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

Title
Learning to Play Video Games with Intuitive Physics Priors

Permalink
https://escholarship.org/uc/item/92f5b1hk

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

Authors
Jaiswal, Abhishek
Srivastava, Nisheeth

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

Learning to Play Video Games with Intuitive Physics Priors

Abhishek Jaiswal (abhijais@cse.iitk.ac.in)
Department of CSE, IIT Kanpur
Kalyanpur, UP 208016 India

Nisheeth Srivastava (nsrivast@cse.iitk.ac.in)
Department of CSE, IIT Kanpur
Kalyanpur, UP 208016 India

Abstract

Video game playing is an extremely structured domain where algorithmic decision-making can be tested without adverse real-world consequences. While prevailing methods rely on image inputs to avoid the problem of hand-crafting state space representations, this approach systematically diverges from the way humans actually learn to play games. In this paper, we design object-based input representations that generalize well across a number of video games. Using these representations, we evaluate an agent's ability to learn games similar to an infant - with limited world experience, employing simple inductive biases derived from intuitive representations of physics from the real world. Using such biases, we construct an object category representation to be used by a Q-learning algorithm and assess how well it learns to play multiple games based on observed object affordances. Our results suggest that a human-like object interaction setup capably learns to play several video games, and demonstrates superior generalizability, particularly for unfamiliar objects. Further exploring such methods will allow machines to learn in a human-centric way, thus incorporating more human-like learning benefits.

Keywords: Category Learning; Object-based Reinforcement Learning; Generalization; Inductive priors; Intuitive physics

Introduction

Deep reinforcement learning (DRL) algorithms have shown professional to superhuman competency in gaming environments such as MuJoCo, and Atari (Shakya, Pillai, & Chakrabarty, 2023; Goodfellow, Shlens, & Szegedy, 2014). But, at the same time, like other black box deep learning models, they can break with even slight modifications of the environment (Justesen et al., 2018; Goodfellow et al., 2014).

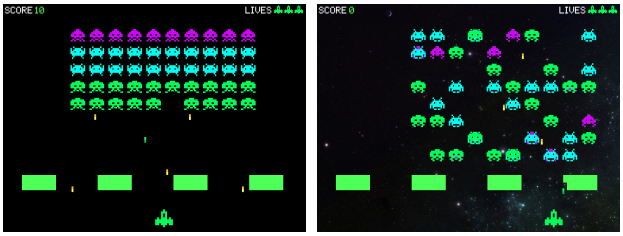

Figure 1: Simple Variations, Crippling Results - Deep Learning Models break even with a slight variation of the environment (Right image - partially randomized enemy positions).

For example, to contrast human and machine-level learning, Figure 1 shows two variants of the space invaders game we tested. DQN was trained on the basic version on the left for one million iterations and tested on the variant with partially randomized enemy positions on the right. The base variant's average score was 510, whereas the right variant could score only 280; right around random performance. On the contrary, humans play through such variants with ease. Also, DRL-based approaches still fail at generalizing and transferring learned knowledge to novel domains (Kansky et al., 2017). Humans demonstrate superior learning trajectories, learning games quickly and also performing well on modifications (Tsividis, Pouncy, Xu, Tenenbaum, & Gershman, 2017).

Attempting to bring the study of video game-playing closer to human cognitive behavior, in this paper, we learn game-playing using common human inductive biases. With this line of work, we aim to leverage the same advantages humans show in generalization and zero-shot transfer on related tasks. To this end, we design generalizable affordance-based representations of a reinforcement learning agent's state space using two primary assumptions. First, we try to incorporate the thinking of a first-time player in game playing using inductive biases drawn from humans' common core knowledge of intuitive physics (Spelke, 1990; Spelke & Kinzler, 2007). Second, we design a state space representation using object categories instead of classically used object-based input representations. We test the value of these representations by training a simple Q-learning agent (Watkins & Dayan, 1992), and comparing its performance against DQN (Mnih et al., 2015). Finally, we show that rather than using standalone objects, describing the game world in terms of object categories offers learning and generalization trends practically unattainable by resource-hungry pixel-based DRL agents.

Using affordances to infer states

Theory-based RL is a form of model-based reinforcement learning where the model is defined in terms of rich ontological symbolic representations pertaining to physical objects, their relations, and interactions. Using various intuitive theories, theory-based RL explicitly tries to incorporate human ways of learning (Tsividis et al., 2021). Such intuitive theories stem from a core knowledge representation of the world visible even in infants who can segregate the visual input into ontological structures such as objects, goals, and physics (Baillargeon, 2004; Spelke, 1990; Spelke & Kinzler, 2007; Csibra, 2008). Humans also have been shown to make internal models using theory representation (Tomov, Tsividis, Pouncy, Tenenbaum, & Gershman, 2023). Similarly, semantic and syntactic biases, such as those used in

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).

theory-based RL, show a strong resemblance to human-like learning (Pouncy & Gershman, 2022). Humans show a wide range of flexibility in adapting to variations within the same task domain. As such, Pouncy, Tsividis, and Gershman (2021) have shown evidence that such flexibility, a hallmark of human intelligence, can arise by representations composed of objects and interactions within a model-based framework. Thus, theory-based RL has shown a promising resemblance to human-like learning. However, being dependent on already possessing a detailed model of the environment, it has significant practical limitations.

Structure representation is considered a key ingredient to human-like learning. For example, Lake, Ullman, Tenenbaum, and Gershman (2017) contrasted a set of key ingredients for more human-like learning against recent developments in deep reinforcement learning. Bapst et al. (2019) stressed structured inputs as a key to better generalization and solving situations beyond the training space. In a series of related works, Doumas et al. (2022) show cross-game generalization by using relational analogies over symbolic representations, whereas we use categorial object affordance for solving game variants and hope to extend the concepts across games.

Gershman and Niv (2010) proposed using observation to infer states and actions through intuitive physics and intuitive psychology. To make sense of these observations, humans utilize various priors that help them explore efficiently. Dubey, Agrawal, Pathak, Griffiths, and Efros (2018) explore and quantify such priors for video gaming tasks. Our work builds upon such principles to learn a working structure of the world.

Tsividis et al. (2021) worked on the idea of making machines learn more like humans starting from early childhood state using strong theories about the working of the world. We take a slightly different approach and learn the affordances from object specifications rather than using pre-defined rules of interactions. Much like them, we also levy inductive biases for this task, which we understand to be a product of evolution, such as agent identification, threat perception, and goal attribution.

The inductive biases we adopt are drawn primarily from the work of Elizabeth Spelke (Spelke & Kinzler, 2007). Spelke (1990) reason that infants perceive objects based on perceptual units moving together, moving separately, interacting on contact, and maintaining their shapes and sizes while in motion. We leverage these visual signals to learn fundamental affordances such as avoid, touch, and block through our Reinforcement Learning (RL) agent's actions trained exclusively on object-specific properties that are interpretable and in alignment with concepts of infant learning.

Thus, in this domain, akin to the work of Ding et al. (2023) in the space of natural languages, we try to answer a simple question - 'can we enable an agent to learn like a small child?' and test the hypothesis in game settings.

Humans look at the world in terms of objects and their interactions; This is one of their core knowledge (Spelke & Kinzler, 2007). Drawing on this insight, we shape the task of object reasoning around basic principles of core knowledge and show that we can achieve game-playing in a more cognitive and less mechanistic manner. Specifically, instead of just looking at objects in isolation, our agents infer meaningful object categories as part of their interactions with the game world.

Learning How to Play

We look at the game screen from the view of a novice player holding a very minimal baggage of experience. Such players would see certain entities stand out on the screen by virtue of their specific forms, colors, or movements but would not know the associated affordances – a task necessary to accomplish their desire to win (Csibra, 2008). The first step in this learning process would be detecting what we control on the screen, i.e., the agent representing the player in the game. Agent detection is one of the key ideas in human-like learning and also a stark differentiator from large-scale machine-like pattern matching (De Freitas et al., 2023). After knowing the where and how of the agent, the next step would be to devise a locomotive strategy, necessitating knowledge of at least a minimal set of affordances associated with other game entities, for which we devise a set of representative object categories and learn category-level affordances . Refer to Figure 2 for an overview of the complete pipeline.

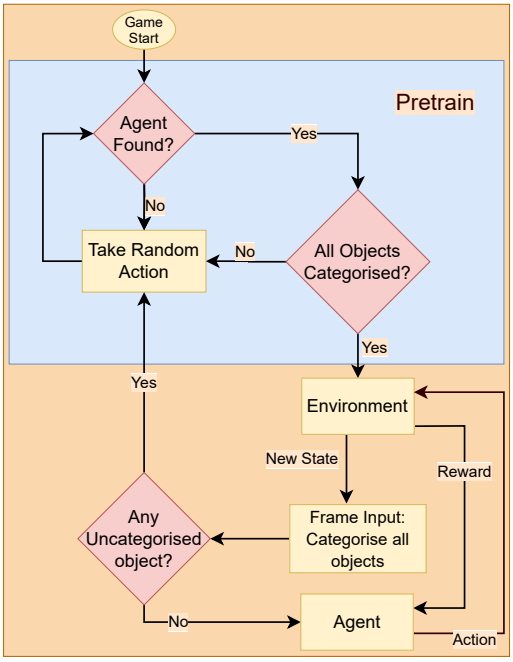

Figure 2: Schematic representation of Agent-Action pipeline based upon intuitive physics priors.

Categories

Theory-based RL methods, even if showing human-like learning traits, use an object-interaction definition known a

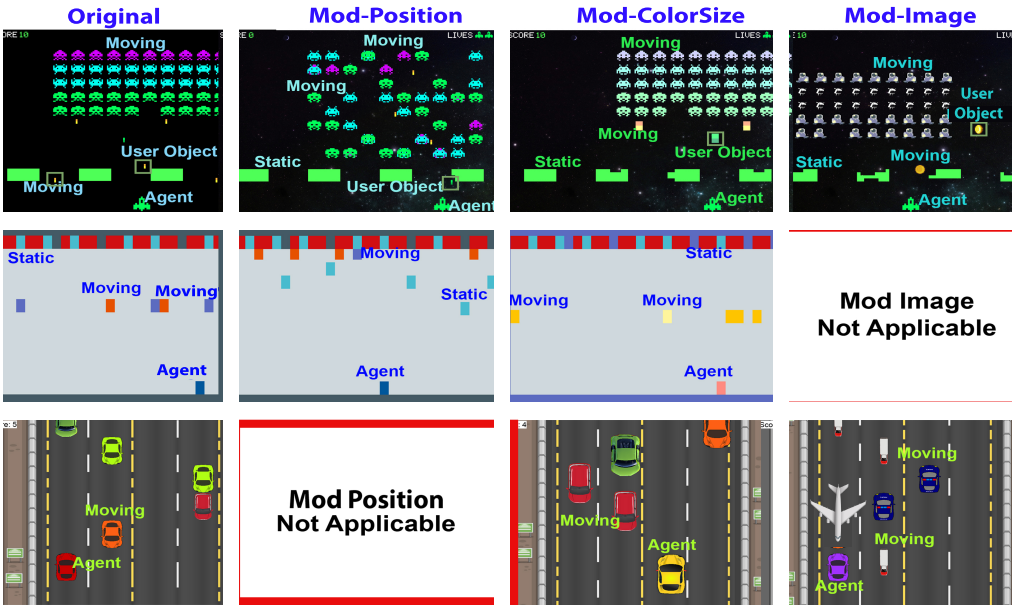

Figure 3: From Left to Right, each column shows the original games, Position Modification, Color and Size Modification, and Image Modification. For GVGAI games, all objects are rectangles with fixed shapes and sizes, and Image modification is not applicable. For Roadrash, enemy cars are randomly spread over the road, and thus Position Modification is not applicable.

priori and focus on exploration and planning with a very strong world model (Tsividis et al., 2021). We take a different route here. Rather than working with the objects directly, we focus on affordance-based object categories inspired from intuitive physics. These categories are motivated by perceptual signals such as identifying objects as static or moving and then the agent learns their attributes from experience. Humans also learn such object categories having similar affordances over isolated entities and tend to generalize strategies from previously learned knowledge to unseen situations (Perfors & Tenenbaum, 2009; Medin, Wattenmaker, & Hampson, 1987).

For all our games, we utilize only these five simple categories:

- **Agent** - Agent detection relies on a minimal set of inductive biases, which varies depending on the complexity of the environment.
- **Static objects** - The positions of these objects remain unchanged in successive frames. In simple games, they could be harmless, offering secondary benefits like protection from bullets. In a more complicated setting, they could be part of a winning precondition.
- **Moving-Good objects** - These objects change their previously occupied positions. They are the interesting and primary interacting entities apart from the agent and may give a positive reward for touching.
- **Moving-Bad objects** - This other moving category represents the prime obstacles in the game. They give negative rewards or kill the agent on touching. As we perceive a threat and move away, the primary affordance associated with this class is to avoid them.
- **Agent objects** - Primarily bullets spawned by the agent.

After a player learns these categories, downstream classification becomes instantaneous. Similar to humans, we store their characteristic properties, such as color and classify them instantaneously as they become visible on the screen.

Identifying the Agent

As discussed in the previous section, of all the object categories, the agent is of primal importance and requires special attention. Based on previous studies on infants (Spelke, 1990; Spelke & Kinzler, 2007), we utilize a set of inductive biases to mimic how a new player would detect the agent in the game.

Inductive biases

Even though identifying objects and the associated properties occurs concurrently and continuously, we try to solve the agent identification problem by utilizing as little information as possible. Thus, we use a sequence of inductive biases for agent identification, stopping at whichever one yields a unique agent representation. In order of priority, these biases are:

Inductive Bias 1 - Uniqueness. This property suggests that the agent is expected to have a unique form. On the game screen, if two objects appear visually similar, they are less likely to be the agent.

Inductive Bias 2 - Permanence. From a gaming perspective, "permanence" refers to the sustained existence of an entity on the game screen. As the game world is centered around

the agent, other objects enter and exit, but the agent is expected to persist at all times.

Inductive Bias 3 - Action-Object Motion binding. The agent is meant for action. As a final conclusive test, we assess all the objects for their mobility with different key presses, the intent being that the agent, as an active principle in the game, would be dynamic rather than passive unless killed by an undesirable interaction. Moreover, as a specific key is pressed repeatedly, only the agent is expected to consistently manifest a repeated action, as outlined by De Freitas et al. (2023).

Agent Action Key Bindings

This involves learning the activities an agent does in response to different key presses. It is a form of reinforcement where the player presses keys to observe the agent's behavior. Through repeated iterations of this exercise, the player gradually discerns the mapping of each action to a specific outcome on the screen. We apply a similar principle in our games by taking random actions and observing the changes in the agent's position to map the action-key bindings. Thus, evaluating movement action key bindings is straightforward. For bullet firing, we check for the generation of a new object near the agent immediately after a keypress. If this occurrence repeats until a specified threshold, we assign the key's action affordance as "Fire."

Implementation

As object detection is a well-researched field, for our tests, we commence with a preexisting list of objects. Subsequently, we categorize these objects into the aforementioned groups solely based on their bounding box and color.

Incorporating the above object definitions, we try to learn game playing using the Q-learning algorithm (Watkins & Dayan, 1992). For Q-learning to work, we need a state representation that is concise and, at the same time, rich enough for the agent to have sufficient winning information. Consequently, we parse all object category details into state representations tailored to each game setting. Specifically, we take 2k+1 relative orientation bits, two bits to denote the left and right boundary, and 1 bit to mark the presence of agent bullet if applicable to the game. The 2k+1 orientation bits represent the time it takes for the agent to reach each bit while stationed at the kth bit and store the time it would take for a moving object to cover the horizontal distance from the agent. Here, k varies depending on the requirements of each game (refer Figure 4 for an example with k = 4).

GVGAI Games

We modify the MyAliens game from GVGAI framework (Perez-Liebana et al., 2019) into two variants to test our hypotheses on human-like learning.

MyAliens - variant 1 (MyAliensV1). In this game, the objective is to avoid getting hit by any moving object falling from the top till timeout, as they all kill the agent on touching. Game Categories - Agent, Static: Enemy Spawn Points, Moving-Bad: Enemies.

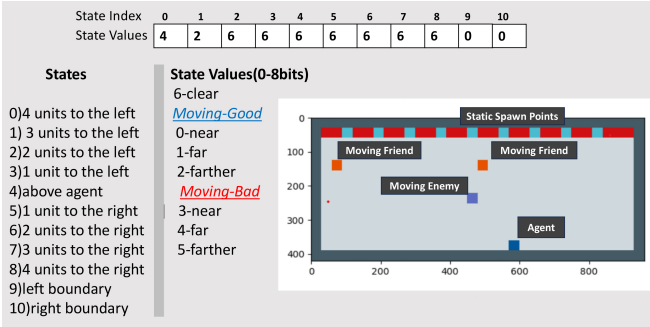

Figure 4: MyAliens State Repsentation.

MyAliens - variant 2 (MyAliensV2). This game has two types of moving objects - one food item giving a positive reward and another enemy killing the agent. The agent has to learn to collect ten positively rewarding objects before timeout to win the game. Game Categories - Agent, Static: Enemy Spawn Points, Moving-Bad: Enemies, Moving-Good: Food items (Figure 4).

Custom Games

Additionally, we also test two custom games to check our hypothesis on more visually exciting games.

Roadrash - Car Driving. In this game, the player car has to avoid crashing into the incoming traffic cars. There are only two categories present - agent and moving bad objects. The vehicles can drive only in 4 lanes, making the game very challenging under heavy traffic (Figure 3). Game Categories - Agent, Moving-Bad: Enemies.

SpaceInvaders. This game is based on the classic Atari Space Invaders and has the same features with better visuals. The enemies travel horizontally and then move a row down while shooting bullets at the agent spaceship. The agent can shoot only one bullet at a time. Game Categories - Agent, Static: Shields, Moving-Bad: Enemy Spaceships and Bullets, Agent-object: Agent Bullets.

Generalization Experiments

Our goal here is not to defeat a Deep Reinforcement learning algorithm but to show that using a methodology like ours has certain benefits that opens up new avenues for mimicking human learning characteristics. For all the tests, we train DQN for 10e6 with linear learning rate decay from 1 to 0.01.

First, we test our games to see their learning capacity compared against a DQN agent for all four games. For this, we plot the normalized average scores over 20 runs for model vs. epoch, where epoch is defined as one game run loop. We normalize the scores as follows:

$$Normalized\ score = \frac{actual\ score}{maximum\ achievable\ score}$$

MyAliensV1 has five levels with different placement of

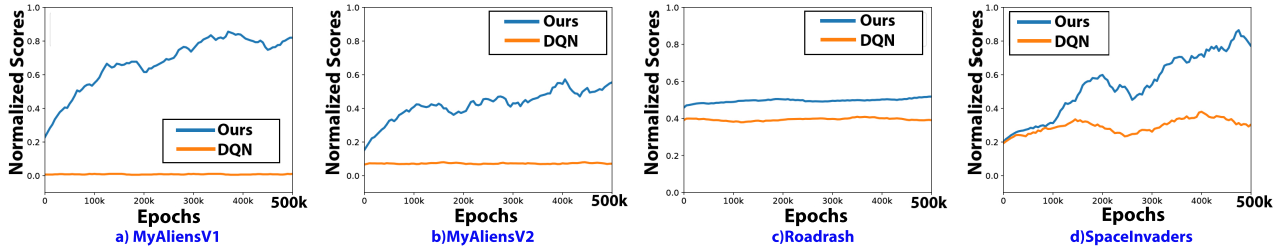

Figure 5: Affordance-based Q-learning (Ours) vs. Image-based DQN Normalized Score per epoch plots. a) MyAliensV1: DQN is probably still exploring as it could not learn any meaningful action. b) MyAliensV2 - Both algorithms found difficulty; Q-learning still fairs better, but DQN could not clear even the first level for both variants of MyAliens. c) Roadrash - Very stochastic game with many occasions where avoiding collision is impossible. Q-learning still does better than DQN. d) SpaceInvaders - our algorithm easily learns gameplay using its object-based representation.

Table 1: Normalized score for DQN vs. our method averaged over 20 runs of the games. All the models are trained for 1 Million epochs. SpaceInvaders has two levels and a maximum achievable score of 1000. MyAliensV1 has five levels with a maximum score of 50 and MyAliensV2 has three levels with a maximum score of 30. Roadrash ends if the agent can avoid collision for 300 steps, and the score is measured in the number of steps survived.

Modifications	**MyAliensV1**		**MyAliensV2**		**Roadrash**		**Space Invaders**	
	DQN	Ours	DQN	Ours	DQN	Ours	DQN	Ours
Random Action	- 0.20		- 0.33		0.27		0.27	
Base-Variant	- 0.08	0.80	- 0.23	0.57	0.45	0.50	0.51	1.0
Mod-Position	- 0.20	0.74	- 0.27	0.52	NA	NA	0.28	0.42
Mod-ColorSize	- 0.20	0.80	- 0.27	0.57	0.40	0.48	0.31	1.0
Mod-Image	NA	NA	NA	NA	0.38	0.47	0.30	1.0

spawn points for moving enemies with a maximum score of 50, +10 for winning each level, and -10 for losing. We test MyAliensV2 for three levels, with a maximum achievable score of 30. The agent receives a -10 reward for losing, and a reward of +1 for collecting a food item. The level is won when ten such items are collected.

In Roadrash, the agent needs to survive the traffic onslaught for 300 steps. SpaceInvaders has two levels with a maximum achievable score of 1000, from +10 received for killing each of the 50 enemy spaceships over the two levels.

Humans, after learning how to play a game, would easily adapt to slight variants of the game. To test the generalized ability of our agent to mimic human-like learning, we run the agent in different variations of the games. For all such cases, we train the game only on the base variant of the game.

We explore three types of game variations (Figure 3):

- **Mod-Position**: Partially random placement of moving enemies (SpaceInvaders) or static enemy spawn points (MyAliens).
- **Mod-ColorSize**: Alters size and color of game objects.
- **Mod-Image**: Substitutes default game images.

GVGAI games do not permit the modification of object sizes, and all in-game objects are constructed using unit-sized colored rectangles. Thus, the Mod-Image variant is not applicable, as it involves using external images, which the GVGAI framework does not support. In Roadrash, where enemy cars are spawned randomly, the Mod-Position variant will not yield any new variation and is not used.

We train the DQN algorithm using a batch size 32 on each game run loop with experience replay. The Q-learning agent is trained only once on each run on the latest experience. Thus, even on the same level of epochs, DQN weights are updated 32 times more than Q-learning.

Observations

To test the efficacy of our category-level representations, we run two kinds of tests. First, we compare DQN and our method under varying training durations. Scores from the trained models from different training epochs are plotted, and we analyze the agent's normalized score (Figure 5). One Epoch is defined as one run of the game loop. Even though Q-learning updates 32 times less than DQN, it is able to learn correct decisions quickly. We plot the results up to 0.5 million epochs, equivalent to approximately 2 hours of gameplay at 60 frames per second, and in most cases, DQN did not show any improvement owing to its sample inefficiency.

Our algorithm demonstrates strong performance in MyAliensV1, successfully winning all five levels. In SpaceInvaders also, it wins both the levels. However, MyAliensV2 presents a more challenging scenario, requiring the agent to distinguish between moving-good and moving-bad categories and collect ten of the good ones before a time-out to win the level. Here also, our method does well, but the performance degrades as compared to MyAliensV1, primar-

ily because our agent has only a 9-bit state space representation, i.e., it can see nearest objects only within a range of four units on both the left and right sides. Given that the moving-good category is dispersed over a broader x-range of thirty bits, the agent often struggles to locate the moving-good category within its narrow field of vision. Consequently, the timer runs out before the agent can collect the required ten items, impacting the overall score.

The escalating difficulty in subsequent levels, coupled with a reduction in the number of Moving-Good spawn points, adds to the complexity of the task. We also tested a broader state representation of 25 bits, but the learning became computationally intractable, and the agent struggled to learn meaningful affordances. Nevertheless, even with a limited view, our agent clears the first two levels and fails only on the final level (Table 1).

For Roadrash, even though our method does better, we do not see substantial performance gain with training. The game has four lanes, with enemies spawning stochastically in any of them. Thus, in many cases, all four lanes get blocked, and a crash becomes unavoidable. In other situations, avoiding accidents requires precise control because of the crowded structure of game objects. So, even a reasonably learned agent could not perform well in this game, and the performance was more-or-less stagnant. Nonetheless, our algorithm still fairs better against the DQN agent.

Our second set of comparisons focuses on the transferability of the acquired knowledge. For deep learning algorithms, object level alterations, such as changing object colors, can have devastating consequences (Lake et al., 2017). On the other hand, humans can easily manage such variations. Our results indicate that, unlike a DQN agent, our category-based method exhibits similar performance scores, aligning more with human-level gameplay (Table 1).

This is primarily because, at the category level, the state representations remain relatively stable despite the aforementioned generalization modifications. Consequently, our algorithm's performance does not degrade with these variations. It's noteworthy that both models in these comparisons are trained for one million epochs. However, for MyAliensV1 and MyAliensV2, DQN is still in its exploration phase, exhibiting minimal performance improvement, and the introduced variations further degrade its performance. This is particularly evident in SpaceInvaders, where the DQN agent while displaying some learning traces in the base variant, regresses to the level of a random agent when faced with varying input pixel combinations. As the Roadrash game is challenging from the start, there is little difference after making a difficult game more difficult.

Among all the alterations, only SpaceInvaders Mod-Position resulted in a substantial decline in Q-learning performance. This is primarily due to position modifications creating new, and previously unseen, state representations. In this setting, as the enemies get randomly arranged, some enemies get placed too close to the agent. As such states are previously unseen, a table-based Q-learning agent struggles to navigate this variation (Table 1). Such instances could potentially be avoided by using techniques to extrapolate for unseen states based on prior experience. Apart from this, other game modifications consistently exhibit performance similar to the unmodified original versions of the games, as is also expected from a human player.

Thus, our comparisons show that an object-based representation, even if applied within a model-free framework, offers much better sample efficiency (Figure 5). This improvement is evident in results with environmental perturbations, such as varying enemy positions and differently shaped enemies, among other variations. The primary factor contributing to this enhanced performance is the category-based representation, in which minor perturbations do not alter the game representation significantly, while causing problems in approaches where objects are treated as separate entities, and also for pixel-based model-free methods like DQN.

Conclusion

Making machines learn and act like humans is an important goal in Artificial General Intelligence(AGI) (Lake et al., 2017; Tsividis et al., 2017; Pouncy, 2022). In this paper, we look at video game playing from the eyes of a novice player discovering gameplay dynamics. Drawing inspiration from Spelke's conception of core physics knowledge (Spelke, 1990; Spelke & Kinzler, 2007), we developed object category representations that transfer well across simple games. Building upon this state representation, we show that machines can exhibit certain similarities to human-like learning in game playing. In contrast with existing approaches that learn to play games using self-learned pixel-based representations (Mnih et al., 2015; Hessel et al., 2018), our approach focuses on using composable affordance-based object representations and shows faster and more robust learning in such games than is seen in foundationally pixel-based approaches.

It is natural, of course, to question the degree to which intuitive physics priors handcrafted to align with game-world task requirements generalize to more natural settings. While answering such questions is a longer-term project, we point out that the semiotics and affordances of video games, particularly classic games of the sort we test, have been optimized for easy comprehension across language and age barriers (Blomberg, 2018). Thus, the success of classic intuitive physics heuristics (Spelke, 1990) in producing affordance-based representations that generalize across games may well be because these heuristics apply well in the real world also.

Due to its model-free nature, our agent is still not as sample-efficient as a human, but it does well on the generalizability task. Theory-based RL approaches take the model as given and explore planning within such a framework (Tsividis et al., 2021; Pouncy, 2022). Future work may extend the use of affordance-based object representations to learn a model of the environment based jointly on core knowledge and experience.

References

Baillargeon, R. (2004). Infants' physical world. *Current directions in psychological science*, *13*(3), 89–94.

Bapst, V., Sanchez-Gonzalez, A., Doersch, C., Stachenfeld, K., Kohli, P., Battaglia, P., & Hamrick, J. (2019). Structured agents for physical construction. In *International conference on machine learning* (pp. 464–474).

Blomberg, J. (2018). The semiotics of the game controller. *Game Studies*, *18*(2), 311–323.

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, *107*(2), 705–717.

De Freitas, J., Uğuralp, A. K., Oğuz-Uğuralp, Z., Paul, L., Tenenbaum, J., & Ullman, T. D. (2023). Self-orienting in human and machine learning. *Nature Human Behaviour*, 1–14.

Ding, M., Xu, Y., Chen, Z., Cox, D. D., Luo, P., Tenenbaum, J. B., & Gan, C. (2023). Embodied concept learner: Self-supervised learning of concepts and mapping through instruction following. In *Conference on robot learning* (pp. 1743–1754).

Doumas, L. A., Puebla, G., Martin, A. E., & Hummel, J. E. (2022). A theory of relation learning and cross-domain generalization. *Psychological review*, *129*(5), 999.

Dubey, R., Agrawal, P., Pathak, D., Griffiths, T. L., & Efros, A. A. (2018). Investigating human priors for playing video games. *arXiv preprint arXiv:1802.10217*.

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: carving nature at its joints. *Current opinion in neurobiology*, *20*(2), 251–256.

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.

Hessel, M., Modayil, J., Van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., ... Silver, D. (2018). Rainbow: Combining improvements in deep reinforcement learning. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 32).

Justesen, N., Rodriguez Torrado, R., Bontrager, P., Khalifa, A., Togelius, J., & Risi, S. (2018). Illuminating generalization in deep reinforcement learning through procedural level generation. In *Neurips workshop on deep reinforcement learning.*

Kansky, K., Silver, T., Mély, D. A., Eldawy, M., Lázaro-Gredilla, M., Lou, X., ... George, D. (2017). Schema networks: Zero-shot transfer with a generative causal model of intuitive physics. In *International conference on machine learning* (pp. 1809–1818).

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences*, *40*, e253.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive psychology*, *19*(2), 242–279.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... others (2015). Human-level control through deep reinforcement learning. *nature*, *518*(7540), 529–533.

Perez-Liebana, D., Lucas, S. M., Gaina, R. D., Togelius, J., Khalifa, A., & Liu, J. (2019). *General video game artificial intelligence* (Vol. 3) (No. 2). Morgan & Claypool Publishers. (`https://gaigresearch.github.io/gvgaibook/`)

Perfors, A., & Tenenbaum, J. (2009). Learning to learn categories..

Pouncy, T. (2022). *Theory-based reinforcement learning: A computational framework for modeling human inductive biases in complex decision making domains*. Unpublished doctoral dissertation, Harvard University.

Pouncy, T., & Gershman, S. J. (2022). Inductive biases in theory-based reinforcement learning. *Cognitive Psychology*, *138*, 101509.

Pouncy, T., Tsividis, P., & Gershman, S. J. (2021). What is the model in model-based planning? *Cognitive Science*, *45*(1), e12928.

Shakya, A. K., Pillai, G., & Chakrabarty, S. (2023). Reinforcement learning algorithms: A brief survey. *Expert Systems with Applications*, 120495.

Spelke, E. S. (1990). Principles of object perception. *Cognitive science*, *14*(1), 29–56.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental science*, *10*(1), 89–96.

Tomov, M. S., Tsividis, P. A., Pouncy, T., Tenenbaum, J. B., & Gershman, S. J. (2023). The neural architecture of theory-based reinforcement learning. *Neuron*, *111*(8), 1331–1344.

Tsividis, P. A., Loula, J., Burga, J., Foss, N., Campero, A., Pouncy, T., ... Tenenbaum, J. B. (2021). Human-level reinforcement learning through theory-based modeling, exploration, and planning. *arXiv preprint arXiv:2107.12544*.

Tsividis, P. A., Pouncy, T., Xu, J. L., Tenenbaum, J. B., & Gershman, S. J. (2017). Human learning in atari. In *2017 aaai spring symposium series.*

Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine learning*, *8*, 279–292.