# A Cognitive Model of Hippocampal Replay as Hypothesis Testing for Efficient Generalization

Mishaal Kandapath

Department of Psychology

University of Toronto

COG403

Dr. Can Mekik

May 3, 2025

# A Cognitive Model of Hippocampal Replay as Hypothesis Testing for Efficient Generalization

Humans learn so much from so little. Explosive progress in deep learning has presented human-level, and sometimes beyond human, performance on various tasks. Nevertheless, an ability for efficient and generalizable abstractions curiously evades modern deep learning systems (Cherian et al., 2023; Lake et al., 2017; Moskvichev et al., 2023). Our capacity for efficient and generalizable abstractions is coveted. Many have proposed that such an ability is rooted in the formation and utilization of rich hierarchical internal models of the world (Behrens et al., 2018; Lake et al., 2015). Such a view requires the resolution of at least two problems: how do we form accurate and hierarchical models from experience, and how might we use them to make sense of the present and imagine the future? Such a question fits naturally into the Model-Based Reinforcement Learning (MBRL) paradigm. MBRL, in addition to training a value function (or policy, or both) as in regular RL, trains an internal model of the world for offline planning and training. In the brain, MBRL engages the Hippocampus (HPC) and the Prefrontal Cortex (PFC) (K. J. Miller et al., 2017; Vikbladh et al., 2019). Of special interest is the phenomenon of replay - sequences of neural activity encoding external states - posited to play a role in memory (Scoville and Milner, 1957) model-based planning (K. J. Miller et al., 2017), and much more.

For planning, hippocampal replay at rest is commonly conceived as sampling from a generative internal model for offline training of a value or policy function (Mattar and Daw, 2018). Thus, inference over existing abstractions is a pre-requisite. However, the space of learned abstractions is extremely large (Schwartenbeck et al., 2023). Interestingly, generative replay has also been implicated in the online inference of suitable abstractions for novel tasks. In Schwartenbeck et al. (2023), participants are asked to infer the relational structure of a certain arrangement of blocks. Replay sequences reflect samples from internal relational models of blocks, iterating through possible transitions and bindings of abstract relations and perceptual entities. Importantly, these sequences reflect possibilities and not necessarily just the current or intended state of the world. Thus, Schwartenbeck et al. (2023) construed generative replay as

hypothesis testing for suitable "brick" models - a search through abstractions. More broadly, Kurth-Nelson et al. (2023) characterizes hippocampal replay as facilitating compositional computation, drawing new inferences from existing abstractions and their relations for use in novel tasks, in turn informing underlying sub-symbolic policy and value functions for the task at hand. However, a formalization of such a proposal has not been realized (Kurth-Nelson et al., 2023). It remains to be seen whether such a model of the HPC-PFC circuit can fully explain empirical replay data and human performance on complex tasks requiring efficient generalization.

Abstractions, as opposed to implicit information, low-level Q-networks, and more, are best described using symbolic/localist descriptions rather than sub-symbolic/connectionist descriptions. Thus, any model of replay must address interactions between symbolic and sub-symbolic representations. Additionally, MBRL is an action-oriented paradigm. On the other hand, internal models of the world represent information stored in memory, i.e., used in both action and non-action-oriented ways. As a result, a model of replay benefits from a theory of how our experiences in memory may be brought to inform our perception and action. The CLARION cognitive architecture (Sun, 2006, 2016) naturally captures interactions between symbolic and sub-symbolic representation and between action and memory systems. It is a perfect candidate for modeling replay as posed in Kurth-Nelson et al. (2023). In this work, we formalize the hypothesis testing in the HPC-PFC circuit from arguments and evidence from Kurth-Nelson et al. (2023) and Schwartenbeck et al. (2023) using the CLARION cognitive architecture. We hypothesize that our formalization of replay as hypothesis testing within CLARION will not only capture accurate and efficient human performance displayed on the block inference task in Schwartenbeck et al. (2023) but also the replay characteristics aiding such inference.

## Model

CLARION consists of four subsystems, of which only three directly relate to our task: the Action-Centered Subsystem (ACS), the Non-Action Centered Subsystem (NACS), and the Metacognitive Subsystem (MCS). Each subsystem has two levels: top (explicit) and bottom (implicit). The bottom level consists of distributed subsymbolic representations. The top level

consists of localist chunk representations. In the ACS and the NACS, nodes at the top and bottom levels are connected such that top-level activations of a particular concept can activate bottom-level microfeatures that make up the concept and vice versa. For example, a brick, *B*, in Schwartenbeck et al. (2023) may be of four different textures and two shapes. *B* may be represented as a localist node connected to a distributed bottom-level pattern of activations for shapes and textures that make up *B*. Similarly, abstractions implicated in hypothesis testing and generalization through replay can be represented within the top level, whereas more implicit mappings, like Q-value function approximators, may be represented as mappings between distributed representations at the bottom level.

Secondly, there is the distinction between the NACS and ACS. This distinction is important as abstractions can be thought of as residing in the NACS. Their transfer to the ACS enables decision-making in complex tasks like those in Schwartenbeck et al. (2023). However, given the complexity of modeling interacting subsystems, we focus on an implementation within the ACS.

Finally, planning has long been understood to involve executive control, especially the active maintenance of current goals (E. K. Miller & Cohen, 2001). The goal module in the MCS and associated goal structures in the ACS facilitate such processing (Sun, 2016).

**Hippocampal Replay in CLARION**

We model hippocampal replay within a block-based construction task. In this task, a particular arrangement of bricks is presented, and the agent has to recreate such an arrangement. The target arrangement (hereby referred to as *target*) of blocks and the agent's progress (*current* arrangement) will be represented using distributed representations in the implicit layer of the ACS. Note that an *italicized* entry denotes representation at the bottom level.

Specifically, we explore a mapping between replay in the brain and AI suggested in Kurth-Nelson et al. (2023), where neural networks prune large abstract search spaces, new knowledge is discovered by search, and a positive feedback loop between search and networks improves both networks and search. This description is largely motivated by its descriptive fit to

the empirical evidence discussed in Kurth-Nelson et al. (2023) and the success of modern RL systems like Schrittwieser et al. (2021). A proper formulation must consider the overarching cognitive processes at play and how these components play a role in these processes.

The goal of the neural network is to determine which moves are worth exploring in a search through block configurations, setting the scene for efficient search in explosive spaces. The PFC has been shown to be involved in executive control, especially the active maintenance of goals vital to planning (Mattar & Lengyel, 2022). Kurth-Nelson et al. (2023) suggests that the PFC plays a role similar to such search-guiding neural networks: the PFC sets the utility of various moves and biases the hippocampus toward the production of sequences adhering to such relations. As a result, we situate our neural network guiding search within the goal module of the MCS to abide by the executive functions of the PFC.

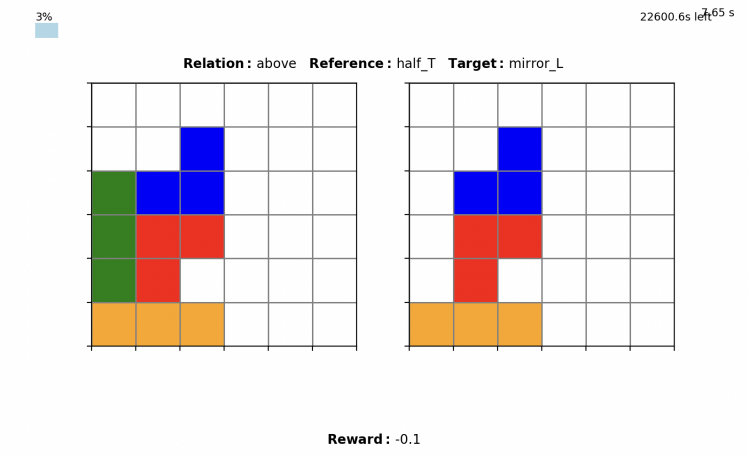


**Figure 1**

*A sample trial with the model constructing its constructions on the right, and the target on the left. The model had previously asked to sample transitions adhering to mirror-L being above half-T. As a result, the model just placed mirror-L above half-T*

The neural network determines two blocks and a relation to explore in achieving the construction in *target* based on the current state in *current*. For example, in Figure 1, we see that the neural network has just guided search to look at configurations of the mirror-L-like blue shape with respect to the half-T-like red shape, where mirror-L comes out on top. Such internal goals

are fed into the ACS as facilitated by goal structures in the ACS (Sun, 2016).

Within the ACS, we have "forward" rules. "Forward" rules specify abstractions. Rules in the top level of the ACS have a condition and an action chunk. A "forward" rule has as its condition some constraints on the *target* and *current* arrangement. These arrangements pertain to three types of information: specific pixel-level configurations of one or two blocks in *target*, specific pixel-level configurations of one or two blocks in *current*, and goal configuration specified by the goal network. Its action node specifies an action following the goal *relation* between two *blocks*: the block being added (in Figure 1 this would be the mirror-L shape), the already present block in relation to which it is being added (in Figure 1 this would be the half-T shape), and the relation between the two blocks (in Figure 1 this would be pixel values adhering to the relation "above"). As is common in CLARION, chosen rules are those that are sampled from the Boltzmann distribution induced by rule activations (Sun, 2016). These activations are influenced by the activations of the condition chunks, positive, and negative match counts (Sun, 2016). To avoid the selection of completely irrelevant rules, match statistics are implemented to possess a value of up to 20% of the maximum possible chunk activation.

When chosen, an action eventually applies onto *current*, replacing the previous *current*. Every rule adds atmost one block to *current*. To reflect the fact that the selection of abstractions is influenced primarily by goal relations, the chunks pertaining to goal structures have much higher weighting than other input structures feeding into "forward" rules. In this way, selecting goals, goals influencing actions, actions influencing the next goal chosen, and so on, we closely adhere to the description in Kurth-Nelson et al. (2023) while remaining close to our understanding of cognitive processes involved in planning. This process terminates when the agent's *target* and *current* arrangements match or a given number of steps, *S*, have elapsed. As the search progresses, rules fire one after another, each adding one block with reference to another according to a particular relation. The rules fired are essentially samples from an internal model of bricks (the set of rules possessed by the model). Hence, the sequence of configurations thus created forms our replay sequence. If indeed the process was a success, we iteratively backtrack to an

empty *current* arrangement, incrementing positive match counts for successful rules. Such a procedure naturally describes reversed replay sequences observed after the presentation of a reward in humans (Liu et al., 2019) and mice (Ambrose et al., 2016).

## Methods

The described model will be evaluated on the simpler non-hierarchical version of the block task described in Schwartenbeck et al. (2023). Each task begins with an initial target arrangement of blocks. In our model, an arrangement of blocks is provided as a bottom-level representation of the 9x9 grid corresponding to that arrangement (grid on the left in Figure 1 presents a sample 9x9 input grid). In a manner similiar to Schwartenbeck et al. (2023), the model is first presented with training trials and then test trials.

Similar to participants in Schwartenbeck et al. (2023), the training phase is for the model to familiarize itself with the block configuration task. Each trial in the training phase can have anywhere between two to four blocks. The model then attempts to construct a target arrangement in the manner described in the Model section. After construction, the model is probed on relations between any two blocks in the target configuration (see Figure 2 for an example). A test trial proceeds in exactly the same manner, with the exception that every trial has an input configuration consisting of exactly three shapes and that it is for a shorter duration (3.5 as opposed to 6 seconds in train trials). In both training and test trials, the model is allowed to use each block exactly once, as in Schwartenbeck et al. (2023). We used the training and test grids presented to participants in Schwartenbeck et al. (2023) to train and test our model.

Instead of explicitly specifying $S$, the number of steps allowed to be taken in a trial, we leverage the trial times specified in Schwartenbeck et al. (2023): 6 seconds for construction in the training trial, and 3.5 seconds for construction in the test phase. For simplicity, we only account for the time it takes for a rule to fire. We specify this as 50ms, following standard timings in many cognitive models such as ACT-R, EPIC, and more (Anderson et al., 1995), and evidence from models using standard neuron membrane and neurotransmitter properties (Stewart & Eliasmith, 2009).

**Construction Rules**

Since we are not attempting to model the formation of abstractions, we manually specify all possible configurations for one and two blocks. Additionally, since our goal network is randomly initialized at the start of the experiment and thus needs to be trained, it might produce incorrect goals. In turn, this may lead to the firing of extremely incorrect rules. To protect against such rules, we have a threshold of activations that "forward" rules must possess in order to be applied. Instead of manually specifying this rule, we have a dummy rule in the ACS that fires whenever the dummy rule possesses the most activation. If this rule fires, the network is made to choose another goal.

This is similar to other roles of the MCS, such as the monitoring of activations in the top and bottom levels of the ACS and deciding the manner in which to combine activations to produce one result.

Finally, due to compute and time constraints, we limit the number of times the model can backtrack to 11 times (chosen so as to finish 50 trials in 50 minutes).

**Goal Setting Network**

We specified our goal-setting network as a simple feedforward network consisting of three hidden layers of 128, 256, and 128, respectively. We trained this network using deep Q-learning as specified in Mnih et al. (2013) and as commonly used to train networks in CLARION (Sun, 2016). Mnih et al. (2013) involves a simple replay buffer consisting of transitions experienced by the agent. In our simulation, a transition consists of the current state, the action taken, the reward experienced, and the resultant next state. The failure of a "forward" rule firing receives a -1 reward, an incorrect construction receives a reward equal to the accuracy of the construction, and every step receives a reward of -0.1 to encourage short searches.

In Schwartenbeck et al. (2023), participants have a rest phase after every 50 training trials for a total of 6 training sessions. However, we only run our model for a total of 150 training trials. which roughly corresponds to a rest period every 20 trials when given 150 trials in total. Hippocampal replay and its relation to MBRL has led many to view replay *at rest* as sampling

from a buffer (or a generative model) to train value functions offline (e.g (Momennejad et al., 2018)). As a result, we train our network using its replay buffer after every 20 trials, for 50 epochs with a batch size of 64.

In addition to this training, a simple online training regime occurs during each trial when the model is provided with an explicit reward. This happens whenever a "forward" rule fails to reach a certain threshold of activation (-1 reward) and when a model takes a particular action (-0.1 reward). Such offline training only occurs for the singular transition that happened at that timestep. We incorporate this strategy in our model too.

Additionally, Schrittwieser et al. (2021) uses an epsilon-greedy policy to help their model learn while it has poor estimates in the beginning. This is similar to insights from theories of self-efficacy trading exploration and exploitations when under impressions of low and high self-efficacy respectively (Bandura, 1977).

**Rules for Responses**

Since both training and test trials involve a probe phase where our model produces yes or no responses, an additional and independent repertoire of rules is added in the ACS. These rules have condition chunks that specify the two blocks and relation provided to the participants, one of which is a reference block, the other is a target block, and the relation is of the target block with respect to the reference block. Additionally, these conditions contain pixel configurations that would correctly adhere to the queried blocks and relations. Again, if the rules meet a certain threshold, a "yes response is fired, otherwise "no". For example, in Figure 2, a no-response fires as the half-T shape is not to the right of the horizontal shape. It is worth noting that response rules are tight, i.e, given a construction is correct, the response will also be so. This is so that we can explain any findings as entirely due to the processes afforded by replay, mainly the interplay between goals and actions.

**Evaluation**

This gives us the required setup to run the model described in Model. We now test the model in the same test settings as participants in Schwartenbeck et al. (2023). In both phases, the

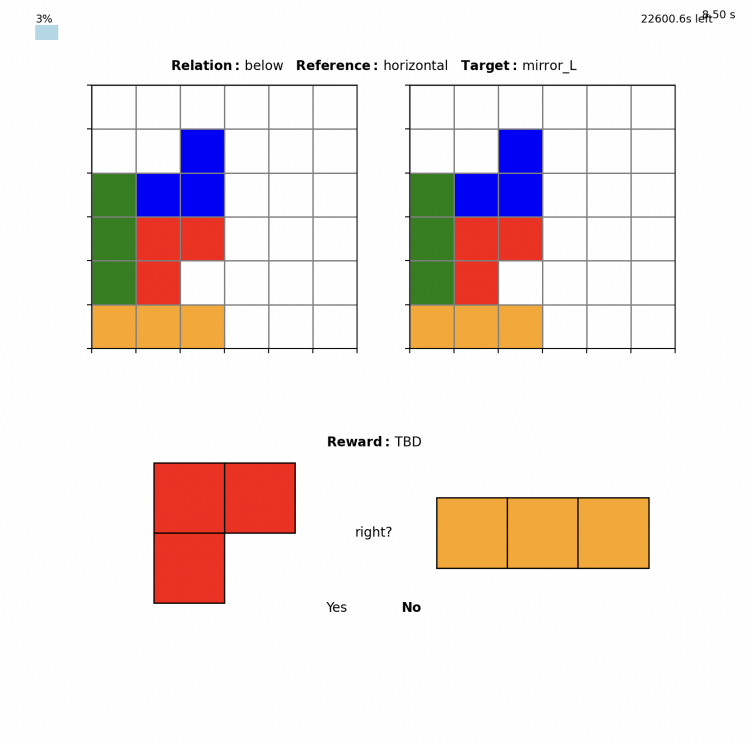


**Figure 2**

*A sample trial with the model having concluded its construction phase, with the model's construction on the right and the target construction on the left. The model is asked whether half-T-like red shape is to the right of the horizontal yellow shape. The model correctly answers "no".*

model will be presented with a *target* and asked to infer the construction of the *target*. Once the search has terminated, we present two blocks on the target grid as a query. A response is selected based on the construction in *current*. For simplicity, we chose to evaluate our model qualitatively, since that is all that is required to compare results against our hypothesis. Specifically, we are looking for a pattern consistent with reduced reaction times and increased correctness with trials. Additionally, we are also looking for a similar temporal evolution of our replay sequences as those presented in Schwartenbeck et al. (2023).

Correctness can be directly judged from the responses to the probe task. Since at each timestep a new goal is chosen (one different from or the same as the previous timestep), the

number of goals chosen gives us a proxy for reaction time. Finally, replay sequences can also be considered as transitions dictated by the goals. Although Kurth-Nelson et al. (2023) specifies replay as only being biased by such goals, for simplicity, our rules are constrained to adhere to the goals. As a result, it suffices to treat the sequence of goals as replay sequences. In Schwartenbeck et al. (2023), replay sequences are characterized as transitions between blocks. For example, in Figure 1, the goal with the reference half-T shape and the target mirror-L shape describes a sequence from half-T to mirror-L. Once the evolution of replay sequences has been obtained, we qualitatively evaluate it for trends similar to those observed in Schwartenbeck et al. (2023). For example, in Schwartenbeck et al. (2023), the first half of the replayed sequence, sequences with blocks in the target are more common than those with at least one block not in the target.

## Results


**Figure 3**

*Training curve of our Q-value network. The loss is the huber loss of the TD-error, as presented in Schrittwieser et al. (2021) so as to guard againt noisy signals*

Figure 3 presents a very noisy training process for the neural network during the training phase of the task. The Q-value approximation network displays no clear reduction in loss. This is

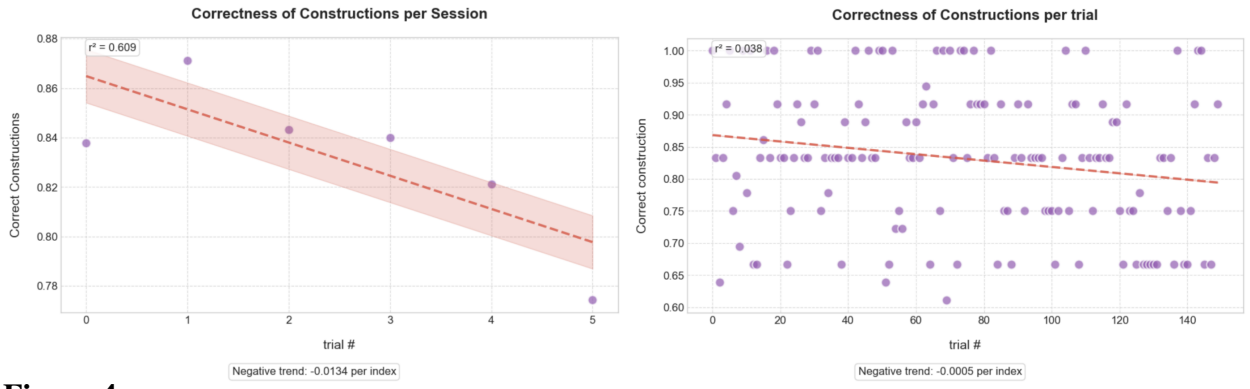


**Figure 4**

*Average accuracy of constructions per session and per trial. Note that a session here is different in duration from Schwartenbeck et al. (2023) due to our model running on fewer overall trials. As a result, we have the same number of sessions, but only 25 (as opposed to 50) trials per session. Note the negative trend in accuracy over trials and sessions.*

expected to some degree, as each block configuration is different from others, leading to large errors arising from experiencing a state for the first time. The periodic well-behaved slopes arise from training the network from the buffer. Overall, this suggests that the network struggles to learn the structure of the task, given that it only receives limited samples of particular brick configurations.

Given this, it is not surprising that the model does not exhibit similar response and correctness characteristics to humans in Schwartenbeck et al. (2023). As is seen in Figure 4, there is a reversal of the expected positive trend in construction correctness. This is because there are a limited number of actions in the task, and due to the presence of an epsilon-greedy policy (Schrittwieser et al., 2021), the model stumbles upon the right answers in the beginning more often than when the choices lean more toward its own largely incorrect scores later on during training. However, we do observe a negative effect of training in the maximum number of goals experienced through train trials (Figure 5, presenting a reduction in reaction times.

However, we see a great qualitative fit for the replay sequences in Figure 6 as compared to those discussed in Schwartenbeck et al. (2023). First we see that there is an early dominance of

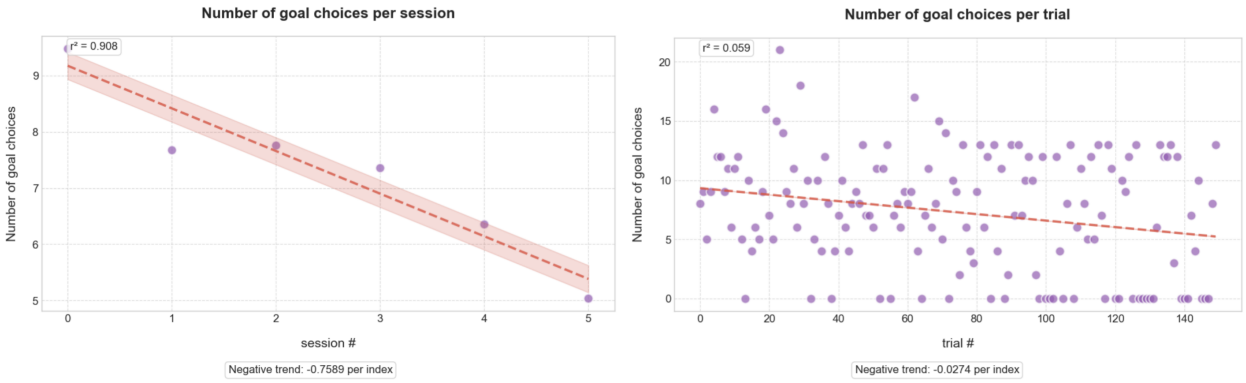


**Figure 5**

*The number of goal choices chosen by the model per session and per trial. Note that a session here is different in duration from Schwartenbeck et al. (2023) due to our model running on fewer overall trials. As a result, we have the same number of sessions, but only 25 (as opposed to 50) trials per session. Note the negative trend in goal choices, our proxy for response times.*

the "Stable to Present" (block placed first to block connected to that which was placed first) transitions (Steps 0-3), just as is observed in Schwartenbeck et al. (2023). Similarly, we see that there is a dominance of "Present to Present" (between blocks present in the shape that are not the first placed block) sequences right after (Steps 3-5), especially when compared against "Present to Absent" (blocks present in the shape that are not "Stable" to the block that is not in the shape) transitions, as is done and observed in Schwartenbeck et al. (2023). Finally, and importantly, we see a dominance of "Present to Stable" transitions towards the later end of the search (Steps 7-10). This last transition is particularly important, as, unlike the other transitions, which can be directly explained to arise by virtue of constructing the shape, it is unclear why the "Present to Stable" transition would arise, given that all the relevant blocks had been iterated through in the sequences that came before. These are the three sequences emphasized in Schwartenbeck et al. (2023), and we observe that we see similar effects in the replay sequences from our model. However, we note that unlike in Schwartenbeck et al. (2023), we do not see a high number of "Stable to Absent" sequences around the time there are a high number of "Present to Present" sequences (i.e, steps 3-5).

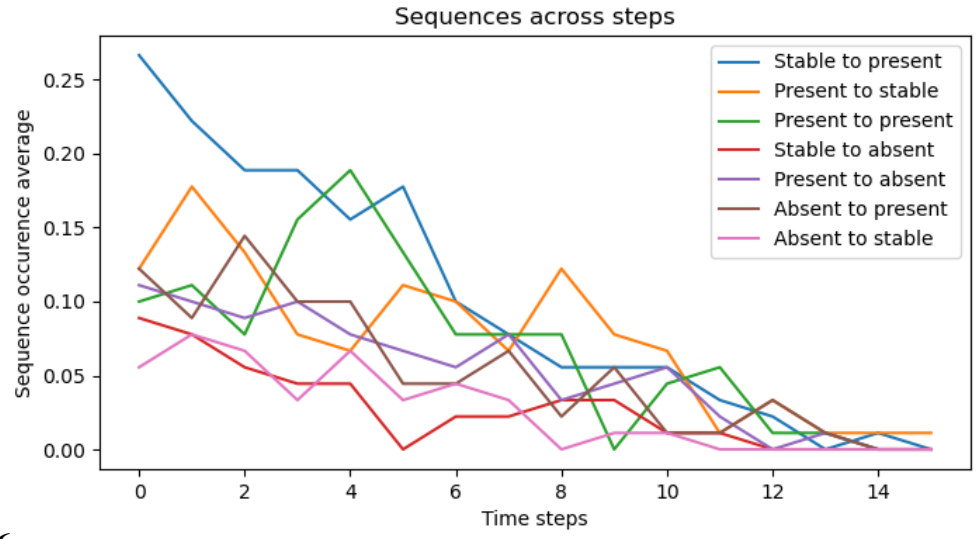


**Figure 6**

*Replay sequences as elicited from the sequences of goal choices chosen by the model. "stable" blocks are those that are placed first, "present" blocks are those that are present in the figure apart from "stable", and "absent" blocks are those that are not present in the target configuration. An early dominance of "stable to present" sequences is seen during steps 0-3, a dominance of "present-present" is seen from steps 3-5, and a dominance of "present to stable" is seen during steps 7-10.*

## Discussion

In this study, we situate a novel hypothesis on hippocampal replay within the CLARION cognitive architecture. Specifically, we attempt to model hippocampal replay as drawing samples from internal generative models (encoded explicitly with rules) for hypothesis testing various abstractions formed from experience. To evaluate this hypothesis, we described a model to capture human performance on a block inference task and the replay sequences recorded in the hippocampus during such tasks. The results indicate a great qualitative fit with the neural data but not the behavioural data observed in Schwartenbeck et al. (2023).

As mentioned in Schwartenbeck et al. (2023), we see that our model adheres by a sensible construction strategy, the same observed in Schwartenbeck et al. (2023): start with the first block,

test which blocks connect to the first block, and then evaluate connections between "Present" blocks. We also observe the latter effect of "Present to Stable" sequences present in Schwartenbeck et al. (2023), suggesting a strong fit of the model to the sequences presented in Schwartenbeck et al. (2023). However, the poor fit of the behavioural data suggests that the sequences of hippocampal replay may not be all there is to forming the solution for the brick task, e.g, if replay sequences form an input to secondary processes for reasoning not fully captured by replay. One way this may be is that the mechanisms for choosing responses in our model may not be reflective of the manner in which the results of replay are utilized in the brain. Further studies exploring the causal role of replay (e.g., through disruption of replay (Kurth-Nelson et al., 2023)) may be required to determine the exact role of replay and other processes in such tasks.

Alternatively, given the performance of the neural network in Figure 3, it may also be that an alternative formulation of our network may lead to better results with respect to response times and correctness. Future work must experiment with whether different RL paradigms for the goal learning network can improve upon this fit qualitatively and quantitatively. For example, MBRL architectures like Dreamer (Hafner et al., 2025) incorporate replay processes to train a generative world model. If an abstract world model could indeed form using replay in such a model, perhaps it can guide low-level actions in the ACS better than the simple form of deep Q-learning deployed in this model.

In conclusion, our fit to the replay sequences presented in Schwartenbeck et al. (2023) lends weight to both the hypothesis of replay as hypothesis testing abstractions (as presented in Schwartenbeck et al. (2023) and Kurth-Nelson et al. (2023)) and guided-tree-search-like processes deployed in the training of neural networks for RL as a promising candidate for explaining the algorithmic motifs by which replay may achieve such operations over abstractions.

## References

Ambrose, R. E., Pfeiffer, B. E., & Foster, D. J. (2016). Reverse replay of hippocampal place cells is uniquely modulated by changing reward. *Neuron*, *91*(5), 1124–1136. https://doi.org/http://dx.doi.org/10.1016/j.neuron.2016.07.047

Anderson, J. R., John, B. E., Just, M. A., Carpenter, P. A., Kieras, D. E., & Meyer, D. (1995). Production system models of complex cognition. *Proceedings of the seventeenth annual conference of the cognitive science society*, 9–12.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological review*, *84*(2), 191. https://doi.org/https://psycnet.apa.org/doi/10.1037/0033-295X.84.2.191

Behrens, T. E., Muller, T. H., Whittington, J. C., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? organizing knowledge for flexible behavior. *Neuron*, *100*(2), 490–509. https://doi.org/https://doi.org/10.1016/j.neuron.2018.10.002

Cherian, A., Peng, K.-C., Lohit, S., Smith, K. A., & Tenenbaum, J. B. (2023). Are deep neural networks smarter than second graders? *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10834–10844.

Hafner, D., Pasukonis, J., Ba, J., & Lillicrap, T. (2025). Mastering diverse control tasks through world models. *Nature*, 1–7. https://doi.org/https://doi.org/10.1038/s41586-025-08744-2

Kurth-Nelson, Z., Behrens, T., Wayne, G., Miller, K., Luettgau, L., Dolan, R., Liu, Y., & Schwartenbeck, P. (2023). Replay and compositional computation. *Neuron*, *111*(4), 454–469. https://doi.org/https://doi.org/10.1016/j.neuron.2022.12.028

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338. https://doi.org/https://doi.org/10.1126/science.aab3050

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences*, *40*, e253. https://doi.org/https://doi.org/10.1017/S0140525X16001837

Liu, Y., Dolan, R. J., Kurth-Nelson, Z., & Behrens, T. E. (2019). Human replay spontaneously reorganizes experience. *Cell*, *178*(3), 640–652. https://doi.org/https://doi.org/10.1016/j.cell.2019.06.012

Mattar, M. G., & Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *Nature neuroscience*, *21*(11), 1609–1617. https://doi.org/https://doi.org/10.1038/s41593-018-0232-z

Mattar, M. G., & Lengyel, M. (2022). Planning in the brain. *Neuron*, *110*(6), 914–934. https://doi.org/https://doi.org/10.1016/j.neuron.2021.12.018

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, *24*(1), 167–202. https://doi.org/https://doi.org/10.1146/annurev.neuro.24.1.167

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2017). Dorsal hippocampus contributes to model-based planning. *Nature neuroscience*, *20*(9), 1269–1276. https://doi.org/https://doi.org/10.1038/nn.4613

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Momennejad, I., Otto, A. R., Daw, N. D., & Norman, K. A. (2018). Offline replay supports planning in human reinforcement learning. *elife*, *7*, e32548. https://doi.org/https://doi.org/10.7554/eLife.32548

Moskvichev, A., Odouard, V. V., & Mitchell, M. (2023). The conceptarc benchmark: Evaluating understanding and generalization in the arc domain. *arXiv preprint arXiv:2305.07141*. https://doi.org/https://doi.org/10.48550/arXiv.2305.07141

Schrittwieser, J., Hubert, T., Mandhane, A., Barekatain, M., Antonoglou, I., & Silver, D. (2021). Online and offline reinforcement learning by planning with a learned model. *Advances in Neural Information Processing Systems*, *34*, 27580–27591.

Schwartenbeck, P., Baram, A., Liu, Y., Mark, S., Muller, T., Dolan, R., Botvinick, M., Kurth-Nelson, Z., & Behrens, T. (2023). Generative replay underlies compositional inference in the hippocampal-prefrontal circuit. *Cell*, *186*(22), 4885–4897. https://doi.org/https://doi.org/10.1016/j.cell.2023.09.004

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, *20*(1), 11. https://doi.org/https://doi.org/10.1136/jnnp.20.1.11

Stewart, T. C., & Eliasmith, C. (2009). Spiking neurons and central executive control: The origin of the 50-millisecond cognitive cycle. *9th International Conference on Cognitive Modelling*, *122*(127), 130–131.

Sun, R. (2006). The clarion cognitive architecture: Extending cognitive modeling to social simulation. *Cognition and multi-agent interaction*, 79–99.

Sun, R. (2016). *Anatomy of the mind: Exploring psychological mechanisms and processes with the clarion cognitive architecture*. Oxford University Press.

Vikbladh, O. M., Meager, M. R., King, J., Blackmon, K., Devinsky, O., Shohamy, D., Burgess, N., & Daw, N. D. (2019). Hippocampal contributions to model-based planning and spatial memory. *Neuron*, *102*(3), 683–693. https://doi.org/https://doi.org/10.1016/j.neuron.2019.02.014