UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Investigating Object Permanence in Deep Reinforcement Learning Agents

Permalink
https://escholarship.org/uc/item/3g6575z9

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Voudouris, Konstantinos
Liu, Jason Darwin
Siwinska, Natasza
et al.

Publication Date
2024

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Investigating Object Permanence in Deep Reinforcement Learning Agents

Konstantinos Voudouris, Jason Darwin Liu, Natasza Siwinska, Lucy Cheke
({kv301, jl2191, nds45, lgc23}@cam.ac.uk)
Department of Psychology, University of Cambridge,
Downing Street, Cambridge, CB2 3EB, United Kingdom

Wout Schellaert
(wschell@vrain.upv.es)
Valencian Research Institute for Artificial Intelligence (VRAIN), Universitat Politècnica de València,
Camí de Vera 14, València E-46022, Spain

Abstract

Object Permanence (OP) is the understanding that objects continue to exist when not directly observable. To date, this ability has proven difficult to build into AI systems, with Deep Reinforcement Learning (DRL) systems performing significantly worse than human children. Here, DRL Agents, PPO and Dreamer-v3 were tested against a number of comparators (Human children, random agents and hard coded Heuristic agents) on three object permanence tasks (OP) and a range of control tasks. As expected, the children performed well across all tasks, while performance of the DRL agents was mixed. Overall the pattern of performance across OP and control tasks did not suggest that any agent tested except children showed evidence of robust OP.

Keywords: Object Permanence, Reinforcement Learning, Animal-AI

Introduction

Object Permanence (OP) is the understanding that objects continue to exist when not directly observable. Human adults use OP to reason about how objects behave and interact in the external world, and the development of OP in infants and children has been a topic of psychological research for over a century (Baillargeon, Spelke, & Wasserman, 1985; Piaget, 1952).

OP has proven difficult to build into AI systems. Deep Reinforcement Learning (DRL) systems perform significantly worse than human children when solving problems involving OP (Voudouris et al., 2022a). Tracking objects under partial occlusion appears to be difficult for modern computer vision methods ((Van Hoorick, Tokmakov, Stent, Li, & Vondrick, 2023)). The need for AI agents with robust OP is important for creating trustworthy embodied AI such as self-driving cars. Furthermore, robust object tracking under occlusion would have many applications in the field of robotics. However, the methods for evaluating whether an agent has OP suffer from a lack of precision, reliability, and validity. Developmental and comparative psychologists have been investigating OP in biological agents for around a century, developing many experimental paradigms along the way.

The **O**bject-**P**ermanence in **A**nimal-**A**i: **GE**neralisable **T**est **S**uites (**O-PIAAGETS**; Voudouris et al., 2022b) is a battery of Object Permanence tasks hosted in the **Animal-AI Environment** (AAI: Crosby et al., 2020; Voudouris et al., 2023), a 3D virtual environment with Euclidean geometry and Newtonian physics built in Unity (Juliani et al., 2020). The goal of any task in AAI is to obtain green and yellow rewarding objects (goals) while avoiding red 'death zones' before time runs out, to maximise final reward. O-PIAAGETS has an internal structure in which certain tasks are designed to test certain aspects of OP understanding. There is also a tailored training curriculum to facilitate out-of-distribution testing, and more direct comparison between biological and artificial machines. This work complements and extend the work of Piloto, Weinstein, Battaglia, and Botvinick (2022), Crosby et al. (2020), and Voudouris et al. (2022a, 2022b).

In this paper, we take a subset of the tasks from the "O-PIAAGETS" Battery (Voudouris et al., 2022b) to conduct a comparative analysis between four kinds of agent: random agents, which take randomly sampled actions; 'heuristic' agents, which follow a set of simple behavioural rules; DRL agents, which learn to take actions in response to observations; and human children aged 4-7 years old.

Three task-types were selected from O-PIAAGETS: Two from the sub-suite inspired by the Primate Cognition Test Battery (Herrmann, Call, Hernández-Lloreda, Hare, & Tomasello, 2007), and one inspired by Chiandetti and Vallortigara (2011) tests of intuitive physics in chicks (*Gallus gallus*). O-PIAAGETS contains versions of these tasks that require OP to be solved (OP Test Tasks), as well as versions that preserve the gross structure of the task, in terms of where obstacles and goals are placed, but do not require OP to be solved (Control Tasks). We use all the tasks from O-PIAAGETS within these experimental paradigms in this study. To test the basic capabilities that are required to interact successfully with AAI, a battery of simple tasks were also included (Basic Tasks). The chosen tests are relatively simple in terms of what is required of the agent, compared to some other tasks in O-PIAAGETS. They also contain a large number of variants, controlling for several alternative hypotheses.

It is hypothesised that while human children will perform well on all tasks (being developmentally at a stage where object permanence should be well established). In contrast, it is hypothesised that DRL agents lack OP, and will therefore only perform well on the Basic and Control Tasks. Previous work on a small number of tests of Object Permanence (without explicit controls) indicates that contemporary DRL agents are not able to learn to track and search for occluded objects (Voudouris et al., 2022a). We establish alternative and null hypotheses using random and "heuristic" agents. We con-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* ©2024 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

sider robust evidence of object permanence ability to be represented not only by outperforming the random agents, but also in outperforming the 'heuristic' agent, which is designed to search for goals while avoiding obstacles, but which lacks any capability to remember objects that have been previously observed, a necessary component of OP.

Materials & Methods

Participants

Random Agents We provide a principled cohort of random agents to provide a diversity of random behaviour on these tasks that act as a "chance" baseline. We used three kinds of *random walker* and two kinds of *random action agent*. *Random walkers* (RWs) take a certain number of steps in the forwards-backwards direction (saccades) followed by a certain number of degrees of rotation in left or right (turn angles), and they repeat this until the end of the episode. The three RWs vary in how saccades and turn angles are selected. *Random action agents* (RAAs) take one of the nine actions available to them in AAI, repeating them for a variable number of steps selected from different probability distributions. The two RAAs use different biases and distributions to sample these actions - for example one is more likely than the other to stick with the same action for a period of time. The performance of these five agents was collapsed into a single population of Random Agents.

Heuristic Agent With the Heuristic Agent we simulate what performance we would expect from an agent that lacks object permanence, but possesses many other capabilities that would lead to success in AAI. This agent follows a simple set of rules: When it sees a goal, it orients and approaches towards it. When it sees a red 'death zone', it orients and moves away. If the agent is stationary or if there is a wall in front of it, the agent navigates around it.

Human Children Children aged 4–7 were recruited through word-of-mouth and social media in the Cambridge, UK area. This age cohort was selected to provide a range of performances on the test set, with a limited likelihood of ceiling effects, while maximising proficiency with computers.

Evaluation Agents: PPO and Dreamer

DRL agents based on two learning algorithms are evaluated in this study: Proximal Policy Optimisation (PPO; Schulman, Wolski, Dhariwal, Radford, & Klimov, 2017) and Dreamer-v3 (Hafner, Pasukonis, Ba, & Lillicrap, 2023). Both algorithms are considered amongst the state-of-the-art in contemporary Deep Reinforcement Learning. While PPO-based agents are evaluated in Voudouris et al. (2022a), to our knowledge, Dreamer-v3-based agents have never been evaluated using cognitive science experiments such as ours.

PPO is a model-free DRL algorithm that has gained prominence for its versatility and its impressive behaviour in game-like environments, including Atari games. It is part of a family of Policy Optimisation algorithms. Where other DRL algorithms, such as Deep-Q-learning seek to calculate the value of taking actions in certain states, converging on a policy that is greedy with respect to those action-values, policy optimisation involves manipulating the policy directly through learning, without calculating state- or action-values. Proximate Policy Optimisation blends policy optimisation with an efficient way to make effective updates of the policy parameters. The objective is to maximise the probability of taking an action that will lead to a reward.

In contrast to PPO, Dreamer-v3 is a model-based DRL algorithm, meaning that it constructs a latent model of the environment and how it expects it to evolve (by approximating the transition function). Dreamer-v3 has out-performed other existing DRL algorithms on numerous standard benchmarks, including Atari games and Minecraft (Hafner et al., 2023). As such, it is considered a state-of-the-art agent. Dreamer-v3 learns four components of the environment. First, it learns compressed representations of its observations of the environment. Second, it learns to predict the expected reward it will receive from each state of the environment. Third, it learns the value of being in particular states. Finally, it learns to generate predicted future states based on a sequence of past states and predictions, leading to an approximation of the transition function. By combining this information using several design innovations, agents trained using the Dreamer-v3 algorithm learn to maximise reward in a diverse range of environments.

Tasks

Basic Tasks Basic Tasks acted as a comparison to confirm whether agents have the basic abilities needed to successfully interact with the environment. We use a series of tasks that involve the basic objects present in the OP tasks in simple combinations. This includes instances where the goals rolls from one side to the other, or is on a blue platform obtainable by navigating up a ramp. There are some tasks that require detouring around opaque and transparent walls and death zones, or to make simple choices (Fig. 1A). Children played six of these simple tasks during a tutorial without a time limit. Agents played the full set of 240 tasks.

Primate Cognition Test Battery (PCTB) Tasks Two PCTB paradigms from O-PIAAGETS were included in this study. The first is called the *Three Cup Task* (Fig. 1B). The agent/player observes a goal fall into one of three occluded areas, and must choose which ramp to climb to retrieve it. The ramps allow the agent to enter the cup but not exit it. The control version of this task mimics this set up, but with transparent or absent walls, such that the goal does not become occluded. Agents played 432 Control Tasks and 1512 OP Test Tasks. Children played 1 Control Task and 10 OP Test Tasks. The second paradigm is called the *Grid Task*(Fig. 1C). Here the agent/player witnesses the goal fall into one of a series of (4, 8, or 12) holes in the floor from atop a large ramp. In the object permanence version of the task, once the goal enters the hole it is occluded. They must then navigate to and drop down into the correct hole. In the control version, the holes

are shallow and goals are not occluded. Agents played 240 Control Tasks and 192 OP Test Tasks. Children played two Control Tasks and ten OP Test Tasks.

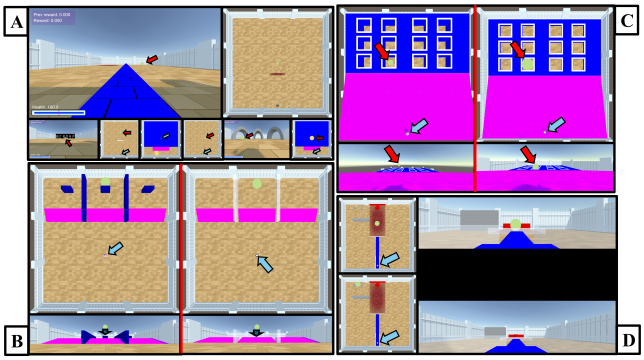

Figure 1: Tasks from O-PIAAGETS: Red arrows denote the location of goals and grey arrows denote the location of the agent. **A:** Basic Tasks. Top Left: A forced choice task—the agent is spawned on a platform with a goal on the right and a death zone on the left. Top Right: An avoidance task—the agent must navigate around the death zone to obtain the goal. Bottom: A selection of further basic tasks. **B:** PCTB Cup Task. Left: An object permanence version of the task. Right: A control version of this task. **C:** PCTB Grid Task. Left: An object permanence version of the task. Right: A control version of this task. **D:** CV Chick Object Permanence Task. Right: Three stages of the task: rolling goal, blackout, occluded reward.

Chiandetti & Vallortigara Tests of Intuitive Physics Henceforth, these tasks are called "CV Chick Tasks". In these tasks, a goal rolls from the centre of the arena to one side, either to be occluded behind a wall (OP Test) or remaining visible (Control) (Fig. 1D). The environments may have a wall on either side, or a wall on only one side. The agent/player is frozen while watching the goal roll, to prevent premature movement. In some tasks ('blackout'), the 'lights' are turned off at the moment where the goal is still equidistant from the two sides of the arena. When the 'lights' come back on, the agent/player must infer the location of the goal based on trajectory information or inference (i.e., if there is only one occluder, or one occluder is too low to occlude the goal). Agents played 1454 Control Tasks and 126 OP Test Tasks. Children played two Control Tasks and 18 OP Test Tasks.

Task Overview

Agents were tested on a total of 4202 tasks (including 1830 OP Test Tasks), while children were tested on a subset of 51 these tasks (including 38 OP Test Tasks), considerably more than the agents and children in Voudouris et al. (2022a; 90 and 4 OP Test tasks respectively). Children could not play all the tasks due to time constraints.

The PCTB tasks and the CV Chick tasks are an important dyad, because they control for distinct hypotheses about behaviour. One sophisticated policy that could lead to good performance on PCTB tasks (as well as many standard OP Test Tasks in O-PIAAGETS) would be to navigate to where the goal was last observed, and resume searching from there. This does not require the agent to *understand* that objects continue to exist when occluded. This would not be a successful strategy for the CV Chick tasks, which use 'blackout' periods to render such a heuristic ineffective. In the other direction, the CV Chick task could be solved by an agent that navigates behind the largest wall in the scene. This policy would lead to success on many (but not all) CV Chick tasks, but it would not lead to success on the PCTB tasks, where there are either 3 walls of equal width in the case of the Three Cup tasks, or no walls at all in the case of the Grid tasks. Thus, using both the PCTB tests and the CV Chick tests is a particularly useful subsection of O-PIAAGETS to focus on.

Procedure

Random & Heuristic Agents The Random Agents and the Heuristic Agent were evaluated on 4202 tasks. Five random number generator seeds were used to initialise each agent.

Dreamer-v3 & PPO We trained Dreamer-v3 and PPO on 5 different curricula, resulting in 5 different trained agents (see Table 1). Dreamer-v3 was trained on a High Performance Computer at the Universitat Politècnica de València on NVIDIA A40 48GB GPUs and PPO was trained on a NVIDIA RTX 4090 GPU through a major cloud computing provider. Both agents were then evaluated on 4202 tasks.

Human Children Data from children was collected between January and March (inclusive) 2023, with a minimum target sample size of 30. Participants conducted the study at the Department of Psychology, University of Cambridge, in the presence of their guardians and researchers. Guardians were provided with an information sheet and the opportunity to provide informed consent. Participants watched a short video introducing the 'Get the Fruit!' game in which they were asked to assist 'Farmer John' in finding all the green and yellow 'apples' in the farmyard, while avoiding 'lava' (death zones).

Participants were then introduced to AAI, using a small hand-held controller. First, they played a 'tutorial' round, consisting of 11 Basic and Control Tasks, which they could play as many times as they wished. They then played the 'test' round, consisting of 38 OP Test tasks. These tasks were ordered randomly, with half of participants playing the fixed random order and the other half playing the reverse of that order. Every 10 tasks, the participant could take a break from the game, and receive stickers as a reward. During the test round, guardians filled out a short survey about the participant, with questions on age, gender, and video-game playing habits. Upon completion of the test round, participants and their guardians were debriefed and offered a £5 book voucher for their participation.

Table 1: The 5 different trained agents for each architecture (Dreamer-v3 & PPO; 10 total), their training curricula, and the number of training steps (in millions). Stratification was done by Paradigm.

Agent Name	Training Curriculum	Training Steps
PPO/Dreamer 1	All Basic tasks	2M
PPO/Dreamer 2	All Basic tasks + a stratified random sample of 300 Control tasks	4M
PPO/Dreamer 3	All Basic tasks + all Control tasks (in 3 randomly sampled batches)	8M
PPO/Dreamer 4	All Basic tasks + a stratified random sample of 300 Control tasks + a stratified random sample of 300 Test tasks	6M
PPO/Dreamer 5	All Basic tasks + all Control tasks (in 3 randomly sampled batches) + a stratified random sample of 300 Test tasks	10M

This study was reviewed and approved by the Cambridge Psychology Research Ethics Committee, Department of Psychology (PRE.2020.024).

Analysis

AAI returns step-by-step information on the agent's position, velocity, and current reward. Given the number and size(s) of the goal(s) present, we determine a *pass mark*, the minimum possible reward that the agent could finish the task with after having collected all goals. This provides a success/failure measure for each task.

We provide results for each of the seven groups of tasks (Basic, CV Chick Control, CV Chick OP Test, PCTB Three Cup Control, PCTB Three Cup OP Test, PCTB Grid Control, PCTB Grid Test). In terms of the proportion of instances each type of agent passed. While we also analysed trajectory information for choices made, reporting of this is beyond the scope of this short paper.

We use generalized linear mixed effects models with logit links to investigate two questions. First, for each type of task, we investigated the significance of the difference in odds of success for each agent compared to the Random Agent (defining 'chance' performance). Second, for each agent on each of the three paradigms, we also investigate the significance of the difference in odds of success on the OP Test Tasks compared to the Control Tasks. Random slopes and random effects for participant were used to capture the within-subjects nature of the data. These models were fitted using the `MixedModels.jl` library. p-values are not corrected for multiple testing. Data and code for agent training, evaluation, and data analysis, along with test statistics, odds ratios, and p-values, can be found at https://github.com/Kinds-of-Intelligence-CFI/comparative-object-permanence.

Results

Data were collected from 5 types of random agent, each run with 5 RNG seeds (n=25), 1 type of simple goal-directed agent run with 5 RNG seeds (n=5), 5 PPO agents, and 5 Dreamer-v3 agents. Data from 4-year-old (n=6), 5-year-old (n=15), 6-year-old (n=6), and 7-year-old (n=3) children (total n=30). Of the children, all were identified as either male (n=21) or female (n=9) by their guardians. A small number of children did not complete the full study (n=3). We include the data for all instances that they did complete.

Across all instances that were played, children (73.63%) passed considerably more than any other agent (next highest, Heuristic agent: 42.79%; Table 2). Breaking down the results from children by age group demonstrates a clear upward performance trajectory across ascending age groups (4yo: 67.91%; 5yo: 72.70%; 6yo: 74.57%; 7yo: 85.96%). There was little difference between genders (females: 74.44%; males: 73.28%).

Basic Tasks On the basic tasks, both children and the heuristic agent performed almost at ceiling. The Dreamer and PPO agents trained explicitly on these tasks (PPO 1, Dreamer 1) also performed well, with significantly higher odds of success than the Random Agents. All other agents have similar or significantly lower odds of success than the Random Agents. Noticeably, agents trained further on more instances from the Control and Test sets failed an appreciably higher number of instances (see Table 2 and Fig. 2).

PCTB Three Cup Tasks On the OP Test Tasks, only children showed high performances, although they appeared to find this task harder than the other two overall (see Fig. 2). Only children and the Heuristic Agents have significantly higher odds of success than the Random Agents. All other agents have similar, or significantly lower, odds of success than the Random Agents (see Table 2).

A similar pattern is seen for the Control Tasks, with only children and the Heuristic Agents having significantly higher odds of success than the Random Agents (see Table 2). While children succeeded on a higher proportion of OP Test Tasks than Control Tasks, the difference in odds of success is not significant. In contrast, the Heuristic Agent has significantly lower odds of success on the OP Test compared to the Control Tasks, reflecting the requirement for OP, which this agent lacks (see Fig. 2).

PCTB Grid Tasks On the OP Test Tasks, children perform the best. All other agents did not have significantly different

Table 2: Percentage of instances of each task passed by each agent, with the significance of the difference in odds of success compared to the Random Agent for the seven task types.

Agent	Overall	Basic		PCTB Cup Task				PCTB Grid Task				CV Chick Task			
				Control		OP		Control		OP		Control		OP	
Random Agent	8.91	34.44	–	12.50	–	6.65	–	8.72	–	3.27	–	7.15	–	3.08	–
PPO 1	9.09	75.20	***	5.43		0.00	–	28.33	***	0.00	–	11.57		0.00	–
PPO 2	3.71	34.55		3.47	***	3.70	***	0.00	–	0.00	–	0.00	–	0.00	–
PPO 3	9.42	37.40		6.48	**	0.33	***	36.67	***	0.00	–	12.45	***	1.59	
PPO 4	1.14	10.57	***	0.00	–	0.07	***	3.75	*	4.69		0.21	***	0.00	–
PPO 5	0.90	5.28	***	0.00	–	0.00	–	4.58		3.13		0.55	***	0.00	–
Dreamer 1	13.30	80.49	***	4.17	***	2.12	***	44.17	***	3.65		12.86	***	6.73	***
Dreamer 2	10.60	9.35	***	3.24	***	0.00	–	42.92	***	3.13		20.15	***	4.76	
Dreamer 3	20.16	18.29	***	1.16	***	0.00	–	70.42	***	5.73	*	40.58	***	21.43	***
Dreamer 4	15.25	10.98	***	6.48	**	7.01		20.83	***	5.73	*	26.62	***	25.40	***
Dreamer 5	1.67	9.76	***	0.00	–	0.00	–	9.58		1.04		0.00	–	0.00	–
Heuristic Agent	42.79	98.54	***	41.76	***	11.46	***	58.00	***	3.85		68.38	***	48.57	***
Children	73.63	98.61	***	47.22	***	60.78	***	70.83	***	69.18	***	70.83	***	68.18	***

'*' (p <0.05) '**' (p <0.005) '***' (p <0.001) '–' (p *undefined*)

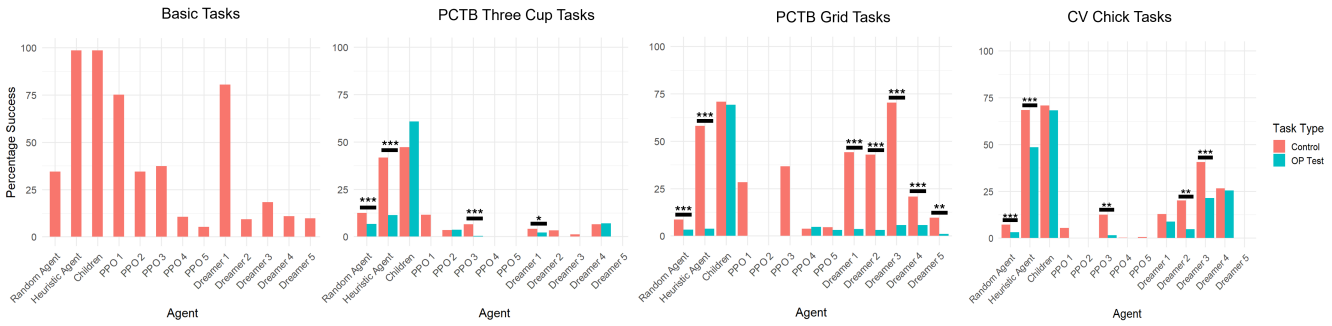

Figure 2: Percentage of tasks that each type of agent passed, with Basic/Control tasks shown in red and OP tasks shown in Blue. The significance of the difference between odds of success on the OP tasks compared to the Control tasks is shown above the bars. '*' (p <0.05) '**' (p <0.005) '***' (p <0.001)

odds of success to the Random Agents, except for Dreamer 3 and Dreamer 4 which have higher odds of success with marginal significance, and PPO 1, PPO 2, and PPO 3, which did not pass any instance of these tasks (see Table 2).

In contrast, on the PCTB Grid Control Tasks, most agents have significantly higher odds of success compared to the Random Agent. Children perform best, succeeding on the majority of instances but closely matched by Dreamer 3, which was trained on all Basic and Control Tasks. All Dreamer agents except Dreamer 5 have higher odds of success than the Random Agents, while Dreamer 5 does not have significantly different odds, similar to PPO 5. PPO 4 has significantly lower odds of success compared to the Random Agents, and PPO 2 fails every instance of this task type (see Table 2).

Most agents except children have significantly lower odds of succeeding on OP Test Tasks compared to Control Tasks. The children performed similarly on both types of task (see Fig. 2).

CV Chick Tasks On the OP Test Tasks, children performed the best. The Heuristic Agents are second in terms of performance. Both of these agents, along with Dreamer 1, Dreamer 3 and Dreamer 4, are significantly more likely to succeed than the Random Agent, even though only Dreamer 4 was trained on instances of this task. All other agents show relatively poor performance on this task type, with similar, or significantly, worse odds of success compared to the Random Agent (Table 2).

On the Control Tasks, the agents tended to perform better. Children again performed the best, significantly better than the Random Agents. The Heuristic Agents performed similarly to the children in terms of obtaining the goal in each instance. Several DRL agents are significantly more likely to succeed in instances of these tasks than the Random Agents. The same agents that performed well at the OP Test Tasks also performed well on these Control Tasks, with the addition

of PPO 3. The remaining agents were significantly less likely to succeed than the Random Agents, or were not significantly different from them (PPO 1; Table 2).

Children performed similarly on the OP Tasks compared to the Control Tasks, whereas the Heuristic Agents have significantly lower odds of success on the OP Tasks. Dreamer 2, Dreamer 3, and PPO 3, three of the best performing DRL agents on these tasks, are significantly less likely to succeed on the OP Tasks than the Controls. The remaining agents do not have significantly different odds of success on the two types of task (Figure 2).

Discussion

Object Permanence is a foundational ability that develops early in human children and has been demonstrated even in day-old chicks (Chiandetti & Vallortigara, 2011). However despite otherwise impressive performance, modern DRL agents struggle to learn to track occluded objects.

Here, agents based on two DRL algorithms (PPO, Dreamer-v3) were compared to human children (aged 4-7) and reference agents (random and heuristic) on three sets of OP tasks taken from the O-PIAAGETS test battery. A chance baseline was established at the level of the performance of a cohort of 5 random action agents that varied in their specific action distributions, but overall behaved stochastically. As expected, children performed consistently well across all tasks, with the exception of the control version of the PCTB Three Cup Task where, while performing poorly, they still outperformed all other agents. The Heuristic agent, which followed rigid rules to approach goals, avoid death zones and navigate around objects, was almost always the second highest performer. Despite performing notably worse on OP tasks compared to their respective controls, confirming this agent's lack of OP, it reliably outperformed the random agents on most tasks, while the same could not be said for the majority of the DRL agents. While there were instances where the DRL agents - particularly Dreamer - outperformed the random agent (and at times the heuristic agent), performance on the OP tasks themselves was consistently very low. Deeper analysis in future work of the behaviour of the agents, including the the paths they take when solving a task, is likely to be informative as to the nature of the decision-making performed by different types of agent.

Overall, these results suggest that while there have been considerable advances in the capabilities of Deep Reinforcement Learning agents, these do not currently extend to Object Permanence, at least not with the parameters and curricula used in this study.

Acknowledgements

This work was funded by an ESRC DTP scholarship (ES/P000738/1), the US DARPA HR00112120007 (RECoG-AI) grant, a Templeton World Charity Foundation grant (TWCF-2020-20539), the "Programa Operativo del Fondo Europeo de Desarrollo Regional (FEDER) de la Comunitat Valenciana 2014-2020" under agreement INNEST/2021/317 (Neurocalçat), and the "programa de ayudas a la formación de doctores en colaboración con empresas" (DOCEMPR21).

References

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*, 191–208.

Chiandetti, C., & Vallortigara, G. (2011). Intuitive physical reasoning about occluded objects by inexperienced chicks. *Proceedings of the Royal Society B: Biological Sciences*, *278*, 2621-2627.

Crosby, M., Beyret, B., Shanahan, M., Hernández-Orallo, J., Cheke, L., & Halina, M. (2020). The animal-ai testbed and competition. In *Neurips competition and demonstration track* (p. 163-176).

Hafner, D., Pasukonis, J., Ba, J., & Lillicrap, T. (2023). Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*.

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *science*, *317*(5843), 1360–1366.

Juliani, A., Berges, V., Teng, E., Cohen, A., Harper, J., Elion, C., ... Lange, D. (2020). Unity: A general platform for intelligent agents,. *Arxiv preprint arXiv:1809.02627*.

Piaget, J. (1952). *The origins of intelligence in children*. W W Norton & Co. (English translation by M. Cook)

Piloto, L. S., Weinstein, A., Battaglia, P., & Botvinick, M. (2022). Intuitive physics learning in a deep-learning model inspired by developmental psychology. *Nature human behaviour*, *6*(9), 1257–1267.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Van Hoorick, B., Tokmakov, P., Stent, S., Li, J., & Vondrick, C. (2023). Tracking through containers and occluders in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13802–13812).

Voudouris, K., Alhas, I., Schellaert, W., Crosby, M., Holmes, J., Burden, J., ... Cheke, L. G. (2023). Animal-AI 3: What's new & why you should care. *arXiv preprint arXiv:2312.11414*.

Voudouris, K., Crosby, M., Beyret, B., Hernandez-Orallo, J., Shanahan, M., Halina, M., & Cheke, L. (2022a). Direct human-AI com- parison in the Animal-AI environment,. *Frontiers in Psychology*, *13*.

Voudouris, K., Donnelly, N., Rutar, D., Burnell, R., Burden, J., Hernandez-Orallo, J., & Cheke, L. (2022b). Evaluating object permanence in embodied agents using the Animal-AI environment. In *IJCAI EBeM Evaluation Beyond Metrics Workshop*.