The Importance of Forgetting:
Recovering Topology from Spike Train Data with Zig Zag Persistence

Brantley Vose

CSE 5339: Week 5

We have learned a lot of theory...

- Neuroscience introduction
- Persistent homology
- Zig-zag and multiparameter persistence

It's time to focus in on an application!

Review

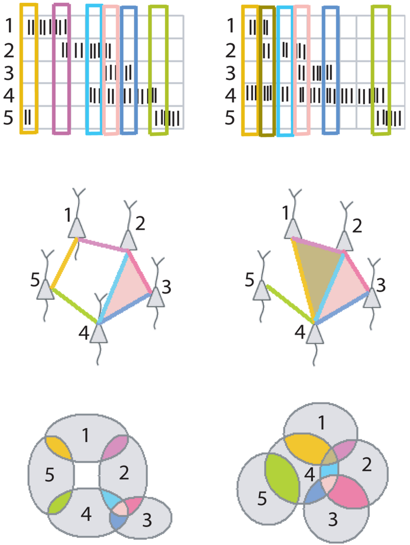
The Space Reconstruction Experiment

Given spike train data from the hippocampus, can we reconstruct geometric or topological aspects of the arena?

First Strategy: Nerve Complex and Persistent Homology

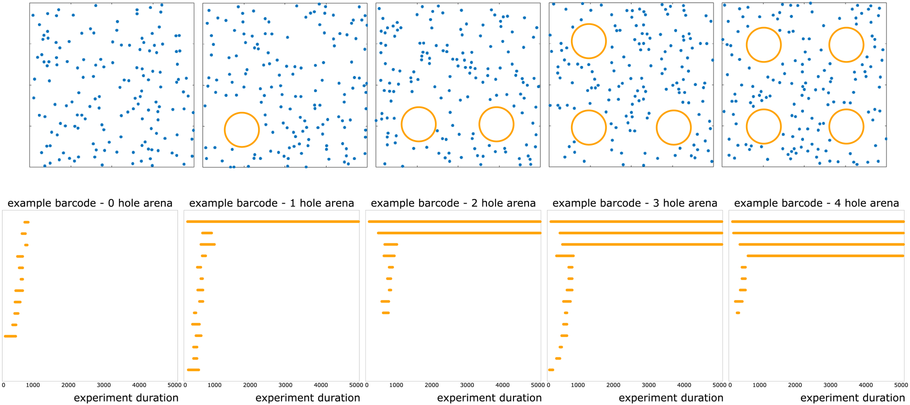
Previous strategy from [CI08]:

1. Collect spike train data from place cell ensemble as mouse explores.
2. Approximate the nerve complex by watching for cofiring.
3. Persistent homology should reflect homology of the arena.

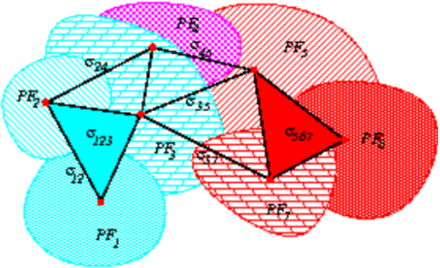


First Strategy: Nerve Complex and Persistent Homology

The hope: the resulting barcodes reflect the topology of the arena.



The Question: Limited Memory

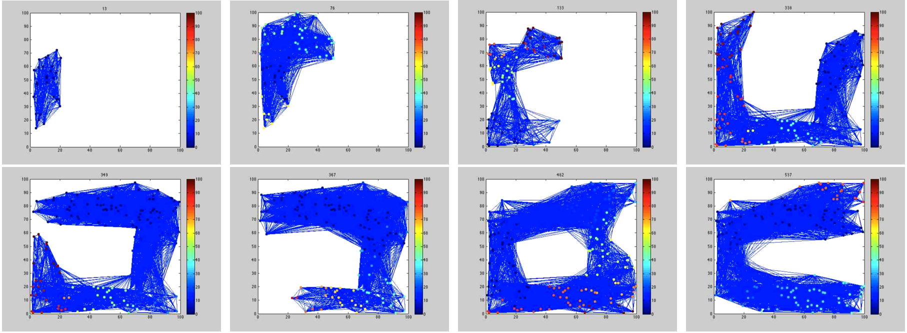


This method uses "perfect memory," i.e. the complex at the end of the experiment includes data gathered at the very beginning.

Question: How does Limited Memory Affect the Reconstruction of the Topology?

One Extreme: Very Short Memory

What would happen if we "forgot" the simplices in our complex after some amount of time?

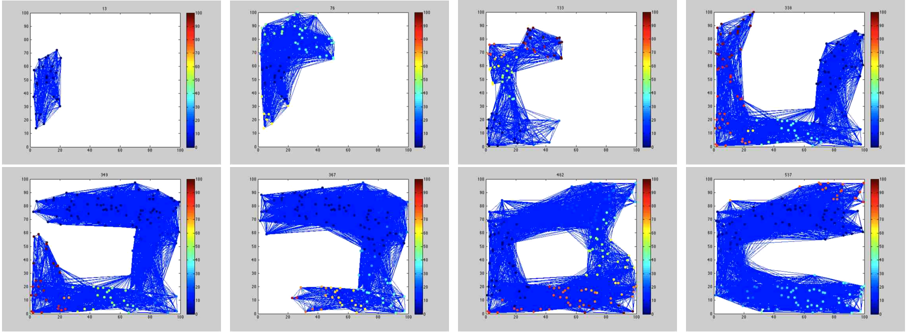


(Not pictured: Obstacle in center of arena, and higher-dimensional simplices.)

If simplices are forgotten too quickly, our complex struggles to capture the topology.

One Extreme: Very Short Memory

What would happen if we "forgot" the simplices in our complex after some amount of time?



(Not pictured: Obstacle in center of arena, and higher-dimensional simplices.)

If simplices are forgotten too quickly, our complex struggles to capture the topology.

Expectation: Longer Memory Should be Better

The Setup

Modeling Limited Memory

Recall the filtered complex with **perfect memory**:

- Fix a window size $\epsilon > 0$ and a threshold $n_0 \in \mathbb{N}$. We say a collection $\{v_0, \ldots, v_k\}$ of cells *cofire* at time t if

$$|s(v_i) \cap [t - \epsilon, t + \epsilon]| \geq n_0$$

  for all $i = 0, 1, \ldots, k$, where $s(v_i)$ is the spike train for place cell i.
- Let V be the collection of observed place fields. Suppose the experiment runs from $t = 0$ to $t = T$. The filtration is given by

$$K_t = \{\sigma \subset V | \text{the cells in } \sigma \text{ have cofired by time } t\}$$

  for integers $t \in [0, T]$

Modeling Limited Memory

What about with **limited memory**? The authors of [CDM18] modify the above model by adding a memory length parameter.

Definition

Fix a memory length $\tau > 0$. The *synaptic potentiation complex with memory length* τ (or simply the *potentiation complex*) is given by

$$K_t^\tau = \{\sigma \subset V | \text{the cells in } \sigma \text{ have cofired at some time in } [t-\tau, t]\}$$

for integers $t \in [0, T]$.

Modeling Limited Memory

What about with **limited memory**? The authors of [CDM18] modify the above model by adding a memory length parameter.

Definition

Fix a memory length $\tau > 0$. The *synaptic potentiation complex with memory length* τ (or simply the *potentiation complex*) is given by

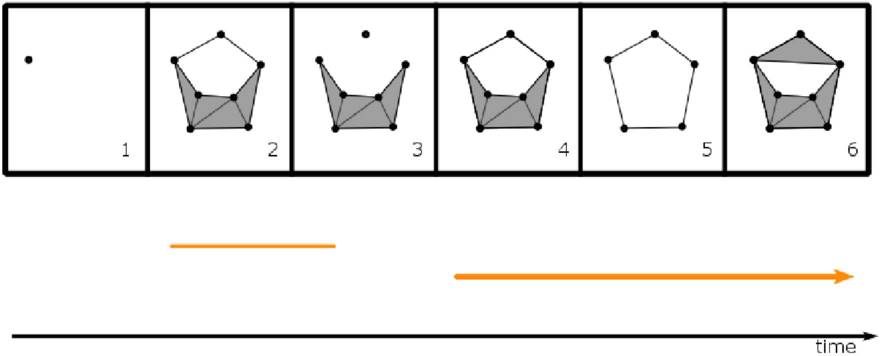
$$K_t^\tau = \{\sigma \subset V | \text{the cells in } \sigma \text{ have cofired at some time in } [t-\tau, t]\}$$

for integers $t \in [0, T]$.

Problem: This is often **not a filtration**! There is no guarantee that $K_t^\tau \subseteq K_{t+1}^\tau$, since some simplices might be **forgotten** between time t and $t+1$.

Using Zig-Zag Persistence

Solution: **Zig-zag persistence**!
The potentiation complex K_t^τ may not be a filtration, but it is a zig-zag filtration! We can still get barcode with zig-zag persistence.



The Experiment

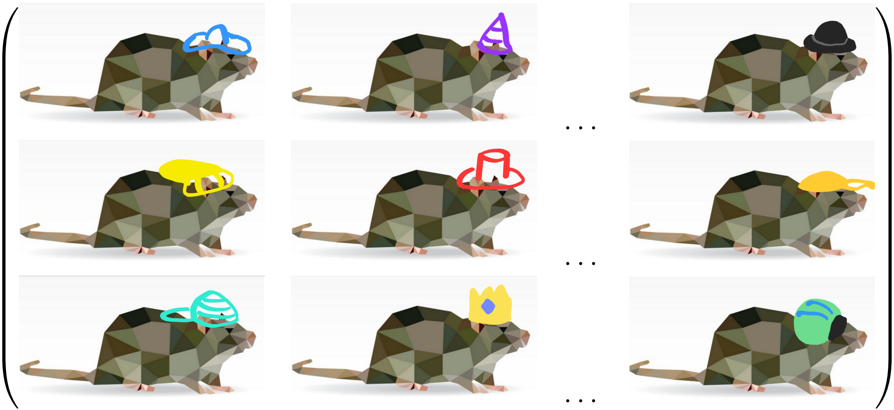
The Experiment

The experiment has three parts.

1. Generate a datasets by simulating mice in arenas.
2. Try to recover the topology of the virtual arenas with potentiation complexes and zig-zag persistence. Try many values of τ.
3. Analyze the results. Do the barcodes tell us which arena our simulated rats were in? **How does limited memory affect the reconstruction of arena topology?**

Step 1: Generating the Dataset (The Mice)

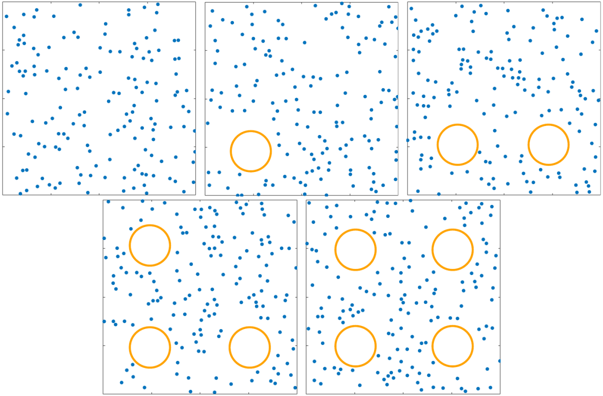
Meet our virtual test subjects:



15 combinations of activated firing rate (12, 14, 16, 18, or 20Hz) and place field radius (14, 15, or 16cm).

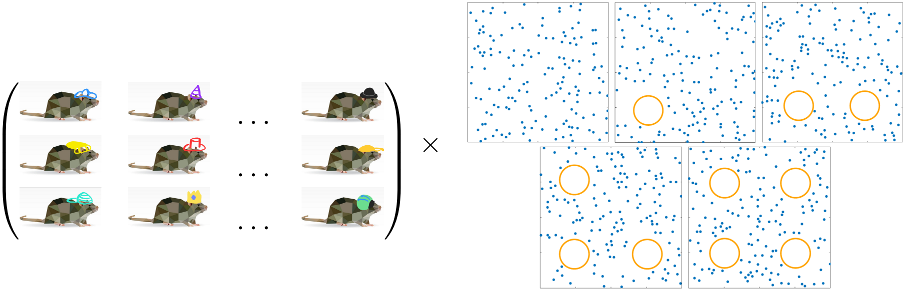
Step 1: Generating the Dataset (The Arenas)

Our virtual laboratory:



5 arenas with different homotopy types.

Step 1: Generating the Dataset



Run each mouse in each arena 10 times and collect simulated spike train data. Total spike train datasets: $15 \times 5 \times 10 = 750$.

Step 2: Reconstruct Topology of Arena with Limited Memory

- Discretize spike trains into 5000 time bins.
- For each memory length $\tau \in [50, 5000]$ in increments of 50, construct potentiation complex $(K_t^\tau)_{t=1}^{5000}$.
- Compute barcodes using zig-zag persistence.

(750 spike trains) $\times$ (100 memory lengths) $= 75,000$ barcodes.

Step 2: Reconstruct Topology of Arena with Limited Memory

- Discretize spike trains into 5000 time bins.
- For each memory length $\tau \in [50, 5000]$ in increments of 50, construct potentiation complex $(K_t^\tau)_{t=1}^{5000}$.
- Compute barcodes using zig-zag persistence.

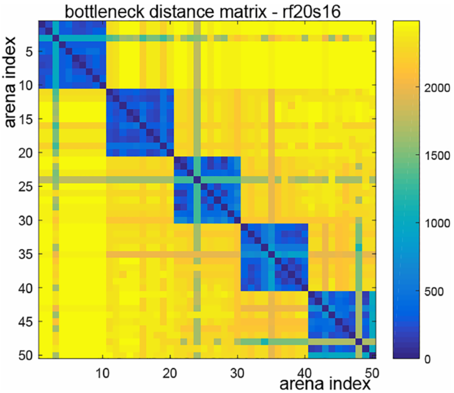
(750 spike trains) $\times$ (100 memory lengths) $= 75,000$ barcodes.

Question: How does Limited Memory Affect the Reconstruction of Topology?

Step 3: Analyzing the Results

Question: How does Limited Memory Affect the Reconstruction of Topology?

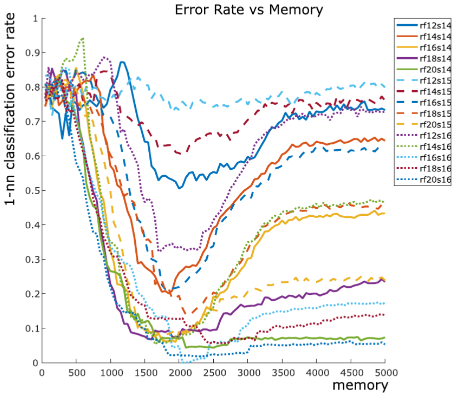
- Pick a mouse and a memory length τ.
- Take the 50 barcodes generated by that mouse at that memory length.
- Calculate the pairwise bottleneck distances between the barcodes.
- Determine the 1-nearest neighbor error rate in predicting the arena.



Step 3: Analyzing the Results

Results:

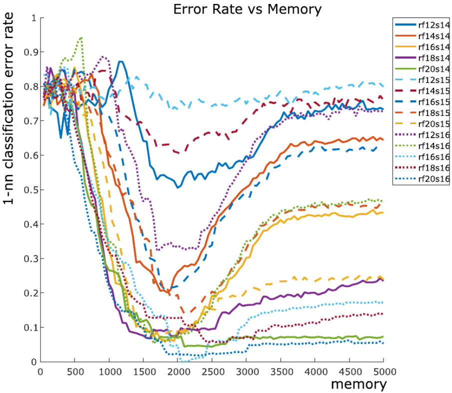
- As predicted, very small τ leads to no topological information.
- Surprisingly, error rate reaches a minimum arount $\tau \approx 2000$, *then rises again*.



Step 3: Analyzing the Results

Results:

- As predicted, very small τ leads to no topological information.
- Surprisingly, error rate reaches a minimum arount $\tau \approx 2000$, *then rises again*.



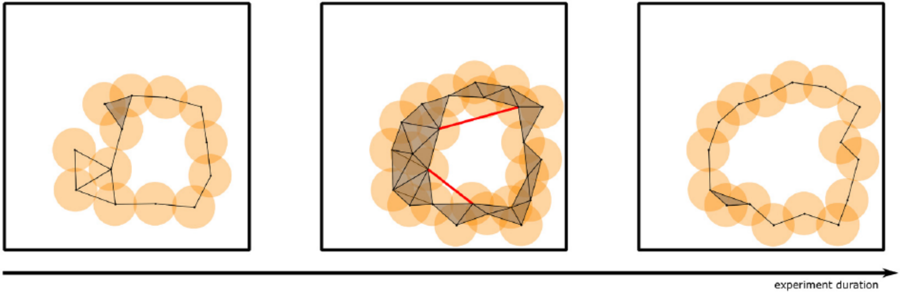
More memory is not always better?

Intermission

Why is More Memory Not Always Better?

We expected more memory to be better, but this is not always the case. Why?

Theory: The problem could be **spurious firing**. Simplices may be added due to noise.



Two ways for mistakes to get fixed:

- Fill in the extra loop with more simplices.
- Forget the mistaken edge.

Testing the Hypothesis

Is spurious firing really the reason more memory is sometimes worse?

As good scientists, we change up the model and see if the effect persists.

Modification 1: Change mouse trajectory model.

- Results so far simulated mouse movement as "billiards".
- Replace with a "random walk" model.

Testing the Hypothesis

Is spurious firing really the reason more memory is sometimes worse?
As good scientists, we change up the model and see if the effect persists.

Modification 1: Change mouse trajectory model.

- Results so far simulated mouse movement as “billiards”.
- Replace with a “random walk” model.

Result: Same effect! Error rate minimized around $\tau \approx 2000$.

Testing the Hypothesis

Modification 2: Remove randomness from firing model.

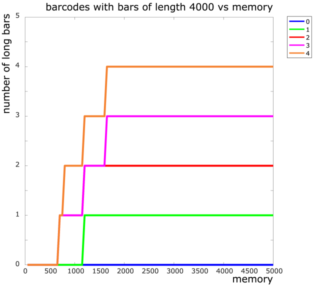
- Firing rate so far modeled with a random Poisson process. A place cell has high firing rate while mouse is inside the place field, low firing rate otherwise.
- Replace with a more deterministic *binary* model. Cell fires while mouse is inside the place field, stops when it leaves.

Testing the Hypothesis

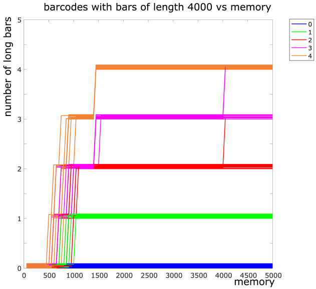
Modification 2: Remove randomness from firing model.

- Firing rate so far modeled with a random Poisson process. A place cell has high firing rate while mouse is inside the place field, low firing rate otherwise.
- Replace with a more deterministic *binary* model. Cell fires while mouse is inside the place field, stops when it leaves.

Result: Effect vanishes!



Deterministic firing model



Random firing model

Testing the Hypothesis

Modification 3: Reintroduce spurious firing.

- Replace firing model with a *fuzzy binary model*.
- Deterministic firing while mouse is in the place field, random firing while outside.

Testing the Hypothesis

Modification 3: Reintroduce spurious firing.

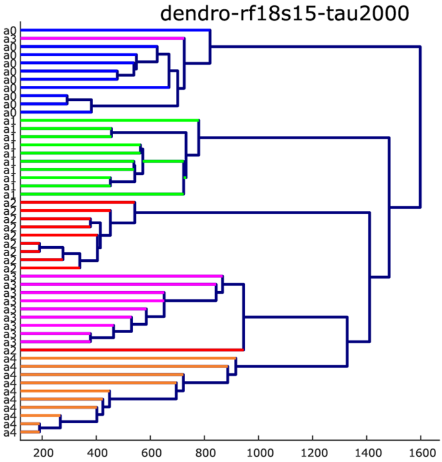
- Replace firing model with a *fuzzy binary model*.
- Deterministic firing while mouse is in the place field, random firing while outside.

Result: Effect reappears!

Conclusion: The error rate "valley" really is caused by spurious firing.

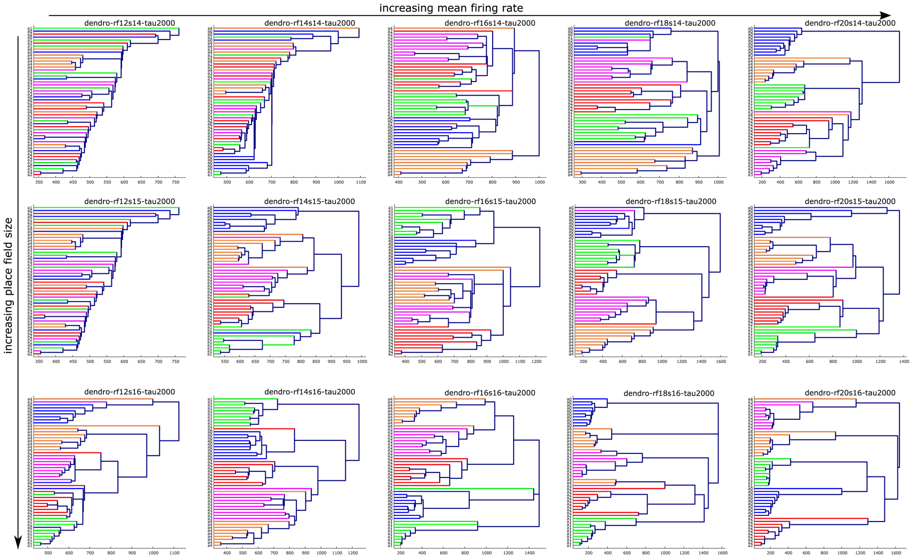
Sanity Checks

Do the barcodes actually form five clusters? Break out the dendrograms.



Sanity Checks

Make one dendrogram for each mouse.



Sanity Checks

Using the single-linkage dendrograms, we can perform statistical tests.

- The barcodes from most mice do fall in 5 statistically significant clusters.
- In all but one case, the clusters do have a statistically significant correspondence with the arenas.
- Clustering is best when firing rate is high and place fields are large.

Table 1. *p*-values for single linkage dendrograms in Fig 8.

		mean firing rate (Hz)				
		12	**14**	**16**	**18**	**20**
place field size (cm)	14	0.200	0.637	0.957	0.011^{*}	0^{*}
	15	0.998	0.956	0^{*}	0^{*}	0^{*}
	16	0.891	$0.001^{\dagger}$	0^{*}	0^{*}	0^{*}

Table 2. 1-nearest neighbor classification error rates at $\tau = 2000$.

		mean firing rate (Hz)				
		12	**14**	**16**	**18**	**20**
place field size (cm)	14	0.729	0.603	0.216	0.192^{*}	0.087^{*}
	15	0.496	0.269	0.090^{*}	0.078^{*}	0.069^{*}
	16	0.340	$0.862^{\dagger}$	0.047^{*}	0.123^{*}	0.022^{*}

Summary

- If we want to model the forgetting of information over time, regular persistent homology is not enough. We need zig-zag persistence.

Summary

- If we want to model the forgetting of information over time, regular persistent homology is not enough. We need zig-zag persistence.
- Zig-zag persistence with an intermediate memory length actually performs better than a perfect memory if place cells can mistakenly fire.

Summary

- If we want to model the forgetting of information over time, regular persistent homology is not enough. We need zig-zag persistence.
- Zig-zag persistence with an intermediate memory length actually performs better than a perfect memory if place cells can mistakenly fire.

Thank You!

Further Reading

Nature article summarizing findings on the importance of forgetting.

MEMORY

The importance of forgetting

Long thought to be a glitch of memory, the ability to forget is actually crucial to how the brain works, researchers are realizing.

BY LAUREN GRAVITZ

Memories make us who we are. They shape our understanding of the world and help us to predict what's coming. For more than a century, researchers have been working to understand how memories are formed and then fixed for recall in the days, weeks or even years that follow. But those scientists might have been looking at only half the picture. To understand how we remember, we must also understand how, and why, we forget.

Until about ten years ago, most researchers thought that forgetting was a passive process in which memories, unused, decay over time like a photograph left in the sunlight. But then a handful of researchers who were investigating memory began to bump up against findings that seemed to contradict that decades-old assumption. They began to put forward the radical idea that the brain is built to forget.

A growing body of work, cultivated in the past decade, suggests that the loss of memories is not a passive process. Rather, forgetting seems to be an active mechanism that is constantly at work in the brain. In some — perhaps even all — animals, the brain's standard state is not to remember, but to forget. And a better understanding of that state could lead to breakthroughs in treatments for conditions such as anxiety, post-traumatic stress disorder (PTSD), and even Alzheimer's disease.

"What is memory without forgetting?" asks Oliver Hardt, a cognitive psychologist studying the neurobiology of memory at McGill University in Montreal, Canada. "It's impossible," he says. "To have proper memory function, you have to have forgetting."

BIOLOGY OF FORGETTING

Different types of memory are created and stored in varying ways, and in various areas of the brain. Researchers are still pinpointing the details, but they know that autobiographical memories — those of events experienced personally — begin to take lasting form in a part of the brain called the hippocampus, in the hours and days that follow the event. Neurons communicate with each other through synapses — junctions between these cells that include a tiny gap across which chemical messengers can be sent. Each neuron can be connected to thousands of others in this way. Through a process known as synaptic plasticity, neurons constantly produce new proteins to remodel parts of the synapse, such as the receptors for these chemicals, which enables the neurons to selectively strengthen their connections with one another. This creates a network of cells that, together, encode a memory. The more often a memory is recalled, the stronger its neural network becomes. Over time, and through consistent recall, the memory becomes encoded in both the hippocampus and the cortex. Eventually, it exists independently in the cortex, where it is put away for long-term storage.

SAM FALCONER

Link on course webpage.

References I

- Samir Chowdhury, Bowen Dai, and Facundo Mémoli, *The importance of forgetting: Limiting memory improves recovery of topological characteristics from neural data*, PloS one **13** (2018), no. 9, e0202561–e0202561 (eng), 30180172[pmid].

- Carina Curto and Vladimir Itskov, *Cell groups reveal structure of stimulus space*, PLOS Computational Biology **4** (2008), no. 10, 1–13.