A Multi-Agent World Model with Self-Driven Curiosity Coordinated Exploration

Annika Treutle, Thomas Kling, Luca Sailer

Multi-Self Model Agents

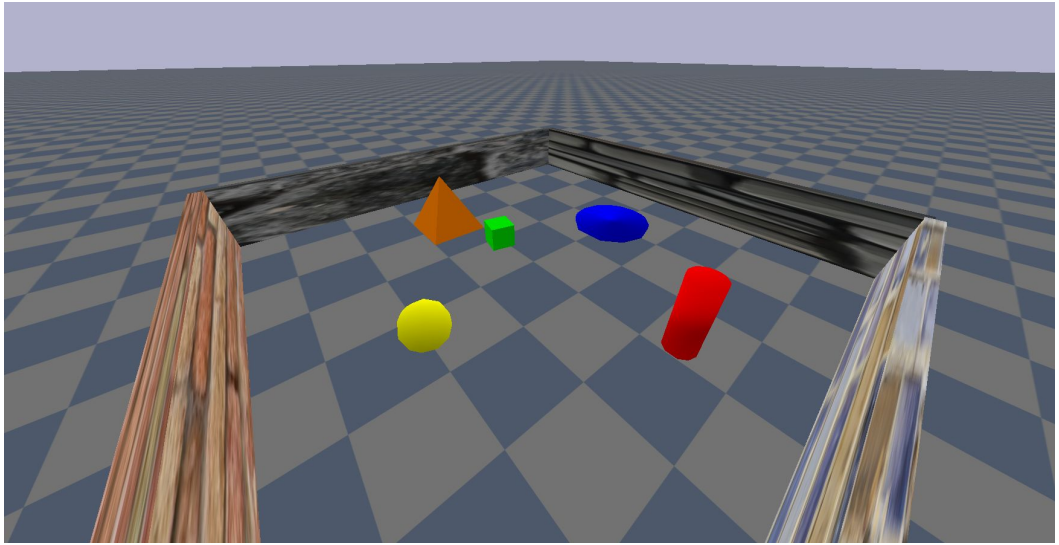
- Learning to Play With Intrinsically-Motivated, Self-Aware Agents

- Goal: A multi-agent environment in which agents learn through curiosity

Gemini (2025)

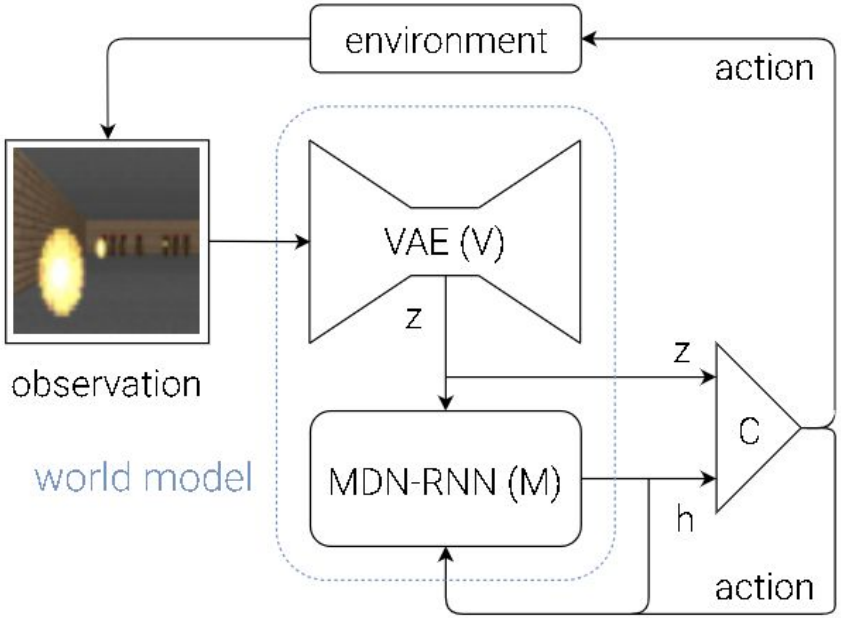
Haber, Nick, et al. "Learning to play with intrinsically-motivated, self-aware agents." *Advances in neural information processing systems* 31 (2018).

Environment



- Environment: A closed room built in PyBullet, a physics simulator.
- Agents: Two cube-shaped agents that can move, turn, and push objects.
- Objects: Four static, pushable objects (pyramid, cylinder, sphere, disk) to create a dynamic environment.

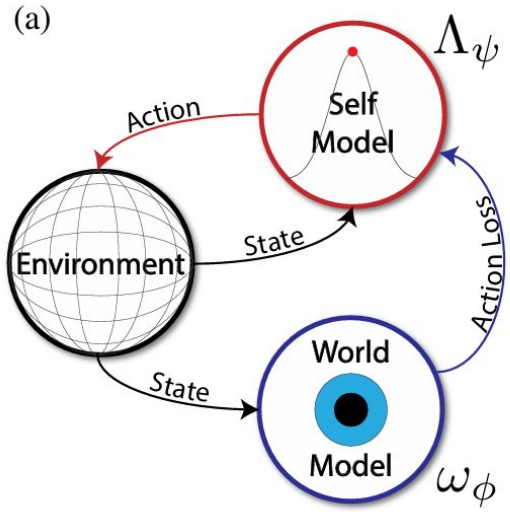
World Models



- **Advantages:** This provides a stable predictive foundation that can be trained efficiently in an unsupervised way.
- **Limitation:** A world model knows how the world works, but it doesn't provide the motivation to explore it.

Ha, D., & Schmidhuber, J. (2018). World models. *arXiv preprint arXiv:1803.10122*.

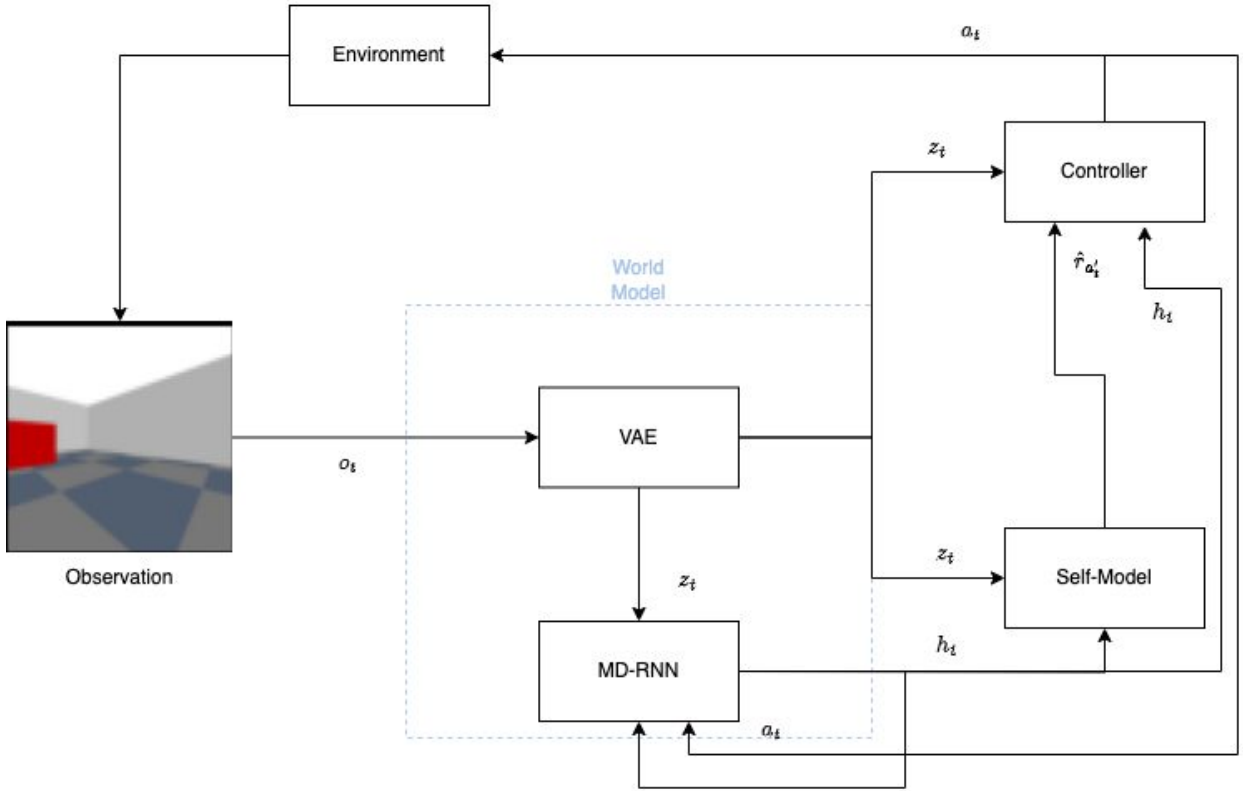
The Curiosity Engine



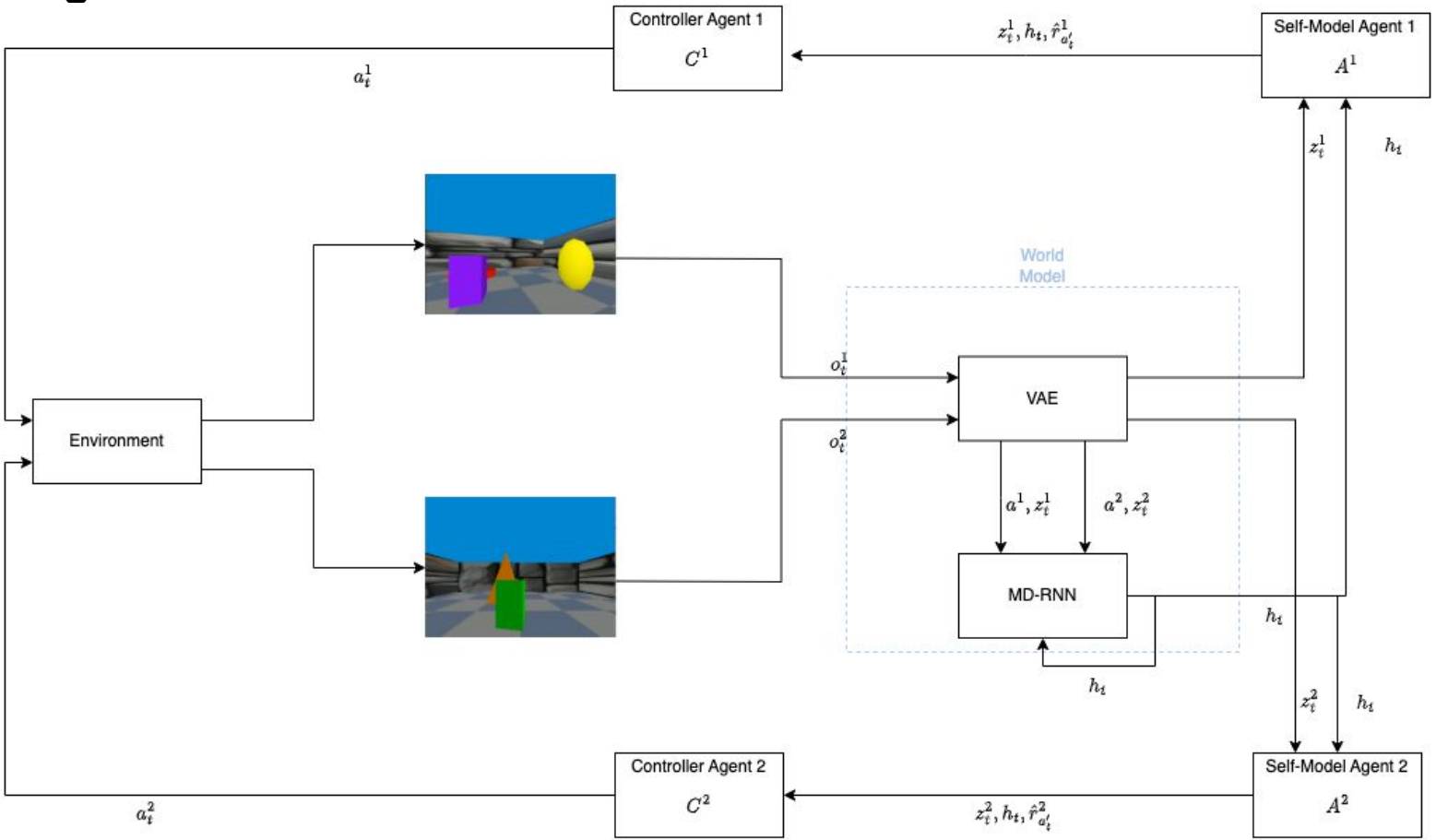
- An agent is "surprised" when its internal World Model fails to predict what will happen next
- The Goal: Encourage the agent to perform actions that will challenge its own understanding of the world

Haber, N., Mrowca, D., Wang, S., Fei-Fei, L., & Yamins, D. L. K. (2018). *Learning to play with intrinsically-motivated, self-aware agents*

Single Agent Architecture

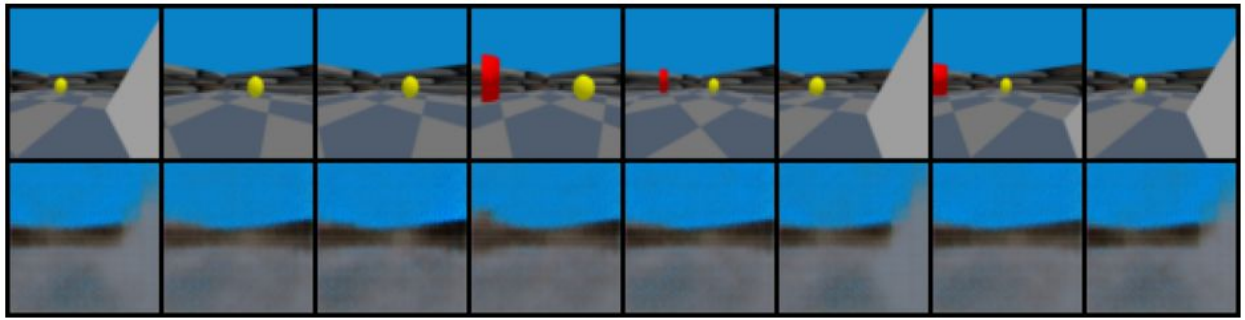


Multi Agent Architecture

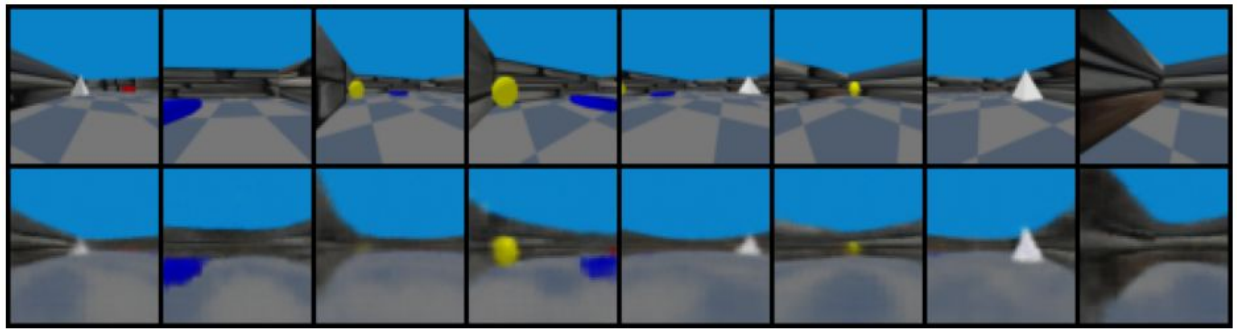


VAE Output

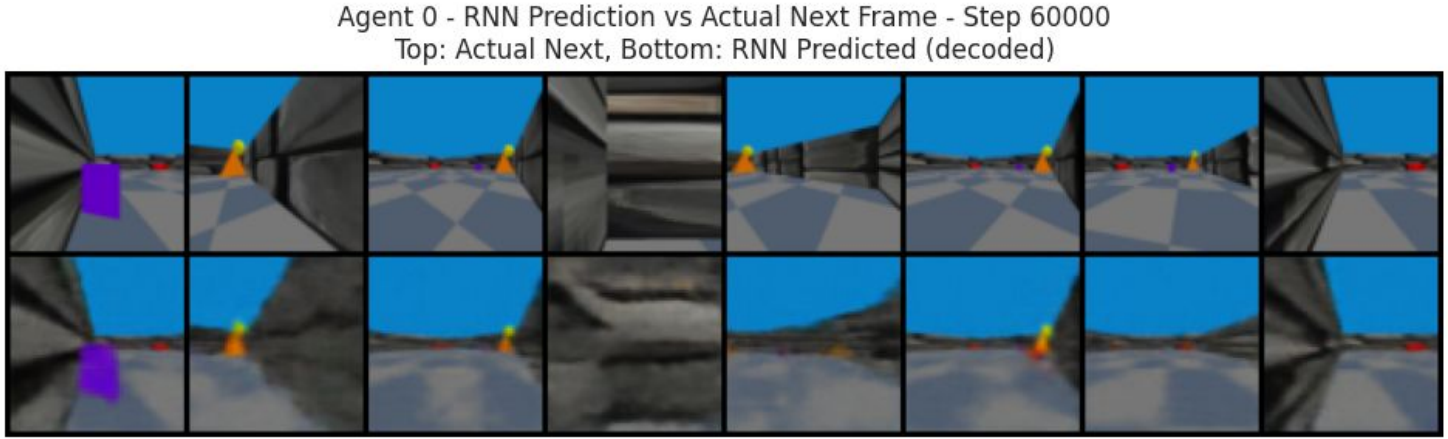
VAE Reconstructions - Step 500
Top: Originals, Bottom: VAE Reconstructions



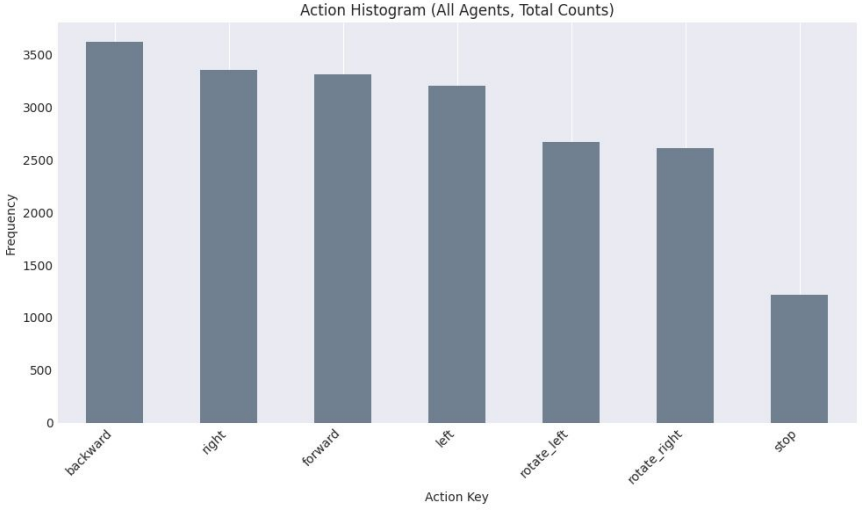
VAE Reconstructions - Step 79000
Top: Originals, Bottom: VAE Reconstructions



Prediction of the Multi Agents



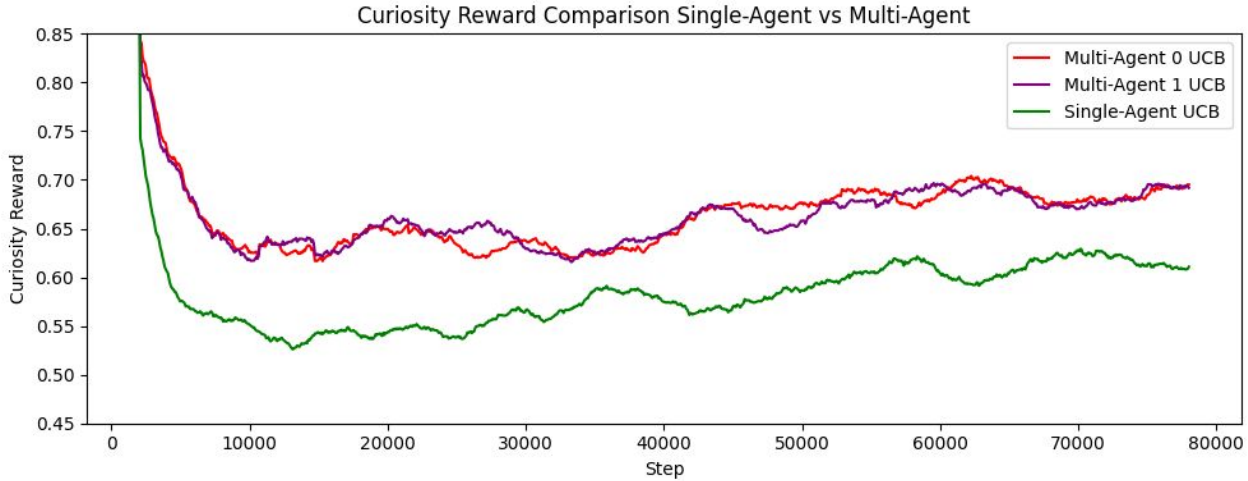
Key Result 1: The agent learns which actions are most relevant



- Policy used: ε-greedy with ε = 0.2
- “Stop” action:
- Agent stays in place for one time step and receives no new observation

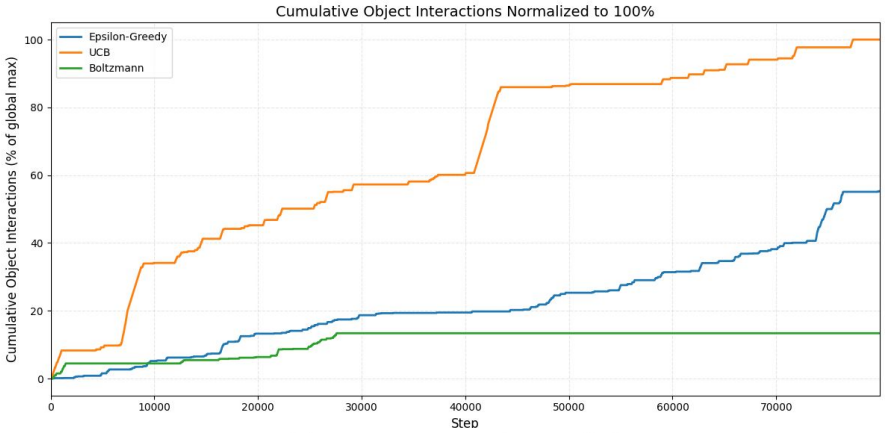
- More object interaction leads to higher rewards

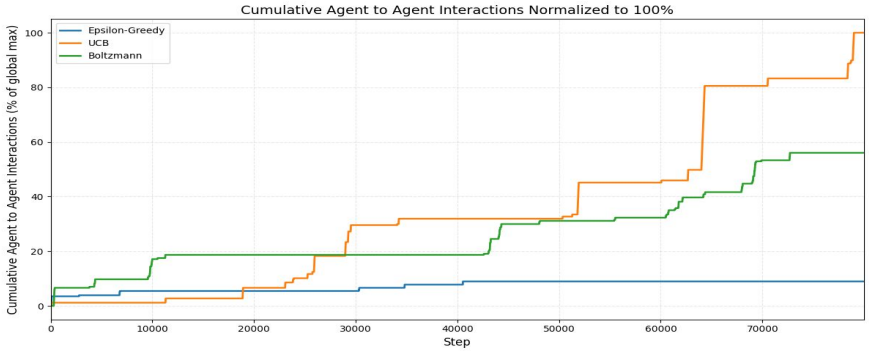
Key Result 2: Multi-Agent Curiosity is Higher



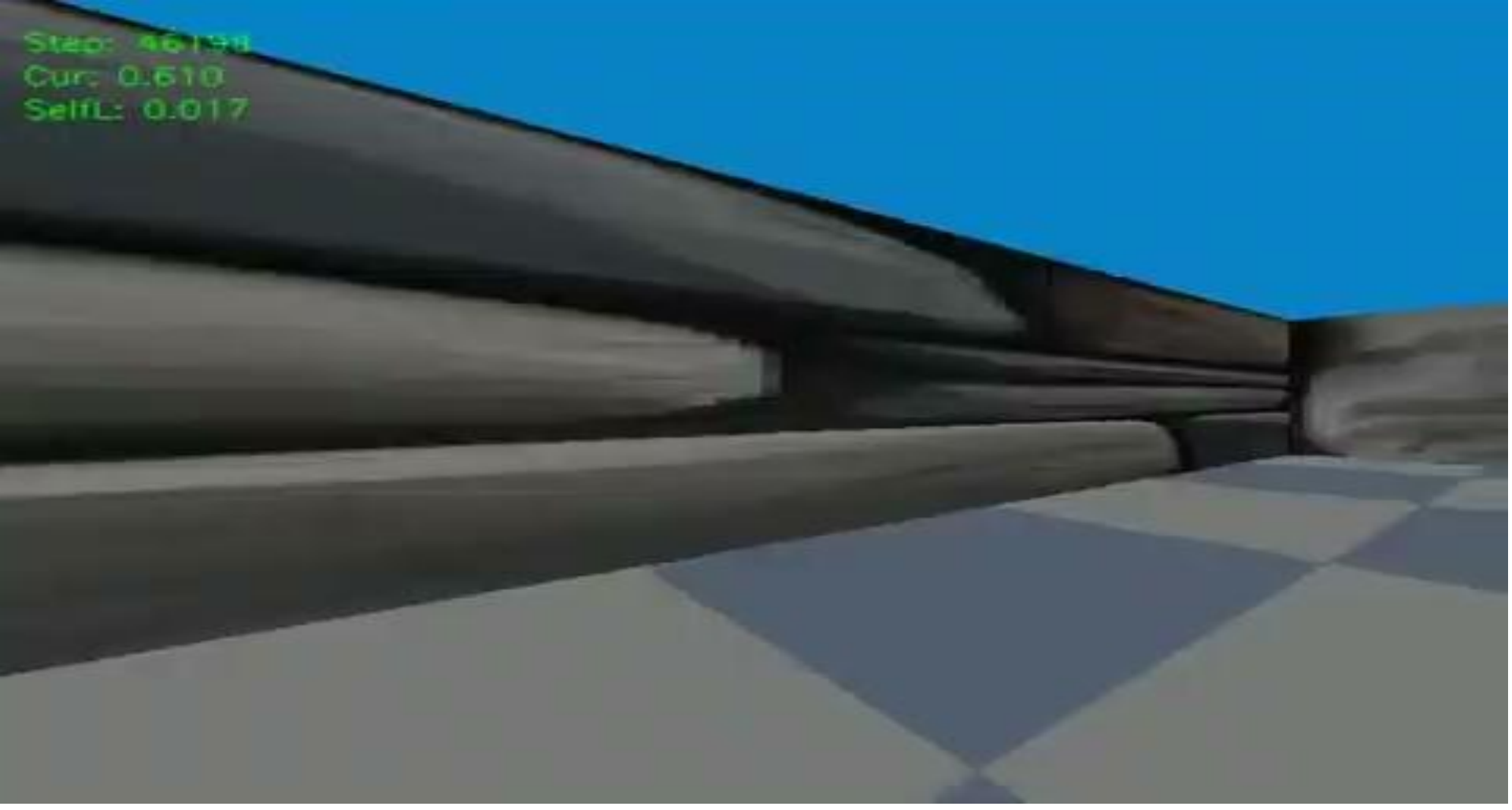
- The multi-agent environment is far more dynamic and unpredictable.
- An agent can be surprised not just by its own actions, but by observing the other agent moving objects.
- As a result, the intrinsic curiosity reward is significantly higher and more sustained in the multi-agent setting compared to the single-agent setting

Key Result 3: Policy Comparison





- We tested three exploration policies: ε-Greedy, Boltzmann, and Upper Confidence Bound (UCB).
- UCB was the most effective. It generated the highest number of both agent-object and agent-agent interactions.
- The Boltzmann policy performed the worst, leading to very few object interactions after an initial period.


Step: 461911
Cur: 0.610
SelfL: 0.017

Conclusion

- Extended a curiosity-driven RL framework to a multi-agent setting
- World model allows agents to learn environmental dynamics collectively
- Individual self-models drive independent, curiosity-based exploration
- Multi-agent environment higher curiosity
- UCB exploration policy proved most effective
- We experimented with multiple agent and environment configurations - varying objects and wall placements - to address issues like the agent's spurious interest in walls caused by shadow effects.