8. Learning and Memory

Memory

- Memory = any persistent effect of experience (not just memorization of facts, events, names, etc.)
- Weights vs. activations
- Gradual, integrative cortical learning and priming effects
- Rapid memorization: The hippocampus
- Active memory: prefrontal cortex

(slide < Frank)

Major Types of Memory: Mechanisms

- Weight-based (changes in synapses)
  — Long lasting, persist over distraction, etc.
  — Very high capacity
- Activation-based (sustained neural firing)
  — Transient, easily lost
  — Very flexible: mental arithmetic, etc.

(slide < O'Reilly)

Weights vs. Activations

- Despite appearances, memory is not unitary
- Weights:
  - Long-lasting
  - Requires re-activation
  - Weights in different brain systems store different types of memories
- Activations:
  - Short-term
  - Already active, can influence processing

(slide < Frank)

Major Types of Memory: Characteristics

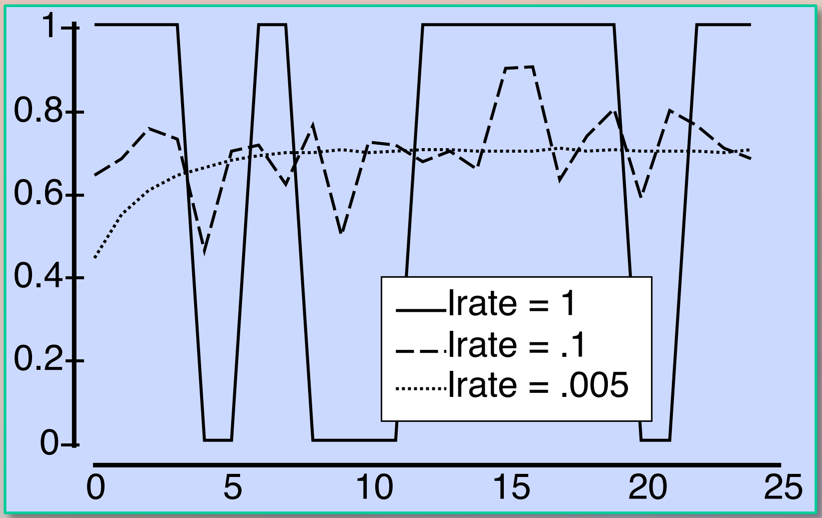
- Episodic Memory: events, facts, etc.
  - — Hippocampus
- Familiarity-based recognition
  - — Perirhinal cortex: “You look familiar, but…”
- Weight-based priming
  - — Subconscious, can be very long-lasting
- Activation-based priming
  - — Also subconscious, but transient…

(slide < O’Reilly)

A. Episodic Memory

Weight-based Memories

- Cortex does gradual, integrative learning
- Cortex can learn arbitrary input-output mappings given: – multiple passes through the training set – a relatively small learning rate
- Rapid weight changes causes interference ☞



- Therefore, two systems needed:
  — slow-learning cortex
  — rapid-learning hippocampus (pattern separation avoids interference)

(slide based on Frank)

Episodic Memory

- Autobiographical memory (life events)
- Arbitrary new memories (lab tasks)
- …

(slide < O'Reilly)

Classic Lab Task: AB-AC

- Humans can rapidly learn overlapping associations without too much interference
- Learn AB paired associates:
  - window-reason
  - bicycle-garbage
  - ...
- Then AC paired associates:
  - window-locomotive
  - bicycle-dishtowel
  - …

(slide based on Frank & O'Reilly)

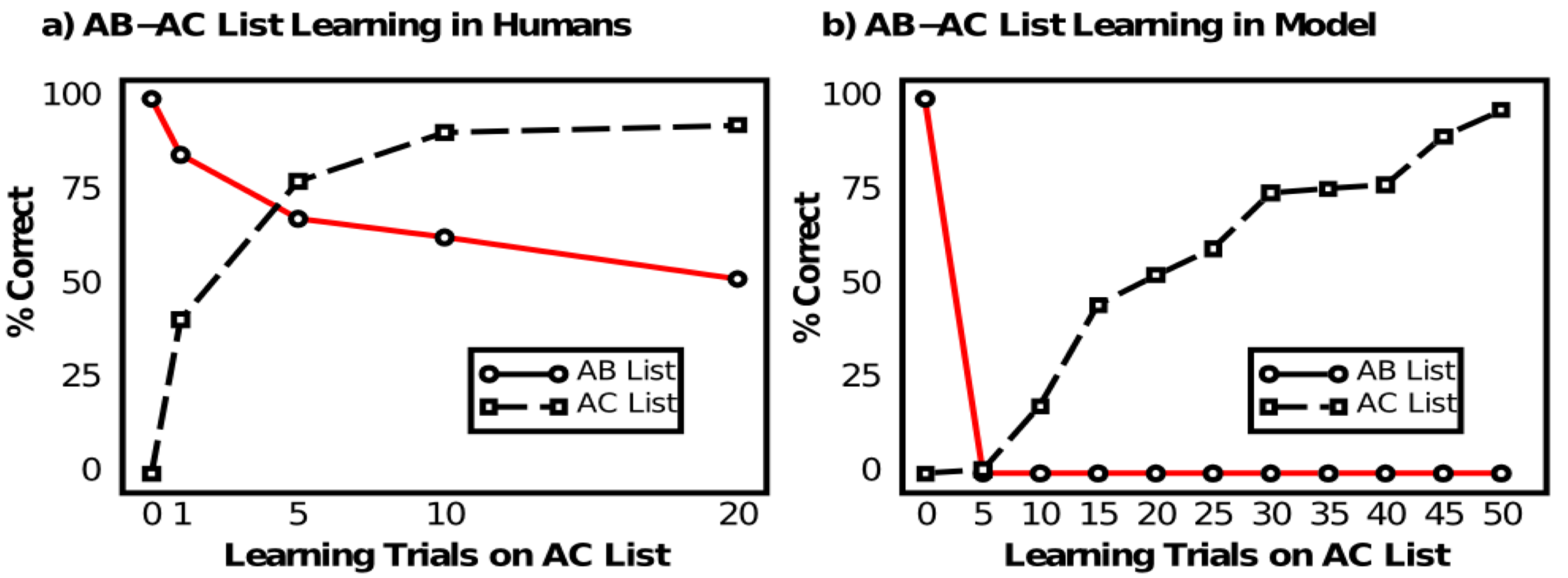
AB-AC

- Test on AB list:
  - Window ?
  - Bicycle ?

- And AC list:
  - Window ?
  - Bicycle ?

(slide < O'Reilly)

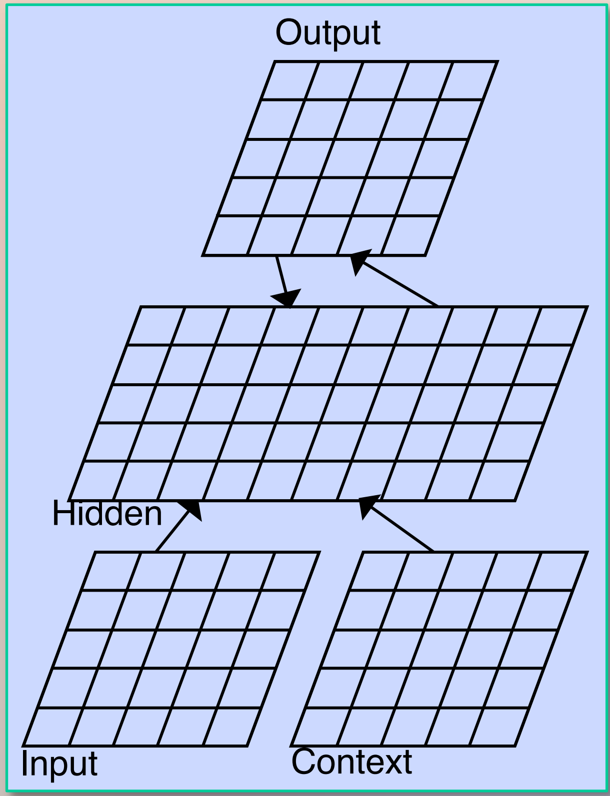
Catastrophic Interference



(slide < O'Reilly)

AB-AC Model

- Input = A
- Output = B, C
- Context differentiates the lists
  — Each list is associated with a different context pattern



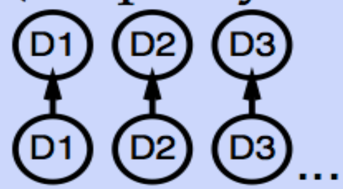
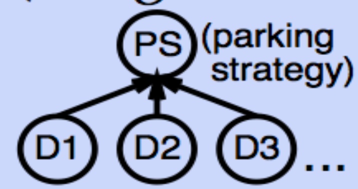
(slide based on Frank)

emergent Demonstration: AB-AC-interference

(slide < O'Reilly)

B. Hippocampus and Pattern Separation

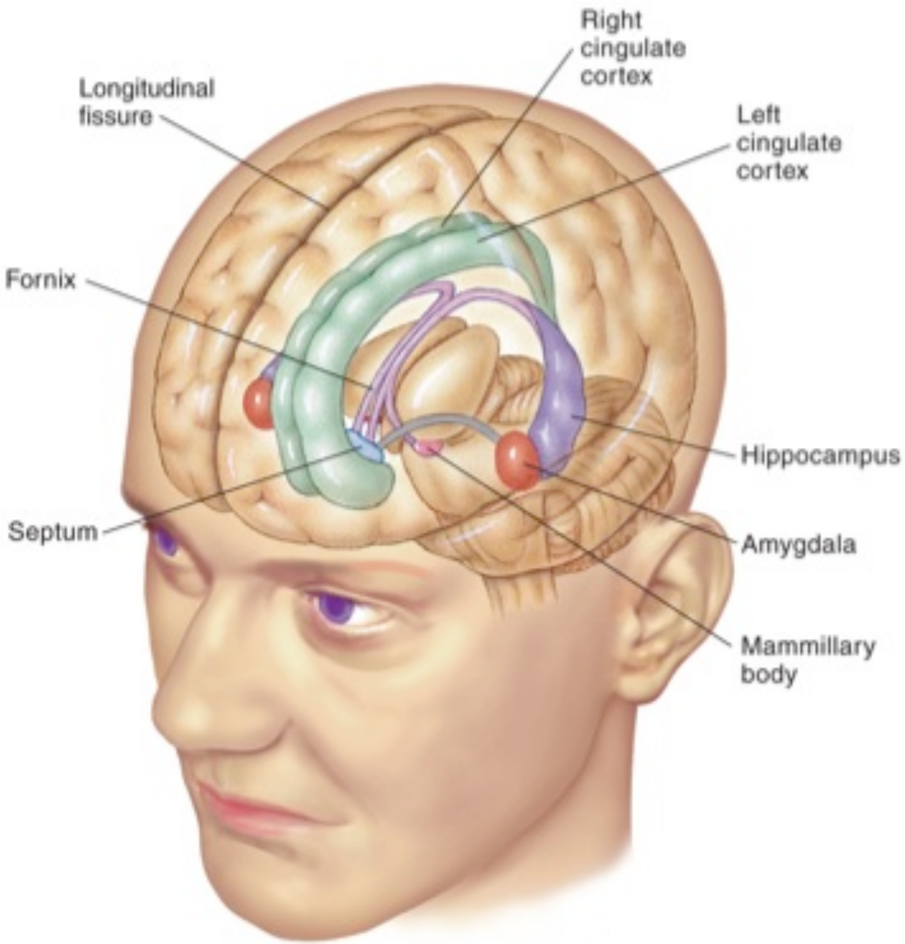
Complementary Learning Systems

Goals:	Remember Specifics	Extract Generalities
Example: Need to:	Where is car parked? Avoid interference	Best parking strategy? Accumulate experience
	Solution:	
1.	Separate reps (keep days separate) 	Overlapping reps (integrate over days) 
2.	Fast learning (encode immediately)	Slow learning (integrate over days)
3.	Learn automatically (encode everything)	Task-driven learning (extract relevant stuff)
	These are incompatible, need two different systems:	
System:	Hippocampus	Neocortex

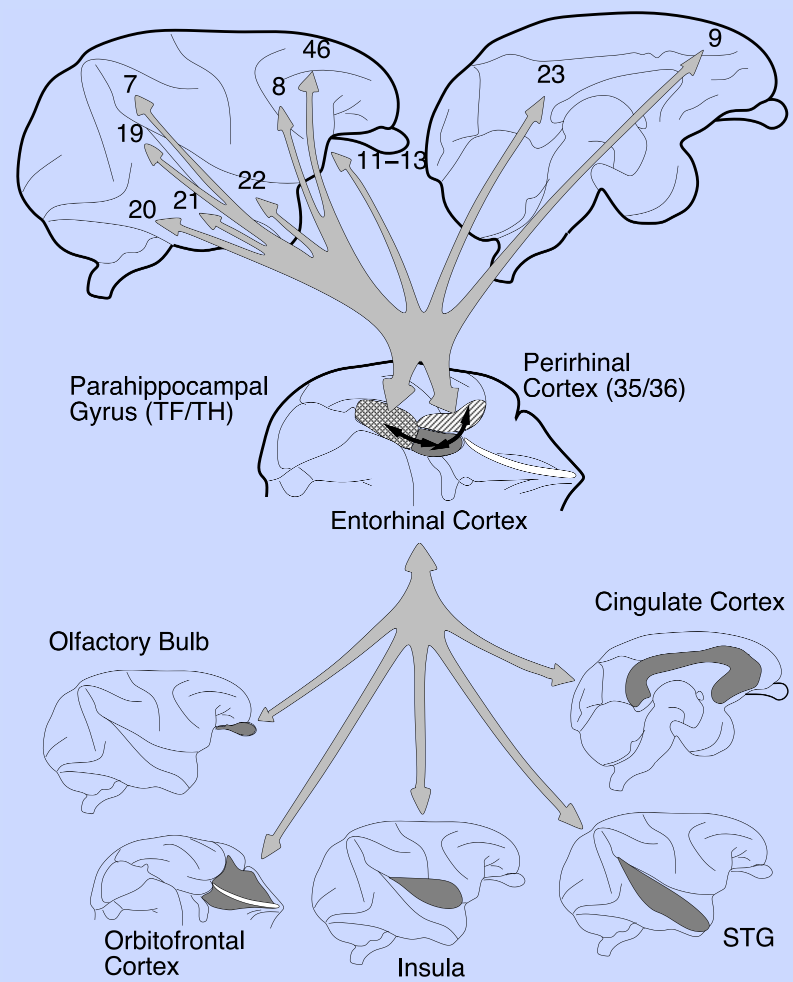
(slide < O'Reilly)

Hippocampus

- 3 or 4 layers
- About 40 million neurons
- Important functions:
  - episodic memory
  - spatial memory & navigation (right HC)
  - memories involving words (left HC)
  - perhaps, “particulars that need to be kept separate”

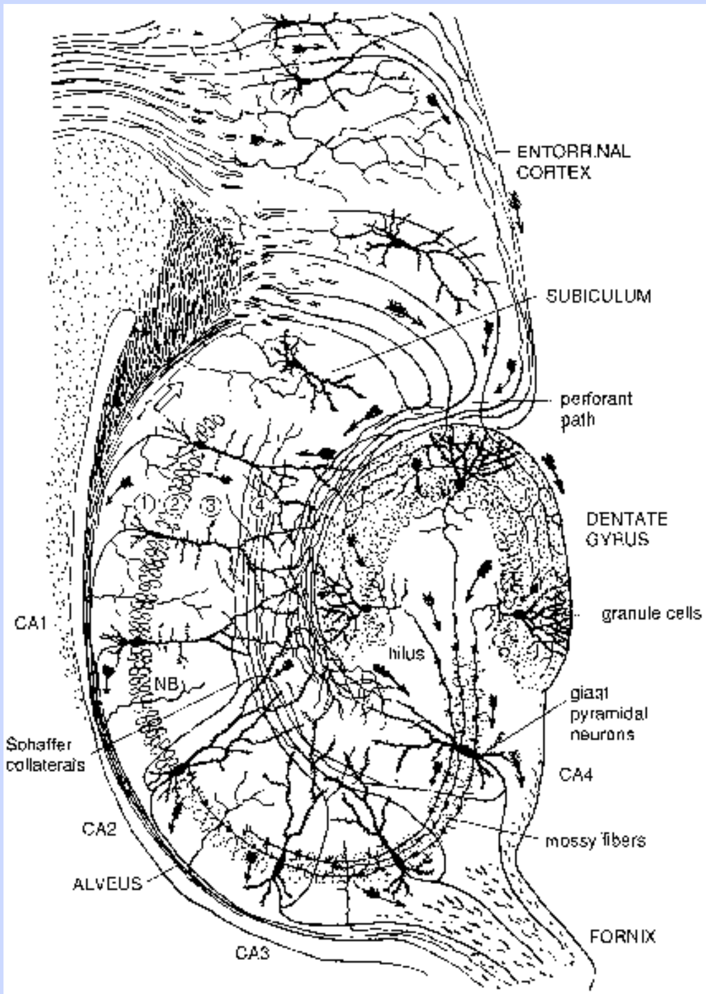


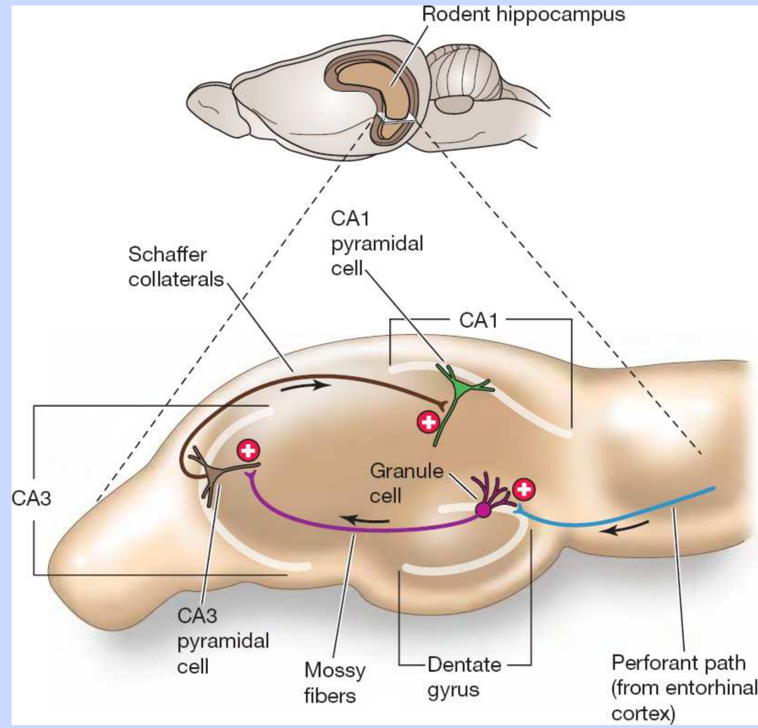
Centrality of Hippocampus

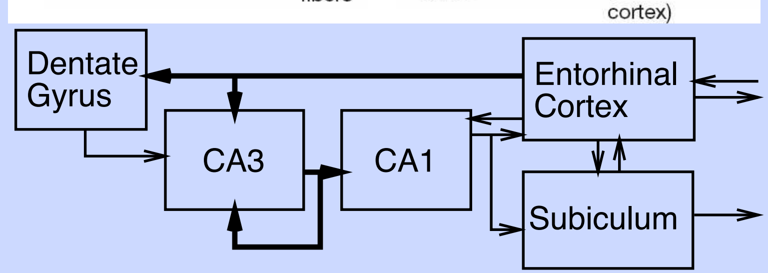


- One of two "summits" in processing hierarchy
  - sensory input areas are "bottoms"
  - other summit is prefrontal cortex
- Access to summary of all brain activity
- Ventral stream connected through perirhinal cortex (PRC)
- Dorsal via parahippocampal cortex (PHC)
- Converge on HC through entorhinal cortex (EC)

Hippocampal Anatomy

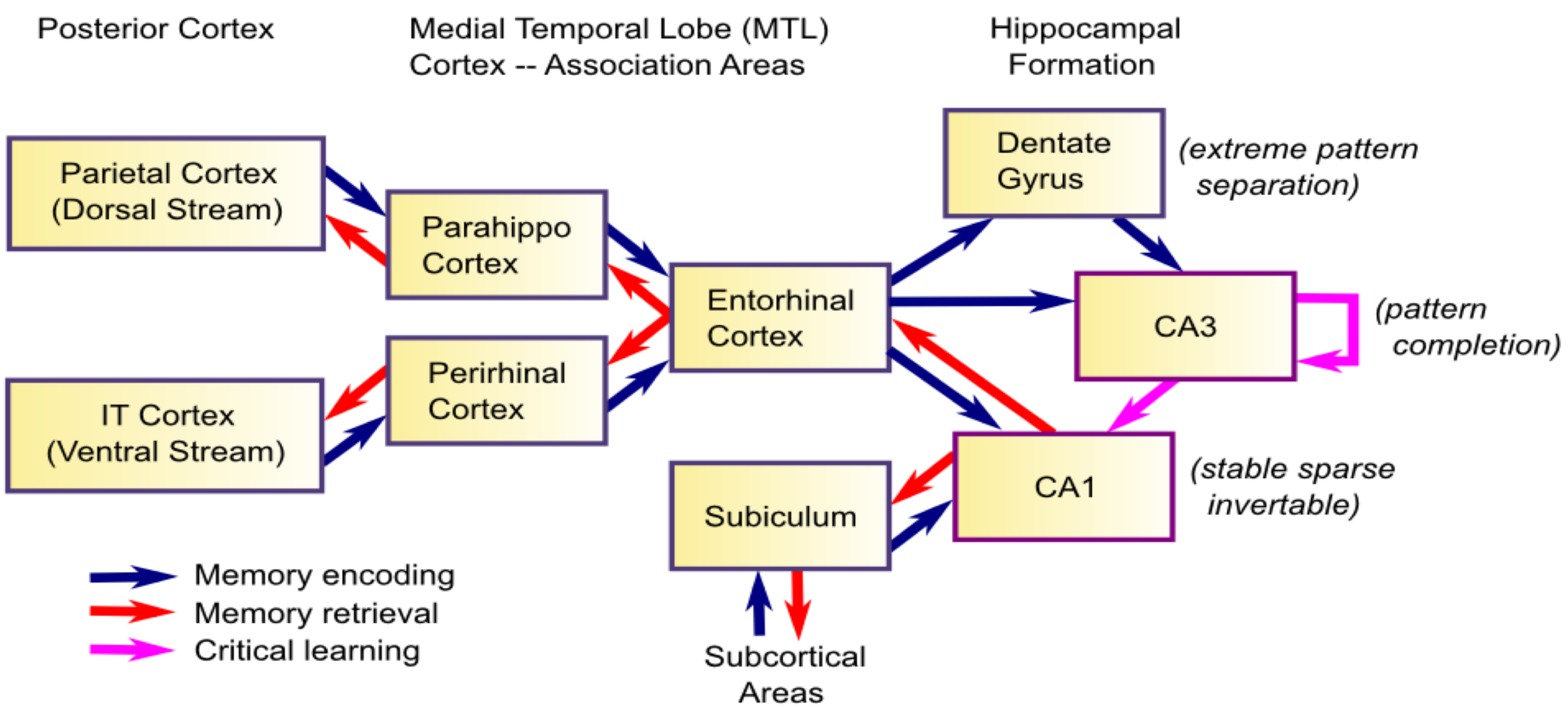






(slide < Frank)

Hippocampal System

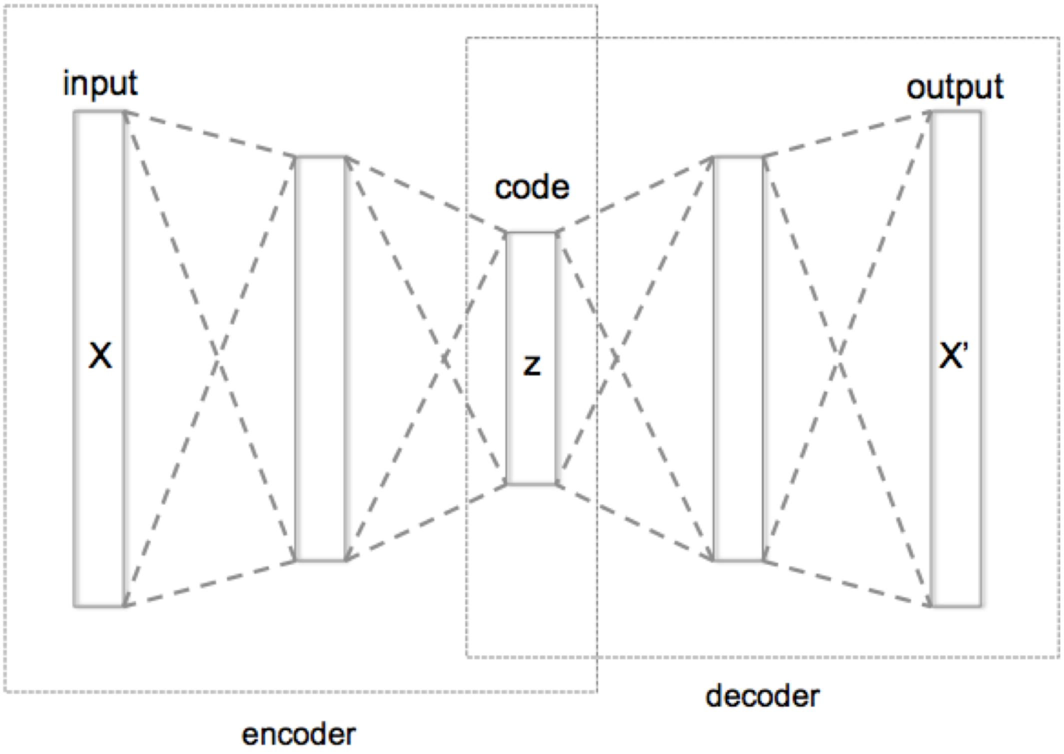


(slide < O'Reilly)

Outline of Episodic Memory Encoding

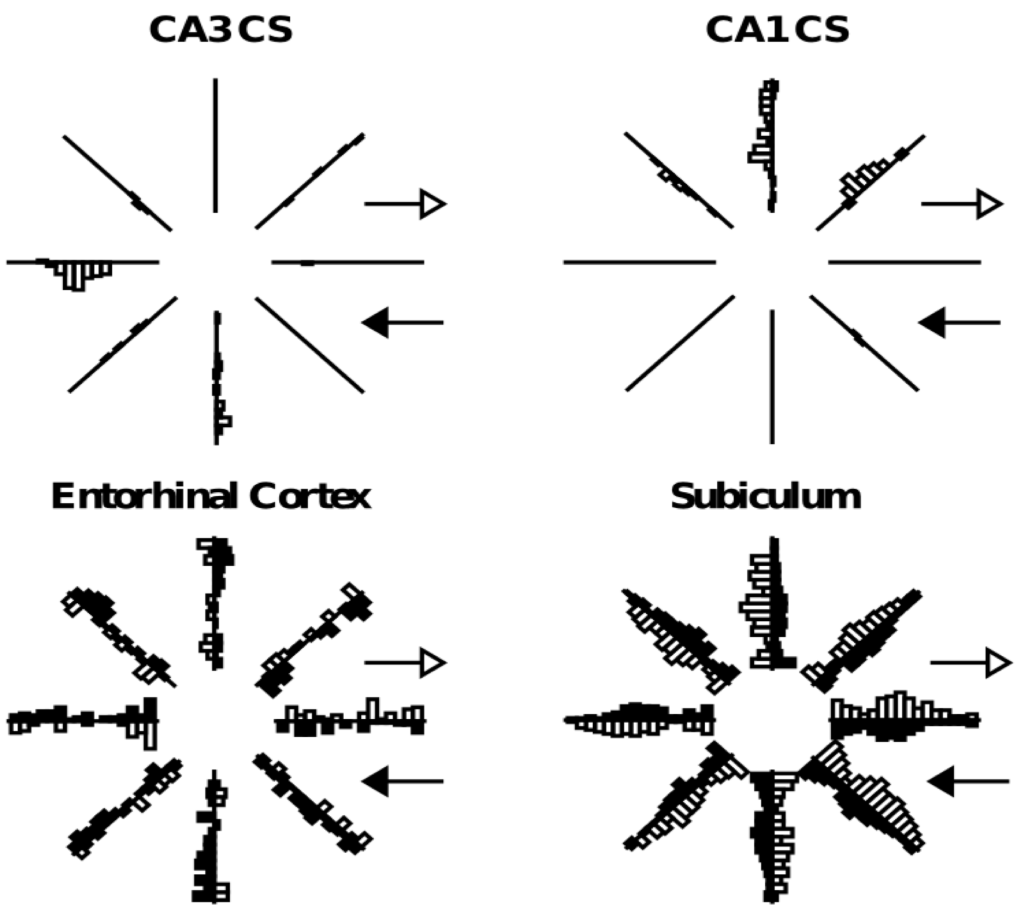
- High-level summary of brain activity in EC
- Drives DG and CA3 via *perforant pathway* resulting in sparse firing pattern in CA3
- EC also drives CA1 via invertible mapping (*autoencoder*)
  - thus CA1 can reactivate the high-level summary in EC
- Activity drives synaptic plasticity
  - among CA3 neurons (in the CA3 recurrent pathway)
  - CA3 to CA1 (the *Schaffer collateral pathway*)
- Binds together components of conjunctive memory so CA3 pattern can activate pattern in CA1, which activates pattern in EC, and thence to neocortex
- Connections are strengthened in all these pathways

Example Auto-encoder Network



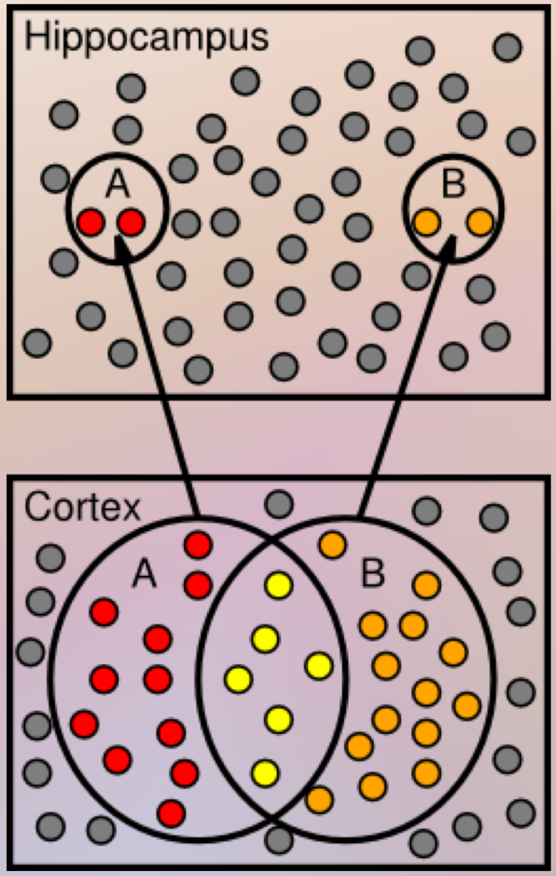
(figure by Chervinskii - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=45555552)

Sparse Activity (Rat Hippocampal Place Cells)



Activation of sample neurons within each area are shown for a rat running on an 8-arm radial maze

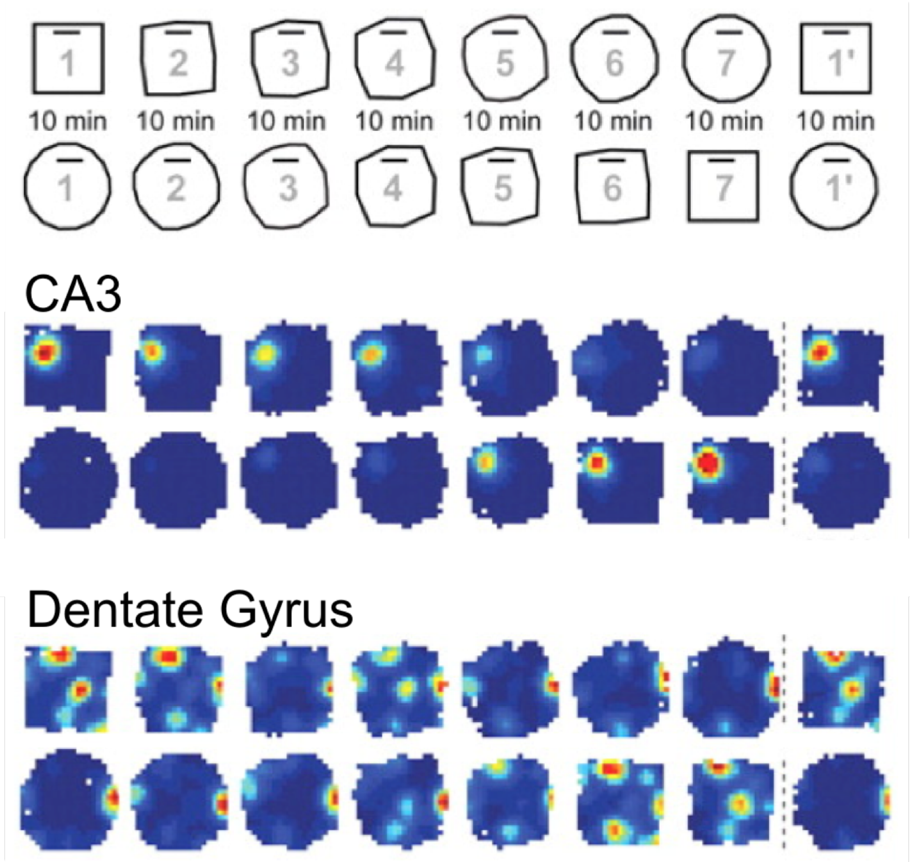
Sparse Representation ⇒ Pattern Separation



- Sparseness thought to result from high levels of GABA inhibition
- Hence, much excitation required to reach threshold
- P = 1% neurons active (typical of DG) in an episode ⇒ P = 0.01% in two random episodes
- P = 25% (typical of cortex) in one ⇒ P = 6.25% in two

Separation vs. Sparseness in DG & CA3

- Activity sparseness in DG (1%) is greater than that in CA3 (2–5%)
- Activity of place cell as shape of environment morphed ☞
- CA3: distinct patterns for square and circular environment
- DG: responds differentially to middle of morph sequence ⇒ greater pattern separation



Cued Recall & Pattern Completion

- Human memory is *content addressable*
- Partial retrieval cue triggers completion of full original pattern
- Pattern completion facilitated by recurrent connections among CA3 neurons
  - glues them together during encoding
  - subset of CA3 neurons can trigger recall of the remainder
- Synaptic changes in perforant pathway during encoding increase likelihood that original DG and CA3 neurons will be reactivated by partial retrieval cue

Pattern Separation vs. Completion

- Tradeoff between pattern separation and pattern completion
- Pattern separation ⇒ likely to treat retrieval cue as novel stimulus
  - — encodes new distinct engram pattern in CA3, instead of completing old one
- System too good at pattern completion ⇒ reactivates old memories instead of encoding new patterns for novel episodes
- Can balance with model parameters
  - — LTP in CA3 supports completion while LTD supports separation
- Hippocampus likely benefits from strategic influences from other brain areas (e.g., PFC executive control)
  - — emphasize either completion or separation depending on whether the current demands require recall or encoding

Spatial Representation in the HC System

- Place cells
- Grid cells
- Head-direction cells
- Border/boundary cells
- Prospective or goal-directed cells
- “Social place cells”
- “Inanimate-object place cells”

Hippocampus Summary

- CA3 stores sparse, pattern-separated representations of cortical input patterns
- Recurrent self-projections in CA3 facilitate recall (pattern completion)
- DG acts as a *removable pattern separation turbocharger*
  - DG uses super-sparse representations, helps increase pattern separation during encoding
  - DG “steps aside” during retrieval
  - Evidence for two modes: theta cycle (e.g., Hasselmo et al, 2002); neuromodulatory control over relative DG effect on CA3
- CA1 helps “translate” sparse, non-overlapping CA3 representations back into overlapping EC representations, by providing an intermediately sparse representation

(slide < Frank)

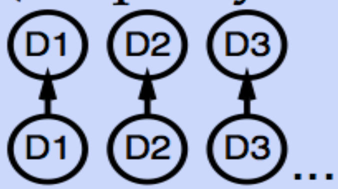
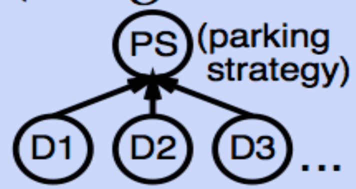
emergent demo: Hip.proj

Applying the hippocampus model to the AB-AC task

C. Memory Consolidation

How memories become (semi-)permanent

Complementary Learning Systems

Goals:	Remember Specifics	Extract Generalities
Example: Need to:	Where is car parked? Avoid interference	Best parking strategy? Accumulate experience
	Solution:	
1.	Separate reps (keep days separate) 	Overlapping reps (integrate over days) 
2.	Fast learning (encode immediately)	Slow learning (integrate over days)
3.	Learn automatically (encode everything)	Task-driven learning (extract relevant stuff)
	These are incompatible, need two different systems:	
System:	Hippocampus	Neocortex

(slide < O'Reilly)

H.M. (Henry Molaison, 1926–2008)

- HC removed in 1957 to treat severe epilepsy
- Developed inability to learn new episodic information (*anterograde amnesia*)
  - some degree of forgetting of previously learned knowledge (*retrograde amnesia*)
  - older memories had somehow become *consolidated* outside of the HC
- Remembered how to talk, meanings of different words and objects, how to ride bike, could learn various new motor skills
- Could learn new semantic information, but relatively slowly and access was more brittle

Hippocampal Amnesia

Hippocampal amnesiacs show:

- Spared implicit memory, skill learning (without recall)
  - small adaptive adjustments in synaptic weights
- Intact repetition priming for existing associations (table-chair) but not for arbitrary novel pairs of words (locomotive-spoon)
  - small cortical adjustments can prime existing representations but not novel conjunctions
- Remote memories spared but recent ones completely forgotten
  - consolidation by reactivation of memories across multiple contexts, sleep, etc.

(slide < Frank)

Memory Consolidation

- Patterns of activity that occur while a rat is running a maze seem to be reactivated when animal is asleep
  - but measured levels of reactivation are relatively weak compared to patterns active during actual behavior
- Humans: slow wave oscillations in non-REM sleep thought to be associated with memory consolidation.
  - external induction of slow wave oscillations during sleep may result in enhanced hippocampal-dependent memories for items encoded just prior to sleep (Marshall et al., 2006)

Effects of Complementary Learning

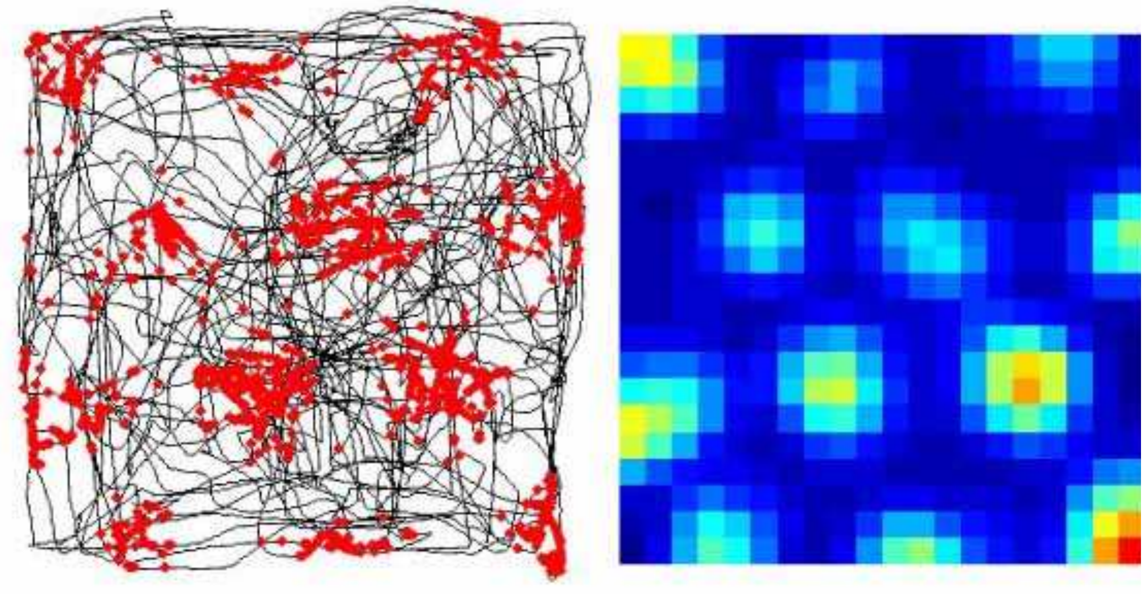
- Information encoded in neocortex of a different character to that initially encoded by hippocampus
- To the extent that episodic memories can be encoded in the neocortex:
  - — will become more "semanticized" and generalized
  - — integrated with other existing memories
- Compare to the more distinct and crisp pattern-separated representations originally encoded in HC

D. Spatial Representation in the Hippocampus

Place, Grid & Head Direction Cells in HC

- Rat hippocampus exhibits robust **place cell** firing
  - individual DG, CA3, CA1 neurons respond to particular locations
  - neuron will have different preferred location in different environments
  - does not appear to be any topography or other systematic organization
  - consistent with random, diffuse nature of perforant pathway projections and effects of pattern separation
- **Grid cells** form regular hexagonal grid over space
  - appear to depend on various forms of oscillations
  - may provide raw spatial information that gets integrated into place cells within the hippocampus proper
- **Head direction cells** in several areas project into hippocampus
  - provide dead reckoning signal about where rat is facing based on accumulation of recent movements

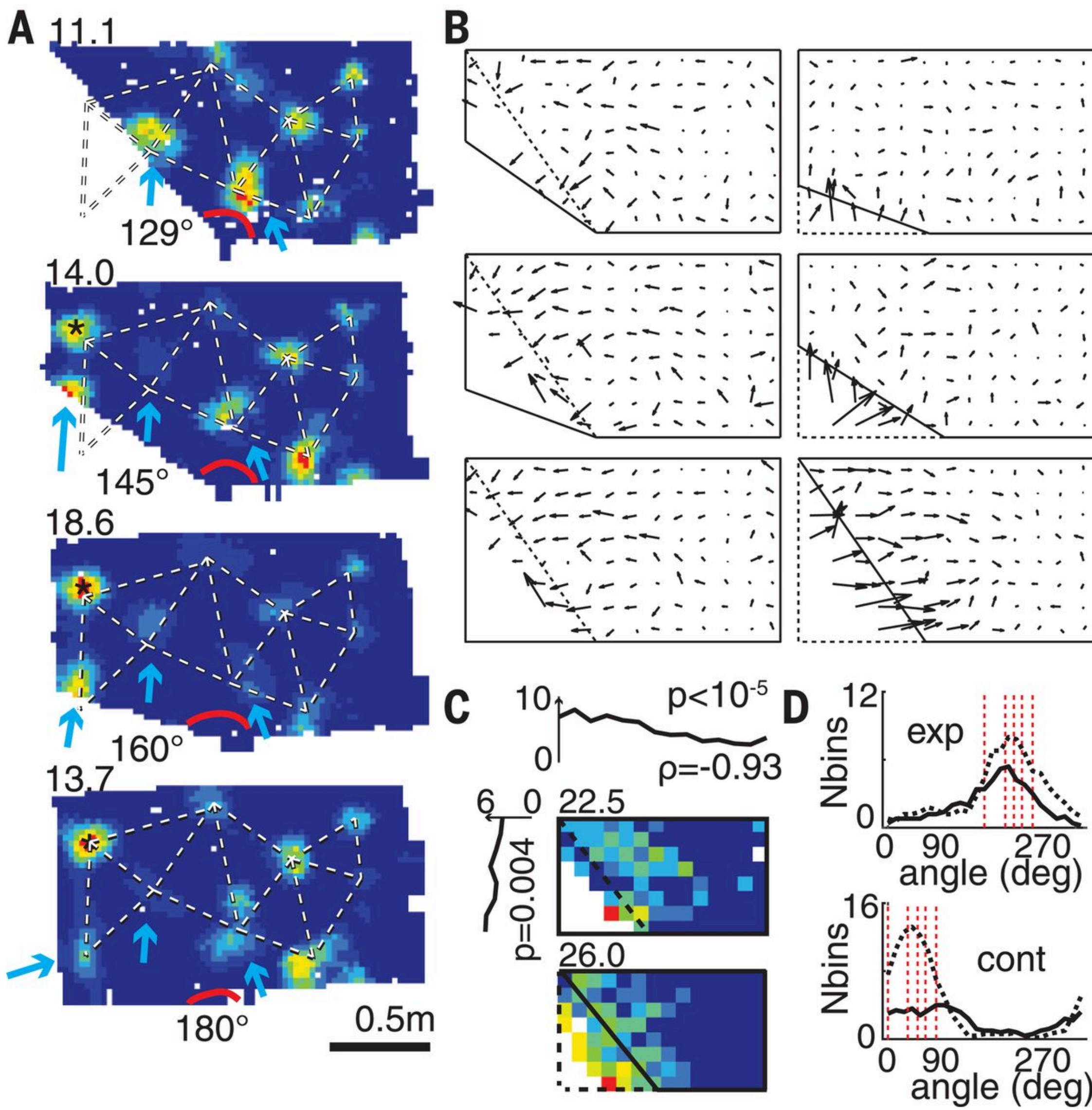
“Hippocampal” Grid Cells



- Grid cells are in medial entorhinal cortex (Haftingetal, 2005), not HC proper
- Hippocampus might integrate location with speed and direction (head direction cells) to perform path integration
- Can be recast as just another example of conjunctive, pattern-separate representations

(slide < Frank)

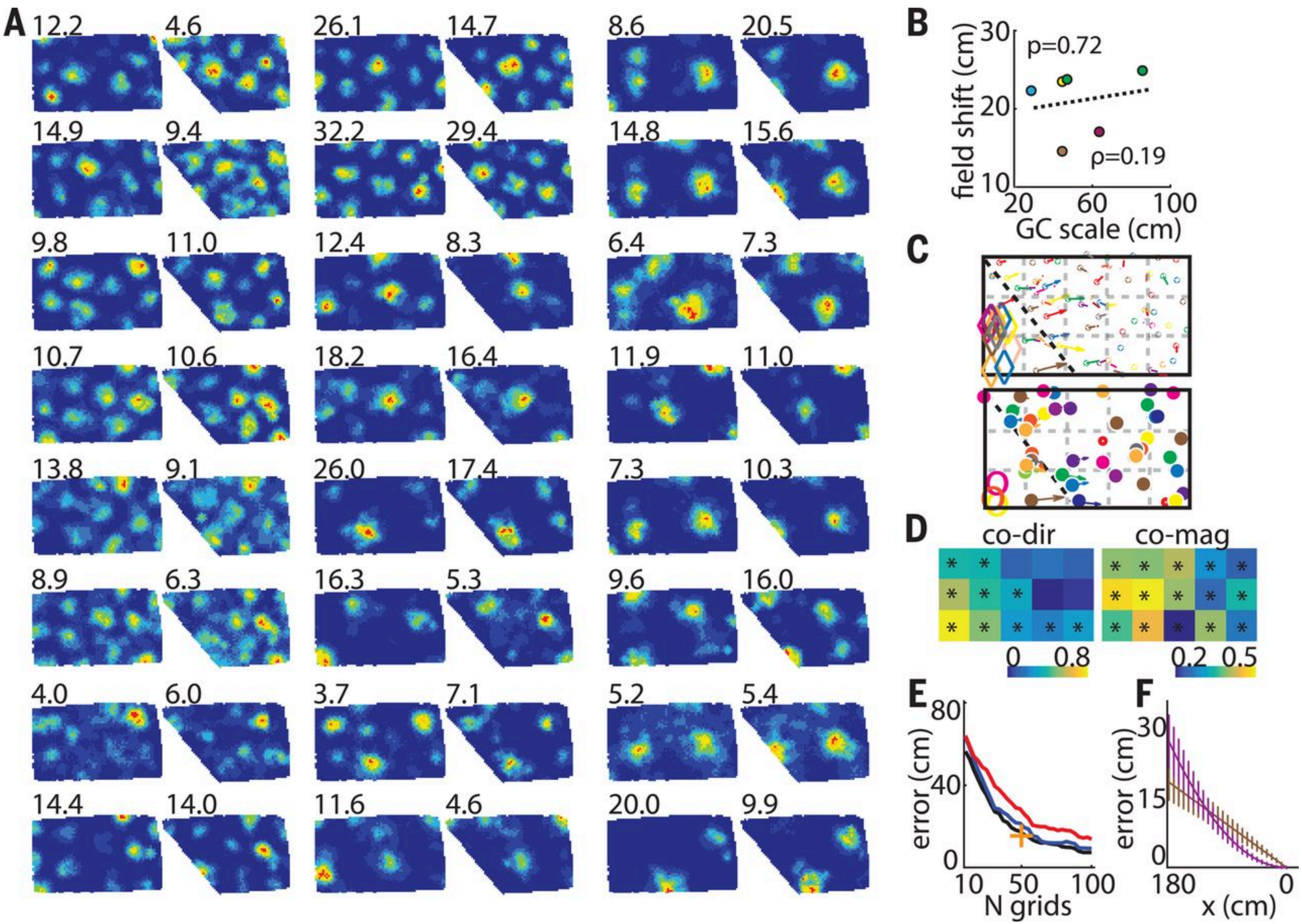
Local transformations of the hippocampal cognitive map



Julija Krupic et al. Science 2018;359:1143-1146

Published by AAAS

Simultaneous changes in grid field positions.



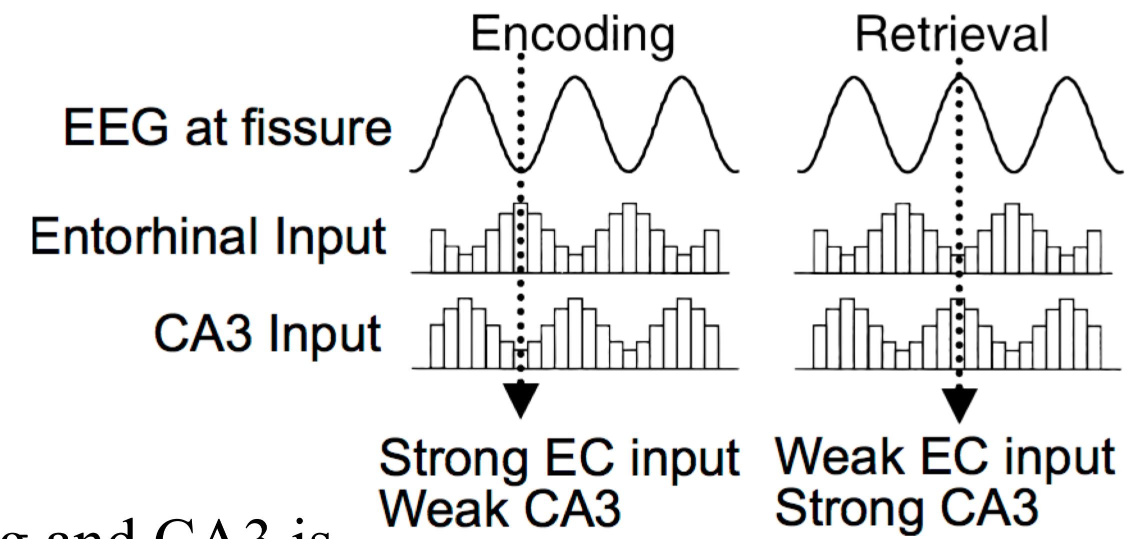
Julija Krupic et al. Science 2018;359:1143-1146

Published by AAAS

Theta Waves

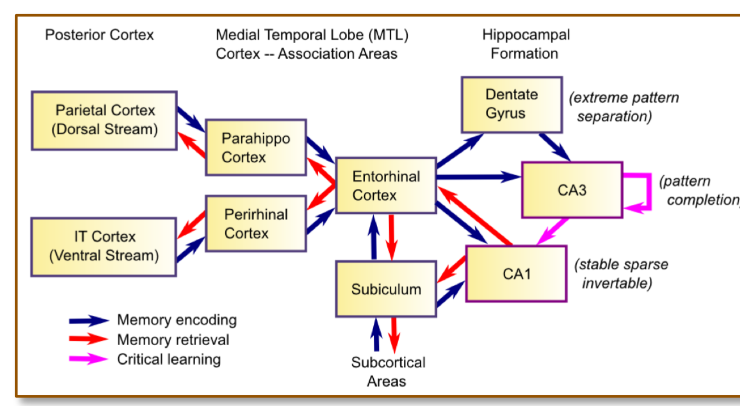
- Hippocampus exhibits oscillation of neuron firing in theta frequency band (8 to 12 Hz in rats)
- Thought to play critical role in grid cell activations in EC
  - perhaps may serve to encode temporal sequence information
  - place-field activity firing shows theta phase procession
  - different place fields fire at different points within unfolding theta wave
- Different areas of hippocampus are out of phase with each other
- Perhaps this phase relationship enables system to alternate between encoding of new information vs. recall of existing information (Hasselmo et al., 2002)
  - alters HC parameters to optimize encoding or retrieval
- Implemented in Hip.prog model

Theta Rhythm Control of Encoding/Retrieval



- Different areas of HC system fire out of phase with respect to theta rhythm, producing dynamics that optimize encoding vs. retrieval.

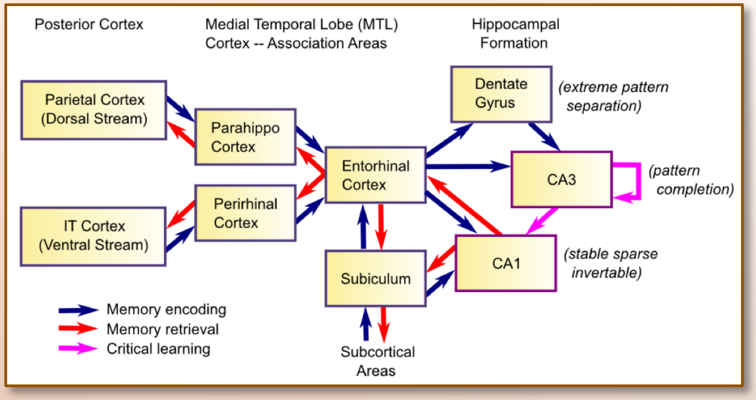
- When the EC input is strong and CA3 is weak, CA1 can learn to encode the EC inputs.
  - — serves as plus phase for error-driven learning dynamic in the Leabra framework.



- When CA3 is strong and EC is weak, the system recalls information driven by prior CA3 ⇒ CA1 learning.
  - — serves as minus phase for Leabra error-driven learning

E. Novelty, Familiarity, and Recognition

Function of the Subiculum



- Relation to HC is analogous to EC, but input-output to subcortical areas ☞
- Might compute relative <u>novelty</u> of a given situation, and communicate to midbrain dopamine systems and thence to basal ganglia
- Novelty can have complex affective consequences:
  - both *anxiogenic* (anxiety producing)
  - and motivational for driving further exploration
  - generally increases overall arousal levels
- HC uniquely capable of determining novelty, taking into account full conjunction of relevant spatial and other contextual information
- Subiculum could compute novelty by comparing CA1 and EC states during recall phase of theta oscillation
- **Conjecture!**

Dual Process Model of Recognition Memory

- Neocortex can support episodic memory traces, but with different properties from those in HC
- Perirhinal cortex (PRC) can produce a familiarity signal
  - — indicates in coarse manner whether a stimulus was experienced recently or not
  - — like a single graded value that varies in intensity depending on how strongly familiar the item is
  - — accessible to consciousness
  - — hypothesis: sharpness of repeated representations in perirhinal cortex due to competition & Hebbian learning
  - — familiarity indicated by average activity of winners (Norman & O'Reilly '03)
- Dual processes: hippocampal recall and perirhinal familiarity

Priming

- Memory increases speed or probability of a particular response
- We are not generally aware of these memories
- Weight-based priming
  - — incremental changes to synapses
  - — persistent (> 1 year from single exposure)
- Activation-based priming
  - — results from residual activation
  - — short-lived

More Robust Activation-Based Memory

In Executive Function Chapter:

- PFC robust active maintenance over seconds to minutes
- BG provides dynamic gating signal for update vs. maintenance
- Used for “working memory,” cognitive control, …

(slide < O’Reilly)

emergent Demonstrations: WtPriming ActPriming