



OPEN

Context-dependent extinction learning emerging from raw sensory inputs: a reinforcement learning approach

Thomas Walther¹, Nicolas Diekmann¹, Sandhiya Vijayabaskaran¹, José R. Donoso¹, Denise Manahan-Vaughan², Laurenz Wiskott¹ & Sen Cheng¹✉

The context-dependence of extinction learning has been well studied and requires the hippocampus. However, the underlying neural mechanisms are still poorly understood. Using memory-driven reinforcement learning and deep neural networks, we developed a model that learns to navigate autonomously in biologically realistic virtual reality environments based on raw camera inputs alone. Neither is context represented explicitly in our model, nor is context change signaled. We find that memory-intact agents learn distinct context representations, and develop ABA renewal, whereas memory-impaired agents do not. These findings reproduce the behavior of control and hippocampal animals, respectively. We therefore propose that the role of the hippocampus in the context-dependence of extinction learning might stem from its function in episodic-like memory and not in context-representation per se. We conclude that context-dependence can emerge from raw visual inputs.

Treatment of anxiety disorders by exposure therapy is often followed by unwanted renewal of the seemingly extinguished fear. A better understanding of the cognitive and neural mechanisms governing extinction learning and fear renewal is therefore needed to develop novel therapies. The study of extinction learning goes back to Ivan Pavlov's research on classical conditioning¹. Pavlov and his colleagues discovered that the conditioned response (CR) to a neutral conditioned stimulus (CS) diminishes over time, if the CS is presented repeatedly without reinforcement by an unconditioned stimulus (US). This phenomenon is called extinction learning.

The arguably best understood extinction learning paradigm is fear extinction, in which an aversive stimulus, e.g. an electric foot shock, is used as US². The conditioned fear response diminishes during extinction, only to return later under certain circumstances³. While such return of fear can occur spontaneously (spontaneous recovery), or can be induced by unsignaled presentation of the US (reinstatement), it can also be controlled by context changes (ABA renewal) as follows: Subjects acquire a fear response to a CS, e.g. a neutral tone, by repeatedly coupling it with an US, e.g. a foot shock, in context A. Then the fear response is extinguished in a different context B, and eventually tested again in context A in the absence of the US. Renewal means that the fear response returns in the test phase despite the absence of the US⁴. ABA renewal demonstrates that extinction learning does not delete the previously learned association, although it might weaken the association³. Furthermore, ABA renewal underscores that extinction learning is strongly context-dependent^{4–10}.

Fear extinction is thought to depend mainly on three cerebral regions: The amygdala, responsible for initial fear learning, the ventral medial prefrontal cortex, guiding the extinction process, and the hippocampus, controlling the context-specific retrieval of extinction¹¹ and/or encoding contextual information¹¹. Here, we focus on the role of the hippocampus in extinction learning tasks. Lesions or pharmacological manipulations of the hippocampus have been reported to induce deficits in context encoding^{3,11}, which in turn impede context disambiguation during extinction learning^{5–8}. In particular, André and Manahan-Vaughan found that intracerebral administration of a dopamine receptor agonist, which alters synaptic plasticity in the hippocampus¹², reduced the context-dependent expression of renewal in an appetitive ABA renewal task in a T-maze⁹. This task depends on the specific involvement of certain subfields of the hippocampus¹³.

However, it remains unclear how behavior becomes context-dependent and what neural mechanisms underlie this learning. The majority of computational models targets exclusively extinction tasks that involve classical

¹Institute for Neural Computation, Ruhr University Bochum, Bochum, Germany. ²Neurophysiology, Medical Faculty, Ruhr University Bochum, Bochum, Germany. ✉email: sen.cheng@rub.de

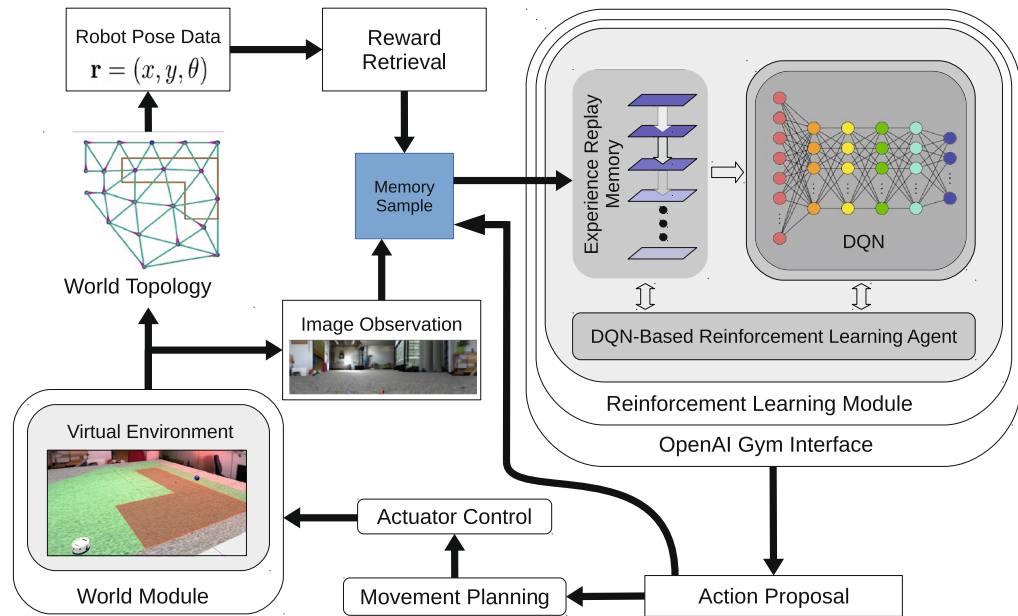


Figure 1. Model overview. Modules are sketched as rounded rectangles, other rectangles represent data and commands, the information/control flow is indicated by arrows. See the “Materials and methods” for an explanation of our model.

conditioning, like the earliest model of conditioning, the Rescorla-Wagner model¹⁴ (for an overview, see Ref.³). However, in its original form this model treats extinction learning as unlearning of previously acquired associations between CS and US and so could not account for the phenomenology of extinction learning, including renewal in particular. Nevertheless, the Rescorla-Wagner model inspired the development of reinforcement learning (RL) and many recent models of extinction learning. For instance, Ludvig et al. added experience replay¹⁵ to the Rescorla-Wagner model to account for a range of phenomena associated with extinction in classical conditioning: spontaneous recovery, latent inhibition, retrospective revaluation, and trial spacing effects¹⁶. However, their approach required explicit signals for cues and context, and did not learn from realistic visual information. Moreover, Ludvig et al. acknowledged that their method required the integration of more advanced RL techniques in order to handle operant conditioning tasks.

Computational models of extinction learning based on operant/instrumental conditioning are rather rare³. One of the few models was developed by Redish et al. and was based on RL driven by temporal differences. Their model emphasized the importance of a basic memory system to successfully model renewal in operant extinction tasks¹⁷. However, their model relied on the existence of dedicated signals for representing cues and contexts, and did not learn from realistic visual input.

Overcoming the caveats of existing approaches, we propose a novel model of operant extinction learning that combines deep neural networks with experience replay memory and reinforcement learning, the deep Q-network (DQN) approach¹⁵. Our model does not require its architecture to be manually tailored to each new scenario and does not rely on the existence of explicit signals representing cues and contexts. It can learn complex spatial navigation tasks based on raw visual information, and can successfully account for instrumental ABA renewal without receiving a context signal from external sources. Our model also shows better performance than the model by Redish et al.: faster acquisition in simple mazes, and more biologically plausible extinction learning in a context different from the acquisition context. In addition, the deep neural network in our model allows for studying the internal representations that are learned by the model. Distinct representations for contexts A and B emerge in our model during extinction learning from raw visual inputs obtained in biologically plausible virtual reality (VR) scenarios and reward contingencies.

Materials and methods

Modeling framework. All simulations in this paper were performed in a VR testbed designed to study biomimetic models of rodent behavior in spatial navigation and operant extinction learning tasks (Fig. 1). The framework consists of multiple modules, including a VR component based on the Blender rendering and simulation engine (The Blender Foundation, <https://www.blender.org/>), in which we built moderately complex and biologically plausible virtual environments. In the VR, the RL agent is represented by a small robot that is equipped with a panoramic camera to simulate the wide field of view of rodents. The topology module defines allowed positions and paths for the robot and provides pose data together with environmental rewards. In conjunction with image observations from the VR module, these data make up a single memory sample that is stored in the experience replay module. The DQN component relies on this memory module to learn its network

weights and generates action proposals that drive the virtual robot. Communication between the VR module and the RL-learning framework is established via a standard OpenAI Gym interface¹⁸ (<https://gym.openai.com/>).

The modeling framework allows for two modes of operation: an online model, in which simulations are run in real-time and physics-based simulations are supported, and an offline mode, in which multiple simulations can be run in parallel on pre-rendered camera images. The latter operating mode increases the computational efficiency of our model, and facilitates the rapid generation of large numbers of learning trials and repetitions. All results in this paper were obtained using the offline mode.

Visually driven reinforcement learning with deep Q-networks. The agent learns autonomously to navigate in an environment using RL¹⁹. Time is discretized in time steps t ; the state of the environment is represented by $S_t \in \mathcal{S}$, where \mathcal{S} is the set of all possible states. The agent observes the state S_t and selects an appropriate action, $A_t \in \mathcal{A}(S_t)$, where $\mathcal{A}(S_t)$ is the set of all possible actions in the current state. In the next time step, the environment evolves to a new state S_{t+1} , and the agent receives a reward R_{t+1} .

In our modeling framework, the states correspond to spatial positions of the agent within the VR. The agent observes these states through visual images captured by the camera from the agent's position (e.g. Fig. 2). The set of possible states, \mathcal{S} , is defined by the discrete nodes in a manually constructed topology graph that spans the virtual environment (Fig. 3). State space discretization accelerates the convergence of RL. The possible actions in each state, $\mathcal{A}(S_t)$, are defined by the set of outgoing connections from the node in the topology graph. Once an action is chosen, the agent moves to the connected node. In the current implementation, the agent does not rotate.

For the learning algorithm we adopt a specific variant of RL called Q-learning, where the agent learns a state-action value function $Q(S, A)$, which represents the value to the agent of performing action A in state S . The agent learns from interactions with the environment by updating the state-action value function according to the following rule¹⁹:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \delta Q_t \quad (1a)$$

$$\delta Q_t = \alpha \left[R_{t+1} + \gamma \max_{a \in \mathcal{A}(S_t)} Q(S_{t+1}, a) - Q(S_t, A_t) \right], \quad (1b)$$

where α is the learning rate and γ is the discount factor that devalues future rewards relative to immediate ones. The Q-learning update rule has been shown mathematically to converge to a local optimum. Since the state-action value function is likely very complex and its analytical model class is unknown, we use a deep neural network to approximate $Q(S, A)$ ¹⁵. In our simulations, the DQN consists of fully connected, dense layers between the input and output layer and the weights are initialized with random values. We use the tanh-function as the activation function.

The weights in the DQN are updated using the backpropagation algorithm²⁰. This algorithm was developed for training multi-layered networks in supervised learning tasks, where a network learns to associate inputs with target outputs. For a given input, the difference between the network's output and the target output is calculated; this is called the error. The network weights are then updated so as to reduce the error. Updating begins in the output layer and is applied successively down to the preceding layer. However, in reinforcement learning, which we use here, the goal is different and there are no target input-output associations to learn. Nevertheless, the backpropagation algorithm can still be applied because the DQN represents the Q function and Eq. (1b) represents the error that should be used to update the Q function after an experience. In other words, in Q learning the target function is not known but the error is, and that is all that is needed to train the DQN. Furthermore, to stabilize the convergence, we use two copies of the DQN, which are updated at different rates: The online network is updated at every step, while the target network is updated every N steps by blending it with the online network.

The input to the network consists of (30,1)-pixel RGB images, which are captured in the virtual environment. The network architecture comprises an input layer (with 90 neurons for 30 pixels \times 3 channels), four hidden layers (64 neurons each), that enable the DQN to learn complex state-action value functions, and an output layer that contains a single node for each action the agent can perform. We use the Keras²¹ and KerasRL²² implementation for the DQN in our model.

To balance between exploration and exploitation our agent uses an ϵ -greedy strategy¹⁹: In each state S_t , the agent either selects the action that yields the highest $Q(S_t, A_t)$ with a probability of $1 - \epsilon$ (exploitation), or chooses a random element from the set of available actions $\mathcal{A}(S_t)$ with a probability of ϵ (exploration). To ensure sufficient exploration, we set $\epsilon = 0.3$ in all simulations, unless otherwise noted.

In our model, experiences are stored in memory to be replayed later. An experience contains information about the state observed, action taken, and rewards received, i.e. $(S_t, A_t, S_{t+1}, R_{t+1})$. The memory module uses first-in-first-out memory, i.e. if the memory's capacity c is exceeded, the oldest experiences are deleted first to make room for novel ones. In each time step, random batches of memory entries are sampled for experience replay¹⁵. The DQN is trained with those replayed experiences; replay memory access is quite fast, and random batches of experiences are sampled from the replay memory in each training cycle in order to drive learning and update the weights of the DQN. Experience replay in RL was inspired by hippocampal replay of sequential neural activity encountered during experience, which is hypothesized to play a critical role in memory consolidation and learning^{23,24}. We therefore suggest that lesions or drug-induced impairments of the hippocampal formation can be simulated by reducing the agent's replay memory capacity. The interleaved use of new and replayed experiences in our model resembles awake replay in animals²⁵ and makes a separate replay/consolidation phase unnecessary since it would not add any new information.

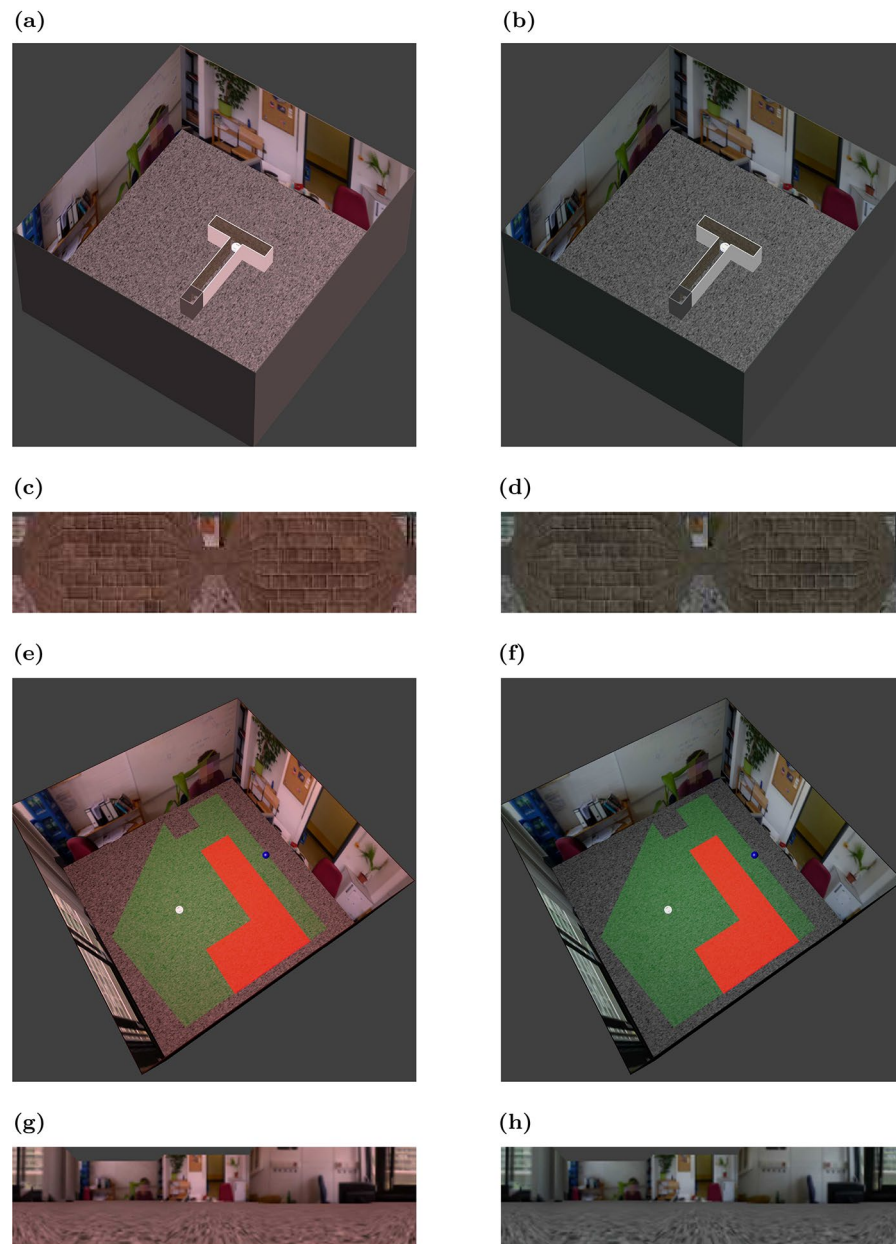


Figure 2. Examples of virtual environments and camera views. **(a,b)** Overview of T-maze in two different contexts. Two different contexts are generated by using different colors for the overall illumination. **(c,d)** Camera views in T-maze from the agent's perspective. **(e,f)** Overview of shock zone maze used to study inhibitory avoidance. **(g,h)** Camera views in the shock zone maze.

Positive or negative rewards drive RL¹⁹. These rewards are task-specific and will be defined separately for each simulation setup below.

Operant learning tasks simulated in our model. *General setup.* All trials used in our simulations had a common structure. Each trial began with the agent located at one of the defined starting nodes. In each following time step, the agent could choose to execute one action that was allowed in the current state as defined by the topology graph. Each time step incurred a negative reward of -1 to encourage the agent to find the shortest path possible from the start to the goal node. Trials ended when either the agent reached the goal node, or the number of simulated time steps exceeded 100 (time-out). The experience replay memory capacity was set to $c = 5000$. Different contexts were defined by the illumination color of the house lights. Context A was illuminated with red house lights and context B with white house lights.

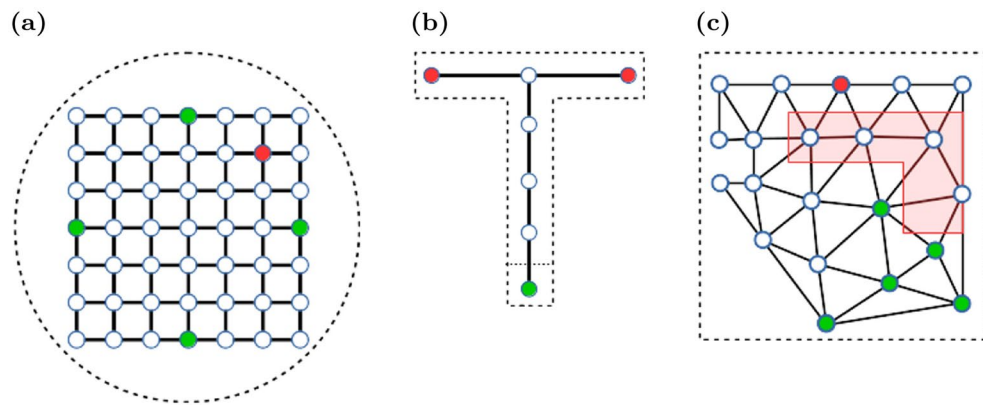


Figure 3. Topology graphs for different simulation environments. **(a)** Morris watermaze. **(b)** T-maze. **(c)** Shock zone maze. The red polygon indicates the shock zone. Nodes in the graph indicate allowed positions, and solid lines allowed transitions between positions. Start nodes are marked green, and (potential) goal nodes are marked red. The dashed lines sketch the maze perimeter.

Linear track. As a proof-of-concept and to facilitate a comparison to Redish et al.'s model, we first studied spatial learning on a simple linear track. The track was represented by a one-dimensional topology graph consisting of $N = 3, \dots, 15$ nodes. Since Redish et al.'s model requires clearly separable state representations, we used binary encoded node state indices in this simulation, instead of the raw camera input as in the other simulations. The starting node was always the first node, while the target node was node N . The agent received a reward of +1.0 for reaching the target node.

Morris watermaze. Next, we simulated spatial navigation in the more challenging Morris watermaze²⁶ to show that our model can learn operant spatial tasks in 2-d from raw visual input. The watermaze was circular with a diameter of 2 m. It was placed in the center of a typical lab environment to provide distal cues. The topology graph for this maze consisted of a rectangular grid that covered the center part of the watermaze (Fig. 3a). Four nodes {N(orth),E(ast),S(outh),W(est)} were used as potential starting nodes. They were selected in analogy to a protocol introduced by Anisman et al.²⁷: Trials were arranged into blocks of 4 trials. Within each block, the sequence of starting locations was a random permutation of the sequence (N, E, S, W). The goal node was always centered in the NE quadrant and yielded an additional reward of + 1.0.

Extinction learning in T-maze. To study the emergence of context-dependent behavior, we simulated the following ABA renewal experiment⁹: In a simple T-Maze, rats were rewarded for running to the end of one of the side arms (target arm). The acquisition context A was defined by a combination of distal and proximal visual cues, as well as odors. Once the animals performed to criterion, the context was changed to B, which was distinct from A, and the extinction phase began. In this phase, animals were no longer rewarded for entering the former target arm. Once the animals reverted to randomly choosing between the two side arms, the context was switched back to A and the test phase started. Control animals showed renewal and chose the former target arm more frequently than the other arm, even though it was no longer rewarded.

In our modeling framework, we maintained the spatial layout of the experimental setup and the richness of visual inputs in a laboratory setting, which is usually oversimplified in other computational models. A skybox showed a standard lab environment with distal cues for global orientation, and maze walls were textured using a standard brick texture (Fig. 2a–d). We made a few minor adjustments to accommodate the physical differences between rats and the simulated robot. For instance, we had to adjust the size of the T-maze and use different illumination to distinguish contexts A and B. In the rat experiments, the two contexts were demarcated by odor cues in addition to visual cues⁹. The simplifications in our model make the distinction between contexts harder, as odor was a powerful context disambiguation cue in the original experiment.

The topology graph in the T-Maze consisted of seven nodes (Fig. 3b). Reaching a node at the end of the side arms yielded a reward of + 1.0 to encourage goal-directed behavior in the agent. The left endpoint was the target node and yielded an additional reward of + 10.0 only in the acquisition phase, not in the extinction and test phases. As in the experiments, a trial was scored as correct, if the agent entered the target arm, and as incorrect, if the other arm was selected or the trial timed out. Finally, to compare our model account of ABA renewal to that of Redish et al.¹⁷, we used the same topology graph, but provided their model with inputs that represented states and contexts as binary encoded indices.

A simulated session consisted of 300 trials in each of the experimental phases: acquisition, extinction and test. So, altogether a session consisted of 900 trials.

Extinction learning in inhibitory avoidance. To model learning in a task that requires complex spatial behavior in addition to dynamically changing reinforcements, we use an inhibitory avoidance task in an ABA renewal design. The agent had to navigate from a variable starting location to a fixed goal location in the shock zone maze

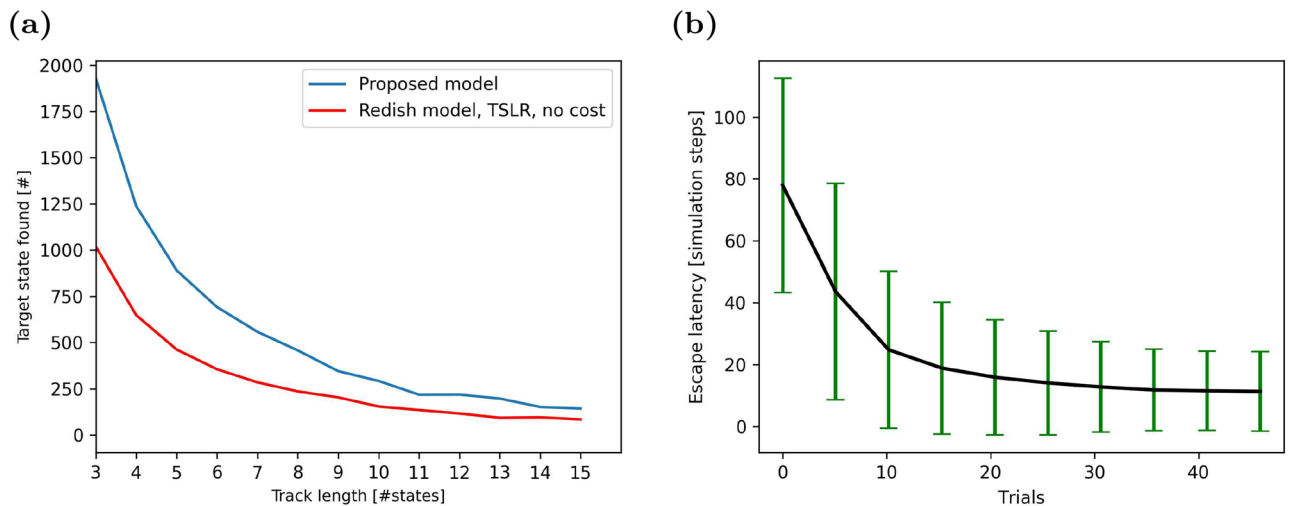


Figure 4. Performance on spatial navigation tasks. **(a)** Number of times the agent finds the goal location given a total of 5000 simulation steps on a linear track in our model (blue line) and comparison model¹⁷ (red line). Results are averaged over 50 simulations for each track length. **(b)** Escape latency in the virtual watermaze decreases with learning trials. Data averaged over 1000 agents, error bars represent standard deviations.

(Fig. 2e–h), while avoiding a certain region of the maze, the shock zone. Distal cues were provided in this environment using the same skybox as in the watermaze simulation. The topology graph was built using Delaunay triangulation (Fig. 3c). The start node was randomly chosen from four nodes (green dots). Reaching the goal node (red dot) earned the agent a reward of +1.0. If the agent landed on a node in the shock zone (red area, Figs. 2e–f, 3c), it would receive a punishment (reward = −20.0). In both the extinction and test phases, the shock zone was inactive and no negative reward was added for landing on a node within the shock zone.

Neural representations of context in ABA renewal tasks. To study the influence of experience replay and model the effect of hippocampal lesions/inactivations, we compared agents with intact memory ($c = 5000$) to memory-impaired agents with $c = 150$.

To study the representations in the DQN layers, we used dimensionality reduction as follows: We defined $I_l(z, \mathbf{x})$ to represent the activation of layer $l \in [0, 1, \dots, 5]$ when the DQN is fed with an input image captured at location \mathbf{x} in context $z \in \{A, B\}$, where $l = 0$ is the input layer, and $l = 5$ the output layer. Let \mathbf{X}_A and \mathbf{X}_B be the sets of 1000 randomly sampled image locations in context A and B, respectively. We expected that the representations $I_l(A, \mathbf{X}_A)$ and $I_l(B, \mathbf{X}_B)$ in the DQN's higher layers form compact, well-separated clusters in activation space by the end of extinction. This would enable agents to disambiguate the contexts and behave differently in the following test phase in context A from how they behaved in context B at the end of extinction, which is exactly renewal. By contrast, for memory-impaired agents we would expect no separate clusters for the representations of the two contexts, which would impair the agent's ability to distinguish between contexts A and B, which in turn would impair renewal. We applied Linear Discriminant Analysis (LDA) to find a clustering that maximally separates $I_l(A, \mathbf{X}_A)$ and $I_l(B, \mathbf{X}_B)$. The cluster distances d_l are related to the context discrimination ability of network layer l .

Results

We developed a computational framework based on reinforcement learning driven by experience memory replay and a deep neural network to model complex operant learning tasks in realistic environments. Unlike previous computational models, our computational modeling framework learns based on raw visual inputs and does not receive pre-processed inputs.

Proof-of-principle studies of spatial learning. As a proof-of-principle and to facilitate a direct comparison to an earlier model, we first studied spatial learning on a simple linear track with simplified inputs, i.e. the positions were represented by binary indices. To vary the difficulty of the task, the track was represented by a 1-d topology graph with variable node count. The first node was always the starting node and the agent was rewarded for reaching the last node. The larger the number of nodes in the graph, the more difficult it was for the agent to solve the task. Hence, the average number of trials, in which the agent reached the goal node before the trial timed out, decreased with the number of nodes N (Fig. 4a, blue line). The reason for this result is simple. The agent had to find the goal node by chance at least once before it could learn a goal-directed behavior. This was less likely to occur in longer tracks with a larger number of intermediate nodes.

Comparing our model to existing approaches turned out to be difficult, since operant extinction learning has not received much attention in the past. The model proposed by Redish et al.¹⁷ was eventually chosen as a reference. Even though this model was developed for different kinds of tasks, it could solve the spatial learning task on the linear track if inputs provided to the model were simplified to represent states as indices (Fig. 4b, red line). However, Redish et al.'s model did not perform as well as our model regardless of the track length.

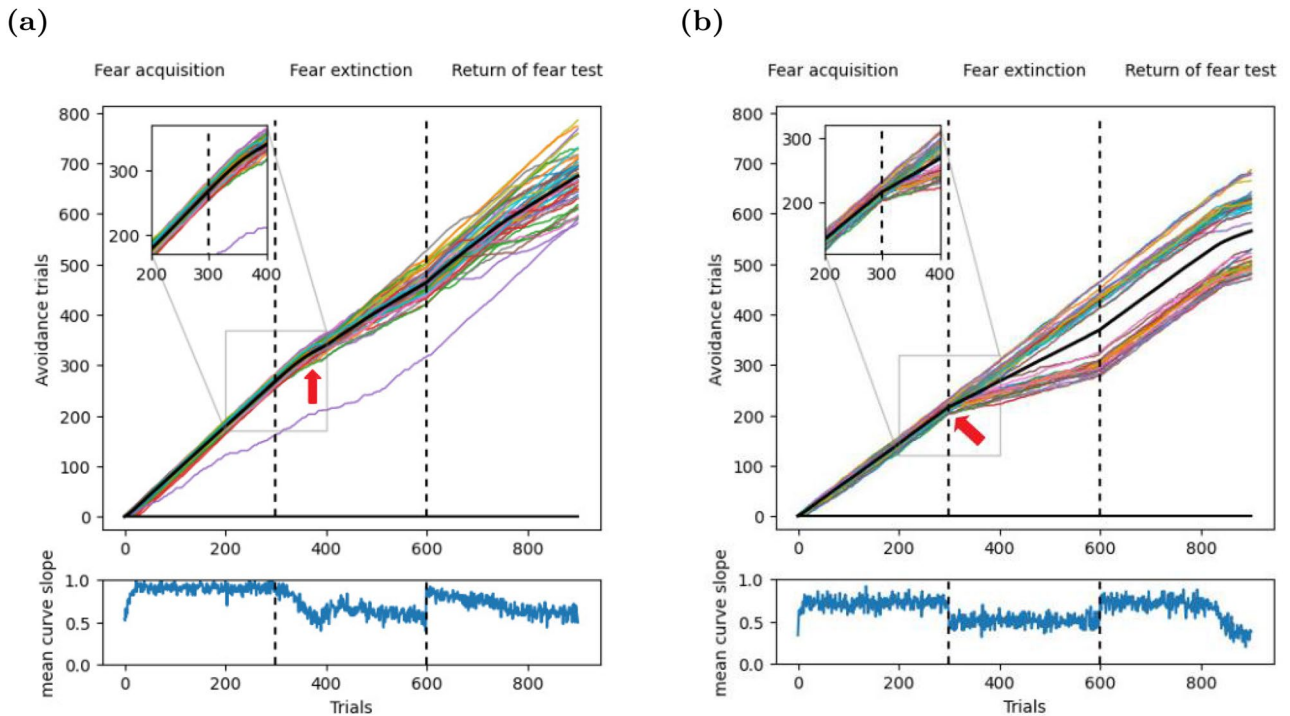


Figure 5. Dynamics of acquisition, extinction, and renewal-testing in T-maze. **(a)** Top: Cumulative response curves (CRCs) in an operant ABA renewal paradigm in the T-maze. Shown are 50 simulated CRCs (colored lines), and the mean CRC (black line). The inset shows the beginning of the extinction phase in context B, where the CRCs continue with the same slope as at the end of the acquisition phase in context A. This indicates that the association learned in context A generalized to context B. The red arrow points where the slopes start to change. Bottom: The slope of the mean CRC reveals rapid acquisition, gradual extinction learning, and renewal. **(b)** Top: In contrast to our model, Redish et al.'s model¹⁷ changes its behavior immediately upon the switch from context A to context B (red arrow). Hence, this model does not generalize the learned association from context A to context B. In addition, one group of CRCs does not exhibit any extinction learning, and hence no renewal. Bottom: The slope of the mean CRC shows the sudden change in behavior at the onset of the extinction phase.

This result is due to experience memory in our model, which could replay samples of successful navigation to speed up learning.

Next, we studied more complex spatial learning in a 2-d watermaze²⁶ with raw camera inputs. Since the agent was rewarded for finding the location of the escape platform as quickly as possible, we quantified learning performance by the escape latency (Fig. 4a). The learning curve indicates clearly that the system's performance improved rapidly with learning trials, i.e. the latency decreased. Our result compared well to those reported in the literature on rodents²⁸.

Extinction and renewal of operant behavior. Having shown that our model agent can learn simple spatial navigation behaviors based on raw camera inputs, we turned our attention to more complex learning experiments such as operant extinction and renewal. Initial tests were run in the T-maze. Performance was measured with cumulative response curves (CRCs). If the agent entered the target arm, the CRC was increased by one. If the other arm was entered or the trial timed out, the CRC was not changed. Hence, consistent correct performance yielded a slope of 1, (random) alternation of correct and incorrect responses yielded a slope of 0.5, and consistent incorrect responses yielded a slope of 0. The CRC for our model looked just as expected for an ABA renewal experiment (Fig. 5a). After a very brief initial learning period, the slope increased to about 1 until the end of the acquisition phase, indicating nearly perfect performance. At the transition from the acquisition to the extinction phase, the slope of the CRCs stayed constant in the first 50 extinction trials, indicating that the agent generalized the learned behavior from context A to context B, even though it no longer received a larger reward for the left turn. After 50 extinction trials, the slope decreased gradually (Fig. 5a, red arrow), indicating extinction learning. Finally, the change from context B back to context A at the onset of the test phase triggered an abrupt rise of the CRC slope, indicating a renewal of the initially learned behavior. The CRC for Redish et al.'s model looked similar to that for our model (Fig. 5b), but exhibited two key differences that were not biologically plausible. First, the CRCs split into one group that showed no extinction, and hence no renewal either, and another group that showed extinction and renewal. Second, the latter group showed a sharp drop of the CRC slope in the very first extinction trial (Fig. 5b, red arrow), indicating an immediate shift in behavior upon the first unrewarded trial.

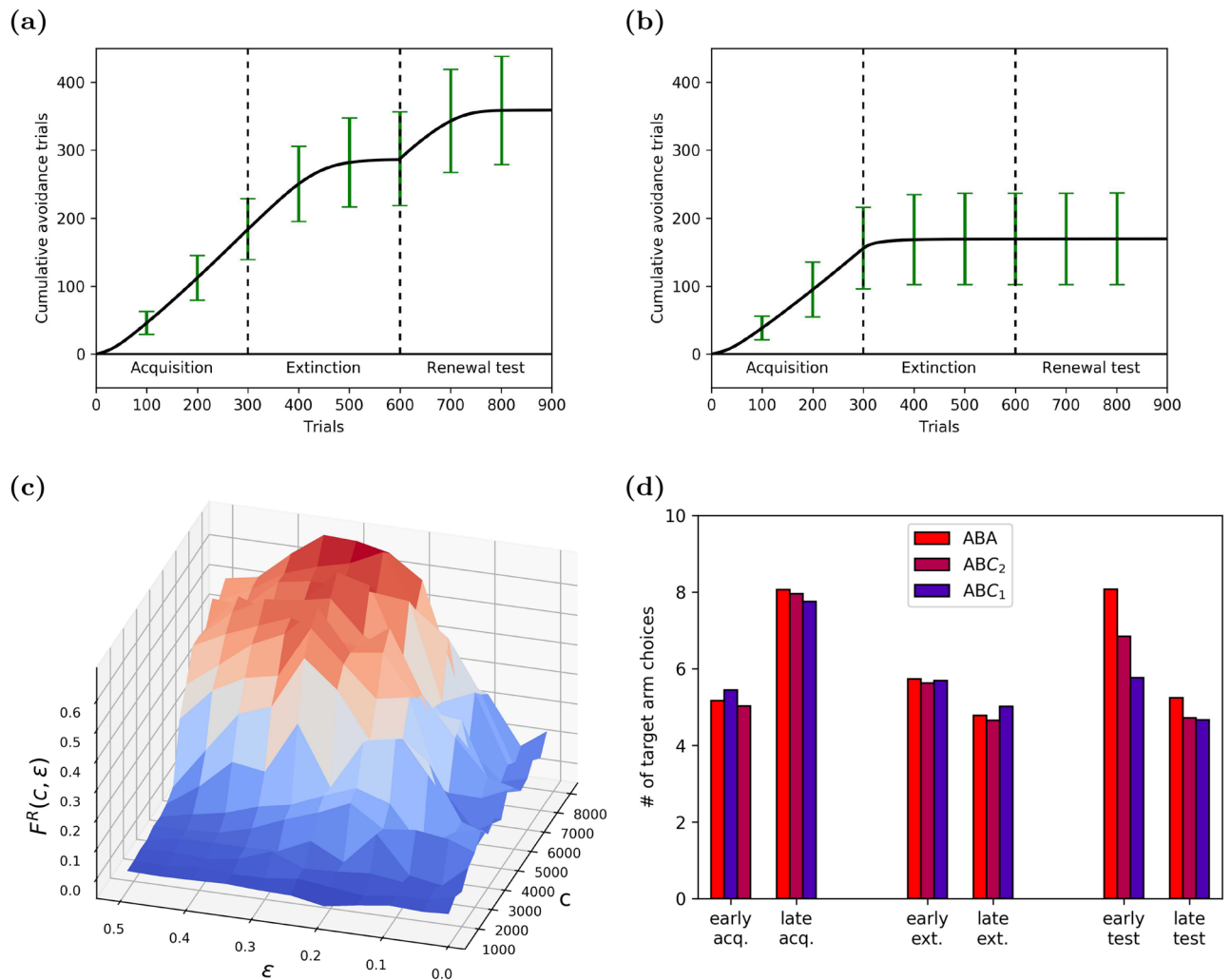


Figure 6. Learning dynamics of inhibitory avoidance in simulated renewal paradigms. Panels (a–c) show results from ABA renewal, panel d compares ABA renewal to ABC renewal. (a) Cumulative response curve (CRC) in the shock zone maze averaged across 1000 memory-intact agents (error bars indicate std). Agents showed clear evidence for extinction, i.e. flattening of CRC at the end of the extinction phase, and renewal, i.e. positive slope at the beginning of the renewal phase. (b) By contrast, memory-impaired agents showed faster extinction (flattening of CRC) and no renewal (flat CRC at onset of renewal testing). (c) Renewal strength $F^R(c, \epsilon)$ as a function of the capacity of the experience replay memory c and the ϵ -greedy exploration parameter. The renewal strength increases with c ; note that the largest renewal effect is observed along the ridge for $c > 4000$ and $0.25 < \epsilon < 0.35$. (d) Renewal strength in the T-Maze in two neutral contexts C_1 and C_2 (ABC renewal), compared to standard ABA renewal strength. (R,G,B) illumination values for the contexts were: (1.0,0.0,0.0) for context A, (0.0,0.0,0.0) for context B, (0.3,0.0,0.7) for context C_1 , and (0.7,0.0,0.3) for C_2 . Bars colored according to the house light color in the renewal test context (A, C_2 or C_1). The strength of ABC renewal decreases as the similarity between context C and A decreases. Results were averaged over 100 runs.

To test our model in an even more challenging task, we used an inhibitory avoidance learning paradigm in the shock zone maze, which combined complex spatial navigation, positive and negative rewards (see “Methods”), in an ABA renewal paradigm. In this task, agents had to navigate to a goal location in a 2-d environment along the shortest path possible, while at the same time avoiding a shock zone with an irregular shape (Fig. 2e,f). A trial was considered correct if the agent reached the goal without entering the shock zone on the way. This definition allowed us to use the CRC to quantify the trial-by-trial performance of simulated agents. The curve increased by one if a trial was correct and remained constant otherwise. That means, a slope of 1 in the CRC indicated perfect avoidance of the shock zone, and zero slope represented no avoidance at all.

Even in this complex learning task, the memory-intact agent in our model showed biologically realistic behavior. The mean CRC started with a slope of about 0.0 in the acquisition phase (Fig. 6a). After ≈ 100 trials, the CRC slope increased to about 0.8, as the agent learned to circumnavigate the shock zone. At the beginning of the extinction phase, the agent continued to avoid the shock zone, even though shocks were no longer applied and even though the context had changed from A to B. However, by the end of the extinction phase, the agent learned to reach the goal faster by traversing the shock zone, as indicated by a CRC slope of about 0.0. When

the context was switched back to context A at the onset of the test phase, the CRC slope rose sharply, indicating renewal (Fig. 6a). Over the course of the test phase, the agents learned that the shock zone was no longer active in context A and that it could be safely traversed, causing a smooth drop of the CRC slope to about 0.0 in the final trials. These results showed that the agents in our model could learn context-dependent behavior even in a highly complex operant learning task in a two-dimensional environment.

To confirm that renewal in our model is indeed triggered by the context change, we tested for ABC renewal in the T-Maze, i.e., acquisition in context A (red house lights), extinction in context B (white house lights), and renewal in one of two neutral contexts, C_1 or C_2 . In the C_1 context, we set the house light to a mixture of 30% red and 70% blue. In C_2 , house lights were set to 70% red and 30% blue. Based on the literature, renewal strength in the ABC paradigm was expected to be diminished compared to ABA renewal. We also expected that renewal would become weaker when context C was more dissimilar to context A, i.e., C_1 should yield a lower renewal strength than C_2 . Our model showed both expected effects (Fig. 6d), confirming that indeed ABA renewal in our model is driven by the context change.

The role of experience replay in learning context-dependent behavior. We next explored the role of experience replay in the emergence of context-dependent behavior in our model agent. To this end, we repeated the above simulation with a memory-impaired agent. The memory-impaired agent learned in much the same way as the memory-intact agent did during acquisition (Fig. 6b), but exhibited markedly different behavior thereafter. First, extinction learning was faster in the memory-impaired agent, consistent with experimental observations. This result might seem unexpected at first glance, but it can be readily explained. With a low memory capacity, experiences from the preceding acquisition phase were not available during extinction and the learning of new reward contingencies during extinction led to catastrophic forgetting in the DQN²⁹. As a result, the previously acquired behavior was suppressed faster during extinction. Second, no renewal occurred in the test phase, i.e., the CRC slope remained about zero at the onset of the test phase. This is the result of catastrophic forgetting during extinction. Our results suggest that ABA renewal critically depends on storing experiences from the acquisition phase in experience memory and making them available during extinction learning. This result is consistent with experimental observations that hippocampal animals do not exhibit renewal and suggests that the role of the hippocampus in memory replay might account for its role in renewal.

The agent's learning behavior in our model crucially depended on two hyperparameters, namely the capacity of the experience replay memory c , and the exploration parameter ϵ . To analyze the impact of these hyperparameters, we performed and averaged 20 runs in the shock zone maze for different parameter combinations of c and ϵ . Let $S^-(c, \epsilon)$ and $S^+(c, \epsilon)$ be the slopes of the mean CRCs in the last 50 steps of extinction and first 50 steps of the test phase, respectively. In the ideal case of extinction followed by renewal: $S^-(c, \epsilon) = 0.0$ and $S^+(c, \epsilon) = 1.0$. We therefore defined the renewal strength for each combination of hyperparameters as

$$F^R(c, \epsilon) = S^+(c, \epsilon) - S^-(c, \epsilon), \quad (2)$$

which is near 1 for ideal renewal and near 0 for no renewal. As expected from our results on memory-impaired agents above, the renewal strength was near zero for low memory capacity (Fig. 6c). It increased smoothly with increasing memory capacity c , so that a memory capacity $c > 4000$ allowed the agents to exhibit renewal reliably. The renewal strength displayed a stable ridge for $c > 4000$ and $0.25 < \epsilon < 0.35$. This not only justified setting $\epsilon = 0.3$ in our other simulations, but also showed that our model is robust against small deviations of ϵ from that value.

The renewal strength was very low for small values of ϵ regardless of memory capacity. This points to an important role of exploration in renewal. With a low rate of exploration in the extinction phase, the model agent persevered and followed the path learned during acquisition. This led to $S^-(c, \epsilon) \approx S^+(c, \epsilon)$, and hence $F^R(c, \epsilon) \approx 0$, and explained why the exploration parameter has to take relatively large values of $0.25 < \epsilon < 0.35$ to generate robust renewal.

Emergence of distinct neural representations for contexts. Having shown that our model learned context-dependent behavior in complex tasks, we turned our attention to the mechanism underlying this behavior in a specific experiment. We simulated an experiment in appetitive, operant extinction learning that found that increasing the dopamine receptor agonists in the hippocampus had a profound effect on ABA renewal in a simple T-maze⁹ (Fig. 7a). The model agent received rewards for reaching the end of either side arm, but the reward was larger in the left (target) arm. Like in the experimental study⁹, we quantified the performance of the simulated agents by the number of trials, out of 10, that the agents chose the target arm at different points of the simulated experiment: {*early acq.*, *late acq.*, *early ext.*, *late ext.*, *early test*, *late test*}. In the acquisition phase in context A, memory-intact agents initially chose randomly between the left and right arm, but by the end of acquisition they preferentially chose the target arm (Fig. 7b, blue filled bars). In the extinction phase in context B, agents initially continued to prefer the target arm, i.e. they generalized the association formed in context A to context B, but this preference attenuated by the end of extinction. Critically, when the context was then switched back to the acquisition context A in the test phase, agents preferred the target arm again, even though it was no longer associated with a larger reward. This is the hallmark of ABA renewal. The renewal score $S^R = (\text{earlytest}/\text{lateext.}) - 1.0$ indicated the relative strength of the renewal effect in the T-maze experiments. $S^R \approx 0.59$ for the memory-intact agent. Since there was no special reward for preferentially choosing the target arm, the agent stopped doing so by the end of the test phase.

The memory-impaired agent showed the same acquisition learning as the memory-intact agent, but exhibited faster extinction learning and no renewal ($S^R \approx -0.14$, Fig. 7b, unfilled bars). These results are consistent with our modeling results in the inhibitory avoidance task discussed above and the experimental results from rats⁹.

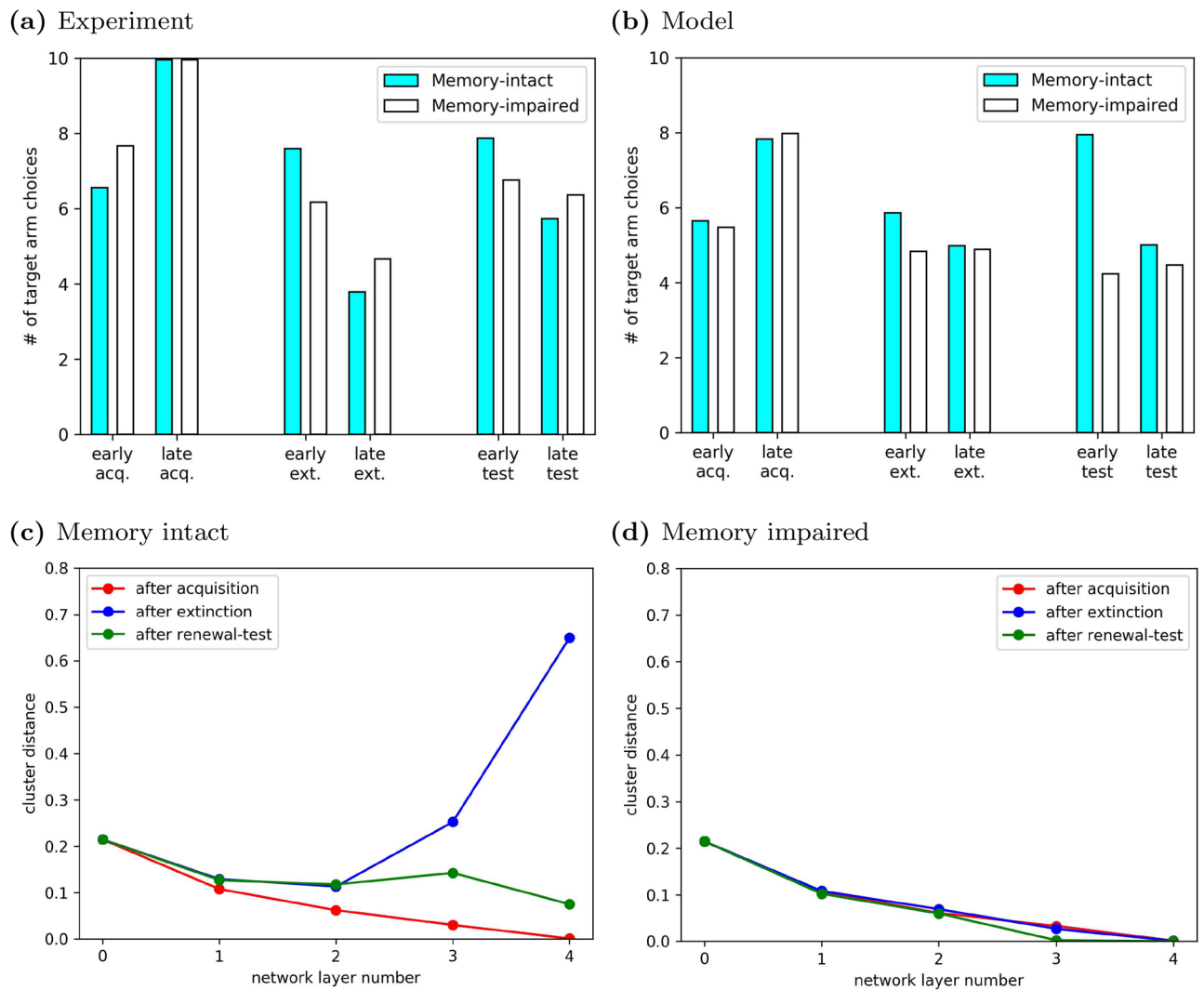


Figure 7. ABA renewal arises because experience replay drives the emergence of two distinct context representations. **(a)** André and Manahan-Vaughan showed that renewal was reduced by a pharmacological manipulation that interfered with synaptic plasticity in the hippocampus. (Adapted from Fig. 1 in Ref.⁹.) Renewal score in experimental animals: $S^R \approx 0.45$, renewal score in control animals: $S^R \approx 1.08$. **(b)** In our model, agents with intact experience replay memory showed ABA renewal in the T-Maze task, i.e. an increase in response rate in early test despite the continued absence of reward ($S^R \approx 0.59$). By contrast, memory-impaired agents did not exhibit renewal and had a lower renewal score ($S^R \approx -0.14$). **(c)** Cluster distances between network representations of context A and B in the T-maze for memory-intact agents. Results were averaged over 1000 runs of our model. The representation of contexts A and B in higher layers of the deep neural network ($l \geq 3$) formed distinct clusters that were well separated. They were learned in the extinction phase, since the distinction was not present at the end of the acquisition phase. **(d)** By contrast, memory-impaired agents did not show distinct context representations in any learning phase.

We hypothesized that context-dependent ABA renewal emerged in memory-intact agents because the higher layers of the DQN developed distinct representations of contexts A and B, which would allow the DQN to associate different reward contingencies with similar inputs. We therefore studied the evolution of context representations in the DQN based on the cluster distances between the representations of the two contexts (see “Methods”). It is important to keep in mind that, in the DQN, different layers have different functions: Lower layers are more closely related to sensory processing, while higher layers (closer to the output layer) fulfill evaluative tasks. Hence, lower layers reflect the properties of the visual inputs to the network. Since images from the context A form a cluster and are different from images from context B (contexts A and B were distinguished by different illumination colors), image representations in lower layers form distinct clusters. Hence, the cluster distance in the input layer, d_0 , was nonzero and remained so through the simulation (Fig. 7c,d). At the end of acquisition (red curve), the agents had been exposed only to visual inputs from context A, so there was no benefit for the DQN to represent the contexts A and B differently. Therefore, the difference in the context representations in the input layer was gradually washed out in higher layers, i.e., $d_1, \dots, d_4 < d_0$. As a result, distinct image representations are mapped onto the same evaluation. In other words, the network generalizes the evaluation it has

learned in context A to context B, just as subjects do. However, after extinction (blue curve), the replay memory contained observations from both contexts A and B, which were associated with different reward contingencies. This required that the higher layers of the DQN discriminate between the contexts, and $d_3, d_4 > d_0$. The network's ability to represent the two contexts differently, and thus maintain different associations during the extinction phase, gave rise to renewal in the subsequent testing phase (green curve), where the experience replay memory encoded novel observations from context A, yet still contained residual experiences from context B. Reward contingencies were identical in both contexts, and there was no longer a benefit to maintaining distinct context representations, hence, $d_3, d_4 \lesssim d_0$. Note that d_3 and d_4 were also much smaller at the end of the test phase than they were at the end of extinction, and would eventually vanish if the renewal test phase was extended.

As expected, cluster distances for the DQN in memory-impaired agents (Fig. 7d) did not change between the phases of the experiment and remained at low values throughout. This was the consequence of the reduced experience replay memory capacity and accounted for the lack of renewal in memory-impaired agents. Since the agents had no memory of the experiences during acquisition in context A, the DQN suffered from catastrophic forgetting²⁹ when exposed to a different reward contingency in context B and therefore the DQN could not develop separate representations of the two contexts during extinction.

Discussion

We have developed a computational model to study the emergence of context-dependent behavior in complex learning tasks within biologically realistic environments based on raw sensory inputs. Learning is modeled as deep Q-learning with experience replay¹⁵. Here, we apply our model to a sparsely researched topic: modeling of operant extinction learning³. We found that our model accounts for many features that were observed in experiments, in particular, in context-dependent ABA renewal, and the importance of hippocampal replay in this processes. Our model has several advantages relative to previous models. Compared to Redish et al.'s model of operant extinction learning¹⁷, our model learned faster on the linear track, and showed more plausible extinction learning in that our model generalized from the acquisition to the extinction context. Further, our model allowed for the investigation of internal representations of the inputs and, hence, the comparison to neural representations. Finally, our model makes two testable predictions: Context-dependent behavior arises due to the learning of clustered-representations of different contexts, and context representations are learned based on memory replay. We discuss these predictions in further detail below.

Learning context representations. One common suggestion in the study of learning is that stimuli can be distinguished a priori into discrete cues and contextual stimuli based on some inherent properties³⁰. For instance, cues are discrete and rapidly changing, whereas the context is diffuse and slowly changing. Prior models of extinction learning, e.g. Refs.^{16,17}, rely on external signals that isolate sensory stimuli and the current context. They maintain different associations for different contexts and switch between associations based on the explicit context signal. Since this explicit context switching is biologically unrealistic, Gershman et al. suggested a latent cause model in which contexts act as latent causes that predict different reward contingencies³¹. When expected rewards do not occur, or unexpected rewards are delivered, the model infers that the latent cause has changed. However, the model still depends on explicit stimulus signals, and requires complex Bayesian inference techniques to create and maintain the latent causes.

An alternative view is that there is no categorical difference between discrete cues and contextual information during associative learning, i.e., all sensory inputs drive learning to some extent¹⁴. Two lines of experimental evidence support this learning hypothesis. First, the learning of context can be modulated by attention and the informational value of contexts³². This finding suggests that contextual information is not fixed and what counts as contextual information depends on task demands³³. Second, the context is associated with the US to some degree³⁴, instead of merely gating the association between CS and US. There are suggestions that the association strength of the context is higher in the early stage of conditioning than later, when it becomes more apparent that the context does not predict the US³⁴. However, when context is predictive of CS-US pairing, the CR remains strongly context-dependent after multiple learning trials³⁵.

In line with the learning view, the internal deep Q-network in our model received only raw camera inputs and no explicit stimuli or context signaling. The model autonomously adapted the network weights to given learning tasks based on reward information collected during exploration and from the experience replay memory. By analyzing the model's internal deep Q-network, we found that distinct context representations emerged spontaneously in the activations of the network's higher layers over the course of learning, if the agent was exposed to two contexts with different reward contingencies. Distinct contextual representations emerge rather slowly in our study due to the role of the context in ABA renewal. If the task had been different, e.g. the agent had been exposed to US in context A and no US in context B in interleaved trials from the outset, distinct contextual representations in the higher layers of the network would have emerged faster. Nevertheless, they would have been learned in our model and not assumed to exist a priori as in almost all previous models.

Finally, we emphasize that, in addition to this task-dependent contextual representation in higher layers, the network's lower layers exhibit task-independent differences in neural activity in different contexts. We do not call the latter "contextual representations" since they are mere reflections of the sensory differences between inputs from different contexts.

The neural representation of context. The hippocampus has been suggested to play a role in the encoding, memorization, and signaling of contexts, especially in extinction learning^{3,11}. However, there is no conclusive evidence to rule out that the hippocampus is involved in the renewal phenomenon primarily due to its other well-established role in general-purpose memory storage and retrieval^{36–38}. In our model, the emergence

of distinct contextual representations critically depended on experience replay, suggesting that experience replay facilitates the formation of stable memory representations—as predicted by neuroscientific studies^{23,39}.

In line with previous suggestions^{6,10}, we therefore propose that contexts might be represented, in brain regions outside the hippocampus and that learning of these representations is facilitated by memory of experiences (inputs, actions and rewards), which are stored with the help of the hippocampus. However, we note that our model does not preclude that the hippocampus generates contextual representations of its own and that these representations drive behavior. Our model only indicates that hippocampal replay of low-level inputs suffices to generate context representations in other regions, such as the PFC. Unlike previous computational models of extinction learning, our model delineates the simulated fast hippocampal replay (experience replay memory) from other cortical regions (deep Q-network) that learn more slowly. As the network learns abstract representations of the visual inputs, the lower layers most likely correspond to visual areas in the cortex, while the higher layers serve functions similar to those served by the prefrontal cortex (PFC). Contextual representations in higher layers of the network emerge slowly in our study because the relevance of the contexts in an ABA renewal paradigm emerges slowly. If the task had been different, e.g., the agent had been exposed to US in context A and no US in context B in interleaved trials from the beginning, like in contextual fear conditioning, the distinct contextual representations in the higher layers of the network would have emerged faster. Nevertheless, they would have been learned in our model, and not assumed to exist a priori as in almost all previous models.

We propose that our hypothesis above could be tested using high-resolution calcium imaging⁴⁰ of the PFC in behaving rodents. Our model predicts that different contexts are encoded by PFC population activity that form well-separated clusters by the end of extinction learning, allowing the animals to disambiguate contexts, and leading to renewal. In hippocampal animals, however, we expect that novel contexts are not encoded in separate PFC clusters, but instead their representations in the PFC population will overlap with those of previously acquired contexts. Furthermore, our model suggests that context representations in the brain are highly dynamic during renewal experiments, adapting to the requirements of each experimental phase. Distinct context representations would therefore arise in control animals over the course of extinction learning, and not necessarily be present from the start, cf.⁴¹.

The role of the hippocampus in renewal. Our hypotheses about contextual representations discussed above naturally beg the question: What role does hippocampus play in renewal? The experimental evidence leaves no doubt that the hippocampus plays a critical role in renewal^{3,7,11,42}. In our model, we focused on the role of experience replay observed in the hippocampus, i.e. we propose that hippocampus is involved in renewal because of its ability to store and replay basic experiences. In our model, replay of experiences from the acquisition phase was critical to learn distinct context representations in the extinction phase. This is consistent with observations in humans⁴³. Only subjects whose activation in the bilateral hippocampus was higher during extinction in a novel context than in the acquisition context showed renewal later. With limited experience memory, learning during extinction in our model overwrites previously acquired representations, also known as catastrophic forgetting. This in turn prevented renewal from occurring when the agent was returned to the acquisition context. If lesioning the hippocampus in animals has a similar effect as limiting the experience memory in our simulation, as we propose, our model predicts that impairing the hippocampus pre-extinction weakens renewal, while impairing the hippocampus post-extinction should have no significant influence on the expression of renewal.

Both of these predictions are supported by some experimental studies, and challenged by others. Regarding the first prediction, some studies have confirmed that hippocampal manipulations pre-extinction have a negative impact on ABA renewal. For instance, ABA renewal was disrupted by pharmacological activation of dopamine receptors prior to extinction learning⁹. By contrast, other studies suggest that disabling the hippocampus pre-extinction has no impact on later renewal. For instance, inactivation of the dorsal hippocampus with muscimol, an agonist of GABA_A receptors⁴⁴, or scopolamine, a muscarinic acetylcholine receptor antagonist⁸, prior to extinction learning did not weaken ABA renewal during testing. The apparent contradiction could be resolved, if the experimental manipulations in Refs.^{8,44} did not interfere with memories encoded during acquisition, as other manipulations of the hippocampus would, but instead interfered with the encoding of extinction learning. As a result, the acquired behavior would persist during the subsequent renewal-test. Indeed, activation of GABA_A receptors, or the inactivation of muscarinic receptors reduce excitation of principle cells, thereby changing signal-to noise ratios in the hippocampus. We hypothesize that, as a result, animals exhibited reduced extinction learning following scopolamine⁸ or muscimol⁴⁴, consistent with impaired encoding of extinction. By contrast, dopamine receptor activation during extinction learning⁹, did not reduce extinction learning, but impaired renewal. The results of the latter study are more consistent with the results of our model when memory capacity is reduced pre-extinction: extinction learning was slightly faster and renewal was absent.

In support of our second prediction, a study showed that antagonism of β -adrenergic receptors, which are essential for long-term hippocampus-dependent memory and synaptic plasticity⁴⁵, post-extinction had no effect on ABA renewal⁴⁶. By contrast, several studies suggested that lesioning the hippocampus post-extinction reduces renewal of context-dependent aversive experience. For instance, ABA renewal in the test phase was abolished by excitotoxic⁴² and electrolytic⁷ lesions of the dorsal hippocampus after the extinction phase. More targeted lesions of dorsal CA1, but not CA3, post-extinction impaired AAB renewal⁶. The authors of these studies generally interpreted their results to indicate that the hippocampus, or more specifically CA1, are involved in retrieving the contextual information necessary to account for renewal. However, damage to these subregions can be expected to have generalized effects on hippocampal information processing of any kind. Another possibility is that post-extinction manipulations, which were made within a day of extinction learning, interfered with systems consolidation. This is very plausible in terms of dopamine receptor activation⁹ as this can be expected to promote

the consolidation of the extinction learning event. To model this account in our framework, one would have to split the learning and replay during extinction phase, which occur in parallel in our current implementation, into two separate stages. Only new learning would occur during the extinction phase, while experience replay (of the acquisition phase) would occur during sleep or wakeful quiescence after the extinction phase, as suggested by the two-stage model of systems consolidation. In this case, the associations formed during acquisition would be initially overwritten during extinction learning, as they were in our simulations without experience memory, and then relearned during the replay/consolidation period^{23,24}. If systems consolidation was altered during this period, extinction learning would still be retained, but the association learned during acquisition would be overwritten and hence there would be no renewal. In line with this interpretation, recent findings suggest that extinction learning and renewal engage similar hippocampal subregions and neuronal populations¹³, thus, suggesting that manipulating the hippocampus can have an impact on prior learned experience and thus, subsequent renewal.

In conclusion, we have shown that, in an ABA renewal paradigm, contextual representations can emerge in a neural network spontaneously driven only by raw visual inputs and reward contingencies. There is no need for external signals that inform the agent about cues and contexts. Since contextual representations might be learned, they might also be much more dynamic than previously thought. Memory is critical in facilitating the emergence of distinct context representations during learning, even if the memory itself does not code for context representation per se.

Received: 6 May 2020; Accepted: 8 December 2020

Published online: 01 February 2021

References

- Pavlov, I. P. & Anrep, G. V. *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex* (Oxford University Press, Humphrey, 1927).
- Auchter, A. M., Shumake, J., Gonzalez-Lima, F. & Monfils, M. H. Preventing the return of fear using reconsolidation updating and methylene blue is differentially dependent on extinction learning. *Sci. Rep.* **7**, 46071 (2017).
- Dunsmoor, J. E., Niv, Y., Daw, N. & Phelps, E. A. Rethinking extinction. *Neuron* **88**, 47–63 (2015).
- Bouton, M. E. Context and behavioral processes in extinction. *Learn. Mem.* **11**, 485–494 (2004).
- Corcoran, K. A. & Maren, S. Factors regulating the effects of hippocampal inactivation on renewal of conditional fear after extinction. *Learn. Mem. (Cold Spring Harbor, N.Y.)* **11**, 598–603 (2004).
- Ji, J. & Maren, S. Differential roles for hippocampal areas CA1 and CA3 in the contextual encoding and retrieval of extinguished fear. *Learn. Mem.* **15**, 244–251 (2008).
- Fujiwara, H. *et al.* Context and the renewal of conditioned taste aversion: The role of rat dorsal hippocampus examined by electrolytic lesion. *Cogn. Neurodyn.* **6**, 399–407 (2012).
- Zelikowsky, M. *et al.* Cholinergic blockade frees fear extinction from its contextual dependency. *Biol. Psychiatry* **73**, 345–352 (2013).
- André, M. A. E. & Manahan-Vaughan, D. Involvement of dopamine D1/D5 and D2 receptors in context-dependent extinction learning and Memory Reinstatement. *Front. Behav. Neurosci.* **9**, 372 (2016).
- Kumaran, D., Hassabis, D. & McClelland, J. L. What Learning systems do 650 intelligent agents need? complementary learning systems theory updated. *Trends Cogn. Sci.* **20**, 512–534. issn: 1364–6613 (2016).
- Maren, S., Phan, K. L. & Liberzon, I. The contextual brain: Implications for fear conditioning, extinction and psychopathology. *Nat. Rev. Neurosci.* **14**, 417–428 (2013).
- Lemon, N. & Manahan-Vaughan, D. Dopamine D1/D5 receptors gate the acquisition of novel information through hippocampal long-term potentiation and long-term depression. *J. Neurosci.* **26**, 7723–7729 (2006).
- Méndez-Couz, M., Becker, J. M. & Manahan-Vaughan, D. Functional compartmentalization of the contribution of hippocampal subfields to context-dependent extinction learning. *Front. Behav. Neurosci.* **13**, 256 (2019).
- Rescorla, R. A. & Wagner, A. R. In *Classical Conditioning II: Current Research and Theory* (eds Black, A. H. & Prokasy, W. F.) 64–99 (Appleton-Century-Crofts, New York, 1972).
- Mnih, V. *et al.* Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
- Ludvig, E. A., Mirian, M. S., Kehoe, E. J. & Sutton, R. S. Associative Learning from Replayed Experience. *bioRxiv* (2017).
- Redish, A. D., Jensen, S., Johnson, A. & Kurth-Nelson, Z. Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling. *Psychol. Rev.* **114**, 784–805 (2007).
- Brockman, G. *et al.* OpenAI Gym. <http://arxiv.org/abs/1606.01540> (2016).
- Sutton, R. S. & Barto, A. G. *Reinforcement Learning—An Introduction* (MIT Press, Cambridge, 1998).
- Rumelhart, D., Hinton, G. & Williams, R. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
- Chollet, F. *et al.* Keras. <https://keras.io> (2015).
- Plappert, M. keras-rl. *GitHub Repo*. <https://github.com/keras-rl/keras-rl> (2016).
- O'Neill, J., Pleydell-Bouverie, B., Dupret, D. & Csicsvari, J. Play it again: Reactivation of waking experience and memory. *Trends Neurosci.* **33**, 220–229 (2010).
- Buhry, L., Azizi, A. H. & Cheng, S. Reactivation, replay, and preplay: How it might all fit together. *Neural Plast.* **2011**, 1–11 (2011).
- Karlsson, M. P. & Frank, L. M. Awake replay of remote experiences in the hippocampus. *Nat. Neurosci.* **12**, 913–918 (2009).
- Morris, R. Developments of a water-maze procedure for studying spatial learning in the rat. *J. Neurosci. Methods* **11**, 47–60 (1984).
- Anisman, H. & McIntyre, D. C. Conceptual, spatial, and cue learning in the morris water maze in fast or slow kindling rats: Attention deficit comorbidity. *J. Neurosci.* **22**, 7809–7817 (2002).
- Hernandez, V. *et al.* Dopamine receptor dysregulation in hippocampus of aged rats underlies chronic pulsatile L-Dopa treatment induced cognitive and emotional alterations. *Neuropharmacology* **82**, (2013).
- Kirkpatrick, J. *et al.* Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci.* **114**, 3521–3526 (2017).
- Nadel, L. & Willner, J. Context and conditioning: A place for space. *Physiol. Psychol.* **8**, 218–228 (1980).
- Gershman, J. S., Norman, A. K. & Niv, Y. Discovering latent causes in reinforcement learning. *Curr. Opin. Behav. Sci.* **5**, (2015).
- Lucke, S., Lachnit, H., Koenig, S. & Uengoer, M. The informational value of contexts affects context-dependent learning. *Learn. Behav.* **41**, 285–297 (2013).
- Eschenko, O. & Mizumori, S. J. Y. Memory influences on hippocampal and striatal neural codes: Effects of a shift between task rules. *Neurobiol. Learn. Mem.* **87**, 495–509 (2016).
- Hall, G. & Honey, R. C. Context-specific conditioning in the conditioned- emotional-response procedure. *J. Exp. Psychol. Anim. Behav. Process.* **16**, 271–278 (1990).

35. Swartzentruber, D. Blocking between occasion setters and contextual stimuli. *J. Exp. Psychol. Anim. Behav. Process.* **17**, 163–173 (1991).
36. Eichenbaum, H. Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron* **44**, 109–120 (2004).
37. Cheng, S. The CRISP theory of hippocampal function in episodic memory. *Front. Neural Circuits* **7**, 88 (2013).
38. Bayati, M. *et al.* Storage fidelity for sequence memory in the hippocampal circuit. en. *PLoS One* **13** (ed Wennekens, T.), e0204685 (2018).
39. Ji, J. & Maren, S. Electrolytic lesions of the dorsal hippocampus disrupt renewal of conditional fear after extinction. *Learn. Mem.* **12**(3), 270–6 (2005).
40. Kim, T. *et al.* Long-term optical access to an estimated one million neurons in the live mouse cortex. *Cell Rep.* **17**, 3385–3394 (2016).
41. Czerniawski, J. & Guzowski, J. F. Acute neuroinflammation impairs context discrimination memory and disrupts pattern separation processes in hippocampus. *J. Neurosci.* **34**, 12470–12480 (2014).
42. Zelikowsky, M., Pham, D. L. & Fanselow, M. S. Temporal factors control hippocampal contributions to fear renewal after extinction. *Hippocampus* **22**, 1096–1106 (2012).
43. Lissek, S., Glaubitz, B., Güntürkün, O. & Tegenthoff, M. Noradrenergic stimulation modulates activation of extinction-related brain regions and enhances contextual extinction learning without affecting renewal. *Front. Behav. Neurosci.* **9**, 34 (2015).
44. Corcoran, K. A., Desmond, T. J., Frey, K. A. & Maren, S. Hippocampal inactivation disrupts the acquisition and contextual encoding of fear extinction. *J. Neurosci.* **25**, 8978–8987 (2005).
45. Hagen, H., Hansen, N. & Manahan-Vaughan, D. β -Adrenergic control of hippocampal function: Subservicing the choreography of synaptic information storage and memory. *Cereb. Cortex* **26**, 1349–64 (2016).
46. André, M., Wolf, O. & Manahan-Vaughan, D. Beta-adrenergic receptors support attention to extinction learning that occurs in the absence, but not the presence, of a context change. *Front. Behav. Neurosci.* **9**, (2015).

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project number 316803389—through SFB 1280, projects A04 (D.M.V.), A14 (S.C.) and F01 (S.C.).

Author contributions

Designed research: T.W., L.W., S.C.; Performed research: T.W., N.D.; Contributed new reagents or analytic tools: T.W., N.C., S.V., J.D.; Analyzed data: T.W., N.D., S.C.; Interpreted data: T.W., D.M.V., L.W., S.C.; Wrote the manuscript: T.W., S.C.; Reviewed the manuscript: all authors.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests


The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021