Efficient replay memory through sensory adaptation

The cortex adapts quickly to repetitive stimuli. Such adaptation suggests that dissimilar (or novel) experiences are more likely to be retained in memory. In contrast, recent deep reinforcement learning models rely on storing every single input into a hippocampal-like episodic memory. Here, inspired by rapid forms of synaptic plasticity – a key neural basis of sensory adaptation – we propose a reinforcement learning algorithm in which only dissimilar enough inputs are stored into the replay memory. We show that our method leads to a more efficient memory representation (reduced memory load), as similar inputs tend to be discarded. In addition, a model in which less experiences are discarded as the agent gradually learns to explore its environment performs similarly to standard replay memory methods. This gradual change in adaptation is akin to the experimentally observed modifications of short-term plasticity over development and learning, and suggests an important role for this phenomenon in systems-level learning. Overall, our work shows how systems models of memory and learning can shed light on the function of synaptic plasticity and sensory adaptation.

Both animals and artificial agents need to efficiently learn to explore their environments. As animals/agents sample the environment, the incoming state tends to be initially highly correlated in time, which leads to poor optimization of non-linear approximators, such as deep neural networks in the context of reinforcement learning. However, our brain seems to deal with this problem efficiently. Here, we focus mainly on the learning implications of temporal filtering mediated by sensory adaptation coupled with hippocampus-like experience replay (ER) (Fig. 1a).

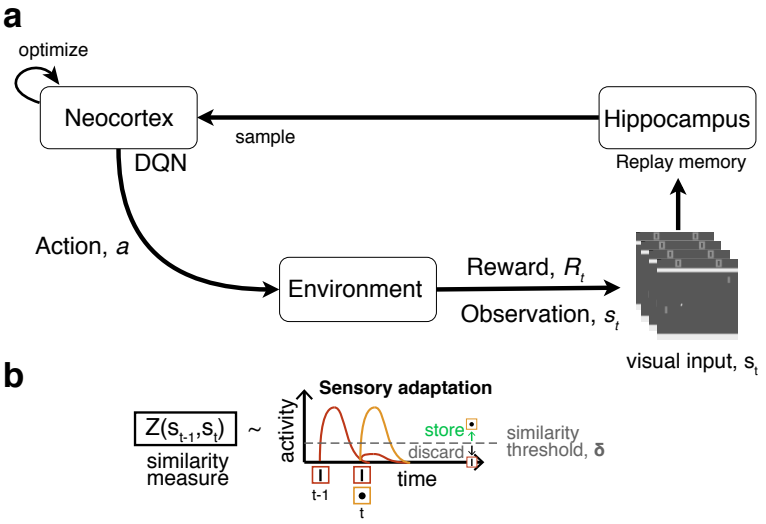

Figure 1: *Hippocampal-neocortical reinforcement learning architecture with sensory adaptation (modelled by similarity measure).*

Experience replay has enabled impressive results in deep RL (Mnih et al, 2015 Nature). On an abstract level, the *experience replay* algorithm (Lin 1992 Machine Learning) can be interpreted as a simplified model of neocortical-hippocampal interactions during memory consolidation (Hassabis et al, 2017 Neuron). The hippocampal network allows for rapid learning of episodic memories, which are in turn gradually replayed to the neocortex for long-term memory storage – systems memory consolidation (Ji et al, 2007 Nature). Following this view the neocortex implements a deep Q network (DQN, Fig. 1a), which is a multi-layered neural network that takes a state s as input and outputs a vector of action-values. By using the parameters d from DQN, with d being the dimension of DQN, this allows us to approximate our function $Q(s, a; \boldsymbol{\theta})$. Hippocampal-like ER allows us to draw past experiences randomly (Fig. 1a), decorrelating experience examples and stabilising the algorithm during the learning phase.

The key insight behind our model is that in many episodes an agent or an animal experiences tend be very similar/correlated. In RL terms, this means high similarity between consecutive states which leads to many similar experiences entering the replay memory (Fig. 1a). We drew inspiration from sensory adaptation to address this by comparing the input states using a similarity measure function as illustrated in Fig. 1b before the samples are added to the replay memory. Additionally, we consider two types of similarity threshold function, one that is constant and another in which the threshold decays over learning. Regarding the decay type, we consider two possible decay schemes, linear and exponential. The decay functions are defined as follows: $D_{\text{linear}}(t) = N_0 - \left(\frac{t}{f}\right)(N_0 - N_f)$ and $D_{\text{exp}}(t) = N_0 e^{-\lambda t}$, where λ controls the decay rate, t is the current step in environment, f is the decay constant, N_0 is the

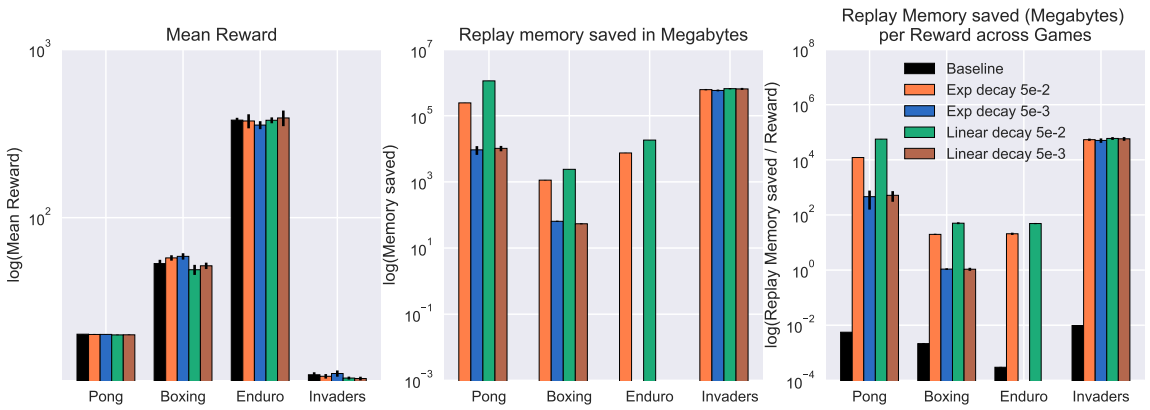

Figure 2: Filtered experience replay (FER; inspired by sensory adaptation) leads to a reduced replay memory load (middle and right) while achieving a similar performance (left) in Atari games when compared to standard experience replay (SER).

initial similarity value and N_f is the final similarity value. In our experiments, we set λ to 0.00001. The constant type has the same value throughout the learning phase, such that $N_0 = D_{\text{constant}}(t) = N_f$. In addition, we used the Normalised Root-Mean Square Error (NRMSE) function due to it being normalised and its ease/speed of computation. NRMSE is defined as $\text{NRMSE}(Ix, Iy) = \frac{\sqrt{\text{MSE}(Ix,Iy)} * \sqrt{n}}{\|Ix\|}$ where Ix and Iy are greyscale images with n pixels and MSE is the Mean Squared Error function. Our algorithm is a simple extension of DQN (Mnih et al, 2015 Nature).

We tested our algorithm on four environments (Pong, Boxing, Enduro and Invaders) of the Arcade Learning Environment (Atari) (Bellemare et al, 2015 IJCAI). The training is initially performed over two million frames. We compared our algorithm against a baseline based on the original DQN algorithm (Mnih et al, 2015 Nature). We used Adam (Kingma et al, 2015 ICLR) as our optimizer with a mini-batch size of 32 and trained our model for two million frames, across five different seed values.

Inspired by changes in short-term plasticity over learning (Costa et al. 2015 eLife) and development (Reyes and Sackman 1999 JNeurosci), we tested scenarios in which the threshold δ varied with learning, using exponentially and linear decays (see above). Our results demonstrate that our method can yield more efficient experience replay (i.e. lower memory load, Fig. 2 middle and right) while retaining a comparable performance (Fig. 2 left). This implies that our model provides a more efficient learning method compared to standard methods.

Finally, we are currently exploring a version of the model that automatically selects which inputs to discard using a simplified form of short-term plasticity at the synaptic level that is tuned during learning.

By combining abstract models of different brain areas and synaptic properties our work provides interesting insights into their interaction and suggests important roles when learning to explore complex environments. Moreover, it predicts that more correlated memories are stored in the hippocampal system as learning progresses, and that this is an important contributor to the exploration-exploitation trade off.