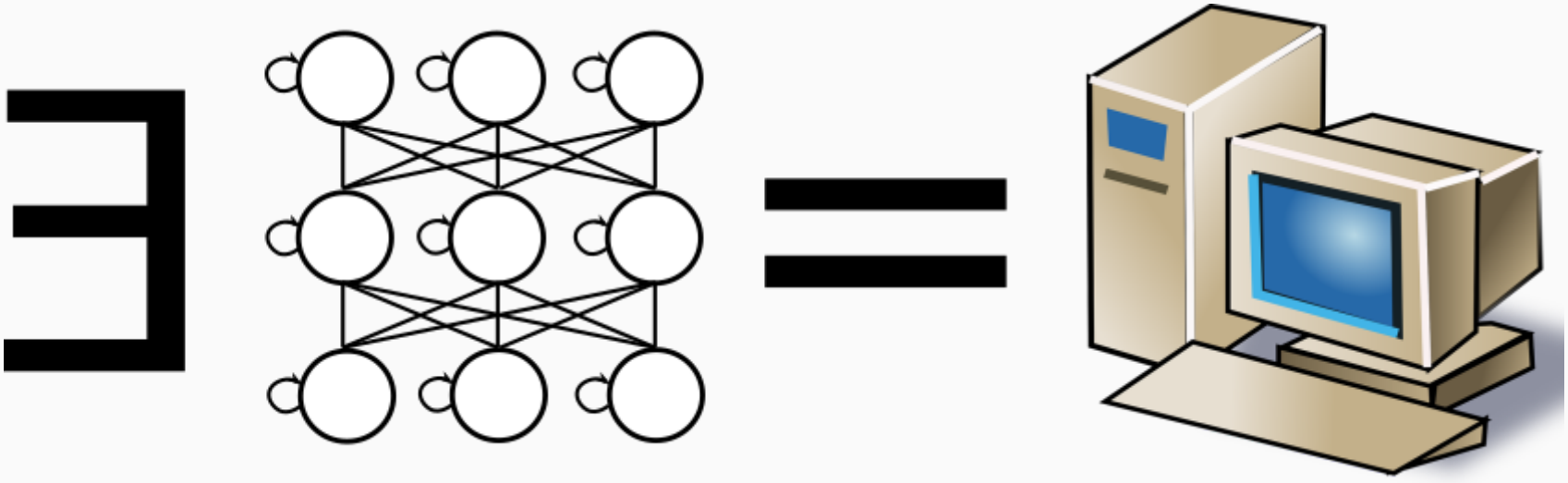


But is it really? (O'reilly et al., 2013)

Fodor on neural nets (not verbatim)



But there are systematic neural networks

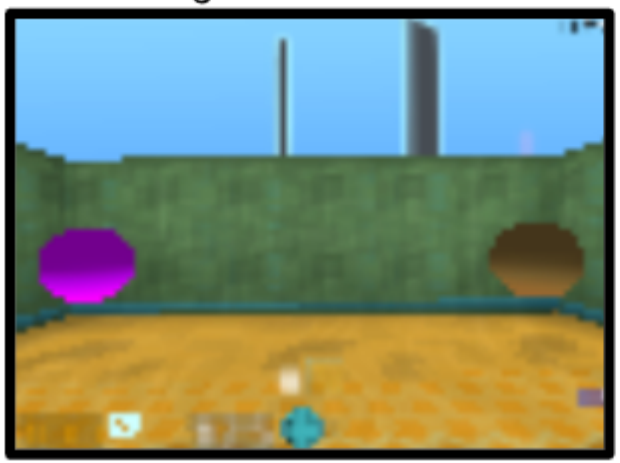


(Siegelman & Sontag, 1992)

So the question is, in practice, when do deep networks learn to generalize systematically?

Color-shape composition

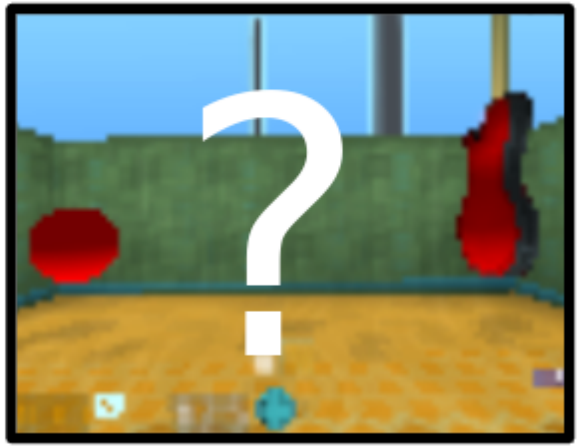
Instr: magenta ball



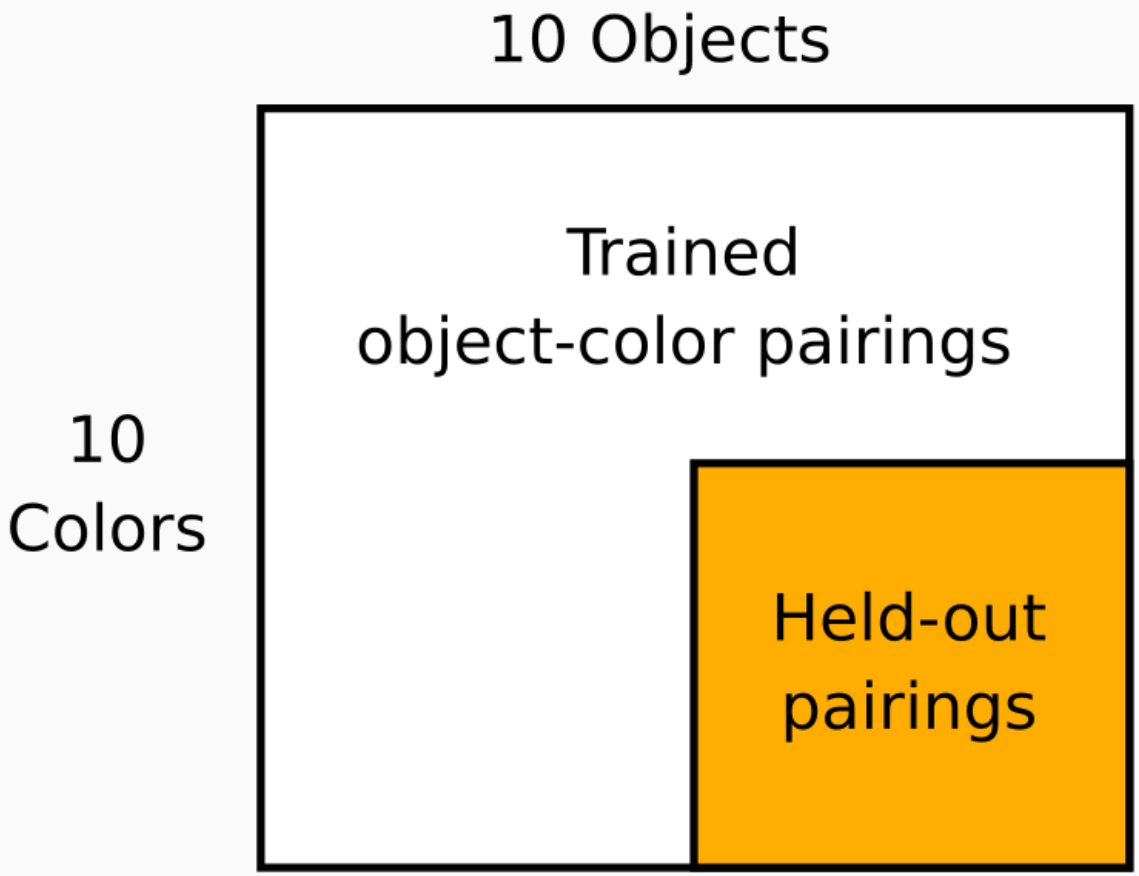
Instr: magenta guitar

Instr: red horse

Instr: red ball

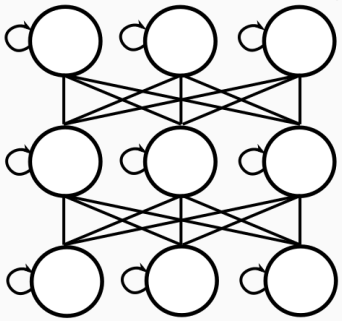


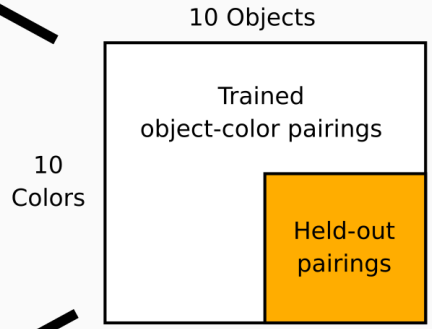
Color-shape composition



Comparing different versions of the same abstract task

Single-frame classifier





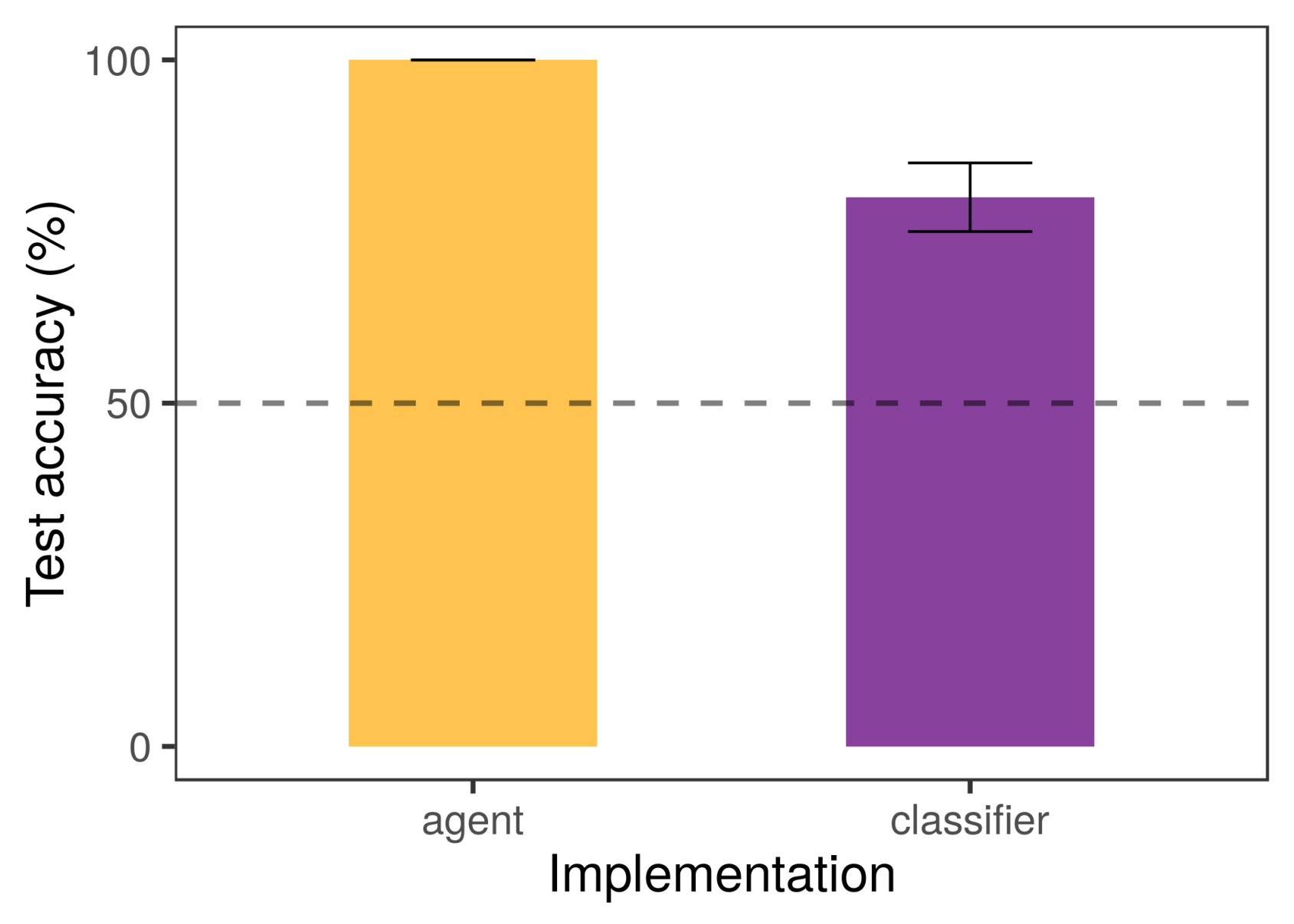
Acting over time

Same archit-ecture

Different instantiations

Same abstract task

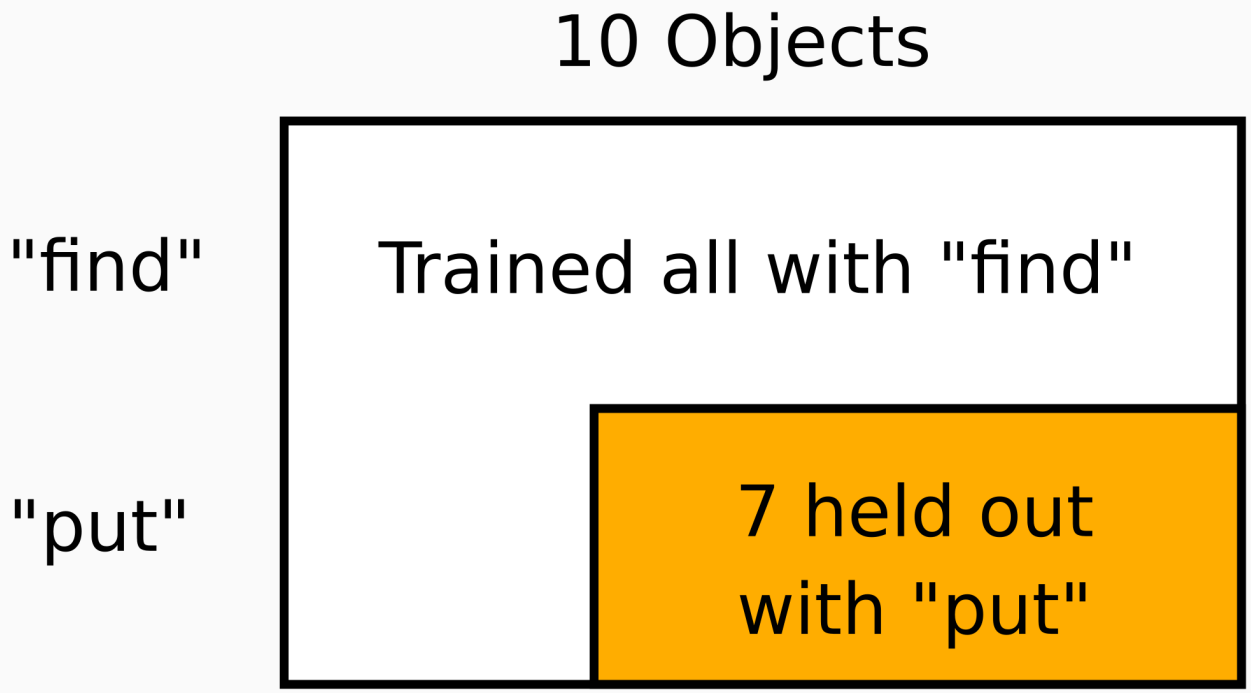
Agent generalizes better!



Generalization is better in a richer environment where the agent acts over time, rather than seeing isolated frames.

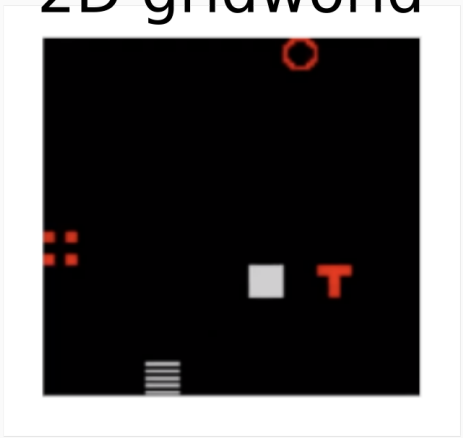
Closer to human experience

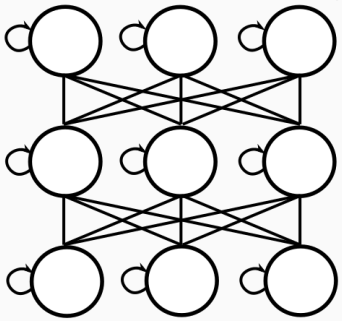
Verb-noun composition



Comparing different versions of the same abstract task

2D gridworld



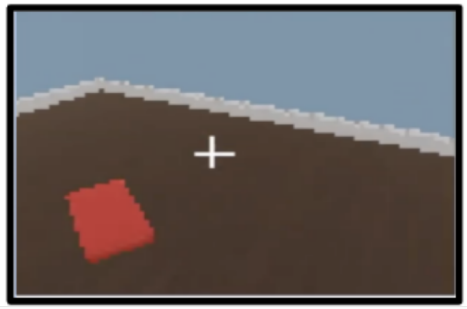


10 Objects

"find" Trained all with "find"

"put" 7 held out with "put"

3D environment

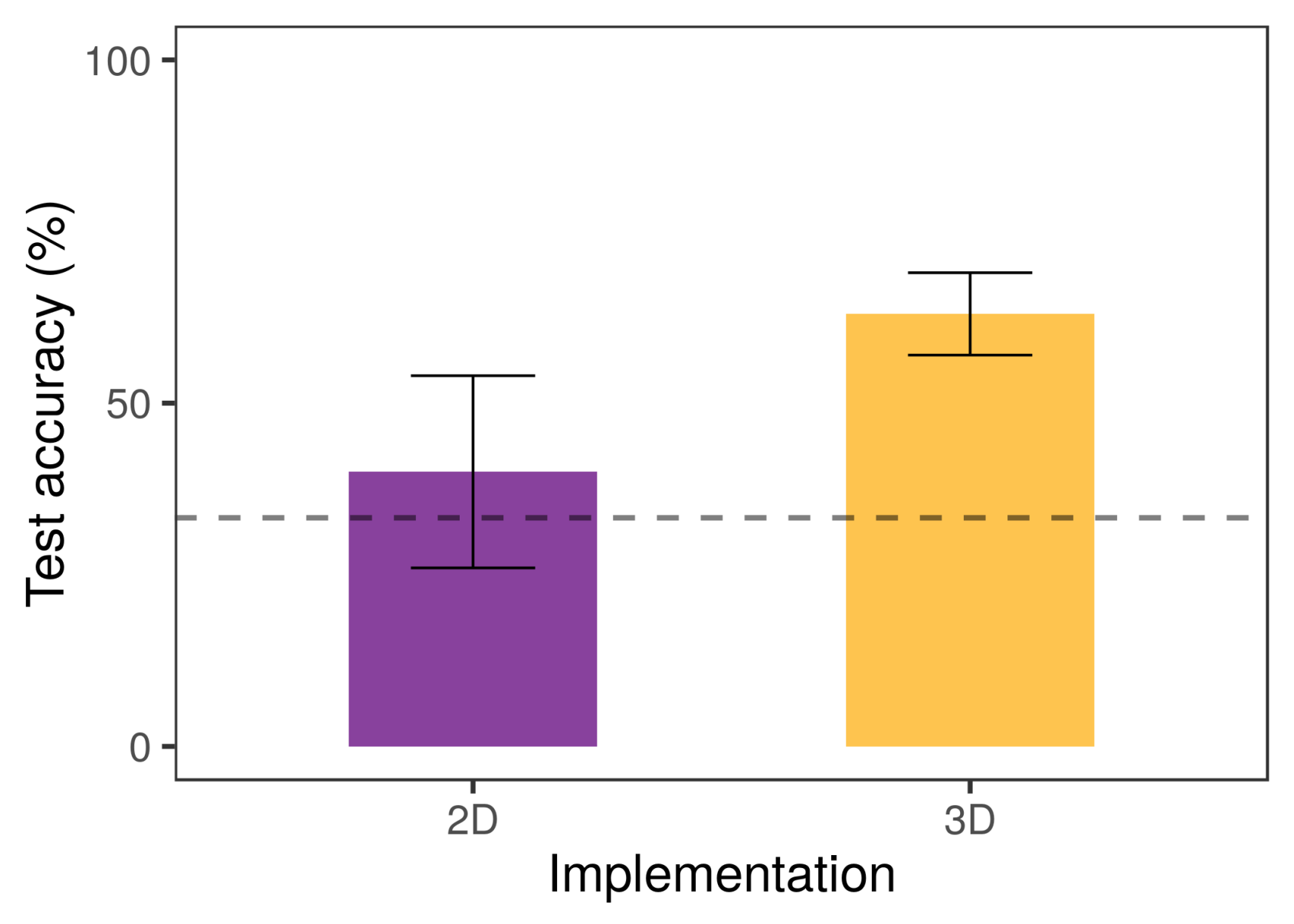


Same archit-ecture

Different instantiations

Same abstract task

Verb-noun composition: 2D vs. 3D

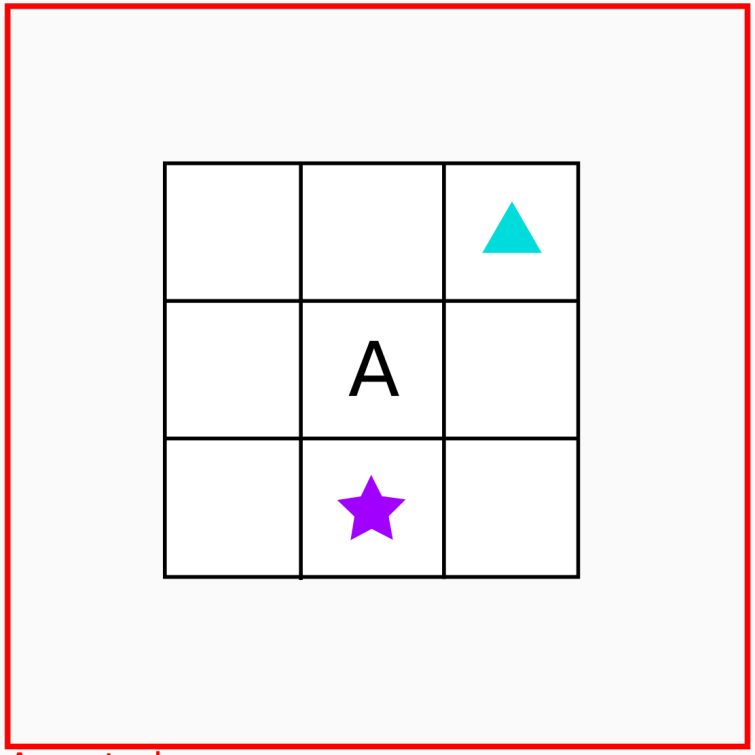


Generalization is better in the 3D environment. Why?

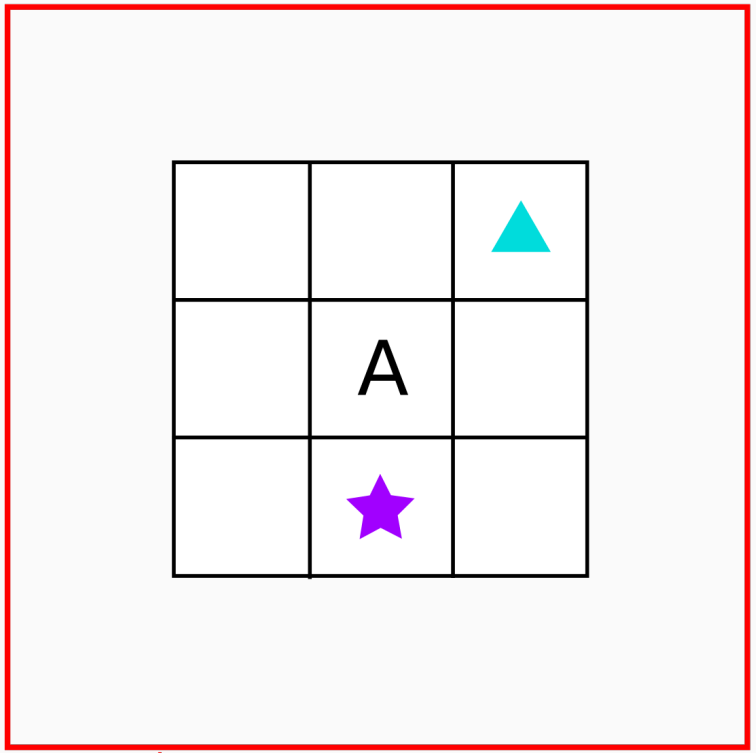
Frames of reference

The first-person perspective of the 3D agent adds a certain amount of invariance, so we compared two frames of reference in 2D:

Allocentric

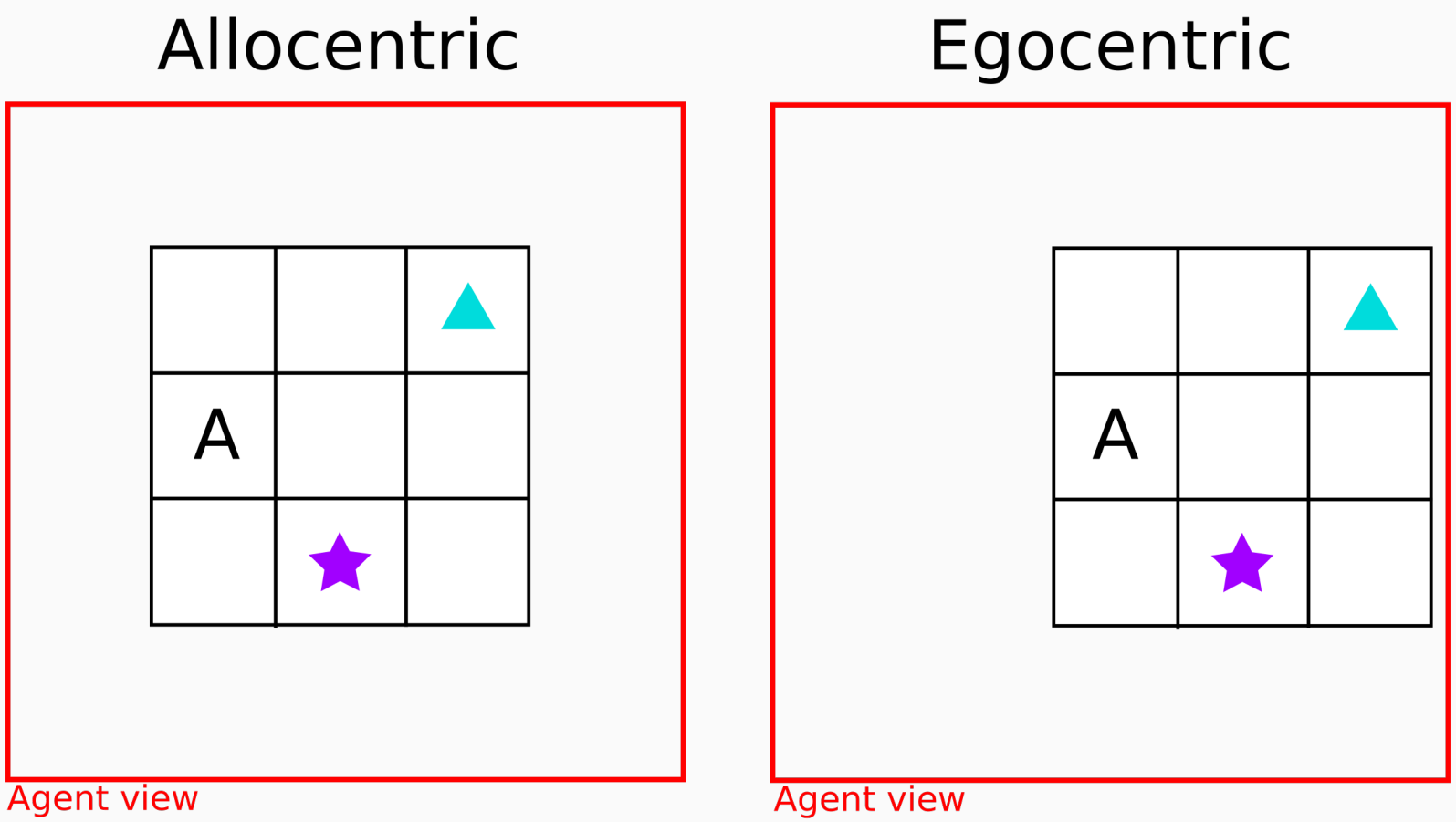


Egocentric



Frames of reference

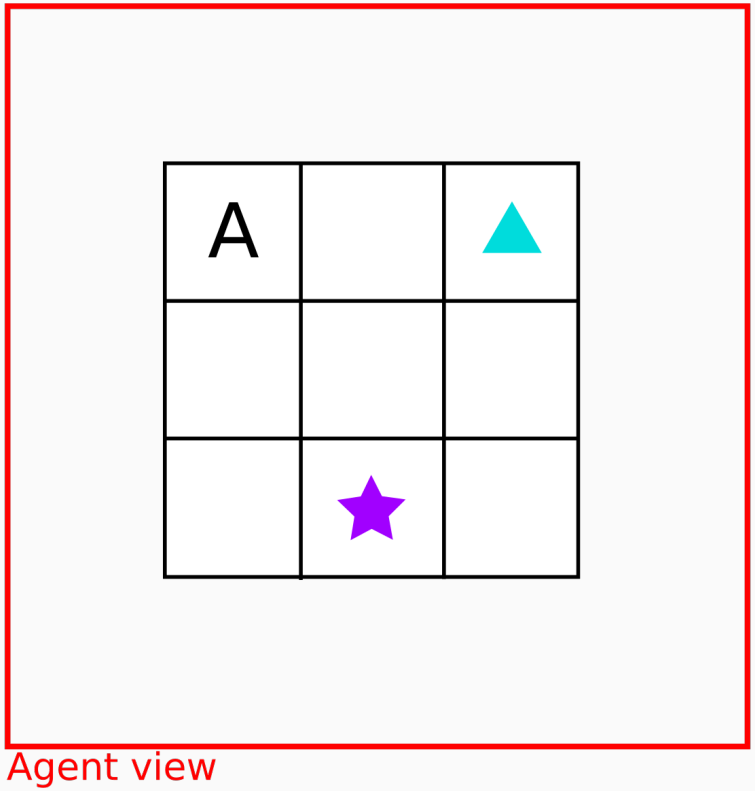
The first-person perspective of the 3D agent adds a certain amount of invariance, so we compared two frames of reference in 2D:



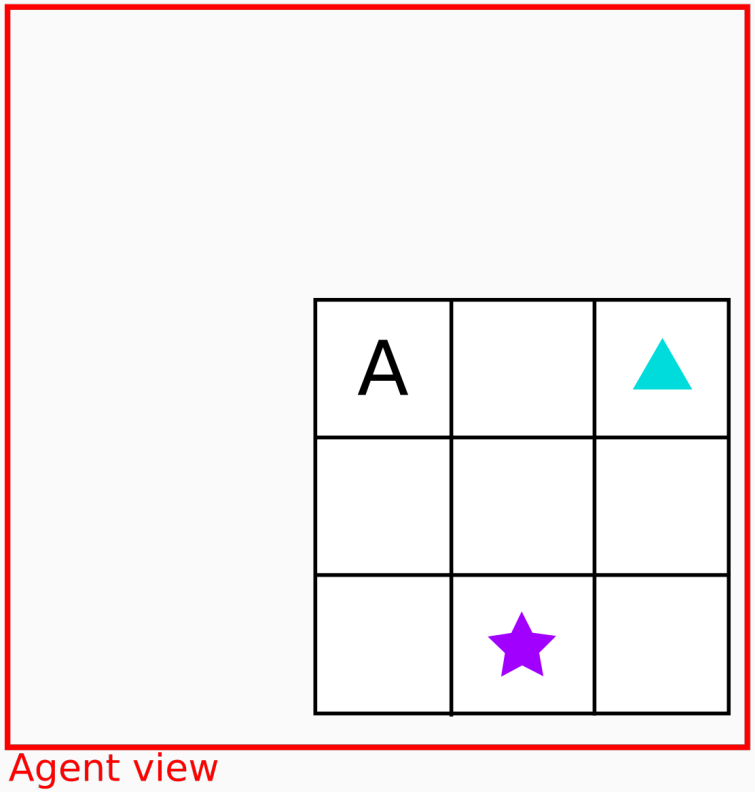
Frames of reference

The first-person perspective of the 3D agent adds a certain amount of invariance, so we compared two frames of reference in 2D:

Allocentric



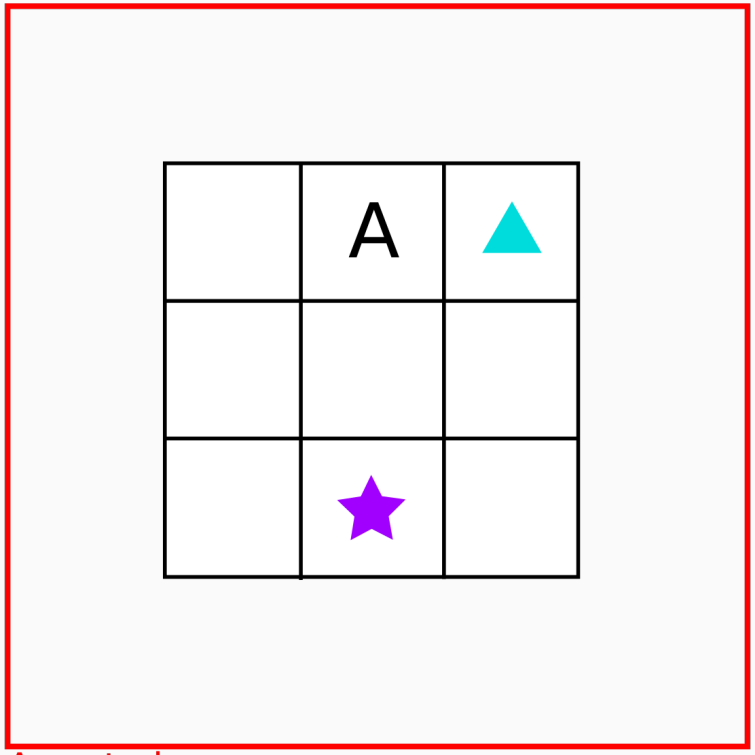
Egocentric



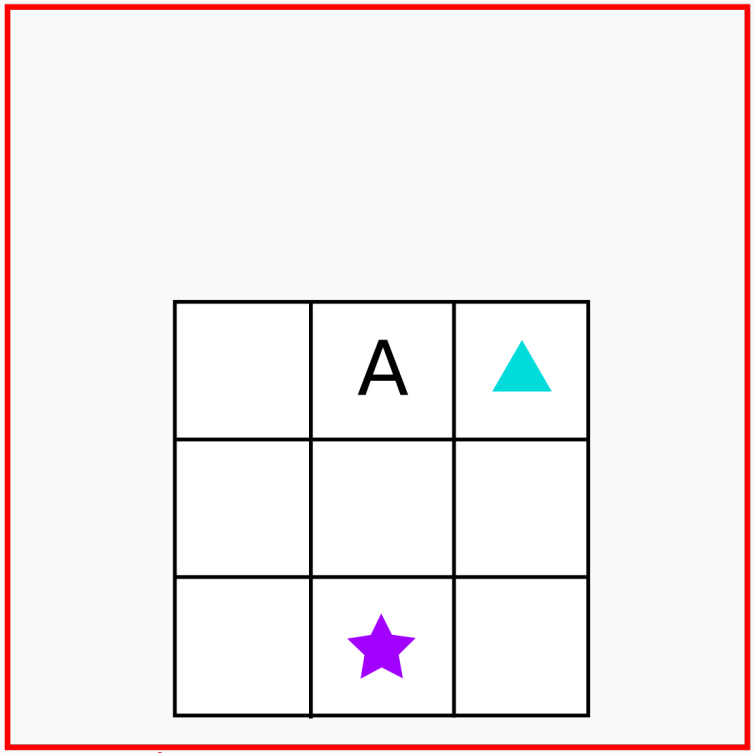
Frames of reference

The first-person perspective of the 3D agent adds a certain amount of invariance, so we compared two frames of reference in 2D:

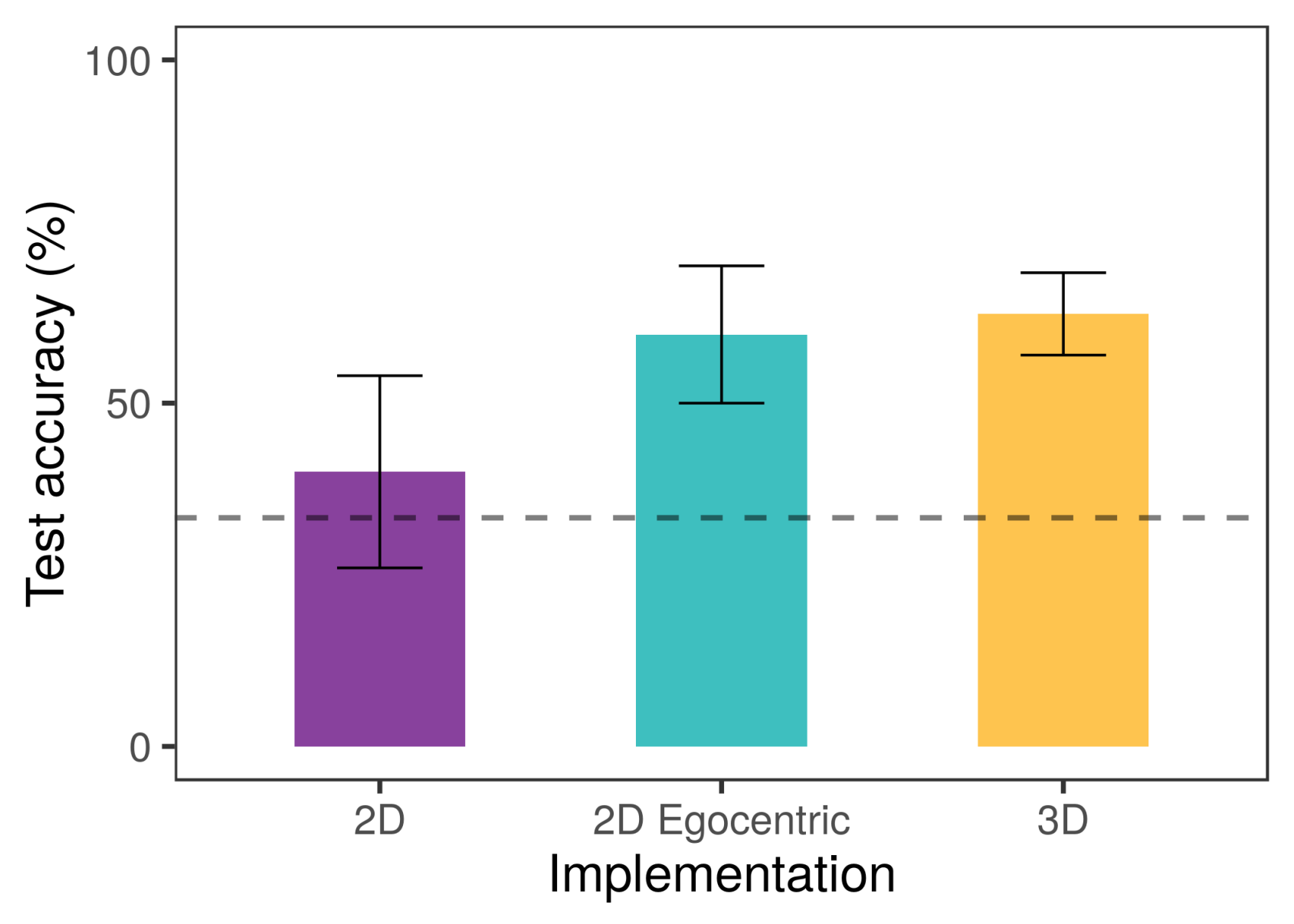
Allocentric



Egocentric



Egocentric perspective helps



Generalization in the 2D environment with an egocentric perspective is closer to 3D!

Again, closer to human experience

What does this mean?

- It's not just the abstract task that matters for generalization, but the specifics of the setting in which it is instantiated.

What does this mean?

- It's not just the abstract task that matters for generalization, but the specifics of the setting in which it is instantiated.
- Embodiment may improve generalization – performance is better when the agent acts rather than classifying a frame, and when it has an egocentric perspective.

What does this mean?

- It's not just the abstract task that matters for generalization, but the specifics of the setting in which it is instantiated.
- Embodiment may improve generalization – performance is better when the agent acts rather than classifying a frame, and when it has an egocentric perspective.
- Both of these changes make the agent's experience more like that of humans (or animals).

What does this mean?

- It's not just the abstract task that matters for generalization, but the specifics of the setting in which it is instantiated.
- Embodiment may improve generalization – performance is better when the agent acts rather than classifying a frame, and when it has an egocentric perspective.
- Both of these changes make the agent's experience more like that of humans (or animals).
- It's not obvious that egocentric perspective should help with verb-noun recomposition – the agent is solving the navigation problem perfectly on the trained "put" tasks either way.

What does this mean?

- It's not just the abstract task that matters for generalization, but the specifics of the setting in which it is instantiated.
- Embodiment may improve generalization – performance is better when the agent acts rather than classifying a frame, and when it has an egocentric perspective.
- Both of these changes make the agent's experience more like that of humans (or animals).
- It's not obvious that egocentric perspective should help with verb-noun recomposition – the agent is solving the navigation problem perfectly on the trained "put" tasks either way.
- We lack a theory for which features affect generalization.

What does this mean?

- It's not just the abstract task that matters for generalization, but the specifics of the setting in which it is instantiated.
- Embodiment may improve generalization – performance is better when the agent acts rather than classifying a frame, and when it has an egocentric perspective.
- Both of these changes make the agent's experience more like that of humans (or animals).
- It's not obvious that egocentric perspective should help with verb-noun recomposition – the agent is solving the navigation problem perfectly on the trained "put" tasks either way.
- We lack a theory for which features affect generalization.
- So we shouldn't dismiss deep RL as a cognitive model just because it fails in unrealistic tasks.

Thanks!

https://arxiv.org/abs/1910.00571