Can sleep protect memories from catastrophic forgetting?

Oscar C. González[a, b], Yury Sokolov[b], Giri P. Krishnan[b], Maxim Bazhenov[a,b]

[a] Neuroscience Graduate Program, [b] Department of Medicine University of California, San Diego, CA 92093

To whom correspondence should be addressed:

Maxim Bazhenov

Department of Medicine

University of California, San Diego

La Jolla, CA 92093

Phone: 858-534-8391

Email: mbazhenov@ucsd.edu

Keywords: Sleep, memory consolidation, continual learning, neural network, catastrophic forgetting

Abstract:

Previously encoded memories can be damaged by encoding of new memories, especially when they are relevant to the new data and hence can be disrupted by new training – a phenomenon called "catastrophic forgetting". Human and animal brains are capable of continual learning, allowing them to learn from past experience and to integrate newly acquired information with previously stored memories. A range of empirical data suggest important role of sleep in consolidation of recent memories and protection of the past knowledge from catastrophic forgetting. To explore potential mechanisms of how sleep can enable continual learning in neuronal networks, we developed a biophysically-realistic thalamocortical network model where we could train multiple memories with different degree of interference. We found that in a wake-like state of the model, training of a "new" memory that overlaps with previously stored "old" memory results in degradation of the old memory. Simulating NREM sleep state immediately after new learning led to replay of both old and new memories - this protected old memory from

forgetting and ultimately enhanced both memories. The effect of sleep was similar to the interleaved training of the old and new memories. The study revealed that the network slow-wave oscillatory activity during simulated deep sleep leads to a complex reorganization of the synaptic connectivity matrix that maximizes separation between groups of synapses responsible for conflicting memories in the overlapping population of neurons. The study predicts that sleep may play a protective role against catastrophic forgetting and enables brain networks to undergo continual learning.

Significance:

Continual learning, free of catastrophic forgetting, remains to be an unsolved problem in artificial neural networks. Biological networks have evolved a mechanism by which they can prevent interference and allow continual learning throughout the life of the organism. Building upon a range of data suggesting importance of sleep in memory and learning, here we test a hypothesis that deep sleep may play a role in continual learning and protecting memories from catastrophic forgetting. Our results revealed that complex patterns of synchronized oscillatory activity in the thalamocortical network during deep sleep reorganize synaptic connectivity to allow for consolidation of interfering memories and to enable continual learning.

Introduction:

Animals and humans are capable of continuous, sequential learning. In contrast, modern artificial neural networks suffer from its inability to perform continual learning (Ratcliff, 1990; French, 1999; Hassabis et al., 2017; Hasselmo, 2017; Kirkpatrick et al., 2017) . Introducing a new memory results in catastrophic forgetting and interference with old memories (Ratcliff, 1990; McClelland et al., 1995; French, 1999; Hasselmo, 2017). Some attempts have been made to overcome catastrophic forgetting such as (1) explicit retraining or interleaved training of all previously learned memories (Hasselmo, 2017) or (2) using generative models to reactivate previous inputs (Kemker and Kanan, 2017) or (3) artificially "freezing" subsets of synapses responsible for old memories (Kirkpatrick et al., 2017). These solutions allow "new" memories to avoid catastrophically interfering with previously stored "old" memories, however may require to explicitly retrain all memories and have limitations on the type of the new memories to be trained and the network architectures (Kemker et al., 2017). Exactly how biological systems avoid catastrophic forgetting and interference of "old" and "new" memories supporting

continuous learning remains to be understood. In this paper, we propose a mechanism of how sleep modifies synaptic connections that encode multiples memories and predict how it may allow for continual learning in biological networks.

Sleep has been suggested to play an important role in memory consolidation (Paller and Voss, 2004; Walker and Stickgold, 2004; Oudiette and Paller, 2013; Rasch and Born, 2013; Stickgold, 2013; Wei et al., 2016; Wei et al., 2018). Specifically, the role of stage 2 (N2) and stage 3 (N3) Non Rapid Eye Movement (NREM) sleep has been shown to help with the consolidation of newly encoded memories (Paller and Voss, 2004; Walker and Stickgold, 2004; Rasch and Born, 2013; Stickgold, 2013). The mechanism by which memory consolidation is influenced by sleep is still debated, however, a number of hypotheses has been put forward. Sleep may enhance memory consolidation through repeated memory reactivation or replay during characteristic sleep rhythms such as spindles and slow oscillations (Paller and Voss, 2004; Clemens et al., 2005; Marshall et al., 2006; Oudiette et al., 2013; Rasch and Born, 2013; Wei et al., 2016; Ladenbauer et al., 2017; Wei et al., 2018). Memory replay during deep N3 sleep could help strengthen previously stored memories and prevent future interference. Previous work using transcranial direct current stimulation (tDCS) showed that increasing neocortical slow oscillations during early stages of NREM sleep resulted in improved consolidation of declarative memories (Marshall et al., 2004; Marshall et al., 2006; Ladenbauer et al., 2017). Similarly, spatial memory consolidation has been shown to improve following cued reactivation of memory traces during slow-wave sleep (Paller and Voss, 2004; Oudiette et al., 2013; Oudiette and Paller, 2013; Papalambros et al., 2017). These studies point to the critical role of NREM sleep in the consolidation of newly encoded memories.

Here we used a biophysically realistic thalamocortical network model to test the hypothesis that NREM sleep, specifically slow-wave (N3) sleep, could play an important role in consolidation of newly encoded memories and preventing forgetting of the “old” memories. Our model predicts that a period of N3 sleep, following training of the interfering memory sequences during wake, can promote replay of both old and new memories thereby preventing interference. We show that N3 sleep results in the fine tuning of synaptic connectivity to allow for the same population of neurons to store competing memories without the need of explicit retraining of the previously stored memory sequences.

Methods and Materials:

Thalamocortical network model:

Network architecture. The thalamocortical network model used in this study has been previously described in detail (Krishnan et al., 2016; Wei et al., 2016; Wei et al., 2018). Briefly, our network was comprised of a thalamus which contained 100 thalamocortical relay neurons (TC) and 100 reticular neurons (RE), and a cortex containing 500 pyramidal neurons (PY) and 100 inhibitory interneurons (IN). Our model contained only local network connectivity as described in figure 1. Excitatory synaptic connections were mediated by AMPA and NMDA connections, while inhibitory synapses were mediated through $GABA_A$ and $GABA_B$. Starting with the thalamus, TC neurons formed AMPA connections onto RE neurons with a connection radius of 8 ($R_{AMPA(TC\text{-}RE)}$ = 8). RE neurons then projected inhibitory $GABA_A$ and $GABA_B$ connections onto TC neurons with $R_{GABA\text{-}A(RE\text{-}TC)}$ = 8 and $R_{GABA\text{-}B(RE\text{-}TC)}$ = 8. Inhibitory connections between RE-RE neurons were mediated by $GABA_A$ connections with $R_{GABA\text{-}A(RE\text{-}RE)}$ = 5. Within the cortex, PY neurons formed AMPA and NMDA connections onto PYs and INs with $R_{AMPA(PY\text{-}PY)}$ = 20, $R_{NMDA(PY\text{-}PY)}$ = 5, $R_{AMPA(PY\text{-}IN)}$ = 1, and $R_{NMDA(PY\text{-}IN)}$ = 1. PY-PY AMPA connections had a 60% connection probability, while all other connections were deterministic. Cortical inhibitory IN-PY connections were mediated by $GABA_A$ with $R_{GABA\text{-}A(IN\text{-}PY)}$ = 5. Finally, thalamocortical connections were mediated by AMPA connections with $R_{AMPA(TC\text{-}PY)}$ = 15, $R_{AMPA(TC\text{-}IN)}$ = 3, $R_{AMPA(PY\text{-}TC)}$ = 10, and $R_{AMPA(PY\text{-}RE)}$ = 8.

Intrinsic currents. All neurons were modeled with Hodgkin-Huxley kinetics. Cortical PY and IN neurons contained dendritic and axo-somatic compartments as previously described (Wei et al., 2018). The membrane potential dynamics were modeled by the following equation:

$$C_m \frac{dV_D}{dt} = -I_D^{Na} - I_D^{NaP} - I_D^{Km} - I_D^{KCa} - ACh_{gkl} I_D^{KL} - I_D^{HVA} - I_D^{L} - g(V_D - V_S) - I^{syn}$$

$$g(V_D - V_S) = -I_S^{Na} - I_S^{NaP} - I_S^{K}$$

where C_m is the membrane capacitance, $V_{D,S}$ are the dendritic and axo-somatic membrane voltages respectively, I^{Na} is the fast sodium (Na^+) current, I^{NaP} is the persistent Na^+ current, I^{Km} is the slow voltage-dependent non-inactivating potassium (K^+) current, I^{KCa} is the slow

calcium (Ca^{2+})-dependent K^+ current, ACh_{gkl} represents the change in K^+ leak current I^{KL} which is dependent on the level of acetylcholine (ACh) during the different stages of wake and sleep, I^{HVA} is the high-threshold Ca^{2+} current, I^L is the chloride (Cl^-) leak current, g is the conductance between the dendritic and axo-somatic compartments, and I^{syn} is the total synaptic current input to the neuron. IN neurons contained all intrinsic currents present in PY with the exception of the I^{NaP}.

Thalamic neurons (TC and RE) were modeled as single compartment neurons with membrane potential dynamics mediated by the following equation:

$$C_m \frac{dV_D}{dt} = -I^{Na} - I^K - ACh_{gkl} I^{KL} - I^T - I^h - I^L - I^{syn}$$

where I^{Na} is the fast Na^+ current, I^K is the fast K^+ current, I^{KL} is the K^+ leak current, I^T is the low-threshold Ca^{2+} current, I^h is the hyperpolarization-activated mixed cation current, I^L is the Cl^- leak current, and I^{syn} is the total synaptic current input to the neurons. The I^h was only expressed in the TC neurons and not the RE neurons. The influence of histamine (HA) on I^h was implemented as a shift in the activation curve by HA_{gh} as described by:

$$m_\infty = 1/(1 + exp((V + 75 + HA_{gh})/5.5))$$

A detailed description of the individual currents can be found in our previous studies (Krishnan et al., 2016; Wei et al., 2018).

Synaptic currents and spike-timing dependent plasticity (STDP). AMPA, NMDA, and $GABA_A$ synaptic current equations were described in detail in our previous studies (Krishnan et al., 2016; Wei et al., 2018). The effects of ACh on $GABA_A$ and AMPA synaptic currents in our model are described by the following equations:

$$I_{syn}^{GABA} = \gamma_{GABA_A} g_{syn} [O](V - E_{syn})$$

$$I_{syn}^{AMPA} = ACh_{AMPA} g_{syn} [O](V - E_{syn})$$

where g_{syn} is the maximal conductance at the synapse, $[O]$ is the fraction of open channels, and E_{syn} is the channel reversal potential ($E_{GABA\text{-}A}$ = -70 mV, E_{AMPA} = 0 mV, and E_{NMDA} = 0 mv). Parameter γ_{GABA_A} modulates the GABA synaptic currents for IN-PY, RE-RE, and RE-TC connections. For IN neurons γ_{GABA_A} was 0.22, 0.264, and 0.44 for awake, N2, and N3 sleep, respectively; γ_{GABA_A} for RE was 0.6, 0.72, 1.2 for awake, N2, and N3 sleep. ACh_{AMPA} defines the influence of ACh levels on AMPA synaptic currents for PY-PY, TC-PY, and TC-IN. ACh_{AMPA} for PY was 0.133, 0.1938, and 0.4332 for awake, N2, and N3 sleep. ACh_{AMPA} for TC is 0.6, 0.72, 1.2 for awake, N2, and N3 sleep.

Potentiation and depression of synaptic weights between PY neurons were regulated by spike-timing dependent plasticity (STDP). The changes in synaptic strength (g_{AMPA}) and the amplitude of miniature EPSPs (A_{mEPSP}) have been described previously (Wei et al., 2018):

$$g_{AMPA} \leftarrow g_{AMPA} + g_{max}\, F(\Delta t)$$

$$A_{mEPSP} \leftarrow A_{mEPSP} + f A_{PY-PY}\, F(\Delta t)$$

where g_{max} is the maximal conductance of g_{AMPA}, and f = 0.01 represents the slower change of STDP on A_{mEPSP} as compared to g_{AMPA}. The STDP function F is dependent on the relative timing (Δt) of the pre- and post-synaptic spikes and is defined by:

$$F(\Delta t) = \begin{cases} A_+\, e^{-|\Delta t|/\tau_+}, if\ \Delta t > 0 \\ -A_-\, e^{-|\Delta t|/\tau_-}, if\ \Delta t < 0 \end{cases}$$

where $A_{+/-}$ set the maximum amplitude of synaptic change. $A_{+/-}$ = 0.002 and $\tau_{+/-}$ = 20 ms. A_- was reduced to 0.001 during training to reflect the effects of changes in acetylcholine concentration during focused attention on synaptic depression during task learning observed experimentally (Blokland, 1995; Shinoe et al., 2005; Sugisaki et al., 2016).

Sequence training and testing. Training and testing of memory sequences was performed similar to our previous study (Wei et al., 2018). Briefly, trained sequences were comprised of 5 groups of 10 sequential PY neurons. Each group of 10 were sequentially activated by a 10 ms DC pulse with 5 ms delay between subsequent group pulses. This activation scheme was applied every 1 s throughout the duration of the training period. Sequence 1 (S1) consisted of PY groups (in order of activation): A(200-209), B(210-219), C(220-229), D(230-239), E(240-249). Sequence 2 (S2) consisted of PY groups (in order of activation): W(360-369), V(350-359),

X(370-379), Y(380-389), Z(390-399) and can be referred as non-linear due to the non-spatially sequential activations of group W, V, and X. Sequence 1* (S1*) was trained over the same population of neurons trained on S1 but in the reverse activation order (i.e. E-D-C-B-A). During testing, the network was presented with only the activation of the first group of a given sequence (A for S1, W for S2, and E for S1*). Performance was measured based on the network's ability to recall/complete the remainder of the sequence (i.e. A-B-C-D-E for S1) within a 350 ms time window. Similar to training, test activation pulses were applied every 1 s throughout the testing period. Training and testing of the sequences occurred sequentially as opposed to simultaneously as in our previous study (Wei et al., 2018).

Data analysis:

All analyses were performed with custom built Matlab and Python scripts. Data are presented as mean ± standard error of the mean (SEM) unless otherwise stated. For each experiment a total of 10 simulations with different random seeds were used for statistical analysis.

Sequence performance measure. A detailed description of the performance measure used during testing can be found in (Wei et al., 2018). Briefly, the performance of the network on recalling a given sequence following activation of the first group of that sequence (see Methods and Materials: *Sequence training and testing*) was measured by the percent of successful sequence recalls. We first detected all spikes within the predefined 350 ms time window for all 5 groups of neurons in a sequence. The firing rate of each group was then smoothed by convolving the average instantaneous firing rate of the group's 10 neurons with a Gaussian kernel with window size of 50 ms. We then sorted the peaks of the smoothed firing rates during the 350 ms window to determine the ordering of group activations. Next, we applied a string match (SM) method to determine the similarity between the detected sequences and an ideal sequence (ie. A-B-C-D-E for S1). SM was calculated using the following equation:

$$SM = 2 * N - \sum_{i=1}^{N} |L(S_{test}, S_{sub}[i]) - i|$$

where N is the sequence length of S_{test}, S_{test} is the test sequence generated by the network during testing, S_{sub} is a subset of the ideal sequence that only contains the same elements of S_{test}, and

$L(S_{test}, S_{sub}[i])$ is the location of the element $S_{sub}[i]$ in sequence S_{test}. SM was then normalized by double the length of the ideal sequence. Finally, the performance was calculated as the percent of recalled sequences with SM≥Th, where Th is the selected threshold (here, Th = 0.8) indicating that the recalled sequence must be at least 80% similar to the ideal sequence to be counted as a successful recall as previously done in (Wei et al., 2018).

Sequence replay during N3 sleep. To find out whether a trained sequence is replayed in the trained region of the network during the upstate of a slow-wave in N3 sleep, we first identified the beginning and the end of each upstate by considering sorted spike times of neurons in each group. For each group, the time instances of consecutive spikes that occur within a 15 ms window were considered as candidate members of an upstate, where the window size was determined to decrease the chance of two spikes of the same neuron within the window. To eliminate spontaneous spiking activity of a group that satisfies the above condition but is not part of an upstate, we additionally require that the period between two upstate was at least 300 ms, which corresponds to a cortical down state. The values of windows durations reported above were identified to maximize the performance of upstate search algorithm.

Once all Up states were determined, we defined the time instances when groups are active in each Up state. A group was defined as active if the number of neurons from the group that spikes during 15 ms exceeded the activation threshold, and the instance when the group is active was defined as the average over spike times of a subgroup of neurons with the size equals to the activation threshold within the 15 ms window. In our study the activation threshold was selected to be half of a group size (i.e. 5 neurons). Using sorted time instances when groups are active, we counted the number of times a possible transition between arbitrary groups, and if all four transitions of a sequence were observed sequentially in the right order then we counted that as a replay of the sequence.

Connectivity thresholding. To track connections in the trained region of a network, we performed thresholding of the underlying network, called weighted graph, at each second of simulation time. Thresholding results in a directed unweighted graph, where directed edges define significant edges of the weighted graph for storing the memories. The initial values of synaptic weights are drawn from the same Gaussian distribution with mean 0.0707 and variance 10% of the mean. A synaptic connection that encodes a sequence should undergo potentiation,

and, hence, its strength should increase in such an event. With this in mind, we selected a threshold to be 0.08 in figures 8A and B that would guarantee that a specific synapse has taken part in storing a memory. To avoid a runaway of synaptic strength over entire simulation protocols, an upper bound on synaptic strength was imposed, which was selected to be twice the mean of initial weight. This upper bound was also used as a threshold. The directed unweighted graph obtained with this threshold value should identify connections which strongly support a memory, and this value was used in figures 8B-D.

Relative densities of significant connections. Relative densities were computed as a number of connections in a particular direction in an unweighted graph obtained after thresholding divided by a uniform constant, squared group size, regardless whether a connection is unidirectional or has a recurrent pair. Connections that have recurrent pairs were omitted from density evaluation of unidirectional connections.

Results:

Thalamocortical network model developed in this study included two main structures (cortex and thalamus), each comprised of excitatory and inhibitory cell populations (figure 1A). Both populations included basic excitatory-inhibitory loop. In the thalamus, excitatory thalamocortical neurons (TCs) received excitatory connections from cortical excitatory neurons (PYs), and inhibitory connections from thalamic reticular neurons (REs). RE neurons were interconnected by inhibitory connections and received excitatory connections from PYs and TCs. PYs in the cortex received excitatory input from thalamus as well as excitatory inputs from other PYs and inhibitory inputs from cortical inhibitory interneurons (INs). INs received excitatory inputs from PYs in addition to the excitatory connections from TCs. PYs had a 60% probability of connecting to neighboring PYs (PY-PY radius was set to 20 neurons). Figure 1D (left) shows the adjacency matrix for the cortical PYs that arise from these two restrictions. The strength of the synaptic connections between PYs was Gaussian distributed thereby making inhomogeneous initial synaptic weights (figure 1D right). The model was able to simulate transitions between awake and sleep (figure 1B/C) by simulating effects of neuromodulators (Krishnan et al., 2016). PY-PY connections were plastic and regulated by STDP that was biased for potentiation during wake-state to model the higher level of acetylcholine (Blokland, 1995; Shinoe et al., 2005; Sugisaki et al., 2016).

Sequential training of spatially separated memory sequences does not lead to interference

We trained two spike sequences sequentially (first S1 and then S2) in the spatially distinct regions of the network as shown in figure 2. Each memory sequence contained 5 sequentially activated groups of 10 neurons per group. The first sequence (S1) was trained in the population of neurons 200-249 (figure 2B) with groups of 10 neurons that were ordered by increasing neuron numbers (A-B-C-D-E). Training S1 resulted in increased synaptic weight strengths of participating neurons (figure 2D, top) and in increased performance of sequence completion immediately after training (figure 2B/C, top). While the strength of synapses in the direction of S1 increased, synapses in the opposite direction showed reduction of strength consistent with the STDP rule used in this network (see *Methods and Materials*). The second memory (S2) was trained for an equal amount of time in population of neurons 350-399 (W-V-X-Y-Z, figure 2B bottom). Training of S2 also resulted in synaptic weight changes (Figure 2D, middle) and improvement in performance immediately after training (figure 2B/C, bottom). Importantly, training of S2 did not interfere with the weight changes observed for S1 because both sequences involved spatially distinct populations of neurons (compare figure 2D, top and middle).

After successful training of both memories the network went through a period of slow-wave (N3) sleep. Synaptic weights for both memory sequences showed strong increases in the direction of their respective patterns and further decreases in the opposing directions (figure 2D, bottom). These increases in synaptic weight strengths for both memories were accompanied by an increase in pattern completion (figure 2B, right) and improved performance (figure 2C, red bar). These results are in line with previous work using thalamocortical models to store memory sequences and showing the beneficial role of N3 sleep in the consolidation of multiple memories (Wei et al., 2018).

Training overlapping memory sequence results in interference

We next tested whether our network model displays interference when a “new” spike sequence (S1*) (figure 3A) was trained in the same population of neurons as the earlier “old” spike sequence (S1). The S1* included the same exactly population of neurons as S1, but was

trained in the opposite direction to S1, that is the stimulation group order was E-D-C-B-A (figure 3B). Sequence S2 was once again trained in a spatially distinct region of the network such that it would not interact with the other two memory sequences (figure 3A/B). Testing of spike sequence completion was performed immediately after each training period. Similar to the previous results, training of S1 resulted in an immediate increase in performance of S1 completion (figure 3C, top/left). Training of S1 also led to decrease in performance for S1* below its baseline level in "naive" network (figure 3C, bottom/left). This was a result of reduced synaptic strength in the direction of S1*, as seen previously in figure 2D (top). (Note, that even before S1* was explicitly trained, a naïve network displayed some above zero probability to complete a sequence that depended on the initial strength of synapses and random network activity). Training of S2 led to an increase in S2 performance without damaging S1. Subsequent training of S1* resulted in both an increase in S1* performance and a significant reduction of S1 performance (figure 3C). To evaluate impact of S1* training on S1 performance, we varied duration of S1* training (figure 3D). Increase in the amount of S1* training led to reduction of S1 performance up to the point when performance of S1 sequence was reduced back to the baseline level (400 s training duration). This suggests that in a wake-like state, sequential training of two competing memories in the same populations of neurons results in memory interference and catastrophic forgetting of the old memory sequence.

Interleaved training can recover "old" memory after damage

One solution to avoid catastrophic forgetting in neural networks is to use a new training data set that includes both old and new training patterns (McClelland et al., 1995). Thus, we next examined if an interleaved presentation of S1 and S1* can reverse the damage and can rescue S1. In the interleaved training protocol, instead of presenting multiple trails of S1 (S1->S1->…) followed by multiple trials of S1* (S1*->S1*->…), we interleaved S1 and S1* stimulation patterns at subsequent trails (S1->S1*->S1->S1*->…) (see figure 4B).

First, using the same protocol as described in the figure 3, we trained the sequences sequentially (S1->S1->…->S2->S2->…->S1*->S1*->…). This, once again, resulted in an initial increase in both S1 and S2 performance followed by a reduction of S1 performance after training of S1* (figure 4C). Analysis of the synaptic connectivity matrix revealed synaptic weights

dynamics behind performance change (figure 4D). After S1 training, synaptic weights between neurons representing S1 showed increase in the direction of S1 training and decrease in the opposite network direction (figure 4D, top). Training of S1* resulted in increase of synaptic weight strengths in the direction of S1* while reducing the strength of synapses in S1 direction (figure 4D, middle). After initial sequential training phase, interleaved training by the data set containing both S1 and S1* was performed (figure 4A, green bar). An example of the interleaved training protocol is shown in figure 4B. Interleaved training led to the further changes of synaptic connectivity and performance increase for both sequences (figure 4C, green). Importantly, we observed an increase of synaptic strength in non-overlapping subsets of synapses representing both S1 and S1* patterns (figure 4D, bottom). In other words, for each two neuronal groups (e.g., A-B), we saw increase of A->B (and decrease of B->A) synapses for some pairs of neurons and opposite effect for the other pairs. These results suggest that in the cortical model, similar to the artificial neural networks, interleaved training of the interfering patterns can enable successful learning of the both patterns.

Sleep enables replay and performance improvement for interfering memories

Humans and animals are capable of continuous/sequential learning without the need to explicitly retrain all the previously learned memories, as, e.g., required with the interleaved training approach. We hypothesized that sleep in the biological organisms may play a role in "retraining" the old memories while consolidating new memories. To test this hypothesis in our model, we simulated slow-wave sleep (N3) after the patterns S1/S2/S3 were trained sequentially (S1->S1->…->S2->S2->…->S1*->S1*->…) (figure 5A, red), as described in the previous sections (figures 3A and 5A). During sleep phase the network dynamic was fully autonomous; no stimulation patterns were applied. Figure 5B shows raster plots of the spiking activity during different testing periods illustrating improvements of a sequence completion after the sleep. We found that sleep was not only able to reverse the damage caused to S1 following S1* training, but it was also able to enhance all the previously trained memory sequences S1/S2/S3 (figure 5C, red bar). These results suggest that, in brain networks, replay during sleep can take place of explicit retraining of the old memory patterns to make possible continual learning without catastrophic forgetting.

In order to understand how sleep affects S1* and S1 memory traces, and how it leads to enhancement of both memories, we analyzed the synaptic weights between neurons within the entire population of neurons participating in the training of the two memory sequences (i.e. neurons 200-249). Figure 6A shows histograms of synaptic weights for synapses pointing in the direction of S1 (top row) and in the direction of S1* (bottom row). (Thus, e.g., if we would have exactly equal reciprocal synapses for each pair of neurons, the distributions in the top and bottom rows would be identical.) Each column indicates which manipulation was applied prior to the synaptic weight analysis (i.e. After S1 training, After S1* training, After Sleep). Blue histograms represent the initial weight distributions prior to the indicated manipulation, while red histograms show the synaptic weights after the manipulation. As shown in figure 6A, prior to any training, synaptic weights in the direction of either memory sequence were Gaussian distributed (blue histogram, left). After S1 training, the weights for S1 strengthened (shifted to the right), while the weights for S1* weakened (shifted to the left). As expected, this trend was reversed when S1* was trained (figure 6A, middle). Finally, sleep resulted in a strengthening of synaptic weights for both memory sequences (figure 6B, red).

These trends in synaptic weight dynamics can also be seen in the scatter plots of S1*/S1 weights (figure 6B), where for each pair of neurons (e.g., A-B) from population of interest, we plotted the weight of A->B synapses on X-axis and the weight of B->A synapses on Y-axis. Therefore, any point belonging to the X- or Y-axis would indicate unidirectional connection between two neurons. The initial Gaussian distribution of weights was pushed towards the bottom right corner of the plot indicating increases in S1 weights and relative decrease of S1* weights. Training of S1* caused an upward/left shift representing strengthening of S1* weights and weakening of S1 weights. Sleep appears to have taken the weights located in the center of the plots (i.e. strongly bidirectional weights) and separated them by pushing them to the extreme corners of the plot. In doing so, sleep converted strongly bidirectional connections between pairs of neurons into strongly unidirectional connections which preferentially contributed to consolidation of one memory sequence and another (figure 6B right). The matrix of synaptic weight strength in figure 6C further shows that after sleep each memory sequences was represent by a unique subset of unidirectional synapses that was sufficient to ensure sequence completion.

We repeated this analysis for the model with interleaved training and results are summarized in the Fig. 7. As seen in the mean projection of recurrent (bidirectional) synaptic

weights (figure 7A), before any training synaptic weights exhibited a Gaussian distribution. This distribution reflects the existence of functional recurrent synaptic connections in both sequence directions for most cell pairs. Following S1 training, the mean of the weight projection shifted in the direction of S1 indicating strengthening of synaptic weights in the direction of S1 and a weakening those in the direction of S1*. Next, S1* training strengthened recurrent weights in the direction of S1* as indicated by the shift in the mean weight projection back towards the center (figure 7A, third panel). After a period of sleep, the mean weight projections show a strong separation of weights with two main populations of weights, those preferential for only S1 or only S1*. It should be noted that the reduced/lack of weights in the middle portion of the projection suggest that most recurrent synaptic connections became functionally unidirectional thereby contributing to the storage of one sequence over the other (figure 7A fourth panel). The average of dynamics of synaptic weight changes in the sliding time window is summarized in figure 7B. The averaged vector plots in figure 7B reflect the trends of changes in synaptic weights which have been described thus far. It is of interest to note that the vector plots specific for sleep state (figure 7B right) clearly show the separation of recurrent synaptic weights. Interestingly, even weak synaptic weights were recruited for the storage of the one sequence or another (figure 7B right). This, however, was not the case for interleaved training. Figure 7C and D show that interleaved training also results in the separation of recurrent synaptic weights. Unlike sleep, interleaved training preserved many more recurrent connections indicated by the thick band of weights seen in the mean recurrent weight projection (figure 7C right). Increasing the duration of interleaved training did not result in better separation of weights (data not shown). Interleaved training also suppressed weak synaptic weights as shown in the vector plots (note arrows pointing to (0,0) in the bottom left area of the figure 7D right). This was in stark contrast to the recruitment of weak synaptic weights during sleep (same are in figure 7B right).

Sleep replay leads to fine tuning of synaptic connectivity to maximize separation of memory traces

The memories in the model were encoded on synaptic level due to STDP. To check whether qualitative similarities between sleep and interleaved training are also reflected in the underlying connectivity of the relevant cortical regions, for short weighted graph, we looked at

the evolution of the graph obtained by thresholding the weighted graph at each second of simulation time (see *Methods and Materials*). We looked separately at the density of connections that can be defined as unidirectional supporting memory S1, unidirectional supporting memory S1* and bidirectional (or recurrent) supporting both memories. As shown in the figure 8A, the relative density of the recurrent connections increased after training of S1* compare to the baseline condition. That, as well as decrease of S1 unidirectional connections, resulted in a decrease of S1 performance after S1* training as reported in figures 4C and 5C. This implies that the network had not been reorganized to allocate still available synapses to uniquely represent newly stored memory S1*, but instead, the increase in the number of relatively strong recurrent connections created interference with previously stored memory S1. To resolve the interference of the two competing memories, the network suppressed relatively strong recurrent connections during sleep (figure 8A left). Similarly, during interleaved training the network dynamics also attempted to reduce the number of recurrent connections (figure 8A right), however, the network was not able to reach the same connectivity state as seen after sleep.

This observation led to hypothesis that sleep attempts to reorganize network connectivity to an extreme configuration, where underlying connectivity is split into two groups: one representing one memory sequence and the other group representing the opposite competing sequence. To test this hypothesis, we first checked if the unweighted graphs obtained after thresholding have multiple maximal strongly connected components (MSCCs). (Two neurons belong to a MSCC if there is a directed path from one neuron to another along the directed edges of the unweighted graph. So, a single MSCC implies that there is a strong path between any two neurons in the network.) We counted the evolution of the number of MSCCs for different threshold values on synaptic weights as shown in the figure 8B. For a relatively high threshold, 0.08, every neuron in the trained region constituted a MSCC (Fig. 8B, red) as the number of relatively strong connections was negligible, because the mean of the initial synaptic weights was smaller than the threshold (see *Methods and Materials*). As training of S1 progressed, relatively strong connections were formed, which decreased of the number of MSCCs. However, by the end of S1 training the number of MSCCs increased back to a value similar to that prior to training. This corresponded to an extreme reorganization of the connectivity in the trained region, where most of the relatively strong connections had direction similar to the direction of S1. After subsequent S1* training, there was a single MSCC again as there were many relatively

strong connections in the network, and this remained to be true after sleep and interleaved training.

For a higher threshold value, which was equal to the upper bound on synaptic strength, the qualitative picture changed. At this level of granulation, the edges present in unweighted graph after thresholding depict synapses that uniquely represented the trained memories. There were not enough of these strong edges in the underlying graph after training of S1 or S1* memories, but they were created during sleep and interleaved training. While the unweighted graph had a small number of MSCCs for this threshold after sleep, the unweighted graph that resulted after interleaved training had a number of MSCCs exceeding the group size of a sequence, leading to segmentation of the connectivity profile into multiple MSCCs. Color-coded MSCC membership of each neuron from the trained region over the whole simulation duration is shown in Figure 8C, and during a significant connectivity reorganization is shown in Figure 8D. Figure 8C ruled out the possibility of connectivity profile split into two subnetworks after sleep as MSCCs consisted of a giant MSCC, where most of the neurons belonged to and a small number of isolated neurons in remote parts of the trained region. That is, our model after sleep reached the state at which memories were stored efficiently, i.e. without destroying communication between almost all neurons in the trained region, and at the same time, avoiding interference. Thus, a presence of a large number of relatively strong recurrent connections along with the segmentation of the connectivity profile formed by the strong edges into multiple MSCCs after interleaved training, and absence of these properties after sleep, may explain the difference in the performance of two networks shown on figures 4C and 5C, respectively.

Discussion:

We report here that biophysically realistic thalamocortical network model with learning rules based on STDP can exhibit catastrophic interference during training by overlapping patterns in awake-like state. Both interleaved training and slow-wave sleep-like activity were able to recover the damage to the “old” memory by new learning and to support consolidation of both old and new memories. Importantly, sleep-like state was able to accomplish this goal by direct reactivation of the old and new memories and without any access to the original training data. Furthermore, sleep was able to strengthen all the memories to the level beyond what interleaved training did. The mechanism by which network intrinsic activity during sleep was

able to recover and to enhance memories was through the reorganization of the synaptic connectivity matrix. In doing so, sleep created distinct (orthogonal) synaptic weight representations of the competing memories. Interleaved training was not able to reorganize the connectivity matrix to the extent of that seen during sleep. Our study suggests that sleep, by being able to reactivate directly memory traces encoded by synaptic weights, can provide a powerful mechanism to prevent catastrophic forgetting and to enable continual learning.

Catastrophic forgetting and biological systems

The work on catastrophic forgetting or interference in connectionist networks was pioneered by McCloskey and Cohen (Mccloskey and Cohen, 1989) and Ratcliff (Ratcliff, 1990). Catastrophic interference is observed when a previously trained network is required to learn new data, e.g., a new set of patterns. When learning new data, the network can suddenly erase the memory of the old, previously learned patterns (French, 1999; Hasselmo, 2017; Kirkpatrick et al., 2017). Such type of forgetting of the previously learned data occurs only after sufficient presentations of the new patterns. Catastrophic interference is thought to be related to so-called "plasticity-stability" problem. The problem comes from the difficulty of creating a network with connection which are plastic enough to learn new data, while stable enough to prevent interference between old and new training sets. Due to the inherent trade-off between plasticity and memory stability, catastrophic interference and forgetting remains to be a difficult problem to overcome in connectionist networks (French, 1999; Hasselmo, 2017; Kirkpatrick et al., 2017).

A number of attempts have been made to overcome this interference in artificial neural networks (French, 1999; Hasselmo, 2017; Kirkpatrick et al., 2017). Early attempts included changes to the backpropagation algorithm, implementation of a "sharpening algorithm" in which a decrease in the overlap of the internal representations was achieved by making hidden-layer representations sparse, or changes to the internal structure of the network (French, 1999; Hasselmo, 2017; Kirkpatrick et al., 2017). These attempts were able to reduce severity of catastrophic interference in specific cases but could not provide a complete and generic solution to the problem. Another popular method for preventing interference or forgetting is to explicitly retrain or rehearse all the previously learned pattern sets while training the network on the new patterns (Hasselmo, 2017).

In this study, we show that following catastrophic interference of the two overlapping spike patterns trained sequentially, a period of interleaved training was capable of recovering the "old" and partially damaged memory. This method, however, does not result in further changes to the synaptic representation of the "old" memory to achieve optimal separation between old and new overlapping traces. Rather, interleaved training was only capable of recovering the "old" memory to extend when damage to both old and new memory was manageable to avoid forgetting (figure 4). Another primary concern with interleaved training is that it becomes increasingly difficult/cumbersome to retrain all the memories as the number of stored memories continues to increase. And the access to the earlier training data may not be available anymore. As previously mentioned, biological systems have evolved a mechanism to prevent this form of forgetting without the need to explicitly retrain the network by the all previously encoded memories. Studying how these systems overcome this issue should provide insights into novel techniques to combat catastrophic forgetting in artificial neural networks.

Until recently it was debated whether humans could exhibit catastrophic interference during new learning (McClelland et al., 1995). Initial lack of evidence for existence of the catastrophic interference in humans can be attributed to specific memory system being probed and the experimental paradigms being used (Merhav et al., 2014). It was hypothesized that the hippocampus may play a prominent role in protecting against neocortical catastrophic interference (McClelland et al., 1995). In the seminal paper (McClelland et al., 1995), McClelland, McNaughton, and O'Reilly suggested that the separation of the hippocampal and neocortical systems allows for a natural way to prevent the interference between newly encoded information and previously stored memories. According to "Complementary Learning System" theory (McClelland et al., 1995), the hippocampus is responsible for the fast acquisition of new information, while the neocortex would more gradually learn a generalized and distributed representation. Indeed, hippocampal memory replay is currently thought to be the primary mechanism by which the hippocampus can train the neocortex on the newly encoded information (French, 1999; Rasch and Born, 2013; Merhav et al., 2014; Feld and Born, 2017).

Though hippocampal-dependent memory consolidation may not clearly exhibit signs of catastrophic interference, studies showed that certain forms of hippocampal independent memory consolidation can demonstrate catastrophic interference (French, 1999; Merhav et al., 2014). Because the hippocampal system is not fully developed in younger children, it has been

suggested that children rely more heavily on hippocampal independent memory consolidation. A learning scheme known as Fast Mapping is thought to occur independently of the hippocampus as amnesic patients with medial temporal lobe damage are capable of learning word-meaning associations through Fast Mapping (Merhav et al., 2014). Recent studies have demonstrated catastrophic interference using the "AB-AC paired associates" task in both healthy and amnesic patients (Merhav et al., 2014; Borragan et al., 2015). In humans, catastrophic interference is comprised of two different phenomena: proactive interference (PI) and retroactive interference (RI). Similar to the type of interference explored in our study, retroactive interference occurs when a subject is required to learn a new pair association (AC) after previously learning an old, overlapping association (AB) (Merhav et al., 2014; Borragan et al., 2015). We found that the model of learning proposed here was capable of capturing this form of biological interference - training of a new sequence pattern that was similar to the old memory led to damage of the previously trained memory. As sleep has been shown to be beneficial for consolidation of newly formed memories, we tested here a hypothesize that sleep may help to recover from catastrophic interference.

Sleep and memory consolidation

Though variety of sleep functions remains to be understood, there is growing evidence for the role of sleep in consolidation of newly encoded memories (Paller and Voss, 2004; Walker and Stickgold, 2004; Oudiette and Paller, 2013; Rasch and Born, 2013; Stickgold, 2013; Wei et al., 2016; Wei et al., 2018). The mechanism by which memory consolidation is influenced by sleep is still largely debated, however a number of hypotheses have been put forward. One such hypothesis is the "Active System Consolidation Hypothesis" (Rasch and Born, 2013). Central to this hypothesis is idea of repeated memory reactivation (Paller and Voss, 2004; Mednick et al., 2013; Oudiette et al., 2013; Oudiette and Paller, 2013; Rasch and Born, 2013; Stickgold, 2013; Wei et al., 2016). The reactivation of a newly encoded memory has been suggested to take place during NREM sleep (Paller and Voss, 2004; Mednick et al., 2013; Oudiette et al., 2013; Oudiette and Paller, 2013; Rasch and Born, 2013; Stickgold, 2013; Wei et al., 2016). It is thought that neocortical slow oscillations may nest thalamocortical spindles as well as hippocampal sharp-wave ripples; coordinated activation of these major sleep rhythms can support repeatable

reactivation of the hippocampal memory traces and their mapping to the neocortex. Indeed, recent studies using transcranial direct current stimulation (tDCS) (Marshall et al., 2004; Marshall et al., 2006) showed that increase in the neocortical slow oscillations during early stages of NREM leads to improvements in declarative memory consolidation (Marshall et al., 2004; Marshall et al., 2006). These ideas were further tested in the computational models predicting that influencing the initiation sites of the cortical slow waves by external stimulation or hippocampal-like input can facilitate sleep replay and strengthen memory trace (Wei et al., 2016; Wei et al., 2018).

It is important to note that a number of sleep disorders was shown to result in deficits in memory consolidation, and that these deficits are correlated with alterations of NREM sleep architecture (Cellini, 2017). Indeed, patients suffering from insomnia or obstructive sleep apnea exhibit fragmented sleep patterns and show either no improvement or worsening scores on word-pair association tasks following a night of sleep (Cellini, 2017; Maski et al., 2017). Narcolepsy is another disorder which alters the sleep architecture and it has been shown to impact the sleep-dependent gain procedural memory consolidation (Mazzetti et al., 2012; Mazzetti et al., 2016; Cellini, 2017). Similar to our previous work (Wei et al., 2016; Wei et al., 2018), here we found that slow-wave sleep was beneficial for the consolidation of multiple memory sequences in the thalamocortical network model.

The interaction between spike-timing dependent plasticity (STDP) and sleep has been previously explored in both theoretical and experimental studies (Wei et al., 2016; Timofeev and Chauvette, 2017; Wei et al., 2018). In the model presented here, the strength of synaptic weights was regulated by a form of STDP. STDP relies on the relative timing between pre- and postsynaptic spikes in order to make changes to synaptic strengths. Due to the traveling wave nature of the slow oscillations during deep slow-wave sleep, cortical Up states are ideal for implementing cortical replay as they will result in sequential activation of neurons thereby initiating STDP-dependent synaptic changes (Wei et al., 2016). It is important to note that we implemented symmetric STDP both during awake and sleep and the influence of changes in neuromodulators during sleep on the STDP curve have not been explored here.

To summarize, our model predicts that slow-wave sleep could help preventing catastrophic forgetting and reverse effects of memory interference through replay of the old and new memory traces. By selectively replaying new and competing old memories during Up states

of slow oscillation, deep sleep not only allows consolidation of the new memory but also provides a mechanism for reorganizing synaptic connectivity encoding the old traces to maximize separation between memories. Thus, sleep may tend to orthogonalize the vectors of synaptic weights to allow successful pattern completion for multiple interfering memories and, therefore, to enable continual learning in the biological systems.

Acknowledgements: This work was supported by DARPA (HR0011-18-2-0021), ONR (N00014-16-1-2829-P00005)

Conflicts of Interest: None.

Figure Legends:

Figure 1. Network architecture and baseline behavior. A, Network schematic showing the basic network architecture (PY: excitatory pyramidal neurons; IN: inhibitory interneurons; TC: excitatory thalamocortical neurons; RE: inhibitory reticular neurons). Excitatory synapses are represented by lines terminating in a dot, while inhibitory synapses are represented by lines terminating in horizontal bars. **B,** Behavior of a control network exhibiting wake-sleep transitions. As a control network, no memories were trained. Color represented the voltage of a neuron at a given time during the simulation. **C,** Zoom-in of a subset of neurons in the network in B (time indicated by arrows). Left and right panels show spontaneous activity during awake-state before and after sleep, respectively. Middle panel shows example slow wave during sleep. **D,** Left panel shows adjacency matrix for the network in B. Black spots represent synaptic connections, while white represents a lack of synaptic connection between the source and target neuron. Right panel shows the initial synaptic weight matrix for the network in B. The x-axis indicates the post-synaptic neuron onto which the connection is made, while the y-axis indicates the relative index of the presynaptic neuron. In this 1D model, the sign of the presynaptic relative index (negative or positive) corresponds to neurons to the left or right of the post-synaptic neuron, respectively. The color in this plot represented the strength of the AMPA connection between neurons, with white indicating lack of synaptic connections.

Figure 2. Two spatially separated memories can be consolidated and strengthened by sleep. A, Network activity showing testing and training of two spatially separated memories and wake-

sleep transitions. Color indicates voltage of neurons at a given time. **B,** Left panels show an example of training of sequence 1 (top) and sequence 2 (bottom). Middle panels show examples of testing of both sequences prior to sleep. Right panels show examples of testing after sleep. Note, after sleep, the sequences show better sequence completion. **C,** Performance of sequence 1 and 2 before any training (baseline), after sequence 1 training (after S1), after sequence 2 training (after S2), and after sleep (red). **D,** Synaptic weight matrices showing changes in synaptic weight strengths in the regions trained for sequence 1 and 2. Top panel shows weights after training sequence 1; middle panel shows weights after training sequence 2; bottom shows weights after sleep. Color indicates strength of AMPA synaptic connections.

Figure 3. Addition of overlapping and competing memory results in catastrophic interference during awake state. A, Network activity showing training and testing periods for three memories during awake-state. Note, sequence 1 and 1* are trained over the same group of neurons. Color indicates the voltage of the neurons at a given time. **B,** Examples of sequence training protocol for S1 (left), S2 (middle), and S1* (right). **C**, Performances showing that training of a S1* leads to a reduction of S1 performance. **D**, Performance of S1 after training of S1* (black) and performance of S1* (red) as function of S1* training duration.

Figure 4. Interleaved training of the "old" and "new" memory recovers "old" memory and consolidates both competing memories. A, Network activity showing training of sequences (blue bars) and interleaved training (green bar) of S1 and S1*. **B,** Example of stimulation protocol used for interleaved training of S1 and S1*. **C,** Performance of S1, S2, and S1* showing increase in performance after interleaved training (green). **D,** Weight matrices showing training and interleaved-dependent changes.

Figure 5. Sleep recovers the "old" memory and consolidates all memories. A, Network activity showing training of sequences (blue bars) and N3 sleep (red bar). **B,** Examples of testing periods for each trained memory at different times during the simulation. The top row corresponds to testing of sequence 1 (S1), middle is testing of sequence 2 (S2), and bottom is testing of Sequence 1* (S1*). **C,** Performance of S1, S2, and S1* showing damage of S1 after training S1*, and increased performance in all sequences after sleep (red bar). **D,** Cumulative sum of full replays during sleep. Solid lines indicate means and broken lines are SEM.

Figure 6. Sleep leads to the strengthening of synaptic weights associated with both the "old" and "new" memories. A, Histograms of bidirectional synaptic weights for neurons contributing to learning of the two memory sequences S1 and S1*. Top row shows synaptic weights contributing to S1 only. Bottom row shows weights contributing to S1* only. Blue shows the starting points for weights, and red shows new weights. **B,** Scatter plots showing bidirectional synaptic weight for neurons contributing to both S1 and S1*. **C,** Weight matrices showing training dependent changes in the region of the network containing S1 and S1*.

Figure 7. Sleep and Interleaved training lead to separation of synaptic weights which preferentially contribute to either the "old" or "new" memory. A/C, Mean projection of recurrent synaptic weights. Shifts in the distribution to the left (right) signify preference of weights for S1 (S1*). Effects of either sleep (A) or interleaved training (C) are shown in the right most panel. **B/D,** Synaptic temporal averaged behavior showing movement of synaptic weights throughout S1 and S1* training and sleep (B) / interleaved training (D).

Figure 8. Cortical connectivity profile reorganization during N3 sleep and interleaved training. Panels on the left correspond to sleep condition, while panels on the right correspond to interleaved training condition. **A,** Relative densities of recurrent and unidirectional connections along S1 and S1*. **B,** The number of maximal strongly connected components (MSCC) in unweighted directed graphs obtained by thresholding the connectivity of the trained region in the cortex at every second of the simulation with thresholds 0.06 (blue), 0.08 (red), and 0.14 (orange). **C,** Neuronal MSCC membership in network color-coded in time through the entire duration of the simulations. **D,** Zoom-in of neuronal MSCC membership showing a drastic decrease in the number of MSCCs.

References

Blokland A (1995) Acetylcholine: a neurotransmitter for learning and memory? Brain Res Brain Res Rev 21:285-300.

Borragan G, Urbain C, Schmitz R, Mary A, Peigneux P (2015) Sleep and memory consolidation: motor performance and proactive interference effects in sequence learning. Brain Cogn 95:54-61.

Cellini N (2017) Memory consolidation in sleep disorders. Sleep Med Rev 35:101-112.

Clemens Z, Fabo D, Halasz P (2005) Overnight verbal memory retention correlates with the number of sleep spindles. Neuroscience 132:529-535.

Feld GB, Born J (2017) Sculpting memory during sleep: concurrent consolidation and forgetting. Curr Opin Neurobiol 44:20-27.

French RM (1999) Catastrophic forgetting in connectionist networks. Trends Cogn Sci 3:128-135.

Hassabis D, Kumaran D, Summerfield C, Botvinick M (2017) Neuroscience-Inspired Artificial Intelligence. Neuron 95:245-258.

Hasselmo ME (2017) Avoiding Catastrophic Forgetting. Trends Cogn Sci 21:407-408.

Kemker R, Kanan C (2017) FearNet: Brain-Inspired Model for Incremental Learning. In: arXiv e-prints.

Kemker R, McClure M, Abitino A, Hayes T, Kanan C (2017) Measuring Catastrophic Forgetting in Neural Networks. In: arXiv e-prints.

Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, Milan K, Quan J, Ramalho T, Grabska-Barwinska A, Hassabis D, Clopath C, Kumaran D, Hadsell R (2017) Overcoming catastrophic forgetting in neural networks. Proc Natl Acad Sci U S A 114:3521-3526.

Krishnan GP, Chauvette S, Shamie I, Soltani S, Timofeev I, Cash SS, Halgren E, Bazhenov M (2016) Cellular and neurochemical basis of sleep stages in the thalamocortical network. Elife 5.

Ladenbauer J, Ladenbauer J, Kulzow N, de Boor R, Avramova E, Grittner U, Floel A (2017) Promoting Sleep Oscillations and Their Functional Coupling by Transcranial Stimulation Enhances Memory Consolidation in Mild Cognitive Impairment. J Neurosci 37:7111-7124.

Marshall L, Molle M, Hallschmid M, Born J (2004) Transcranial direct current stimulation during sleep improves declarative memory. J Neurosci 24:9985-9992.

Marshall L, Helgadottir H, Molle M, Born J (2006) Boosting slow oscillations during sleep potentiates memory. Nature 444:610-613.

Maski K, Steinhart E, Holbrook H, Katz ES, Kapur K, Stickgold R (2017) Impaired memory consolidation in children with obstructive sleep disordered breathing. PLoS One 12:e0186915.

Mazzetti M, Bellucci C, Cipolli C, Pizza F, Russo PM, Tuozzi G, Vandi S, Plazzi G (2016) Age-related differences in sleep-dependent consolidation of motor skills in patients with narcolepsy type 1. Sleep Med 24:80-86.

Mazzetti M, Plazzi G, Campi C, Cicchella A, Mattarozzi K, Tuozzi G, Vandi S, Vignatelli L, Cipolli C (2012) Sleep-dependent consolidation of motor skills in patients with narcolepsy-cataplexy. Arch Ital Biol 150:185-193.

McClelland JL, McNaughton BL, O'Reilly RC (1995) Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev 102:419-457.

Mccloskey M, Cohen NJ (1989) CATASTROPHIC INTERFERENCE IN CONNECTIONIST NETWORKS: THE SEQUENTIAL LEARNING PROBLEM. The Psychology of Learning and Motivation 24:109-165.

Mednick SC, McDevitt EA, Walsh JK, Wamsley E, Paulus M, Kanady JC, Drummond SP (2013) The critical role of sleep spindles in hippocampal-dependent memory: a pharmacology study. J Neurosci 33:4494-4504.

Merhav M, Karni A, Gilboa A (2014) Neocortical catastrophic interference in healthy and amnesic adults: a paradoxical matter of time. Hippocampus 24:1653-1662.

Oudiette D, Paller KA (2013) Upgrading the sleeping brain with targeted memory reactivation. Trends Cogn Sci 17:142-149.

Oudiette D, Antony JW, Creery JD, Paller KA (2013) The role of memory reactivation during wakefulness and sleep in determining which memories endure. J Neurosci 33:6672-6678.

Paller KA, Voss JL (2004) Memory reactivation and consolidation during sleep. Learn Mem 11:664-670.

Papalambros NA, Santostasi G, Malkani RG, Braun R, Weintraub S, Paller KA, Zee PC (2017) Acoustic Enhancement of Sleep Slow Oscillations and Concomitant Memory Improvement in Older Adults. Front Hum Neurosci 11:109.

Rasch B, Born J (2013) About sleep's role in memory. Physiol Rev 93:681-766.

Ratcliff R (1990) Connectionist models of recognition memory: constraints imposed by learning and forgetting functions. Psychol Rev 97:285-308.

Shinoe T, Matsui M, Taketo MM, Manabe T (2005) Modulation of synaptic plasticity by physiological activation of M1 muscarinic acetylcholine receptors in the mouse hippocampus. J Neurosci 25:11194-11200.

Stickgold R (2013) Parsing the role of sleep in memory processing. Curr Opin Neurobiol 23:847-853.

Sugisaki E, Fukushima Y, Fujii S, Yamazaki Y, Aihara T (2016) The effect of coactivation of muscarinic and nicotinic acetylcholine receptors on LTD in the hippocampal CA1 network. Brain Res 1649:44-52.

Timofeev I, Chauvette S (2017) Sleep slow oscillation and plasticity. Curr Opin Neurobiol 44:116-126.

Walker MP, Stickgold R (2004) Sleep-dependent learning and memory consolidation. Neuron 44:121-133.

Wei Y, Krishnan GP, Bazhenov M (2016) Synaptic Mechanisms of Memory Consolidation during Sleep Slow Oscillations. J Neurosci 36:4231-4247.

Wei Y, Krishnan GP, Komarov M, Bazhenov M (2018) Differential roles of sleep spindles and sleep slow oscillations in memory consolidation. PLoS Comput Biol 14:e1006322.

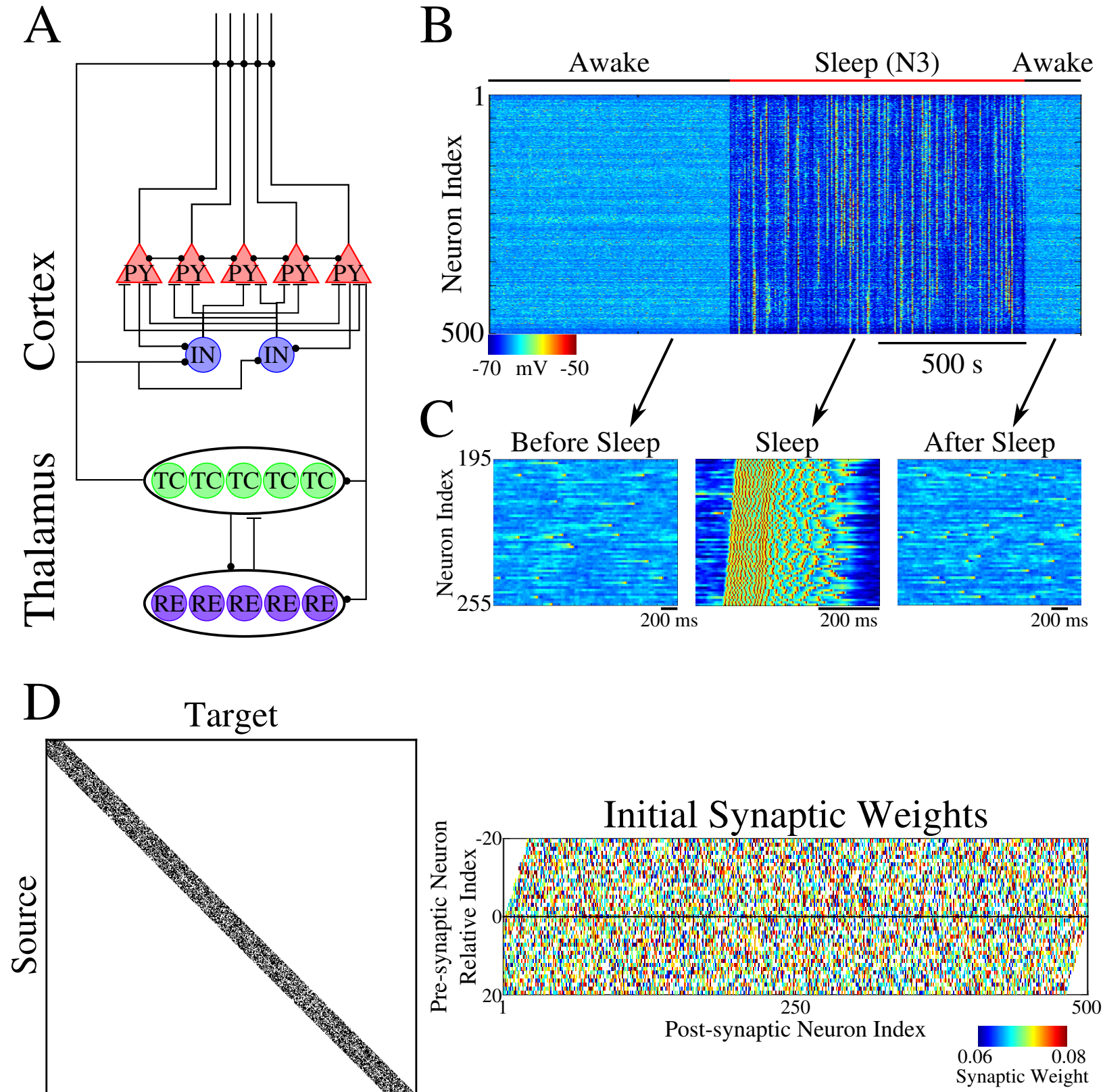


Figure 1.

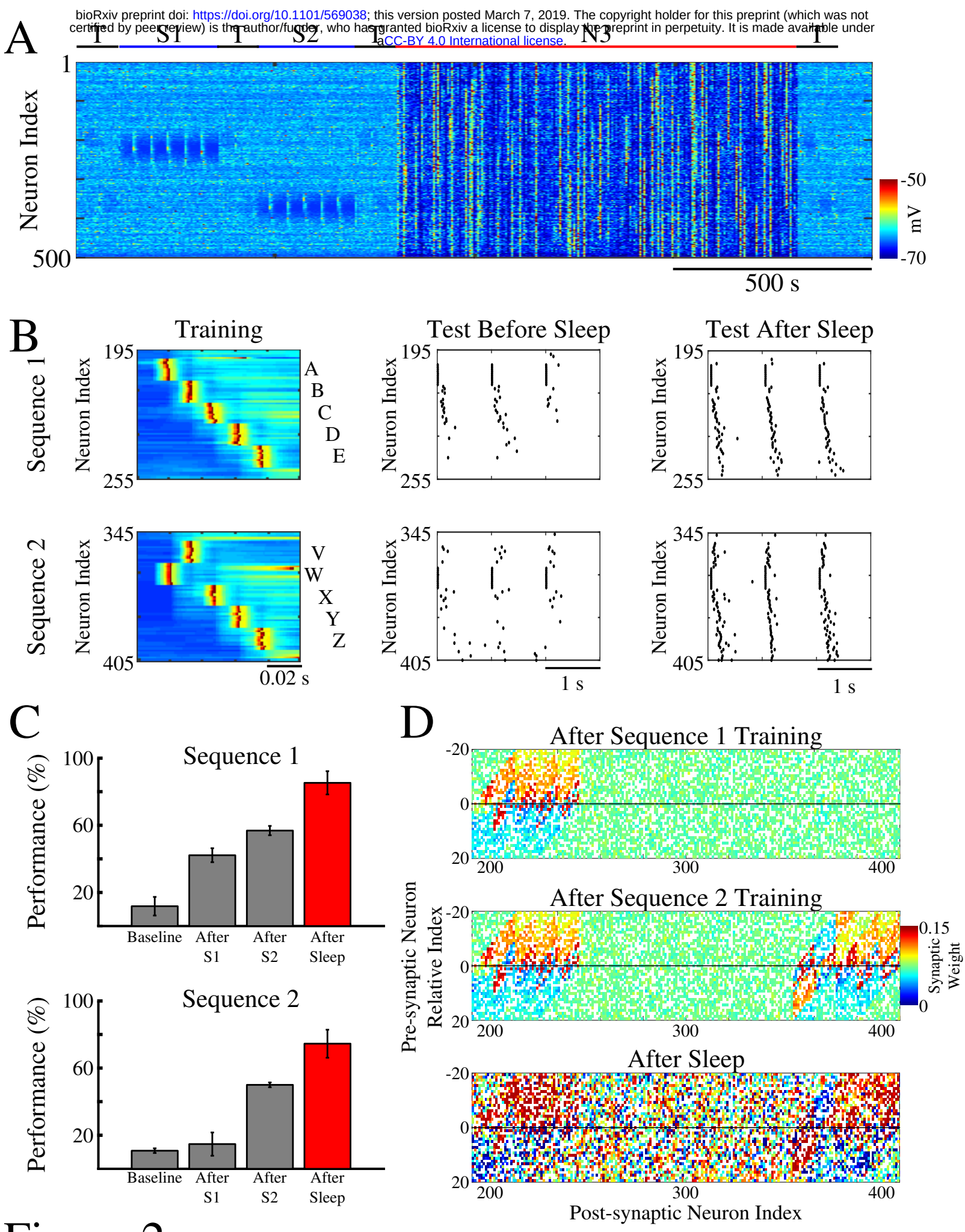
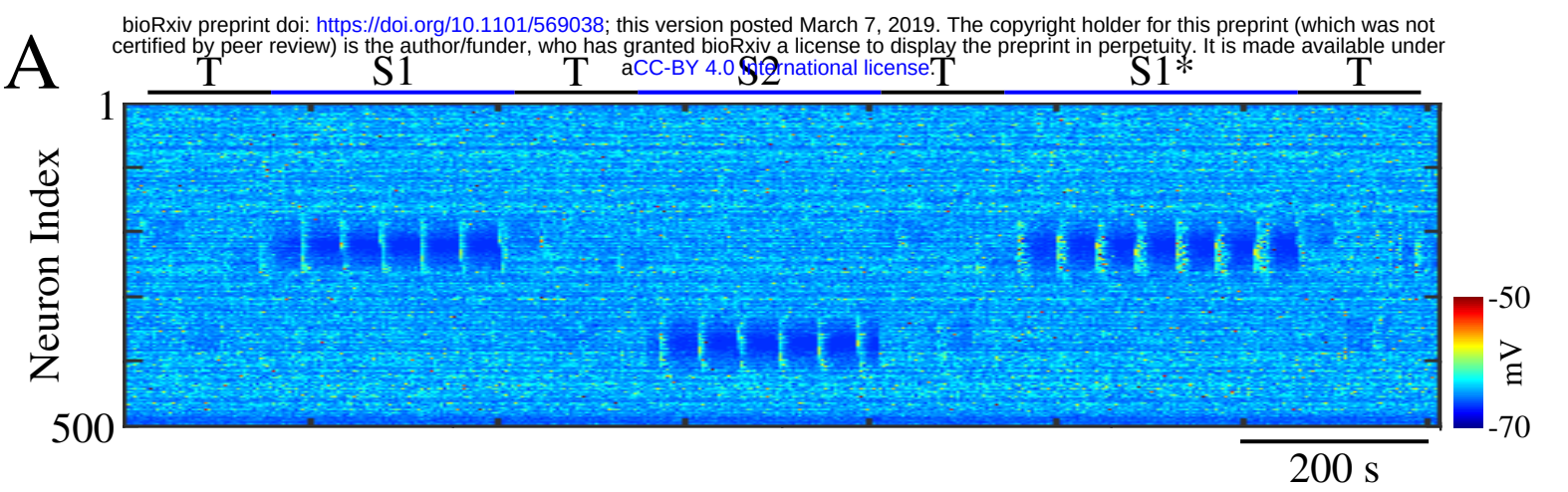
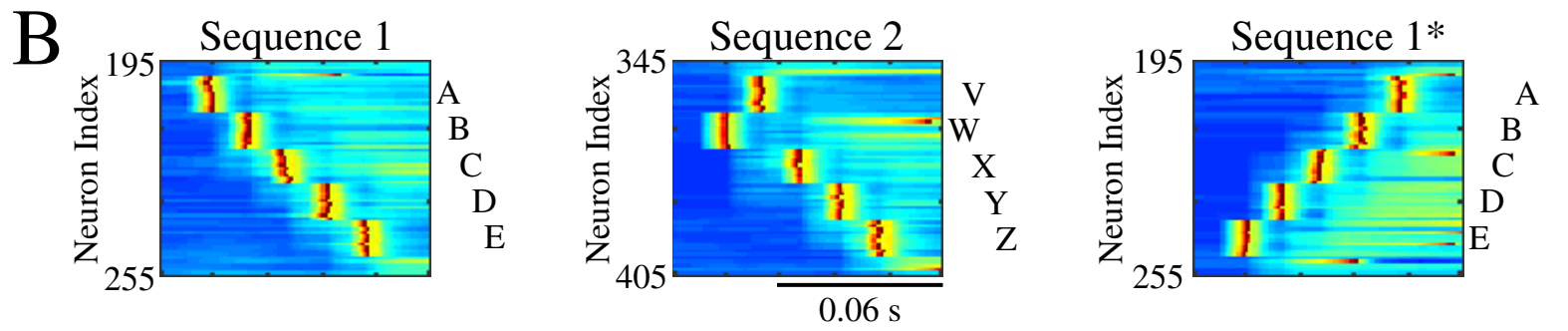


Figure 2.





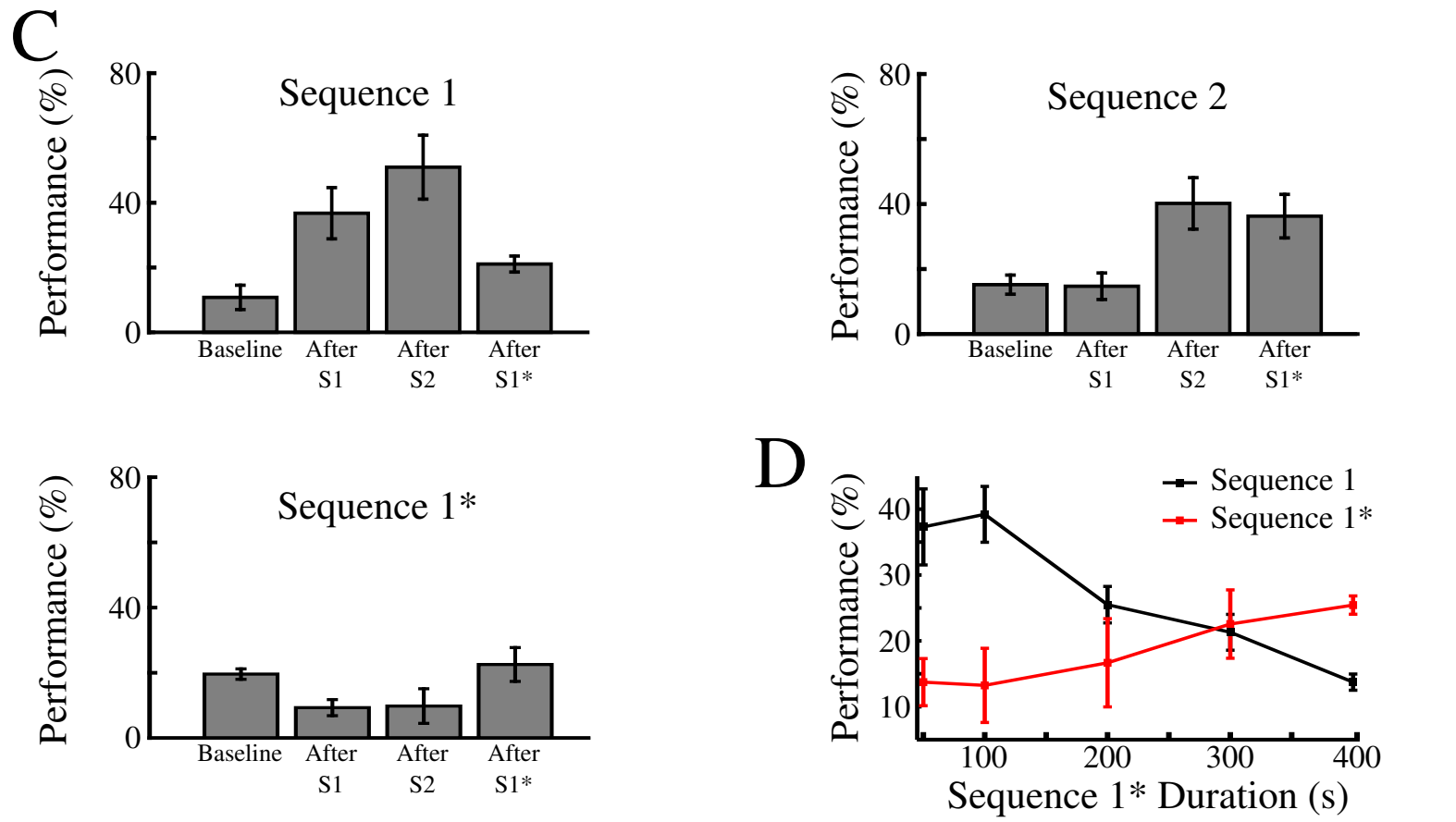


Figure 3.

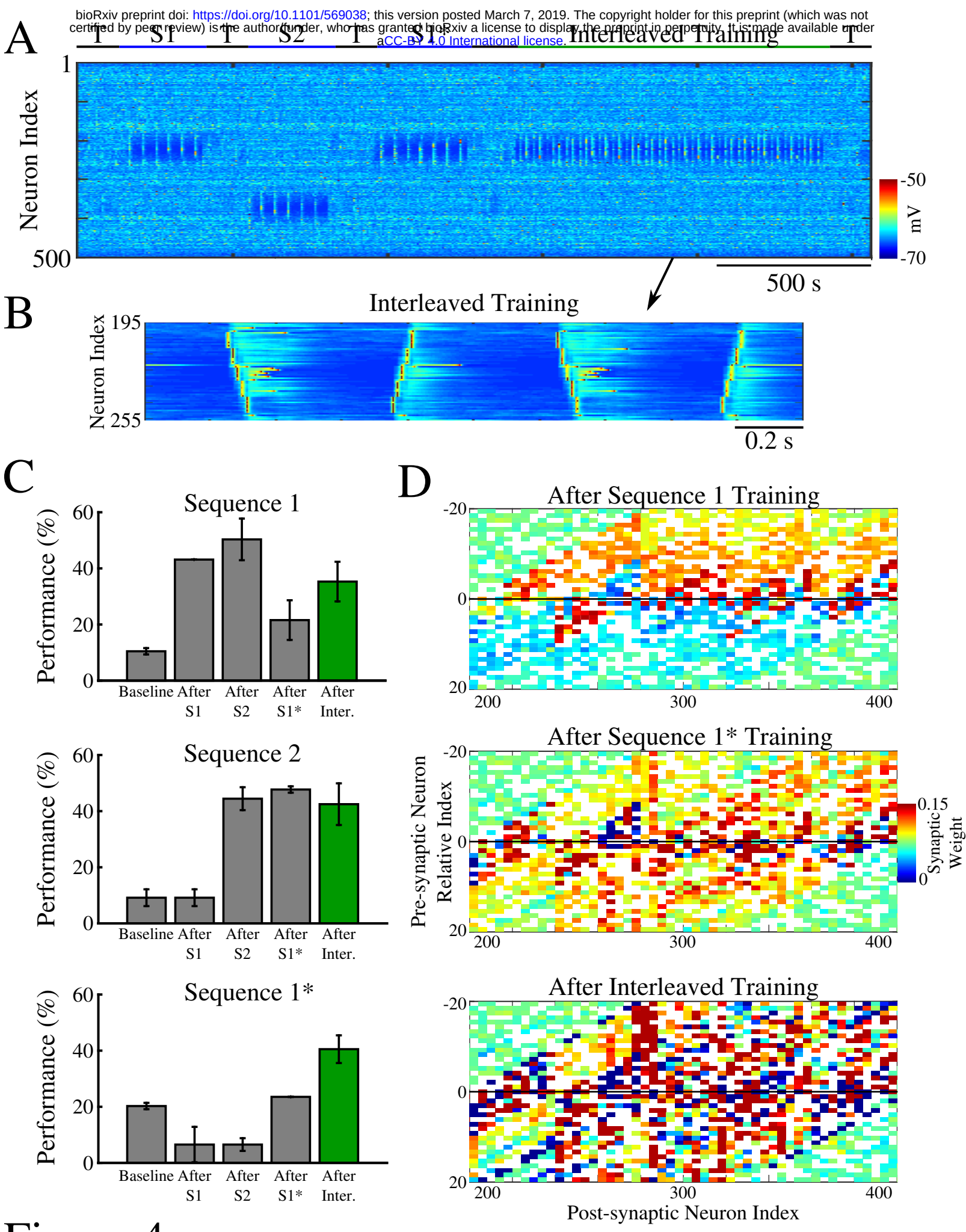


Figure 4.

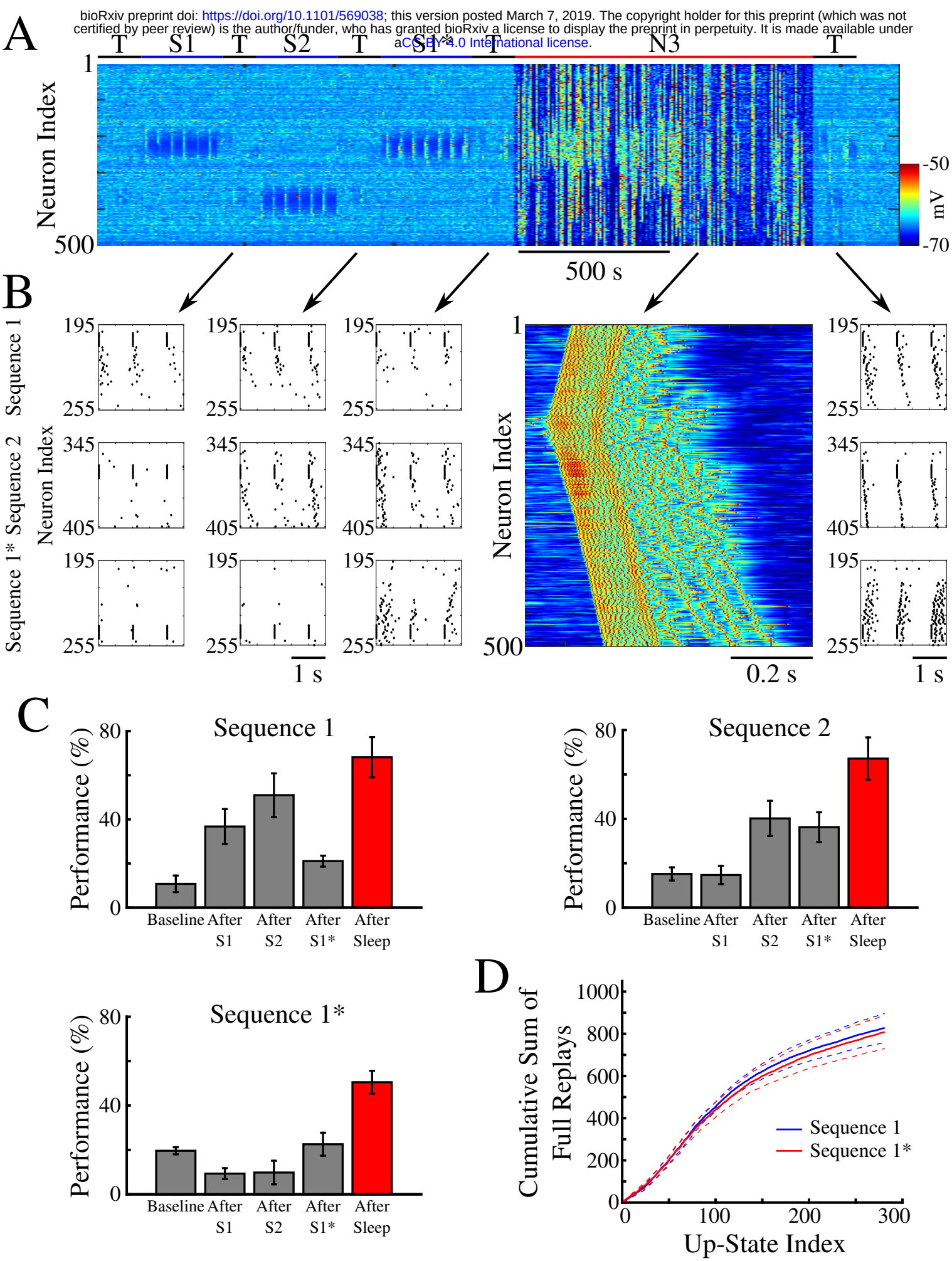


Figure 5.

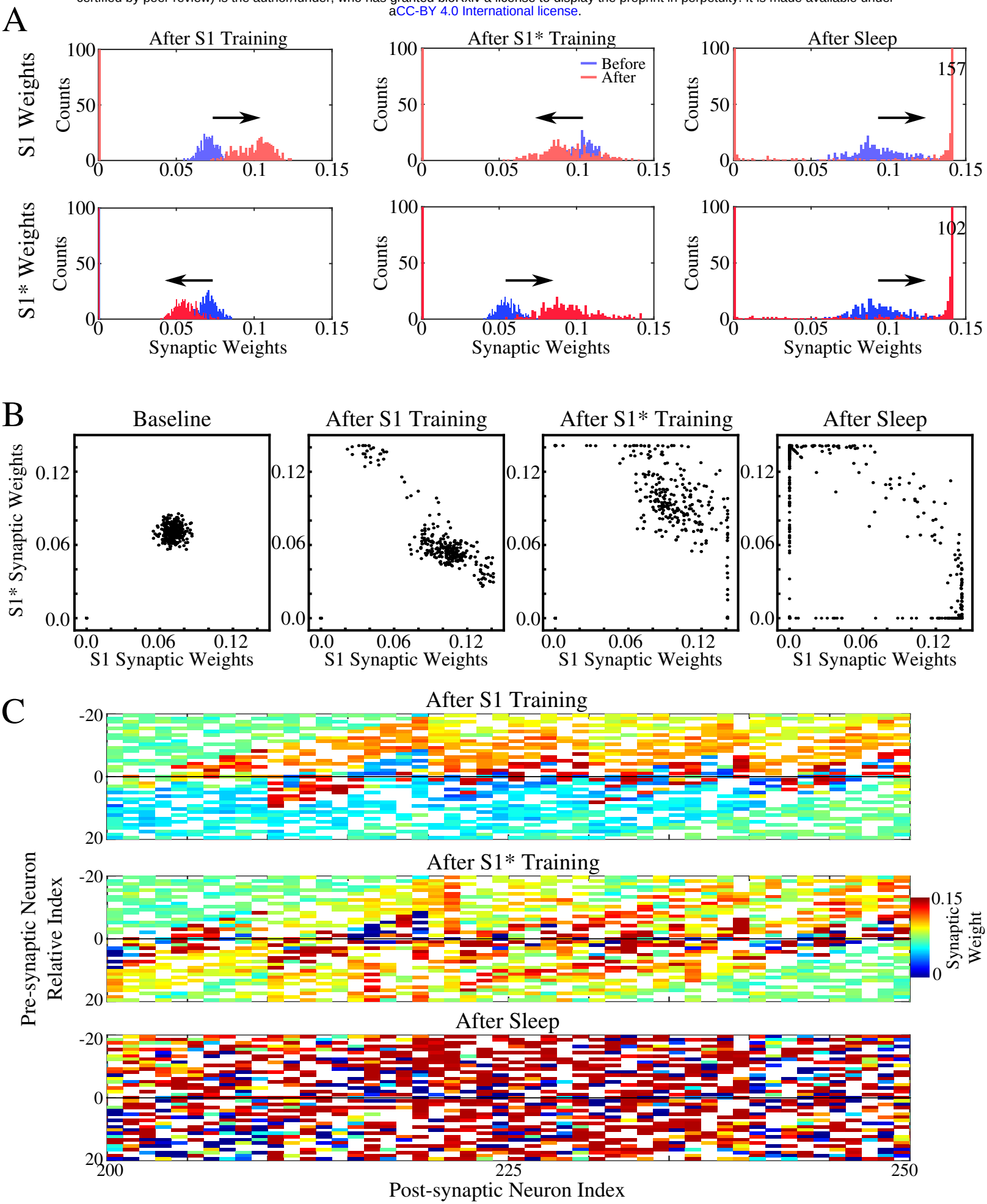
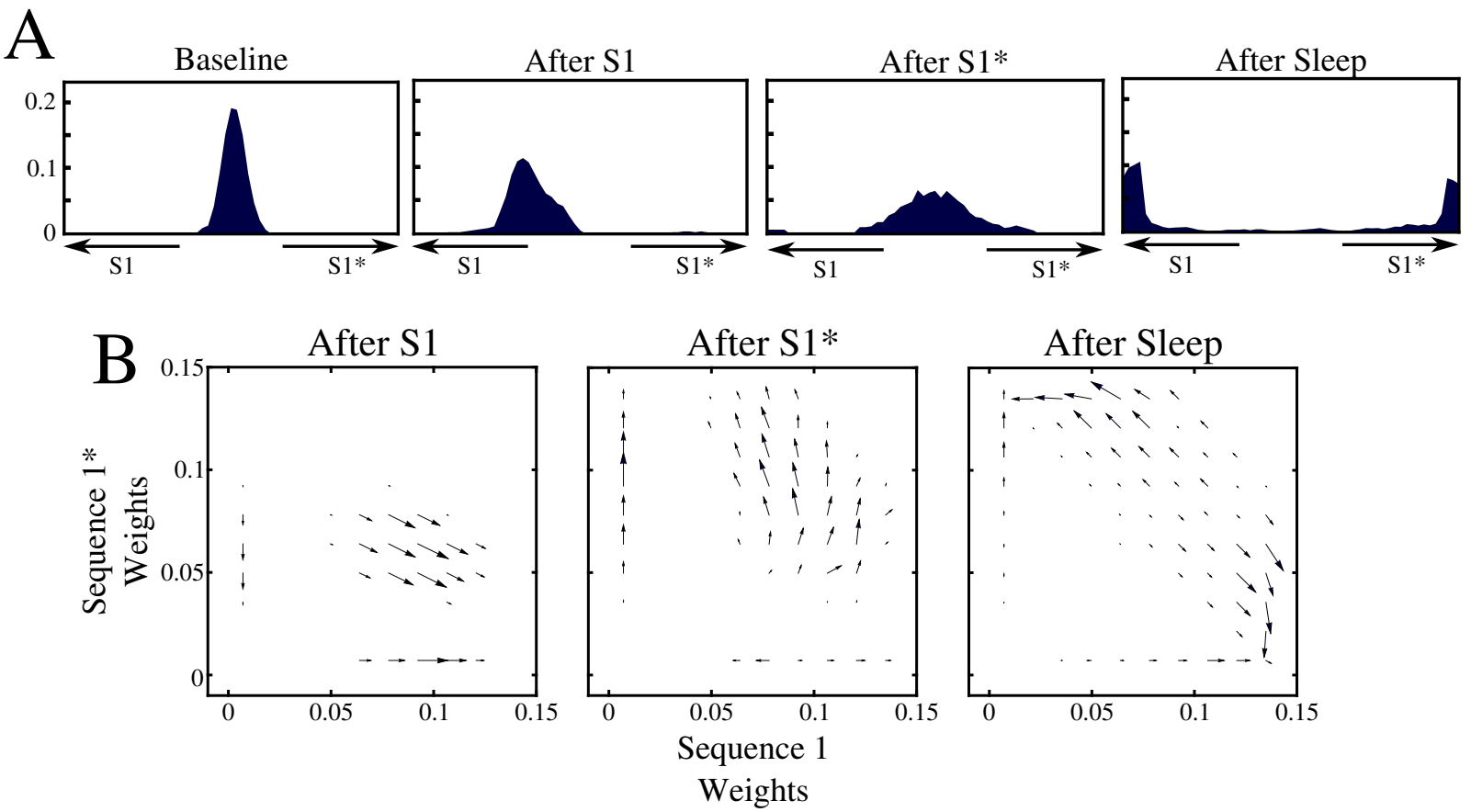


Figure 6.



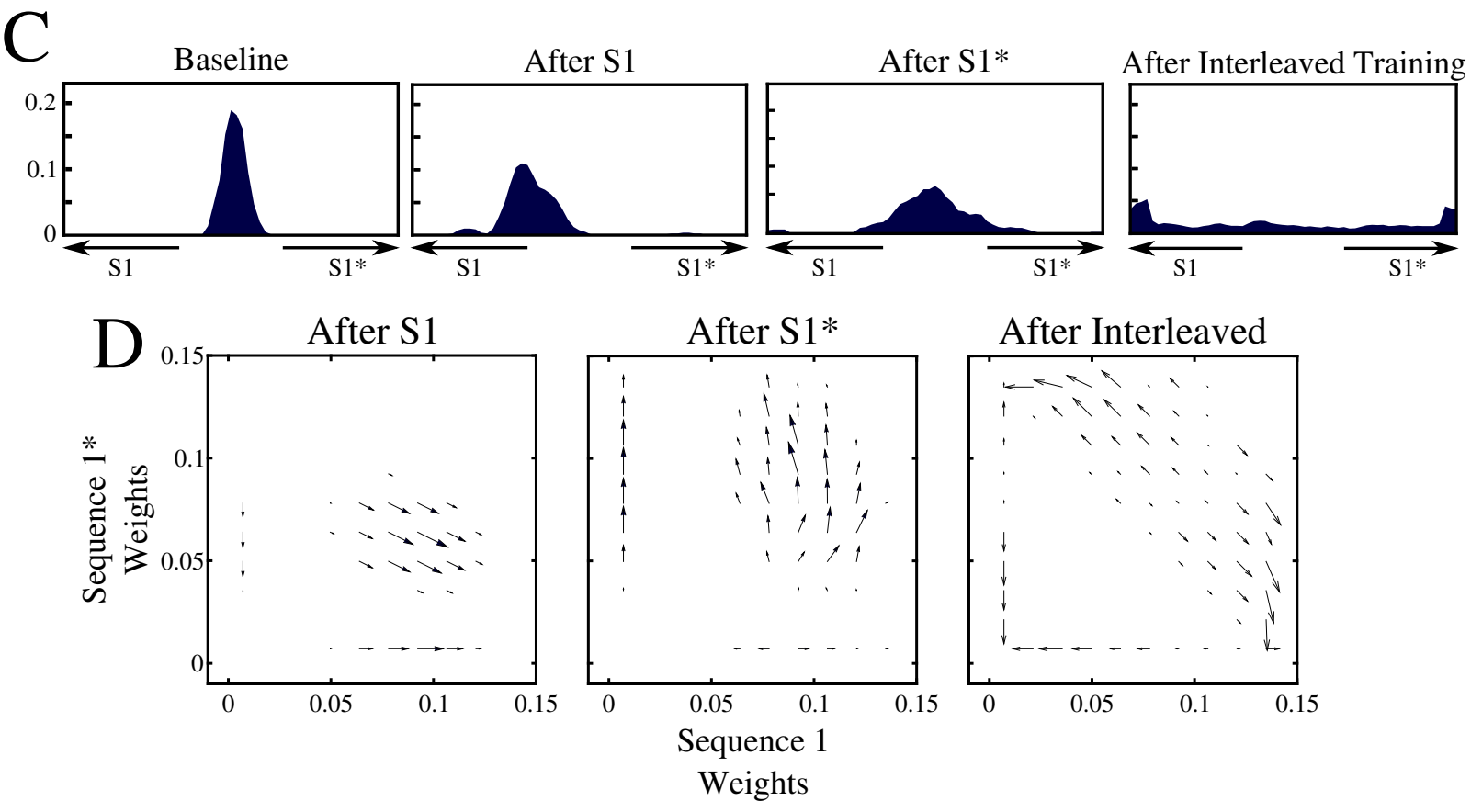


Figure 7.

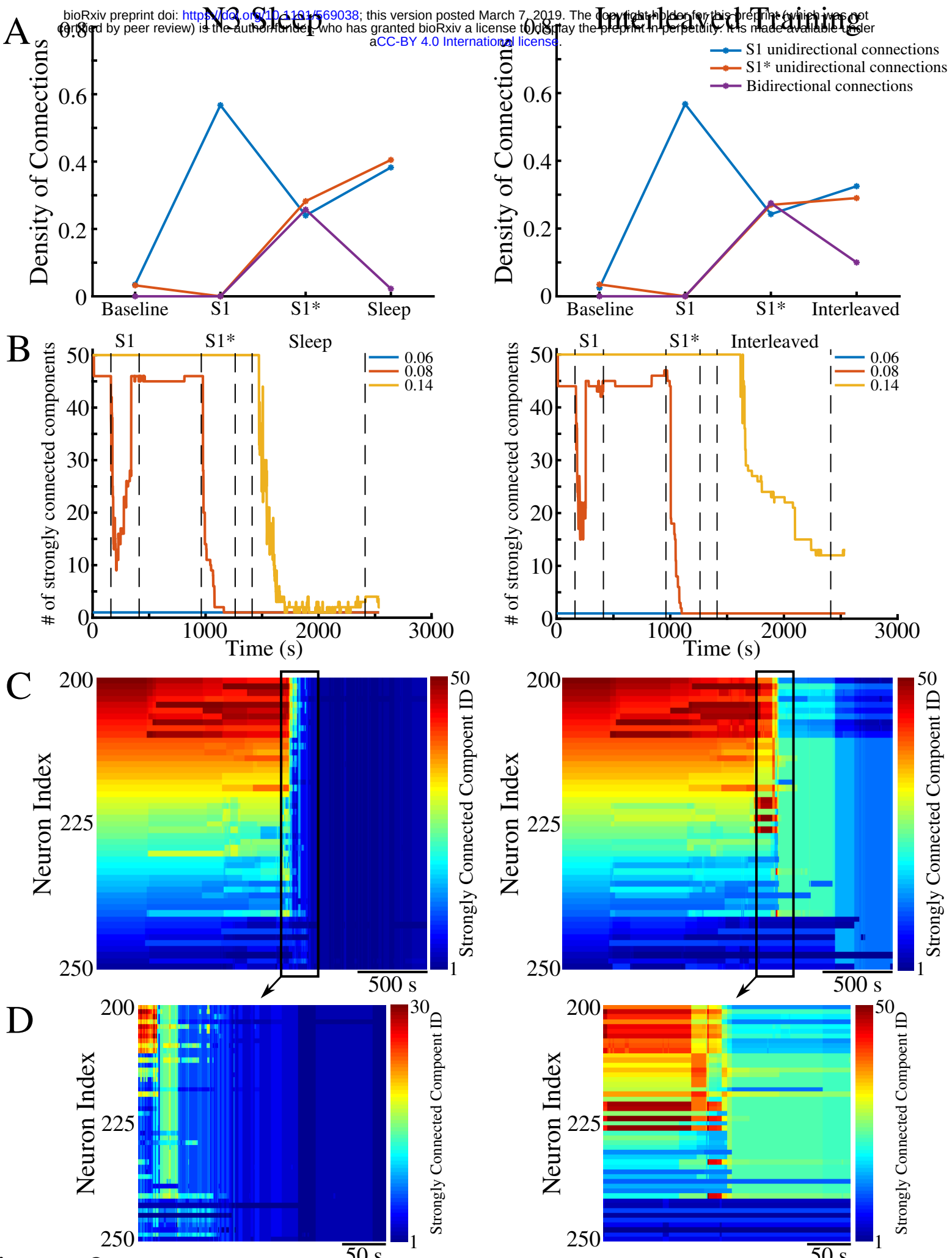


Figure 8.