INVESTIGATING HUMAN PRIORS FOR PLAYING VIDEO GAMES

Anonymous authors
Paper under double-blind review

ABSTRACT

Deep reinforcement learning algorithms have recently achieved impressive results on a range of video games, yet they remain much less efficient than an average human player at learning a new game. What makes humans so good at solving these video games? Here, we study one aspect critical to human gameplay – their use of strong priors that enables efficient decision making and problem solving. We created a sample video game and conducted various experiments to quantify the kinds of prior knowledge humans bring in while playing such games. We do this by modifying the video game environment to systematically remove different types of visual information that could be used by humans as priors. We find that human performance degrades drastically once prior information has been removed, while that of an RL agent does not change. Interestingly, we also find that general priors about objects that humans learn when they are as little as 2 months old are some of the most critical priors that help in human gameplay. Based on these findings, we then propose a taxonomy of object priors people employ when solving video games that can potentially serve as a benchmark for future reinforcement learning algorithms aiming to incorporate human-like representations in their systems.

1 INTRODUCTION

Consider the following scenario: you are tasked to play an unfamiliar computer game shown in Figure 1(a). No manual or instructions are provided. You do not know the goal or even what sprite on the game screen is your character. How quickly can you finish this game? We recruited forty subjects to play this game and found that subjects solved the game comfortably (taking just 1600 actions, corresponding to < 1 minute of gameplay, c.f. Figure 1(c)). Now consider the scenario in which the same game is rendered with new textures as shown in Figure 1(b). How easy is it to play this game? A separate group of forty human subjects took *twice* as many actions to solve the second game as compared to the first game (Figure 1(c)). Even though the placement of different object sprites and the reward structure is preserved across games, humans find it harder to complete the second game.

In a second suite of experiments, we trained a state-of-the-art reinforcement learning (RL) agent (ICM-A3C; Pathak et al. (2017)) on both these games and found that the RL agent took close to four million steps to solve both the games (Figure 1(c)). In stark contrast to humans, the RL agents were virtually unaffected by rendering differences between the two games.

Why is it the case that humans become significantly worse in playing the second game, whereas the RL agents are indifferent? In the first game, it is easy for humans to infer that the goal of the game is to move the robot sprite towards the princess by stepping on the brick-like objects and using ladders to reach the higher platforms while also avoiding the angry purple sprite and the fire object. On the other hand, in the second game, it is impossible for a human to infer just by looking at the game screen, which part of the screen shows the player sprite, let alone figuring out the goal of the game. In contrast to human players, the current state-of-the-art RL agents come *tabula rasa* i.e. without any prior knowledge about the world and thus both these games were more or less the same from the perspective of the RL agent leading to no change in performance across games.

This simple experiment highlights the importance of prior knowledge that humans draw upon to quickly solve tasks given to them. Building RL algorithms that can quickly solve tasks (i.e., sample

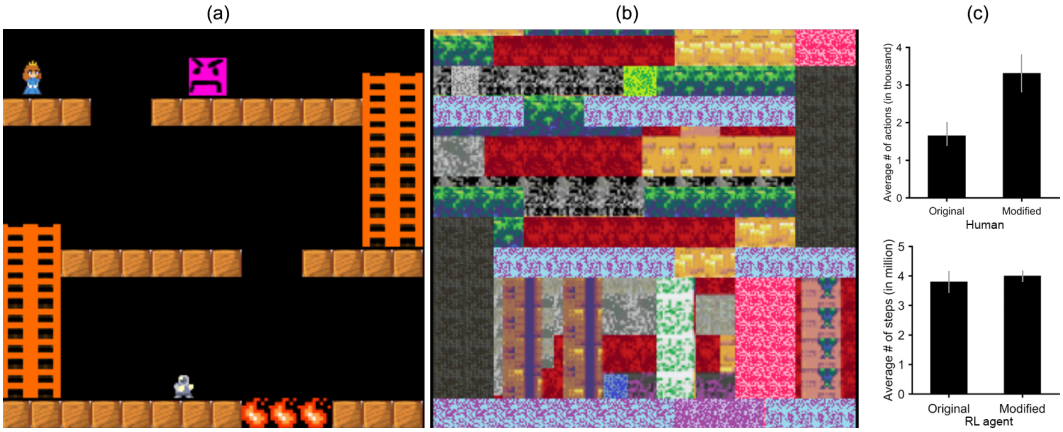

Figure 1: **Prior knowledge affects humans but not RL agents**. (a) A simple platformer game, (b) The same game modified by re-rendering the textures, and (c) Human players and RL agent performance in the two games. Error bars denote standard errors of the mean. Human players took close to 1600 actions to solve the first game (time = 1 minute) and 3300 actions to solve the second game (time = 2 minutes). The RL agent took 4 million steps to solve both the games.

efficient algorithms) is an active area of research, and further progress is inevitable. In addition to developing better optimization methods, we believe that instead of always initializing learning from scratch, either incorporating prior knowledge directly or constructing mechanisms for condensing experience into re-usable knowledge (i.e., learning priors) might be critical for building RL agents with human like efficiency.

What priors should we incorporate in our RL agents? For inspiration, it is natural to study what priors humans use. Developmental psychologists have been documenting the prior knowledge that children draw upon in learning about the world (Spelke & Kinzler, 2007; Carey, 2009). However, these studies have not explicitly quantified how vital are different priors for problem-solving.

In this work, we have systematically quantified the importance of various priors humans bring to use while solving one particular kind of problems – video games. We chose video games as the task for our investigation because it is easy to systematically change the game to include or mask different kinds of knowledge, run large-scale human studies, and video games such as ATARI are a popular choice in the reinforcement learning community. One of the findings of our investigation is that while knowledge of the form that ladders are to be climbed, keys are used to open doors, jumping on spikes is dangerous is important for humans to quickly solve games, more general priors of the form that the objects are subgoals for exploration and things that look the same behave the same are even more critical. Although we use video games as our experimental test bed, such priors are more generally applicable even outside the domain of video games.

2 METHOD

To investigate the aspects of visual information that enable humans to efficiently solve video games, we designed a browser based platform game consisting of a human sprite that could be controlled, platforms, ladders, slimy pink sprites that killed the agent, spikes that were dangerous to jump on, a key, and a door (see Figure 2 (a)). The human sprite could be moved with help of arrow keys and the agent obtained a reward of +1 when it reached the door after taking the key thereby terminating the game. The game was reset whenever the agent touched the enemy, jumped on the spike, or fell below the lowermost platform. We made this game to resemble the exploration challenges faced in the classic ATARI game of Montezuma's revenge that has proven to be very challenging for state of deep reinforcement learning techniques (Bellemare et al., 2016; Mnih et al., 2015).

We systematically created different versions of this game by re-rendering various entities such as ladders, enemies, keys, platforms etc. using alternate textures (see Figure 2). These textures were chosen to mask various forms of prior knowledge that are described in the experiments section. Our

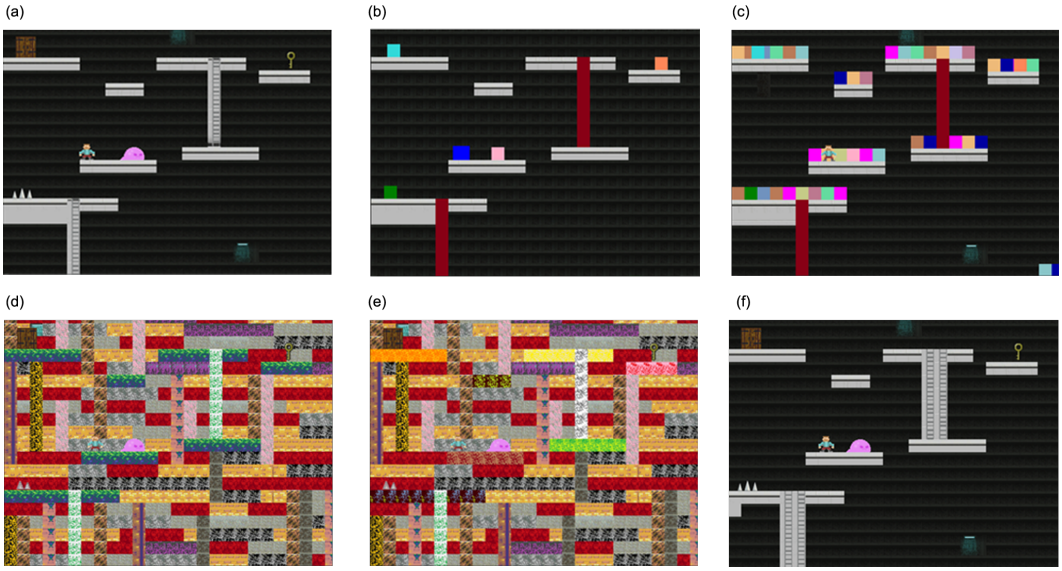

Figure 2: **Various game manipulations**. (a) Original version of the game. (b) Game with masked objects to lesion semantics prior . (c) Game with masked objects and distractor objects to lesion concept of object. (d) Game with background textures to lesion affordance prior. (e) Game with background textures and different colors for all platforms to lesion similarity prior. (f) Game with modified ladder to hinder participant's prior about ladder properties.

experiment style draws inspiration from the neuroscience literature wherein researchers study aspects about the human brain by performing lesion studies (Müller & Knight (2006), Shi & Davis (1999)). For the purposes of our experiment, since it was not possible to go directly inside participant's brain in order to study the importance of various priors, we did the next best thing possible - mask those priors.

For each version of the game created, we quantified human performance by recruiting 120 participants from Amazon Mechanical Turk. Each participant was instructed to use the arrow keys to move and finish the game as soon as possible. No information about the goals or the reward structure of the game was communicated to the participants. Each participant was paid $1 for successfully finishing the game. The maximum time for allowed for playing the game was set to 30 minutes. For each participant we recorded the (x, y) position of the player at every step of the game, the total time taken by the participant to finish the game (i.e. achieve the reward of +1) and the total number of deaths prior to finishing the game. We used this data to quantify participant's performance. Note that, we didn't repeat participants and thus no participant played more than one instance of the game.

3 Quantifying the importance of object priors

The first version of the game is shown in Figure 2(a),game link. From a single glance at the game, human players can employ their prior knowledge to interpret that the game agent can climb on ladders, it is supported by platforms, the pink slimy sprite is dangerous, spikes are to be avoided and probably the goal of the game is to take the key to open the door. As expected such interpretations enable humans to quickly solve the game. Figure 3(a) shows that the average time taken to complete the game is 1.8 minutes (blue bar) and the average number of deaths (orange bar) and game states visited by humans (yellow bar) are quite small.

3.1 Semantics

To study importance of prior knowledge about object semantics, we rendered objects and ladders with blocks of uniform color as shown in Figure 2(b),game link. Thus, in this game manipulation, the appearance of objects conveys no information about their semantics. Results in Figure 3(b),

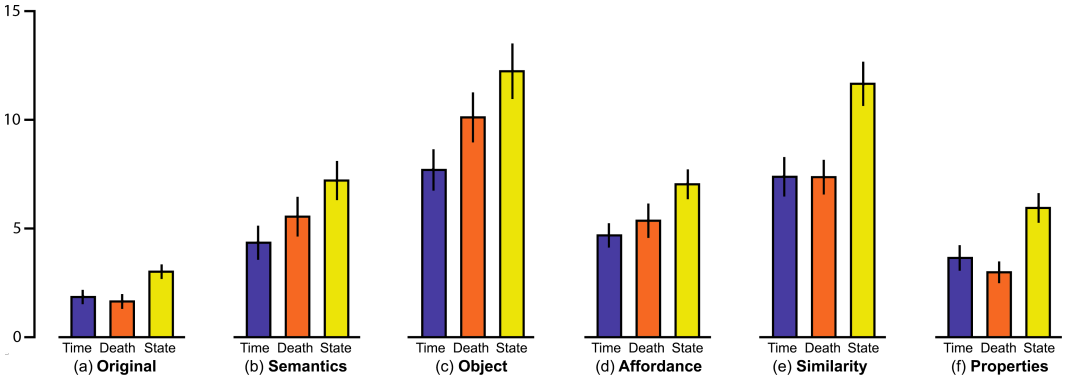

Figure 3: **Quantifying the influence of various object priors.** Blue bar shows average time taken by humans to solve the various games, orange bar shows average number of deaths in the games, and yellow bar shows number of unique states visited by players in the various games. For visualization purposes the number of deaths is divided by 2 and the number of states is divided by 1000 respectively.

show that human players take more than twice the time, have higher number of deaths, and explore significantly larger number of states (p-value < 0.01 for all measures) as compared to the original version of the game clearly demonstrating that lesioning semantics hurts human performance.

A natural next question is how do humans make use of semantic information? One hypothesis is that knowledge of semantics enables humans to infer the latent reward structure of the game. If this indeed is the case, then in the original game players should first visit the key and then go to the door, while in the version of the game without semantics, players should not exhibit any such bias. We found that in the original game, where key and door were both visible, almost all 120 participants reached the key first, while in the version with masked semantics only 42 out of 120 participants reached the key before the door (see Figure 4(a)). Further investigation into time taken by human players to reach the door after taking the key revealed that they take significantly larger amount of time when the semantics are masked (see Figure 4(b)). This provides further evidence that humans are unable to infer the reward structure and consequently significantly increase their exploration when semantics are masked. Note that, to rule out the possibility that increase in time is simply due to the fact players take more time to finish the game without semantics, the time to reach the door after taking the key was normalized by the total amount of the time that was spent by the player to complete the game.

To further quantify importance of semantics, instead of simply masking, we reversed the semantics. This condition further deteriorated human performance and the results are detailed in the appendix.

3.2 Objects as Subgoals for Exploration

While blocks of uniform color in game shown in Figure 2(b) convey no semantics, they are distinct from the background and seem to attract human attention. It is possible that humans infer these distinct entities (or objects) as subgoals, which results in more efficient exploration than random search. This leads to the hypothesis that humans have a prior to treat visually distinct entities as subgoals to guide exploration.

In order to test this, we modified the game to cover each space on the platform with a block of different color to hide where the objects are (see Figure 2(c), game link). Note that most colored blocks are placebos and do not correspond to any object and the actual objects have the same color and form as in the previous version of the game without semantics (i.e. Figure 2(b)). If the prior knowledge about entities that are visibly distinct are interesting to explore is critical, this manipulation in the game structure should lead to a significant change in human performance.

Results in Figure 3(c) show that masking where objects are leads to drastic deterioration in performance. The average time taken by human players to solve the game is nearly four times, number of deaths is nearly six times and humans explore four times as many game states as compared to the

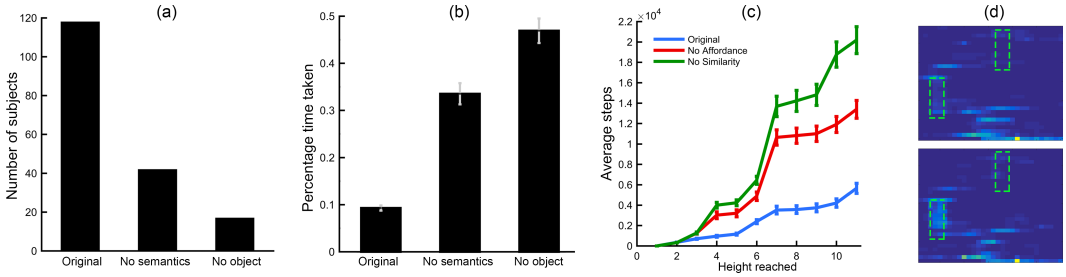

Figure 4: **Change in behavior upon lesion of various priors**. (a) Graph comparing number of participants that reached the key before the door in the original version, game without semantics, and game without object prior. (b) Graph showing amount of time taken by participants to reach the door once they obtained the key. (c) Graph showing average number of steps taken by participants to reach various vertical levels in original version, game without affordance, and game without similarity. (d) Heatmap comparing exploration trajectories of participants in original version of game (top) with respect to game with zigzag ladders (bottom). Ladders are highlighted via the green dashed boxes.

original game (Figure 3(c)). When compared to game version in which only semantic information was removed, the time taken, number of deaths, and number of states are all significantly greater (p-value < 0.01). When only semantics are masked, after encountering one object the human player is aware of what possible locations might be interesting to explore next. However, when objects are also masked it is unclear what to explore next. This effect can be seen by the increase in normalized time taken to reach the door from the key as compared to the game where only semantics are masked (Figure 4(b)). All these results suggest that knowing that visibly distinct entities are interesting and can be used as subgoals for exploration is a more important prior than knowledge of semantics.

3.3 Affordance

Uptil now, we manipulated objects in ways that made inferring the underlying reward structure of the game non-trivial. However, in these games it was obvious for humans that *platforms* can support agent sprites, *ladders* could be climbed to reach different platforms (even when the ladders were colored in uniform red in games shown in Figure 2(b,c), the connectivity pattern revealed what ladders were) and black parts of the game constitutes *free space*. Such knowledge about the use of an entity is referred to as affordance of the entity(Gibson (2014)). Note that we have purposefully constructed a difference between entities such as key, door, enemy, spike which cannot directly be used by the agent but convey the task structure and entities such as platforms, ladders and free space which do not necessarily convey the reward structure but are used to explore the environment. In the next set of experiments we manipulated the game to mask the affordance prior.

One way to mask affordances is to render the free space with random textures, which are visually similar to textures used for rendering ladders and platforms. Such rendering makes it difficult for humans to infer what parts of the game screen belong to platforms or ladders (see Figure 2(d), game link). Note that in this game manipulation, objects and their semantics are clearly observable. When tasked to play this game, humans require significantly more time, visit larger number of states, and die more often (p-value < 0.01) as compared to the original game. On the other hand, there is no significant different in performance of humans in this game when compared to performance of humans in the game without semantics i.e. Figure 2(b), implying that the affordance prior is as important as the semantics prior in our setup.

3.4 Things that look similar behave similar

In the previous game, although we masked affordance information, once the player realizes that it is possible to stand on a particular texture and climb a specific texture, it is easy to use color/texture similarity to identify other platforms and ladders in the game. Similarly, in the game with masked semantics (Figure 2(b)), visual similarity can be used to identify other enemies and spikes. These

considerations suggest that a general prior of the form that things that look the same act the same might help humans efficiently explore environments where semantics or affordances are hidden.

We tested this hypothesis by modifying the masked affordance game in a way that none of the platforms and ladders had the same visual signature (Figure 2(e), game link). Such rendering prevented human players from using the similarity prior. Figure 3(e)) shows that performance of humans was significantly worse in comparison to the original game (Figure 2(a)), the game with masked semantics (Figure 2(b)) and the game with masked affordances (Figure 2(d)) (p-value < 0.01). When compared to the game with no object information (Figure 2(c)), the time to complete the game and the number of states explored by players were similar, but the number of deaths was significantly lower (p-value < 0.01). These results suggests that visual similarity is the second most important prior used by humans in game play after the knowledge of directing exploration towards objects.

In order to gain insight into how this prior knowledge effects humans, we investigated the exploration pattern of human players. In the game when all information is visible we expect that progress of humans would be uniform in time. In the case when affordances are removed, the human players would initially take sometime to figure out what visual pattern corresponds to what entity and then quickly make progress in the game. Finally, in the case when the similarity prior is removed, we would expect human players to be unable to generalize any knowledge across the game and take large amounts of time exploring the environment even towards the end. We investigated if this indeed was true by computing the time taken by each player to reach different vertical distances in the game for the first time. Note that the door is on the top of the game, so the moving up corresponds to getting closer to solving the game. The results of this analysis are shown in Figure 4(c). The x-axis shows the height reached by the player and the y-axis show the average time taken by the players. As the figure shows, the results confirm our hypothesis.

3.5 How to interact with objects

Until now we have analyzed prior knowledge used by humans to interpret the visual structure in the game. However, interpretation of visual structure is only useful if the player understands what to do with the interpretation. Humans seem to possess prior knowledge about how to interact with different objects. For e.g., monsters can be avoided by jumping over them, ladders can be climbed by pressing the up key repeatedly etc. Deep reinforcement learning agents on the other hand do not possess such priors and must learn how to interact with objects by mere hit and trial.

To test how critical is such prior knowledge, we created a version of the game in which the ladders couldn't be climbed by simply pressing the up key. Instead, the ladders were zigzag in nature and in order to climb the ladder players had to press the up key, followed by alternating presses between the right and left key. Note that the ladders in this version looked like normal ladders, so players couldn't infer the properties of the ladder by simply looking at them (see Figure 2(f), game link). As shown in Figure 3(f), changing the property of the ladder increases the time taken, number of deaths, and states explored when compared to the original game (p-value < 0.01). The time spent by the agent in different parts of the game visualized in the original game (top row of Figure 4(d)) and this game reveals (bottom row of Figure 4(d)) that humans spend significantly more amount of time in the first ladder in the modified version of the game. However, once they learn about how to use the ladder they are able to quickly climb the second ladder.

When compared to game versions without semantics (Figure 2(b)) and without affordance (Figure 2(d)), we note that the number of deaths and states explored are significantly lesser ($p < 0.01$). This finding suggests that while prior knowledge about object properties plays a critical role in human gameplay, knowledge about semantics and affordances may be more important than this prior.

4 Taxonomy of object priors

In previous sections, we studied how different priors about objects affect human performance one at a time. We next sought to quantify human performance when all object priors investigated so far are simultaneously masked. This led to creation of the game shown in Figure 5(a) that hid all information about objects, semantics, affordance, and similarity(game link). As shown in Figure 5(b), human performance was extremely poor in this version of the game. The average time taken to solve the game increased to 20 minutes and the average number of deaths rose sharply to 40. Remarkably,

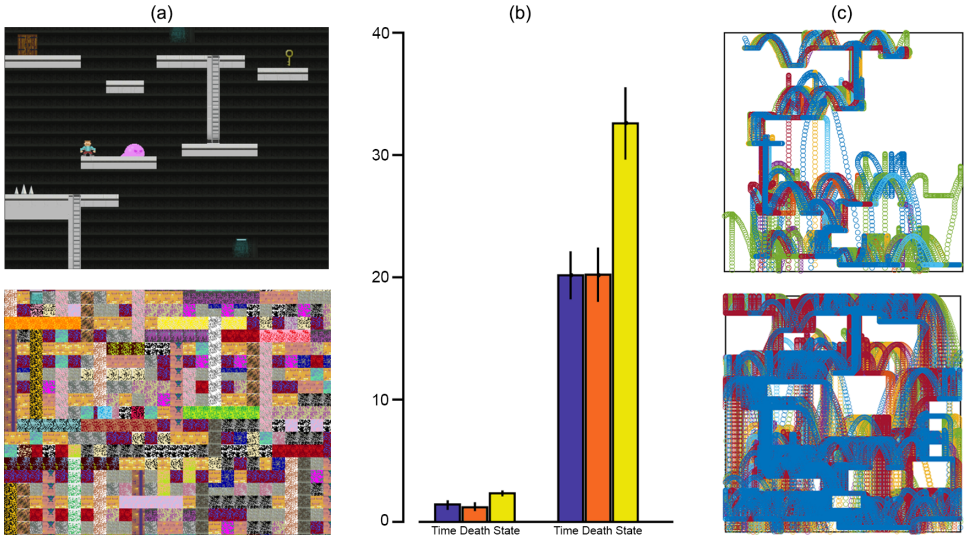

Figure 5: **Masking all object priors drastically affects human performance**. (a) Original version of the game (top) and version of game without any object priors (bottom). (b) Graph depicting difference in participant's performance for both the games. (c) Exploration trajectory for original version (top) vs no object prior version (bottom).

the exploration trajectory of humans is now almost completely random (refer to Figure 5(c)) with the number of unique states visited by the human players increasing by a factor of 9. Due to difficult in completing the game, we noticed a high dropout of human participants before they finished the game. We had to increase the pay to $2.25 to incentivize participants to not quit. Many participants noted that they could only solve the game by memorizing it.

Even though we preserved priors related to physics (e.g. objects fall down) and motor control (e.g. pressing left key moves the agent sprite to the left), just by rendering the game in a way that makes it impossible to use prior knowledge about how to visually interpret the game screen, makes the game extremely hard to play. To further test the limits of human ability, we designed a harder game where we also reversed gravity and randomly re-mapped the key presses to how it affect's the motion of agent's sprite. We, the creators of the game, having played previous version of the game hundred of times had an extremely hard time trying to complete this version of the game. This game placed us in the shoes of reinforcement learning (RL) agents that start off without the immense prior knowledge that humans possess. While improvements in performance of RL agents with better algorithms and better computational resources is inevitable, our results make a strong case for developing algorithms to incorporate prior knowledge as a way for improving the performance of artificial agents.

While there are many possible directions on how to incorporate priors in RL and more generally AI agents, it is informative to study how humans acquire such priors. Studies in developmental psychology suggest that human infants as young as two months old possess primitive notion of objects and expect them to move as connected and bounded wholes that allows them to perceive object boundaries and therefore possibly distinguish them from background (Spelke (1990); Spelke & Kinzler (2007)). At this stage, infants do not reason about object categories. By the age of 3-5 months, infants start exhibiting categorization behavior based on similarity and familiarity. (Mandler (1998), Mareschal & Quinn (2001)). The ability to recognize individual objects rapidly and accurately emerges comparatively late in development (usually by the time babies are 18-24 months old, Pereira & Smith (2009)). Similarly, while young infants exhibit some knowledge about affordances early during development, the ability to distinguish a walkable step from a cliff emerges only by the time they are 18 months old (Kretch & Adolph (2013)).

These results in infant development suggest that starting with a primitive notion of objects, infants gradually learn about visual similarity and eventually about object semantics and affordances. It is quite interesting to note that the order in which infants increase their knowledge matches the importance of different object priors such as existence of objects as sub-goals for exploration, visual

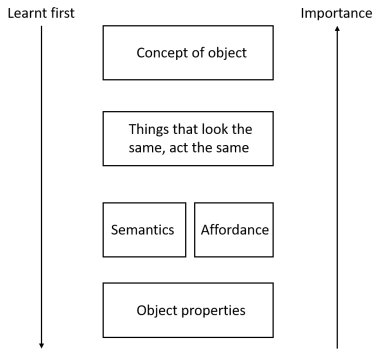

Figure 6: **Taxonomy of object priors.** The earlier an object prior is obtained during childhood, the more critical that object prior is in human problem solving in video games.

similarity, object semantics, and affordances. Based on these results, we suggest a possible taxonomy and ranking of object priors in Figure 6. We here put object properties at the bottom as in the context of our problem, knowledge about how to interact with specific objects can be only learnt once recognition is performed.

5 PRIOR KNOWLEDGE IS NOT ALWAYS DESIRABLE

For many interesting real world tasks, for pragmatic reasons it is often only possible to provide agents with a terminal reward when they succeed and they receive no external rewards otherwise. Success in such scenarios critically depends on the agent's ability to explore its environment and then quickly learn from its success (i.e. exploitation). While understanding what enables an agent to efficiently exploit is an interesting question, without a good exploration strategy no exploitation is possible. It therefore naturally follows that agents that can efficiently explore their environment will be good at completing tasks with sparse rewards. In this vein, our results demonstrate the importance of prior knowledge in helping humans explore efficiently in these sparse reward environments.

However, that being said, being equipped with strong prior knowledge may not be beneficial with regards to reward optimization in all kinds of environments. To illustrate this, we again recruited participants from Mechanical Turk ($n = 30$) to play a short game that simply consisted of a player and a princess at a short distance away from the player (Figure 7.a). Unknown to the participants, the game consisted of 10 hidden rewards (shown in yellow for illustration purposes) and the participants were given a bonus upon discovering them. As shown in (Figure 7.b), human players do not explore this environment and end up with suboptimal rewards. Upon entering the game, the players saw the princess and mostly inferred that as the goal and immediately reached her, thereby terminating the game. In contrast, a random agent (30 seeds of episode count=1 to simulate human experiments) ends up obtaining almost 4 times the rewards than human players.

Figure 7: **Prior information constrains human exploration.** (Left) A very simple game with hidden rewards (shown in dashed yellow). (Right) Average rewards accumulated by human players vs a random agent.

Research in developmental psychology has also demonstrated such instances wherein children have been shown to be better learners than adults in some cases (Lucas et al. (2014)). Thus, while incorporating prior knowledge in RL agents has many potential benefits, it is also important to consider if that could lead to inflexibility in an algorithm leading to inefficient exploration.

6 CONCLUSION

While there is no doubt that the performance of recent deep RL algorithms is impressive, there is much to be learned from human cognition if our goal is to enable RL agents to solve sparse reward tasks with human-like efficiency. Humans have the amazing ability to bring to bear their past knowledge (i.e., priors) to solve new tasks quickly. Our work takes one of the first steps to quantify the importance of various priors that humans employ in solving sparse reward tasks and in understanding how prior knowledge makes humans good at reinforcement learning tasks. We believe that our results will inspire researchers to think about different mechanisms of incorporating prior knowledge in the design of RL agents instead of starting from scratch. We also hope that our experimental platform of video games, available in open-source, will fuel more detailed studies investigating human priors and a benchmark for quantifying the efficacy of different mechanisms of incorporating prior knowledge into RL agents.

REFERENCES

Renee Baillargeon. How do infants learn about the physical world? *Current Directions in Psychological Science*, 3(5):133–140, 1994.

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

Marc Bellemare, Sriram Srinivasan, Georg Ostrovski, Tom Schaul, David Saxton, and Remi Munos. Unifying count-based exploration and intrinsic motivation. In *NIPS*, 2016.

Susan Carey. *The origin of concepts*. Oxford University Press, 2009.

James J Gibson. *The ecological approach to visual perception: classic edition*. Psychology Press, 2014.

Susan J Hespos, Alissa L Ferry, and Lance J Rips. Five-month-old infants have different expectations for solids and liquids. *Psychological Science*, 20(5):603–611, 2009.

Kari S Kretch and Karen E Adolph. Cliff or step? posture-specific learning at the edge of a drop-off. *Child Development*, 84(1):226–240, 2013.

Christopher G Lucas, Sophie Bridgers, Thomas L Griffiths, and Alison Gopnik. When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, 131(2):284–299, 2014.

Jean M Mandler. Representation. 1998.

Denis Mareschal and Paul C Quinn. Categorization in infancy. *Trends in cognitive sciences*, 5(10): 443–450, 2001.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 2015.

NG Müller and RT Knight. The functional neuroanatomy of working memory: contributions of human brain lesion studies. *Neuroscience*, 139(1):51–58, 2006.

Deepak Pathak, Pulkit Agrawal, Alexei A Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. *arXiv preprint arXiv:1705.05363*, 2017.

Alfredo F Pereira and Linda B Smith. Developmental changes in visual object recognition between 18 and 24 months of age. *Developmental science*, 12(1):67–80, 2009.

Changjun Shi and Michael Davis. Pain pathways involved in fear conditioning measured with fear-potentiated startle: lesion studies. *Journal of Neuroscience*, 19(1):420–430, 1999.

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

Elizabeth S Spelke and Katherine D Kinzler. Core knowledge. *Developmental science*, 10(1):89–96, 2007.

Daniel M Wolpert and Zoubin Ghahramani. Computational principles of movement neuroscience. *Nature neuroscience*, 3:1212–1217, 2000.

A FURTHER EXPERIMENTS ON SEMANTICS

A.1 REVERSING SEMANTIC INFORMATION

In Section 3.1, we masked semantic information by recoloring objects with plain colors. An alternate mechanism to manipulate the semantic prior is by reversing the semantic of different entities (i.e. objects that people associate as good are bad, and vice versa)). We created this version by replacing the pink enemy and spikes by coins and ice-cream sprite respectively which have a positive connotation, the ladder by fire, the key and the door by spikes and slimes which have negative connotations (Figure 8(a)).

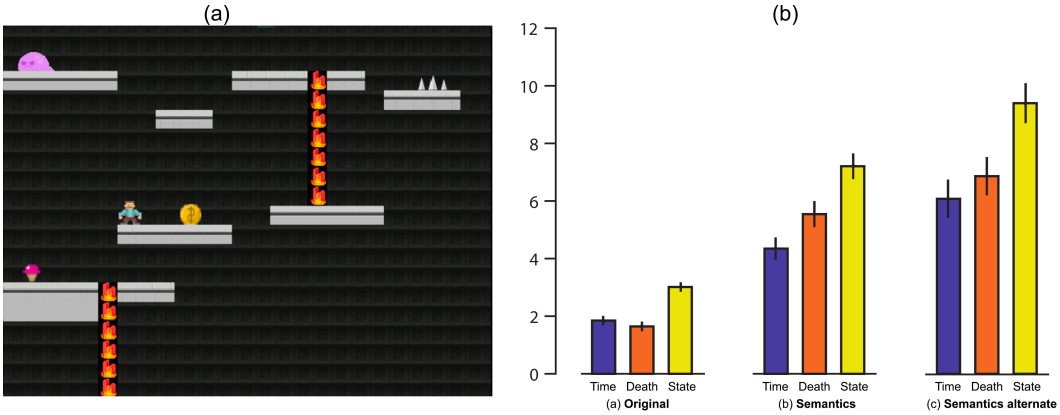

Figure 8: **Quantifying the importance of semantics.** (a) Game with reversed associations as an alternate way to lesion semantics prior. (b) Graph comparing performance of participants with respect to the original game and game with masked semantics.

As shown in Figure 8(b), participants took longer to solve this game compared to the original version with average time taken equal to 6 minutes (p-value < 0.05). The average number of deaths was also significantly greater and participants explored more compared to the original version (p-value $<$ 0.01 for both). Interestingly, participants also took longer to solve this game when compared to the masked semantics version (p-value < 0.05) implying that when we reverse semantic information, humans find the game even tougher to solve. This experiment further demonstrates that in absence of semantics (or reversal of semantics as in this case), human players performance in video games drops significantly.

B PHYSICS AND MOTOR CONTROL PRIORS

In addition to prior knowledge about objects, humans also bring in rich prior knowledge about intuitive physics as well as bring strong motor control priors when they approach a new task (Hespos et al. (2009), Baillargeon (2004), Wolpert & Ghahramani (2000), Baillargeon (1994)). Here, we have taken some initial steps to explore the importance of such priors in context of human gameplay.

B.1 GRAVITY

One of the most obvious knowledge that we have about the physical world is with regards to gravity i.e. things fall from up to down. To mask this prior, we created a version of the game in which the whole game window was rotated 90°. In this way, the gravity was reversed from left to right (as opposed to up to down).

As shown in Figure 9, participants spent more time to solve this game compared to the original version with average time taken close to 3 minutes (p-value < 0.01). The average number of deaths and number of states explored was also significantly larger than the original version (p-value $<$ 0.01).

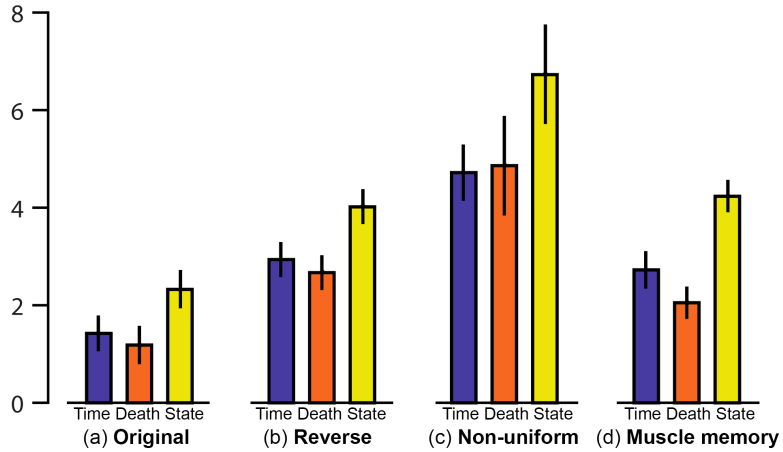

Figure 9: **Quantifying physics and motor control priors.** Graph comparing performance of participants in original version, game with gravity reversed, gave with non-uniform gravity, and game with key controls reversed. For visualization purposes the number of deaths is divided by 2 and the number of states is divided by 1000 respectively.

B.2 NON-UNIFORM GRAVITY

In the previous game, although we manipulated the gravity prior by reversing the gravity, participants still had access to more general notions about gravity such as gravity in the game is uniform and constant. We hypothesized that such a general prior about gravity might guide human exploration in an environment even when the gravity is reversed.

To test this, we modified the original game such that different platforms in the game had different gravity. This meant that some platforms had a very strong gravity so that the agent sprite couldn't jump on these platforms, some platforms had a very weak gravity so that the agent sprite could jump significantly higher, and some platforms had moderate gravity. Thus, in this version, participants had to learn about the dynamics of the game (related to gravity and jumping) from scratch. As shown in Figure 9, participants took a significantly longer time to solve this game compared to the version with reversed gravity with average time taken close to 5 minutes (p-value < 0.01). The average number of deaths and number of states explored was also significantly larger than the version with reverse gravity (p-value < 0.01). This suggests that similar to our results on object priors, general priors related to physics (such as uniform gravity) serve a prominent role in guiding efficient human gameplay.

B.3 MUSCLE MEMORY

Human players also come with knowledge of the form such as pressing arrow keys moves the agent sprite in the corresponding directions (i.e. pressing up makes the agent sprite jump, pressing left makes the agent sprite go left and so forth). We created a version of the game in which we reversed the arrow key controls. Thus, pressing left arrow key made the agent sprite go right, pressing right key moved the sprite left, pressing down key made the player jump (or go up the stairs), and pressing up key made the player go down the stairs.

Participants again took longer to solve this game compared to the original version with average time taken close to 3 minutes (refer to Figure 9). The average number of deaths and number of states explored was also significantly larger than the original version (p-value < 0.01). Interestingly, the performance of players when the gravity was reversed and key controls were reversed is similar with no significant difference between the two conditions.

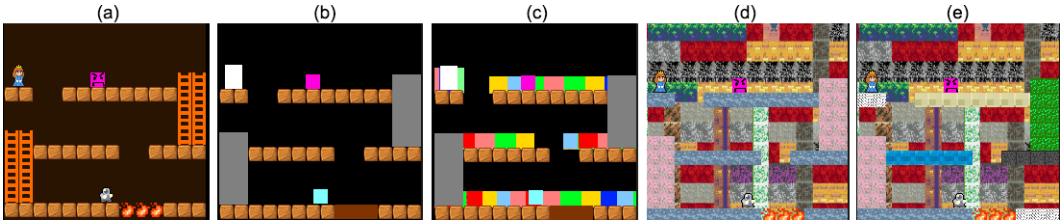

Figure 10: **Various game manipulations on which the RL agent was run.** (a) Original version. (b) Game without semantic information. (c) Game with masked and distractor objects to lesion concept of objects. (d) Game without affordance information. (e) Game without similarity information.

C PERFORMANCE OF RL AGENT ON VARIOUS GAME MANIPULATIONS

In this section, we investigated how the RL agent (ICM-A3C; Pathak et al. (2017)) performed in each of the lesioned settings we investigated with humans. While deep RL agents don't come with any prior knowledge, they can at least find and exploit regularities in data. Thus the experiments in this section can help shed light on how statistical regularities in the data influence deep RL agents.

To do this, we systematically created different versions of the game in Figure 1(a) to mask semantics, concept of object, affordance as well as similarity (refer to Figure 10) and ran 10 random seeds of the RL agent on each of the game versions. Note that, we modified the game in Figure 1(a) and not Figure 2(a) to run the RL experiments as the game in Figure 2(a) was too hard for the RL agent to solve.

As shown in Figure 11, the RL agent is unaffected by the removal of semantics, concept of objects as well as affordance information – there is no significant difference between the group means of the RL agent on these games and the original version. The performance of the RL agent on game without object information (Figure 10(c)) is especially interesting because this prior information is extremely critical to human gameplay. Interestingly, the RL agent is affected by removal of similarity information as it takes nearly two times to solve that version of the game implying that RL agents do exploit visual similarity in the data. Future work aims to investigate how this visual similarity is automatically learned.

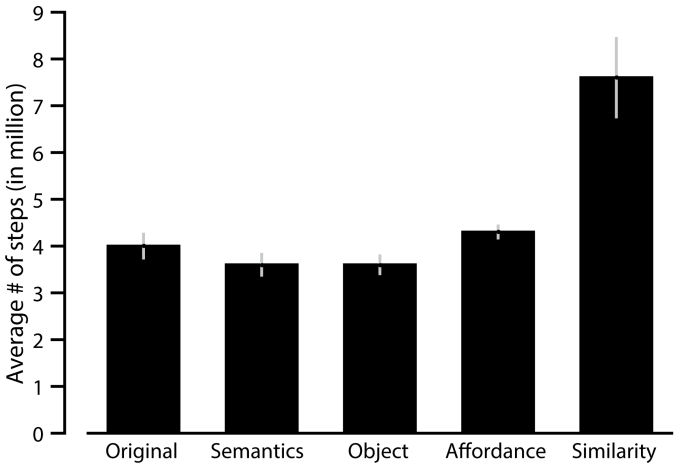

Figure 11: **Quantifying the performance of RL agent.** Graph comparing performance of RL agent on various game manipulations. Error bars indicate standard error of mean for the 10 random seeds. The RL agent performs similarly on all games except for the one without similarity.