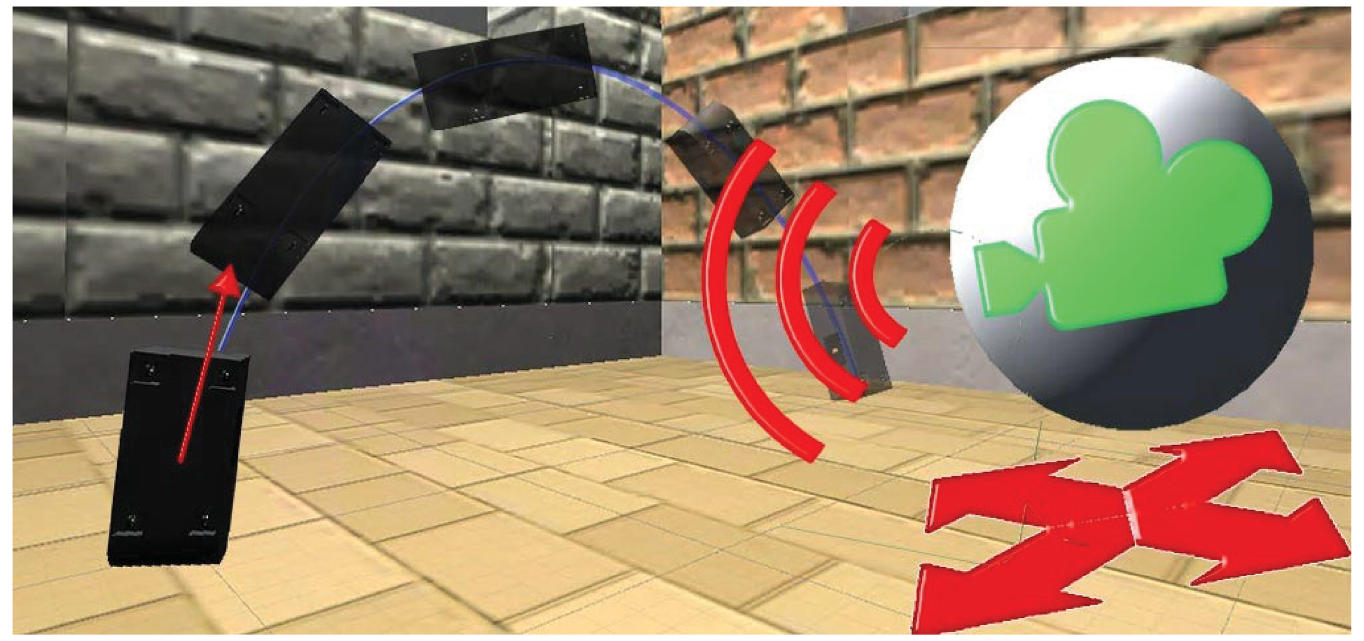
Learning to Play with Intrinsically-Motivated Self-Aware Agents

Stanford Neuro AI Lab

Nick Haber[1,2,3]*, Damian Mrowca[4]*, Stephanie Wang[4], Li Fei-Fei[4], Daniel L. K. Yamins[1,4,5]

Department of Psychology[1], Pediatrics[2], Biomedical Data Science[3], Computer Science[4], and Wu Tsai Neurosciences Institute[5], Stanford, CA 94305

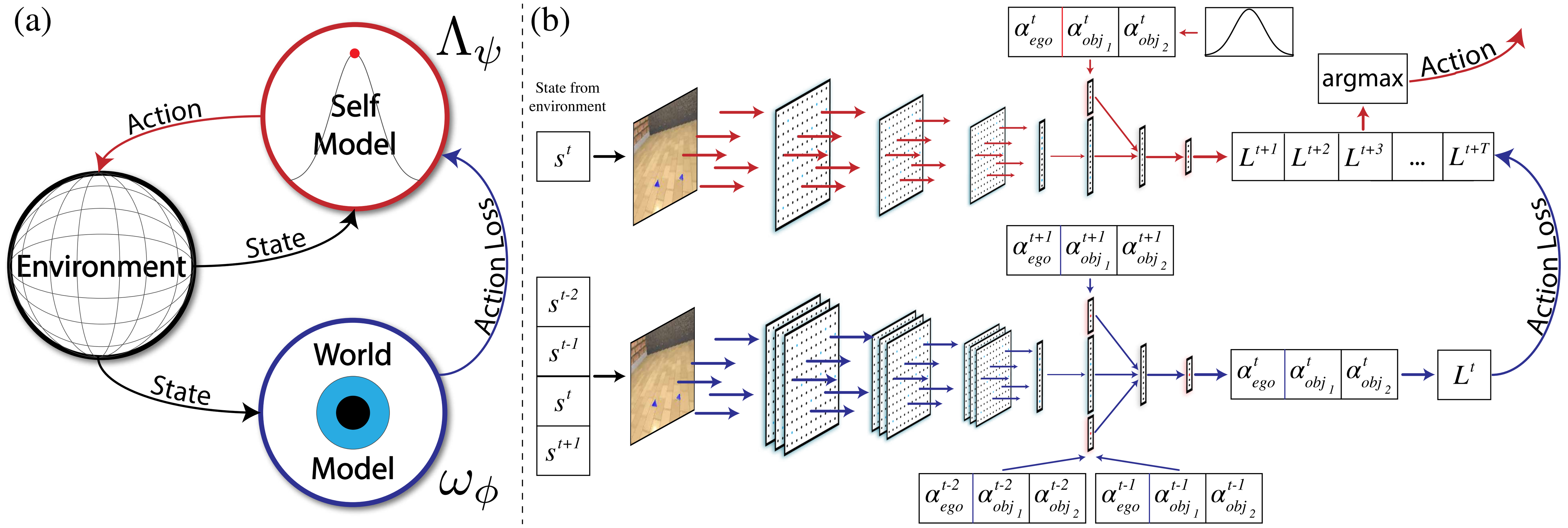
Introduction



Learning in a physically realistic embodied environment. Agents move around and interact with objects.

Artificial intelligent robots cannot learn in unstructured environment without supervision. ↔ Babies learn in unstructured environments through intrinsically motivated curious playing.

Approach: Self-Model Vs World-Model



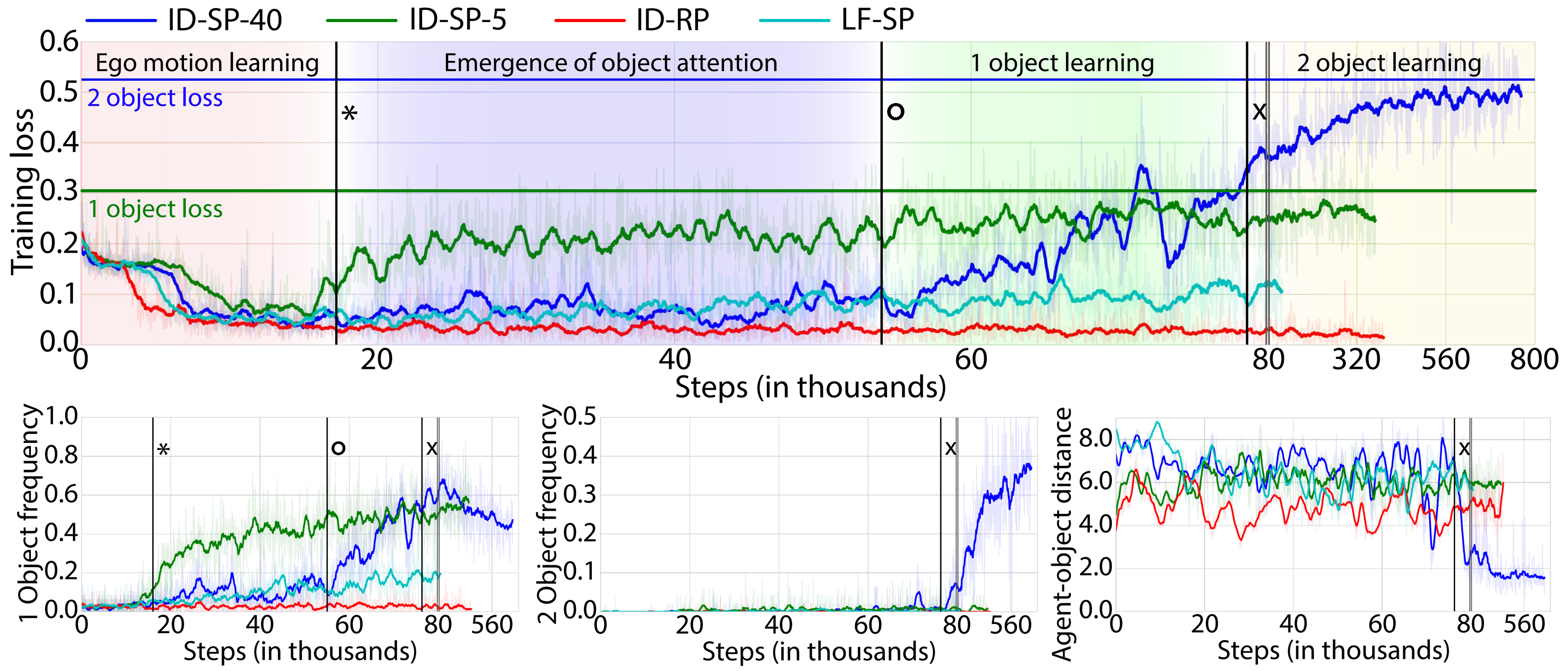
Intrinsically-motivated self-aware architecture.

Curiosity-based learning mechanism

- The world model (blue) solves a dynamics prediction problem.
- The self-model (red) seeks to predict the world-model's loss and is learned simultaneously.
- Actions are chosen to antagonize the world-model, leading to novel and surprising events in the **environment (black)**.

- This creates a virtuous cycle in which the agent chooses novel but predictable actions.
- Playful behavior emerges as the agent pushes the boundaries of what its world-model-prediction systems can achieve.
- As world-modeling capacity improves, what used to be novel becomes old hat, and the cycle repeats.

Emergent Behaviors



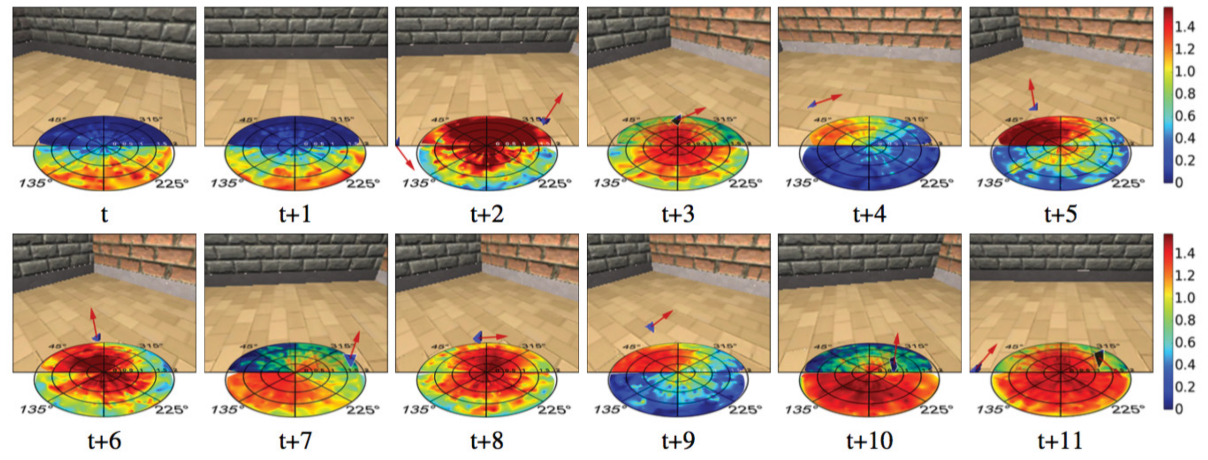
Emerging behaviors across different models. Different combinations of world-model tasks and policy mechanisms are compared.

ID = Inverse Dynamics prediction; RW = Random World model; RP = Random Policy; SP = Self-aware Policy; LF = Latent Future prediction

A sequence of behaviors emerges:

Ego motion prediction → Object attention → Navigation towards ob- → Multi-object gathering → Improved object dynamics predic- → Improvement on task transfers

Planning

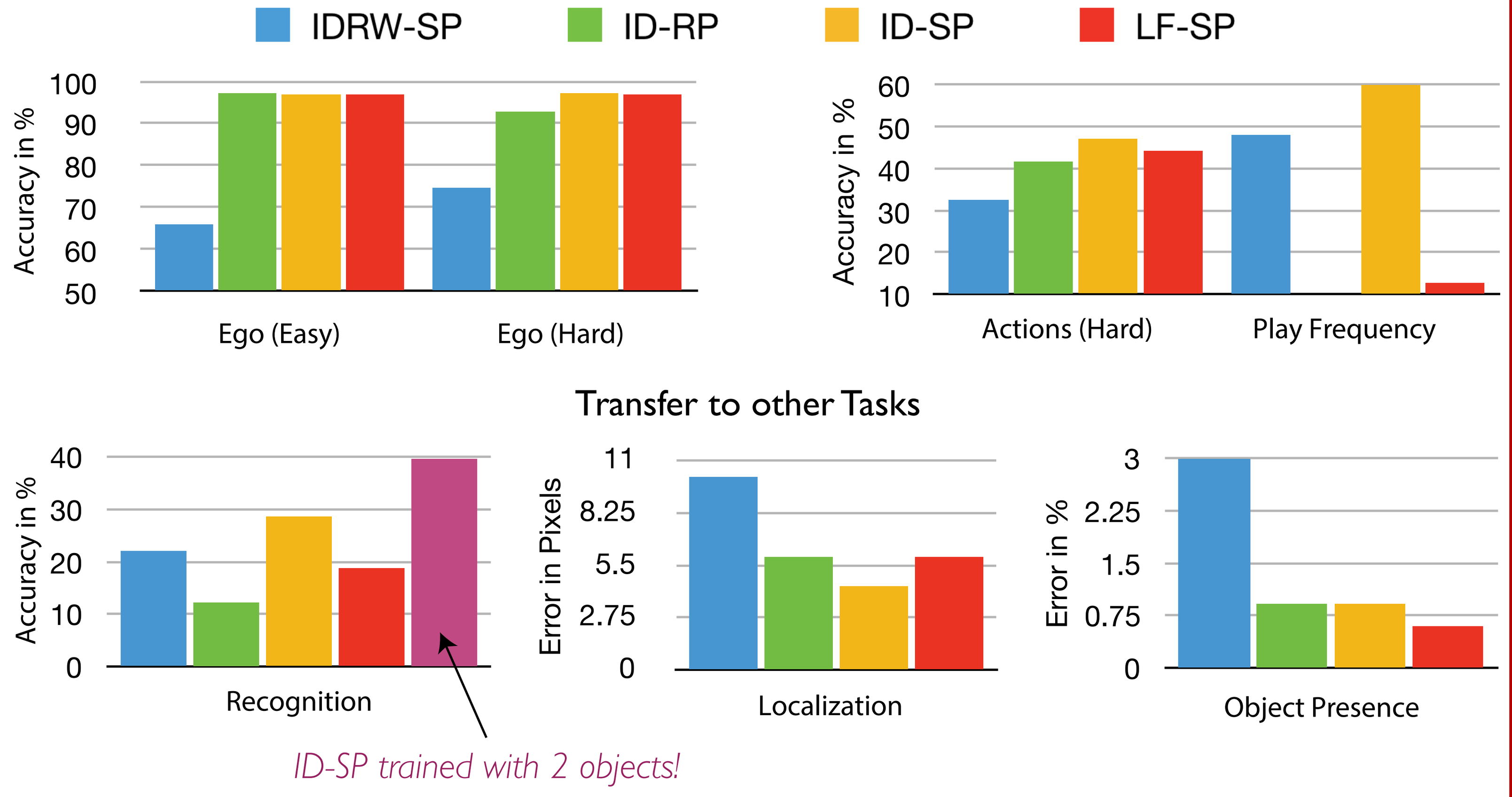


Navigation and planning behavior for 12 consecutive time steps. Red force vectors on the objects depict the predicted actions. Ego-motion self prediction maps are drawn at the agents position.

The agent starts off without seeing an object and turns around to explore for an object. Once an object is in view the agent approaches or turns towards the object to keep it in view.

This work was supported by grants from the James S. McDonnell Foundation, Simons Foundation, and Sloan Foundation (DLKY), a Berry Foundation postdoctoral fellowship and Stanford Department of Biomedical Data Science NLM T-15 LM007033-35 (NH), ONR - MURI (Stanford Lead) N00014-16-1-2127 and ONR - MURI (UCLA Lead) 1015 G TA275 (LF).

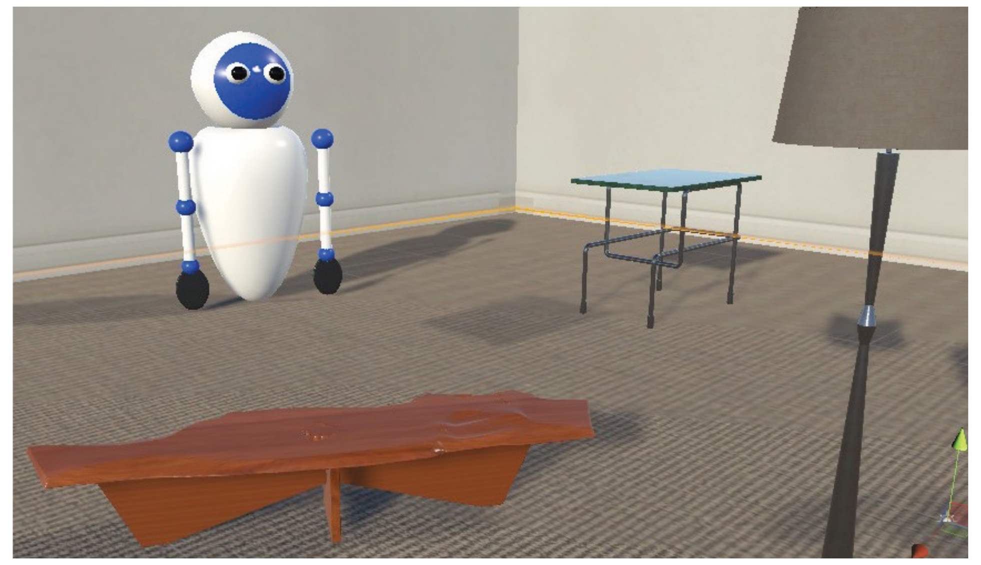
Task Transfers



Model comparison on task transfers including object presence (binary classification), localization (pixel-wise 2D centroid position), and recognition (16-way categorization). Linear estimators were trained on top off the output features of each model.

ID = Inverse Dynamics; RW = Random World model; RP = Random Policy; SP = Self-aware Policy; LF = Latent Future

Future Work



Phyiscally and visually realistic simulation environment.

Next steps include

- more photorealistic simulation
- more realistic agent embodiment
- curiosity towards the novel but learnable
- animate attention and theory of mind
- comparision to human developmental data