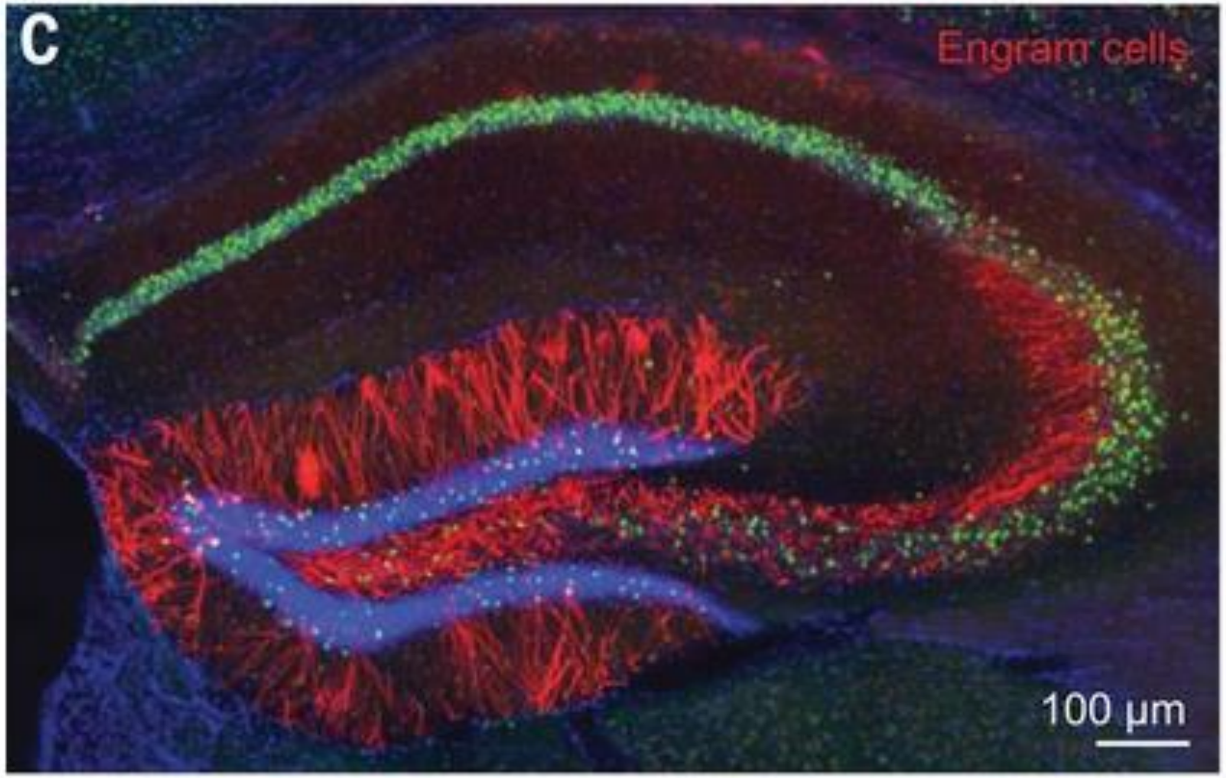
Models of Memory

Angus Chadwick, ANC
School of Informatics, University of Edinburgh, UK
angus.chadwick@ed.ac.uk

Computational Cognitive Neuroscience (Lecture 15, 2023/2024)

Outline of Lecture

- Overview of memory
- The neurobiological basis of memory
- The Hopfield network (associative memory)
- Spatial memory and cognitive maps
- Working memory



Cells in the hippocampus of a mouse representing a fear memory (in red). *Josselyn and Tonegawa, Science (2020)*.

What is Memory?

- Psychologists classify memory in several ways: short vs long term, implicit vs explicit, etc.

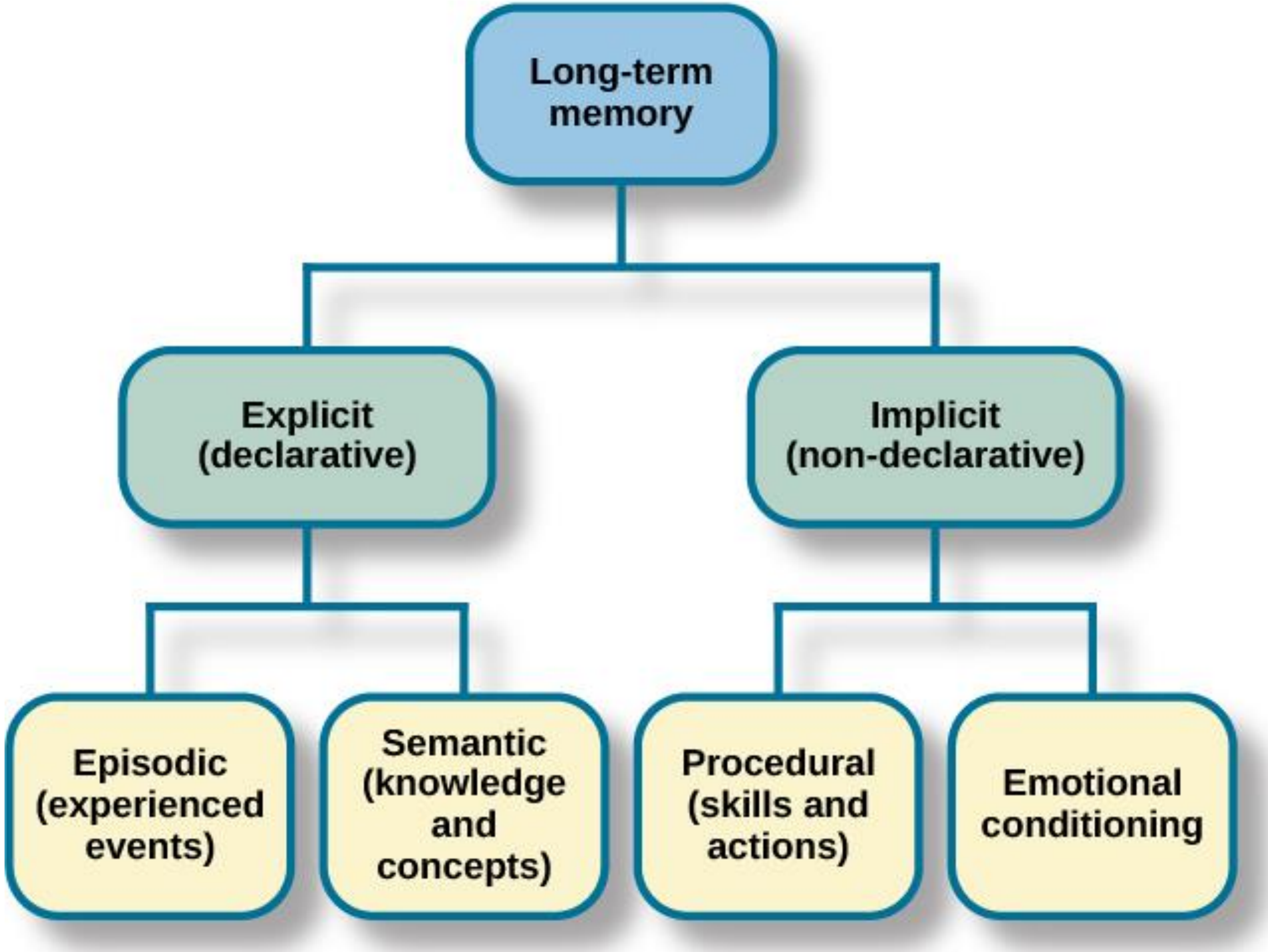
- Explicit memory (episodic/declarative/semantic) – conscious recollection of events and facts ("what did I have for lunch yesterday?", "where is the Eiffel Tower?") *[timescale – days, months, years]*

- Implicit/procedural memory – knowledge used unconsciously, e.g., how to tie your shoes *[timescale – long term]*

- Associative memory – learn and remember relationships between items (e.g., a persons name) *[timescale – long term]* (a form of explicit memory!)

- Working memory – store items in mind for short period (e.g., a phone number or shopping list) *[timescale – seconds]*

What is Memory?



How are Memories Stored in the Brain?

- Very short term memories are maintained dynamically in distributed network activity patterns ("reverberatory" or persistent activity – first postulated by Karl Lashley [1930s] and Donald Hebb [1940s])

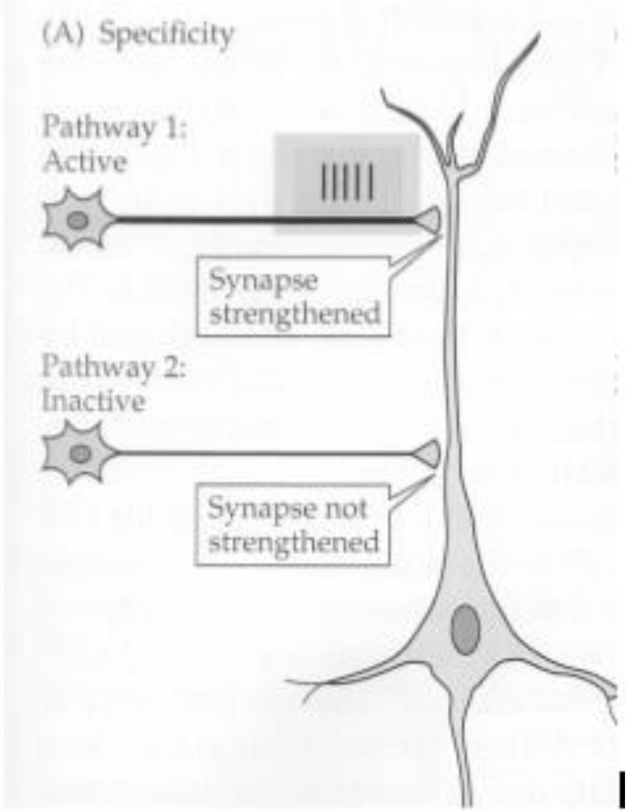
- Storage of long-term memories requires structural changes, especially synaptic plasticity which forms an "engram" or memory trace

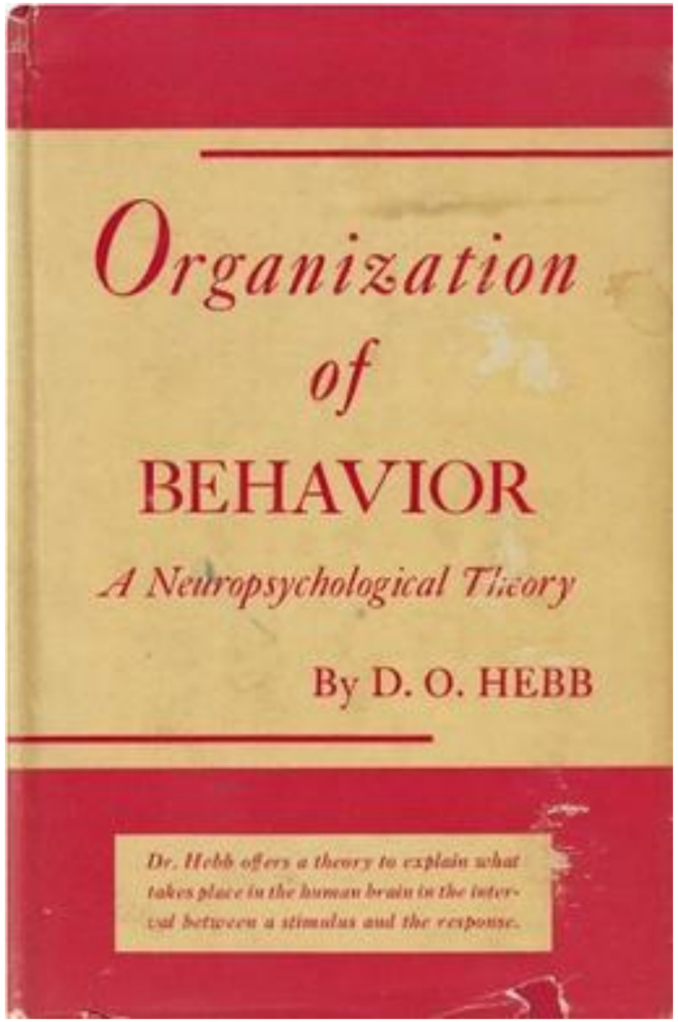
- Memory involves orchestrated interaction between multiple brain areas, most notably hippocampus and cortex

Donald Hebb – Hebbian Learning and Cell Assemblies (1948)

- *“When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased”*

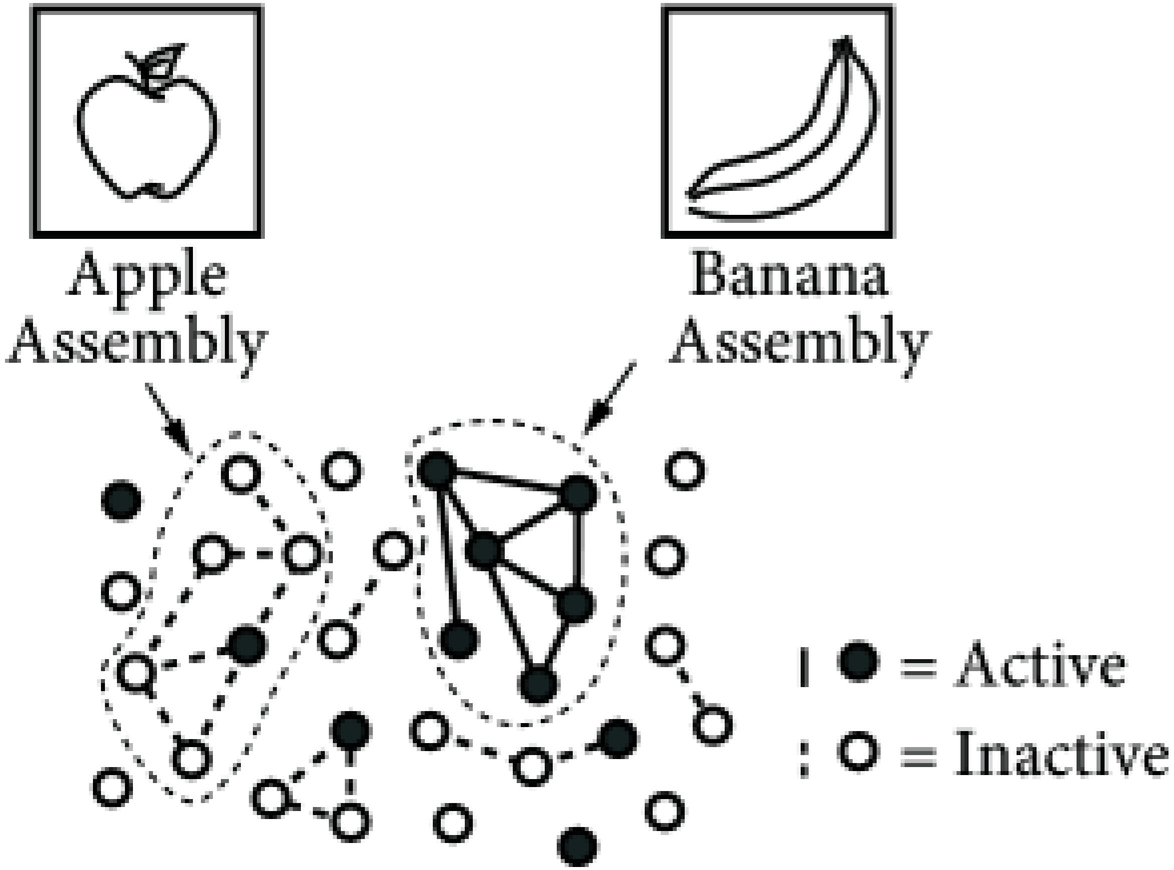
- More commonly put: *“Cells which fire together, wire together”*





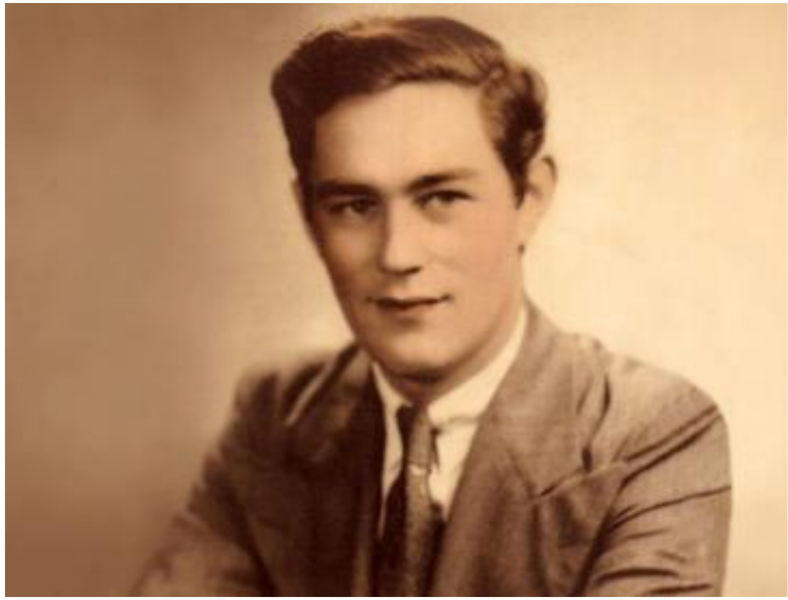
Donald Hebb – Hebbian Learning and Cell Assemblies (1948)

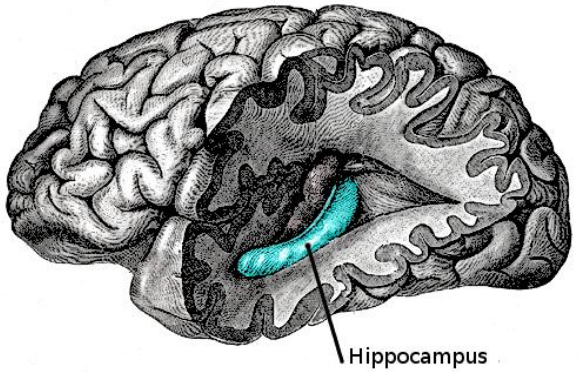
- Coactivation of a group of cells causes connections to form/strengthen (Hebbian learning)
- Subsequent activation of a subset of those cells will reactivate the whole group (a “cell assembly”)
- A mechanism for forming and retrieving associations/memories



Patient H.M. (1927-2008)

- Had his hippocampus bilaterally removed in 1953 (to treat epilepsy)
- Suffered permanent anterograde amnesia – could not form *new* episodic/declarative/semantic memories
- Relatively intact cognition otherwise – procedural memory, working memory, previously stored episodic memory etc.
- Conclusion: hippocampus is required for storing new episodic memories?





“Hippocampus” means seahorse!

Standard Model of Systems Consolidation

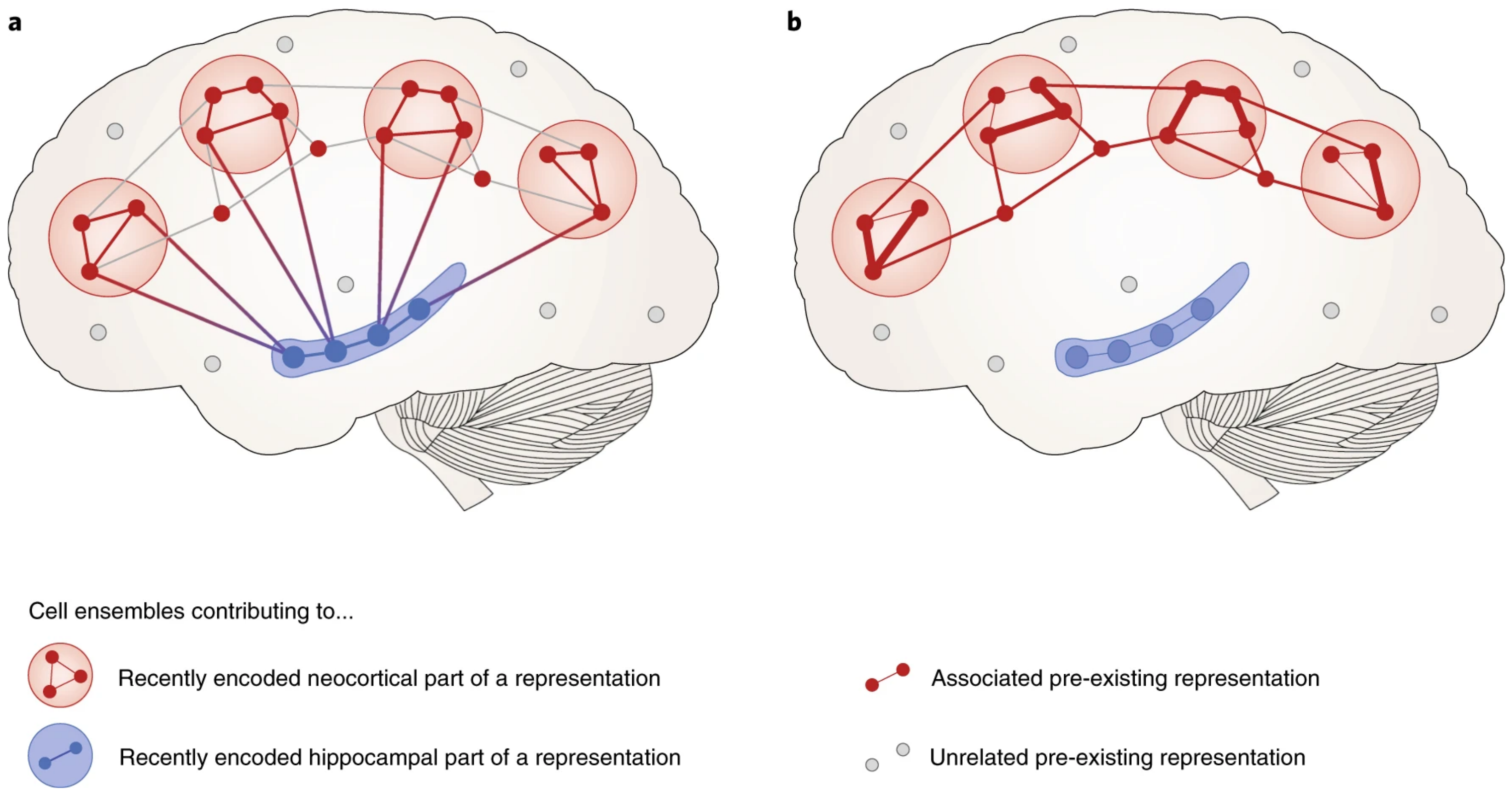
- Memory initially formed in hippocampus

  (one-shot learning)

- Memory slowly transferred from hippocampus to cortex (during sleep)

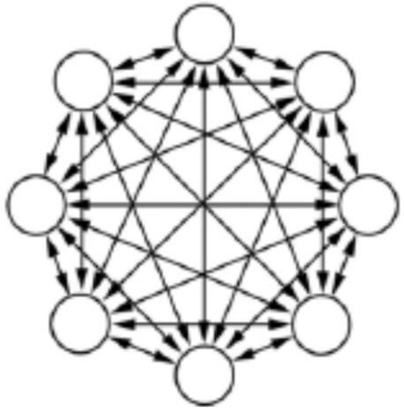
- Eventually, memories become *hippocampus-independent*

- Explains why H.M. could recall previous memories but not store new ones



The Hopfield Network (1982)

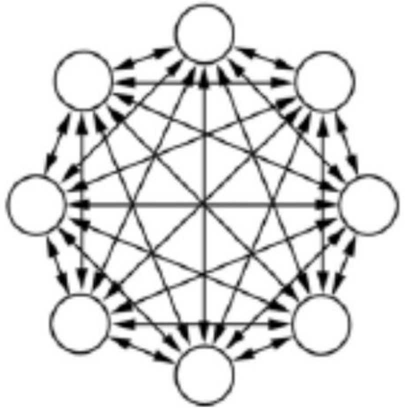
- How many memories can be stored in a network? How should the synaptic weights be set?
- The Hopfield network comprises *N* binary neurons (i.e., with states *s* = -1 or +1) connected via a *symmetric* coupling matrix *J:*



$$s_i(t) = \begin{cases} +1 & \text{if } \sum_j J_{ij} s_j(t-1) > \theta_i \\ -1 & \text{if } \sum_j J_{ij} s_j(t-1) \leq \theta_i \end{cases}$$

The Hopfield Network (1982)

- How many memories can be stored in a network? How should the synaptic weights be set?
- The Hopfield network comprises *N* binary neurons (i.e., with states *s* = -1 or +1) connected via a *symmetric* coupling matrix *J:*



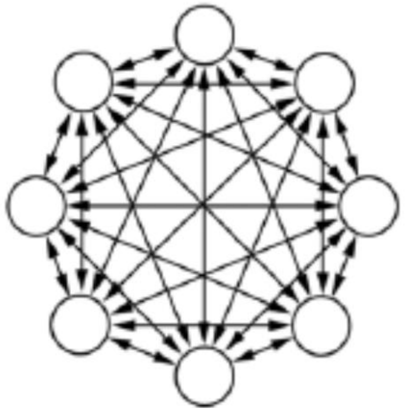
$$s_i(t) = \begin{cases} +1 & \text{if } \sum_j J_{ij} s_j(t-1) > \theta_i \\ -1 & \text{if } \sum_j J_{ij} s_j(t-1) \leq \theta_i \end{cases}$$

- We can define the following "Energy function" (or Lyapunov function) for the Hopfield network:

$$E = -\frac{1}{2} \sum_{i<j} J_{ij} s_i s_j + \sum_i s_i \theta_i$$

The Hopfield Network (1982)

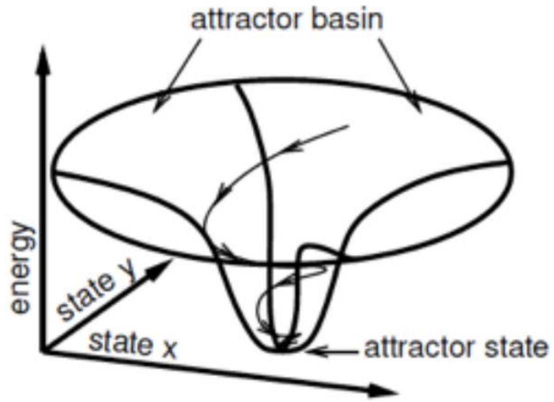
- How many memories can be stored in a network? How should the synaptic weights be set?
- The Hopfield network comprises *N* binary neurons (i.e., with states *s* = -1 or +1) connected via a *symmetric* coupling matrix *J:*



$$s_i(t) = \begin{cases} +1 & \text{if } \sum_j J_{ij} s_j(t-1) > \theta_i \\ -1 & \text{if } \sum_j J_{ij} s_j(t-1) \leq \theta_i \end{cases}$$

- We can define the following "Energy function" (or Lyapunov function) for the Hopfield network:

$$E = -\frac{1}{2} \sum_{i<j} J_{ij} s_i s_j + \sum_i s_i \theta_i$$

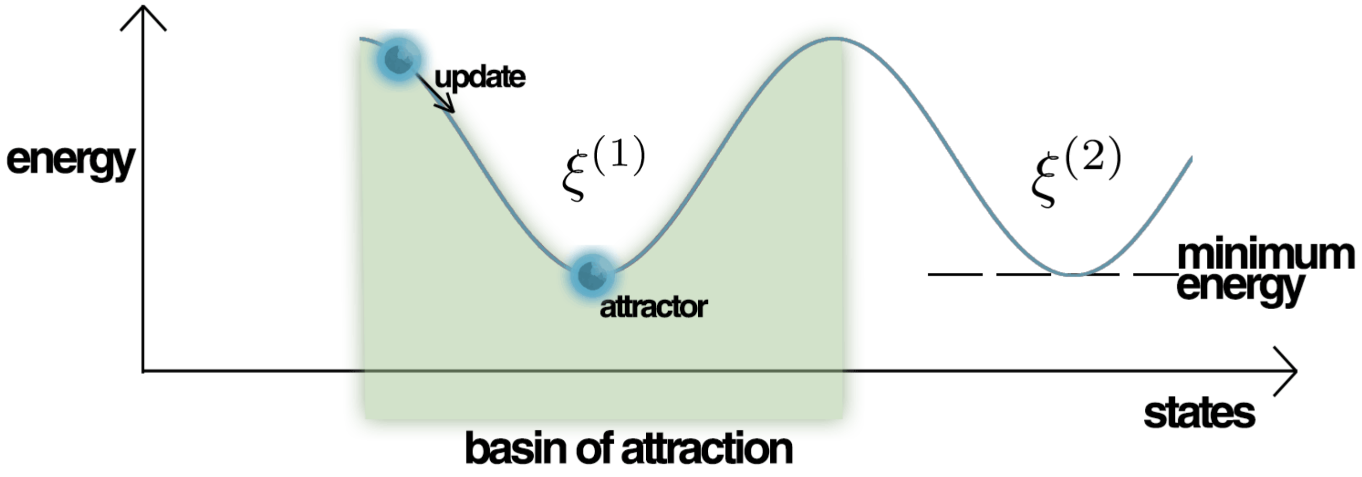


- Updates always decrease *E* – this means that the activity *s* will eventually reach a stable local minimum (a "memory" of the network)

Hebbian Learning in the Hopfield Network

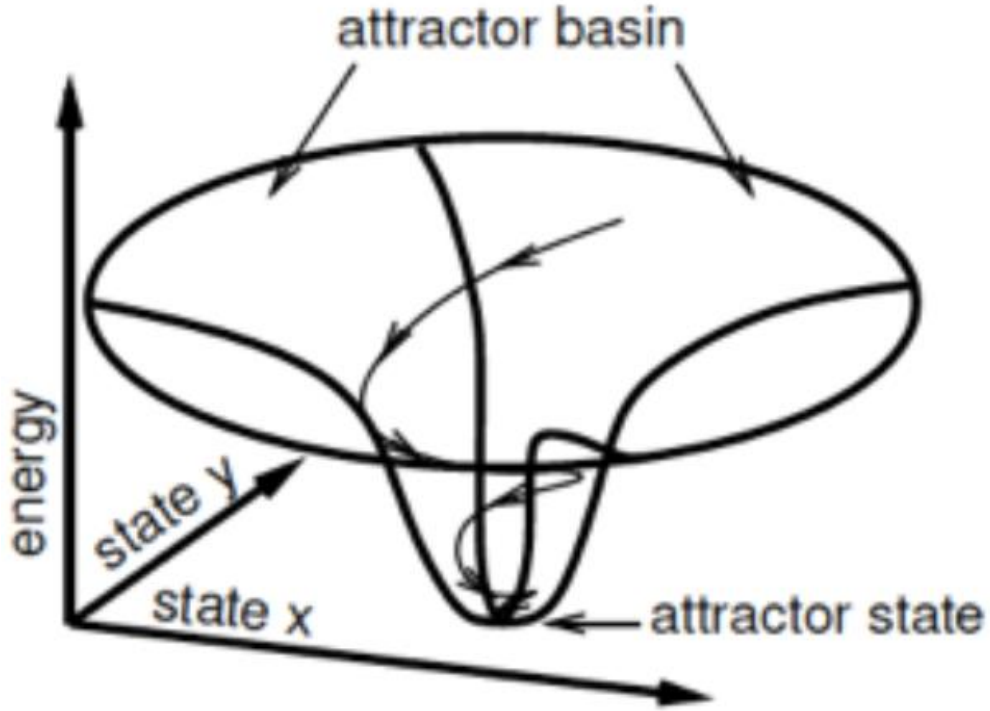
- If we start with a pattern that is sufficiently close to a local minimum, the network will converge to that pattern (pattern completion)

- To store *K* patterns $s_i = \xi_i^{(k)}$ in the network, we set the couplings *J as:* $$J_{ij} = \frac{1}{K} \sum_{k=1}^{K} \xi_i^{(k)} \xi_j^{(k)}$$

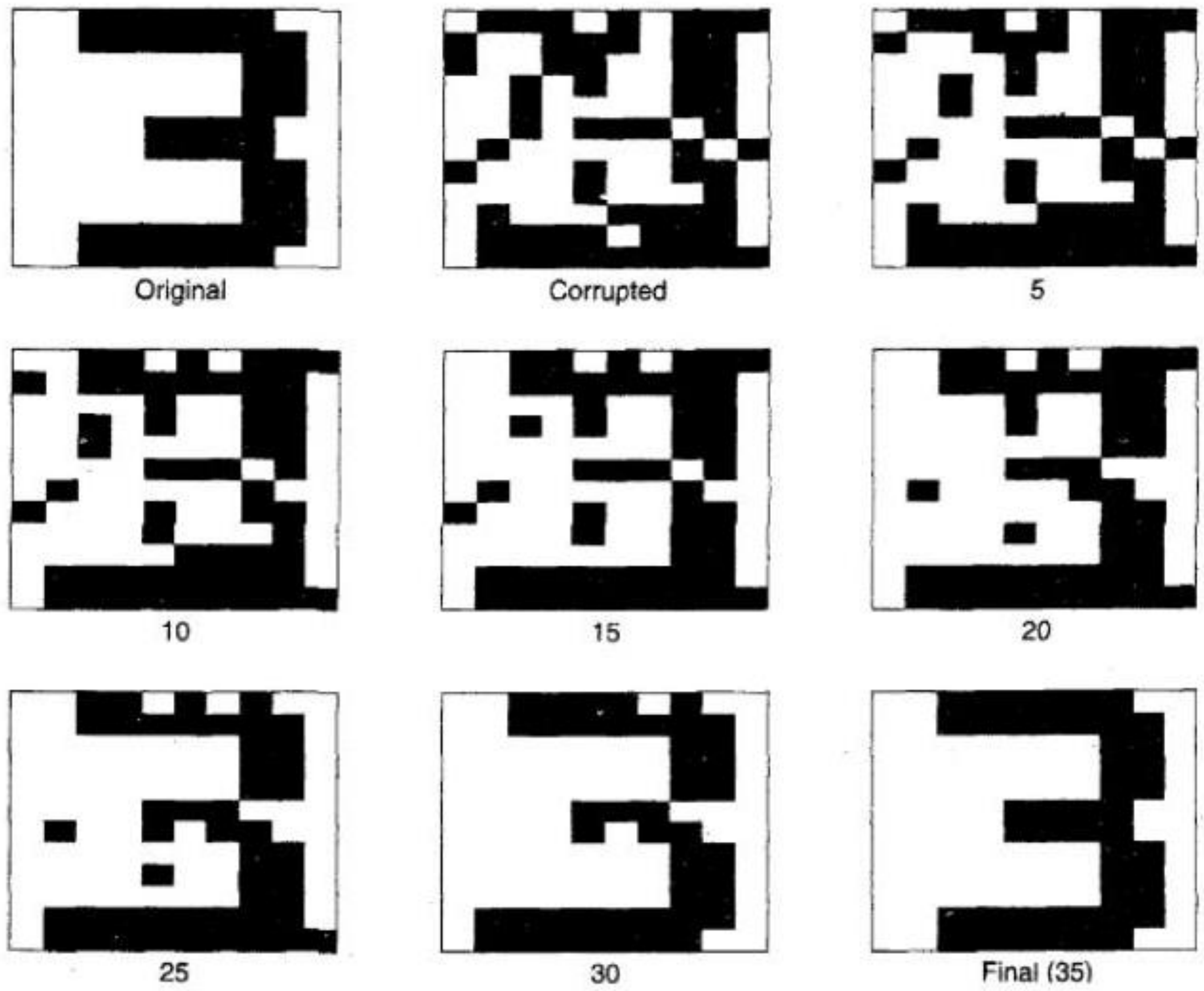


- As we store more patterns, their basins of attraction will become smaller causing catastrophic interference
- There is also the creation of spurious patterns – extra local minima that don't correspond to any stored pattern. The more patterns we store, the more spurious patterns appear.

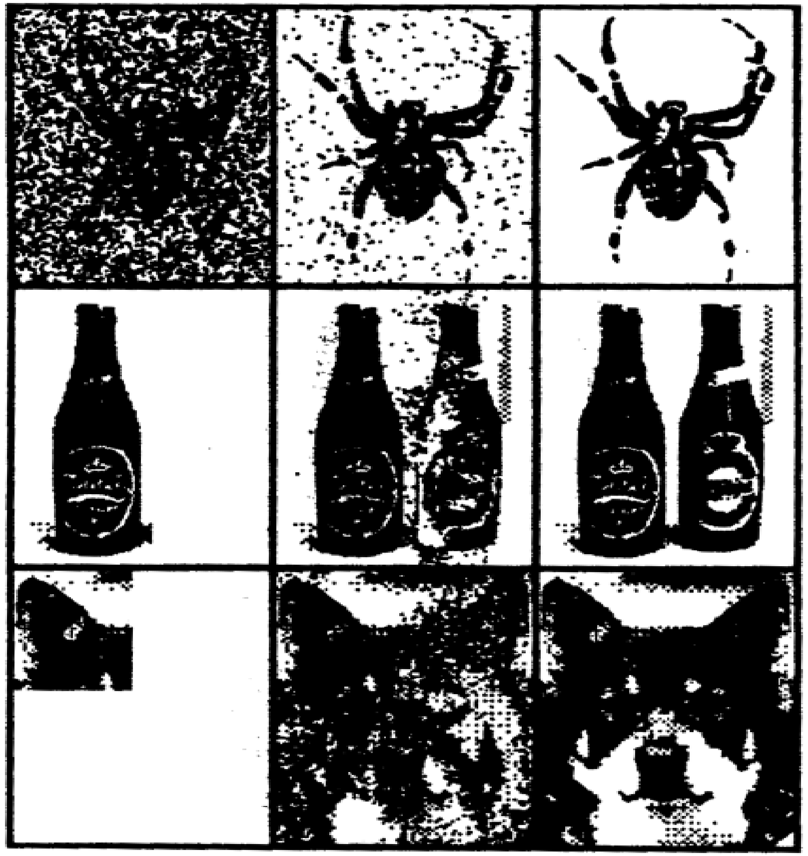
Hebbian Learning in the Hopfield Network

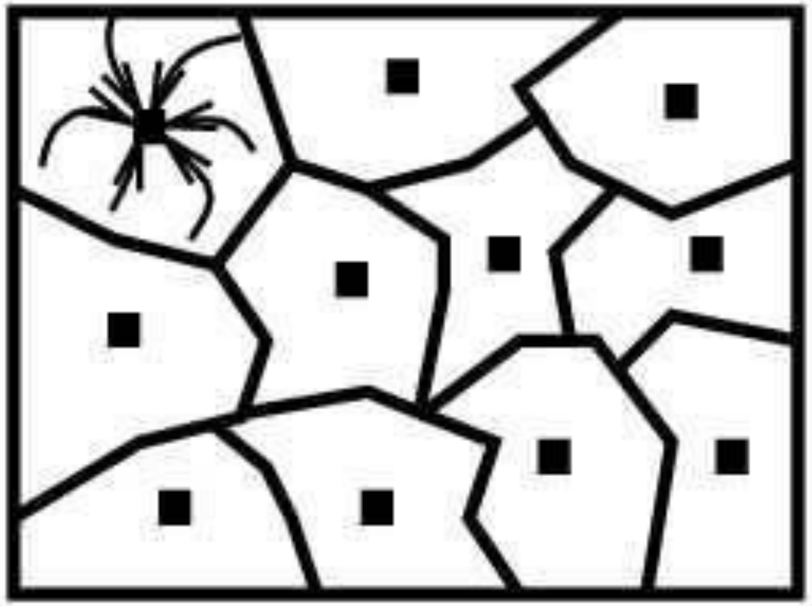


Convergence to stored pattern



Associative Memory Recall

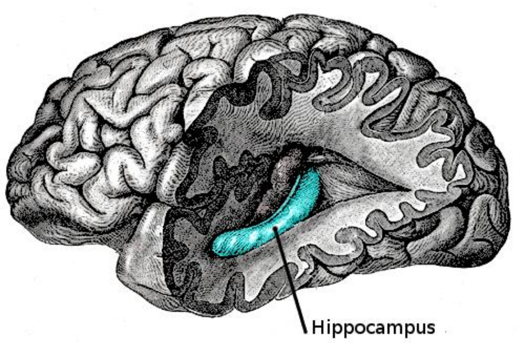


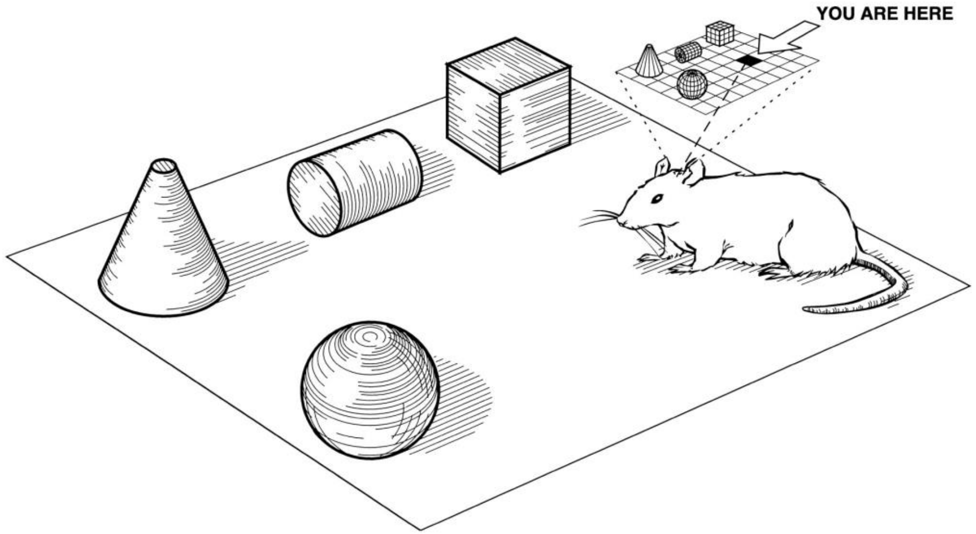


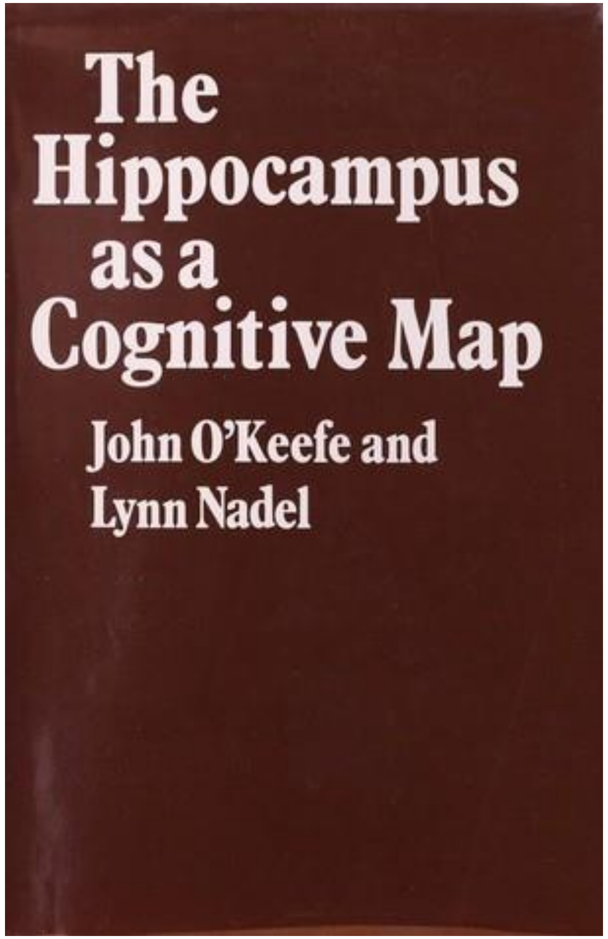
Hopfield Network Summary

1) A model for associative memory
2) Patterns stored in couplings between neurons
3) Uses a Hebbian learning rule
4) These couplings cause activity to converge to a stored pattern (a memory)
5) Storing too many memories causes interference
6) A model for pattern completion/memory in hippocampus

The Hippocampus – Memory or Cognitive Map?







Cell Types in the Cognitive Map

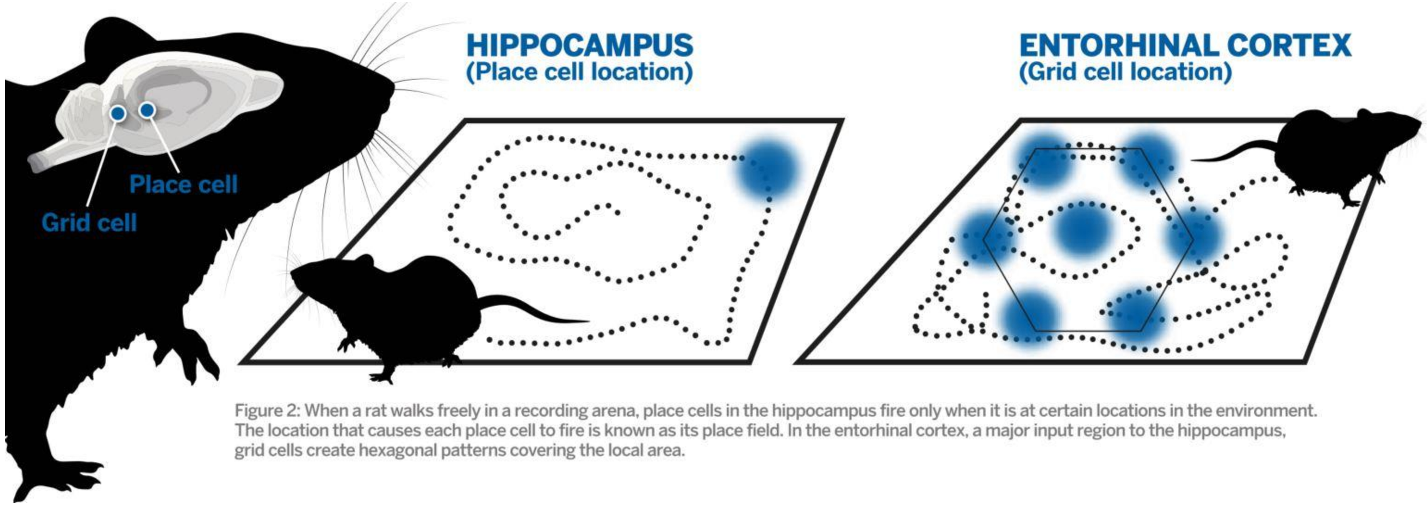
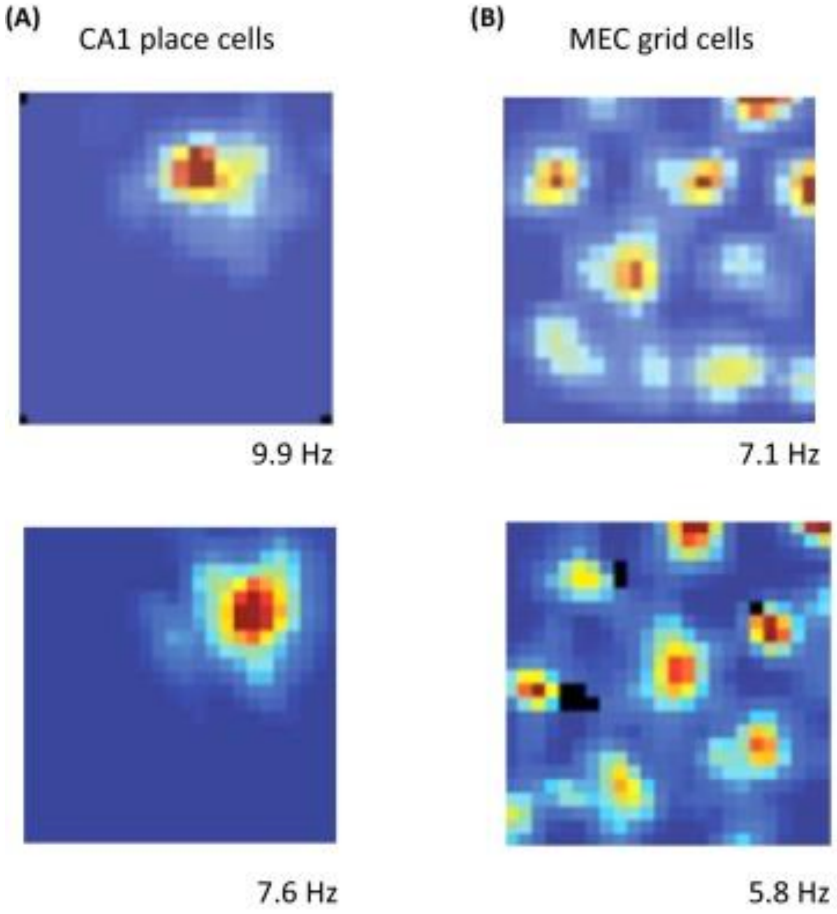


Figure 2: When a rat walks freely in a recording arena, place cells in the hippocampus fire only when it is at certain locations in the environment. The location that causes each place cell to fire is known as its place field. In the entorhinal cortex, a major input region to the hippocampus, grid cells create hexagonal patterns covering the local area.

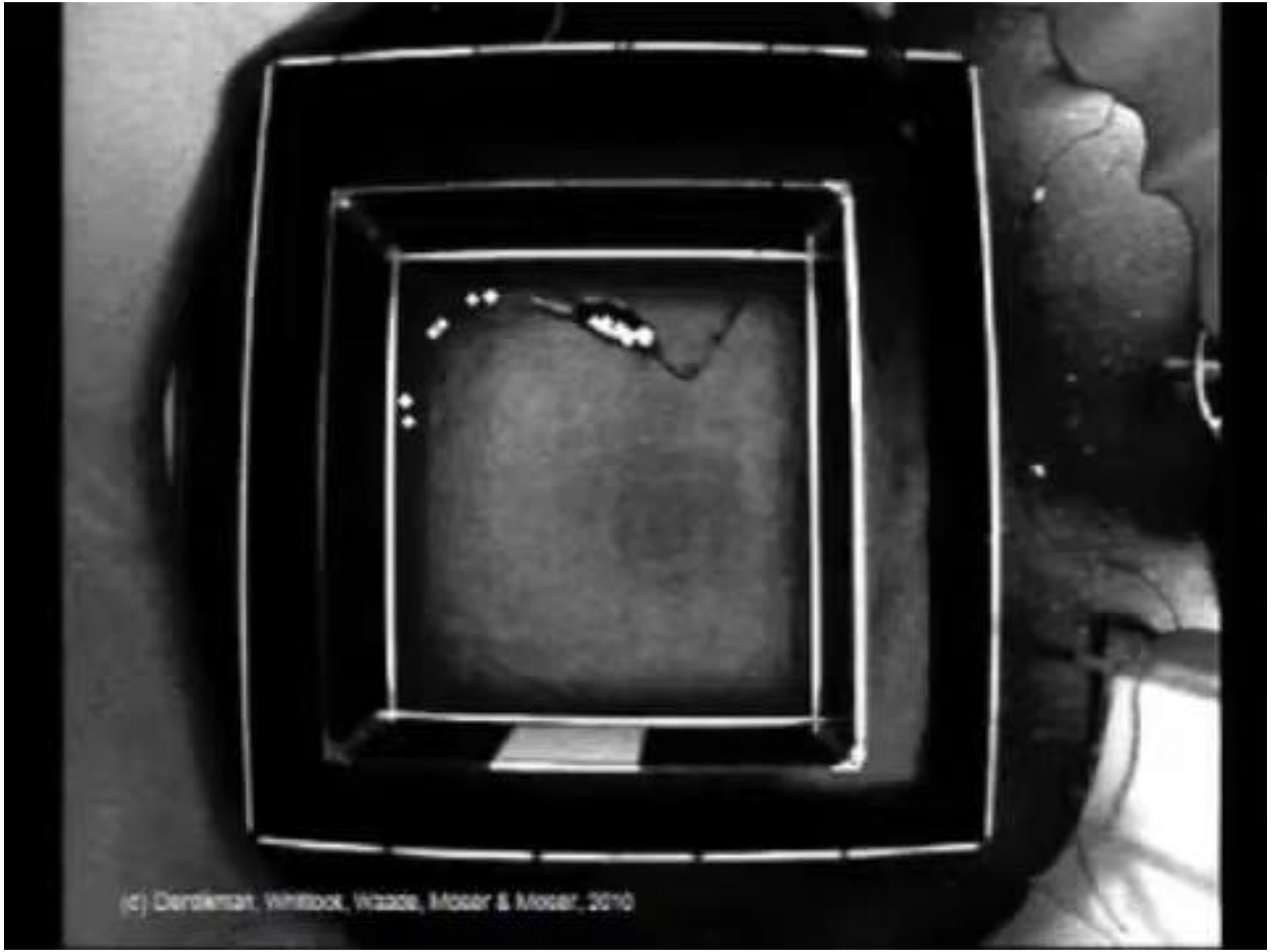


Place cells – fire when the animal visits one location

Grid cells – fire when the animal visits a set of locations

Head direction cells – fire when animal faces a particular direction

Many others spatial cell types…


(c) Derdikman, Whitlock, Waade, Moser & Moser, 2010

Cell Types in the Cognitive Map

The Nobel Prize in Physiology or Medicine 2014

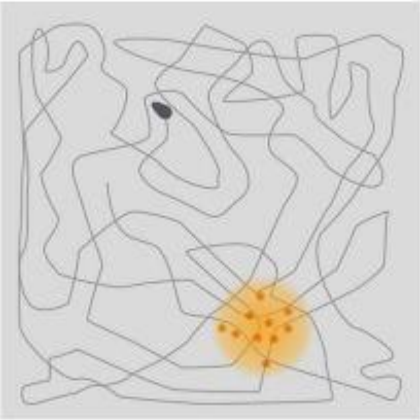


Fig. 1

John O'Keefe

John O'Keefe discovered, in 1971, that certain nerve cells in the brain were activated when a rat assumed a particular place in the environment. Other nerve cells were activated at other places. He proposed that these "place cells" build up an inner map of the environment. Place cells are located in a part of the brain called the hippocampus.

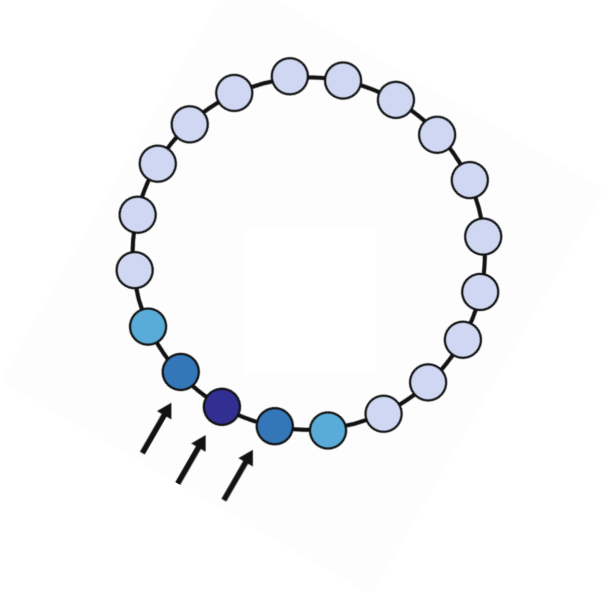
May-Britt Moser and Edvard I. Moser

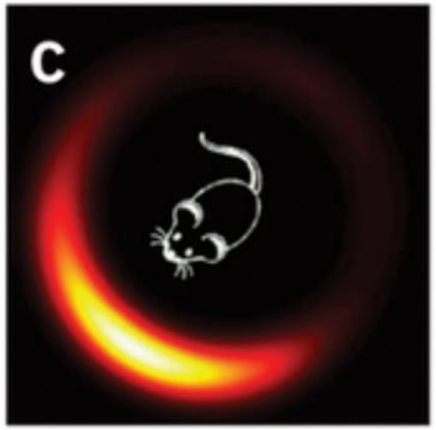
May-Britt och Edvard I. Moser discovered in 2005 that other nerve cells in a nearby part of the brain, the entorhinal cortex, were activated when the rat passed certain locations. Together, these locations formed a hexagonal grid, each "grid cell" reacting in a unique spatial pattern. Collectively, these grid cells form a coordinate system that allows for spatial navigation.

Fig. 2

Attractor Models for Place Cells and Grid Cells

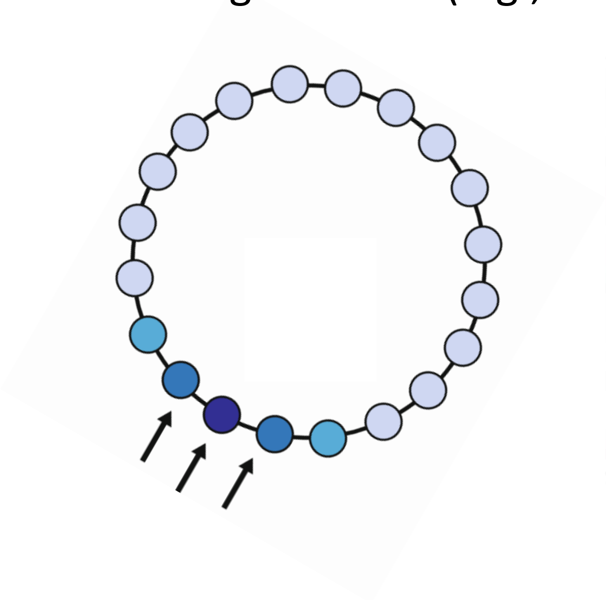
Ring attractor (e.g., head direction cells)

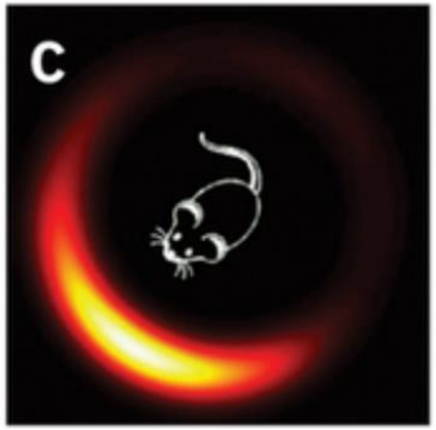




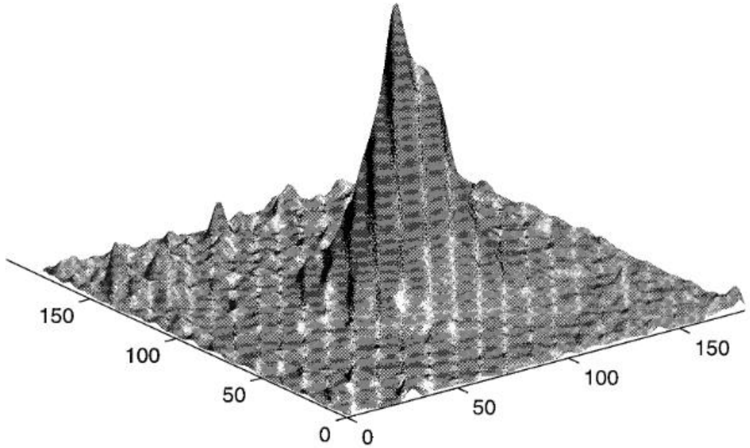
Attractor Models for Place Cells and Grid Cells

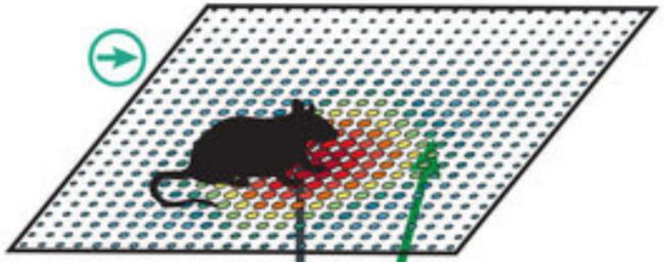
Ring attractor (e.g., head direction cells)





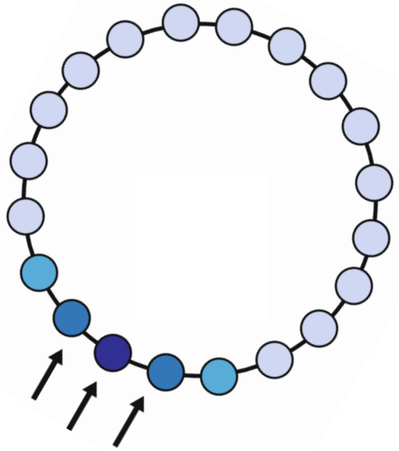
Sheet Attractor (e.g., place cells)

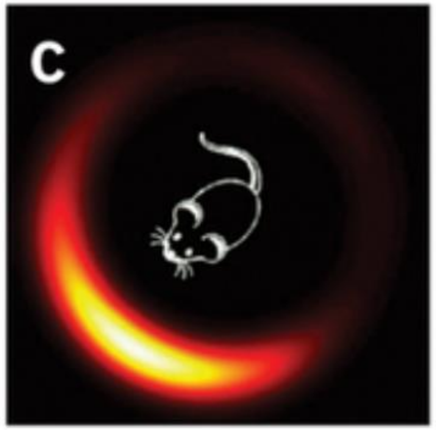




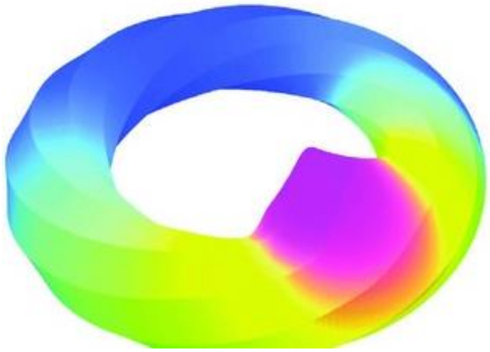
Attractor Models for Place Cells and Grid Cells

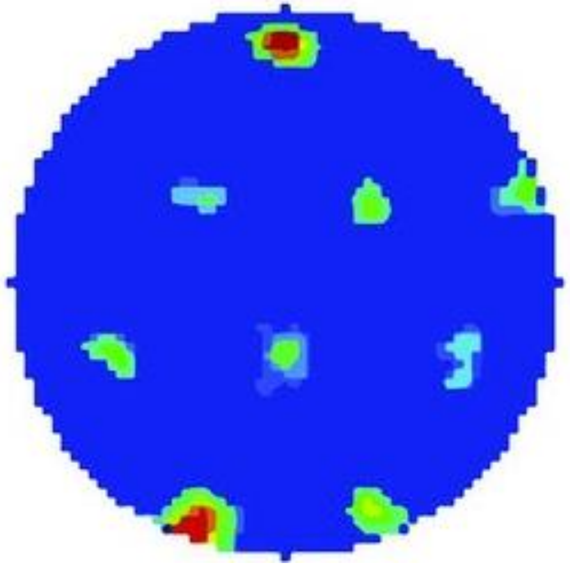
Ring attractor (e.g., head direction cells)



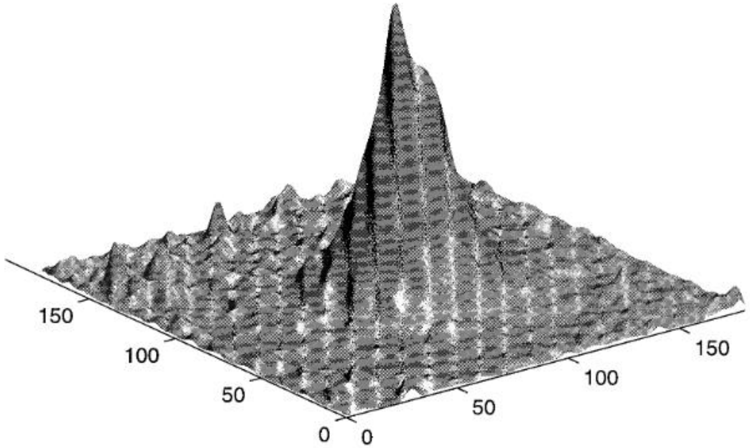


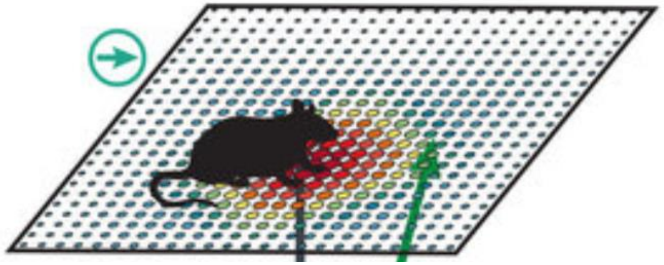
Torus attractor (e.g., grid cells)





Sheet Attractor (e.g., place cells)





Spatial Memory Summary

1) Hippocampus contains representations of space/location

2) Can be understood as a form of spatial memory (memory for different places, contexts, etc.)

3) The exact role of hippocampus in navigation, memory, etc. is still debated

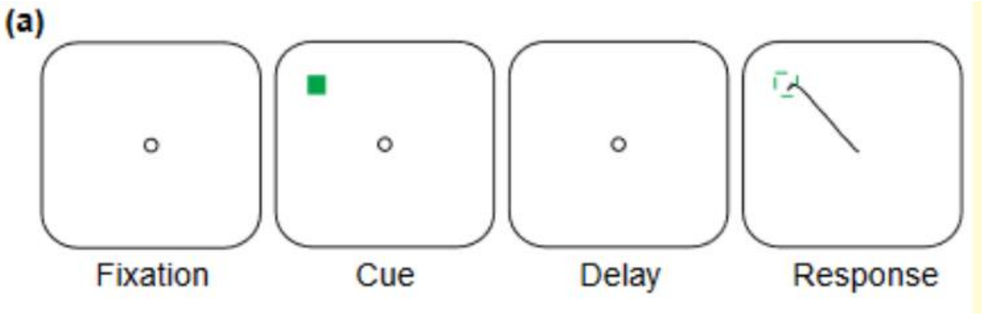
Working Memory

Review TRENDS in Cognitive Sciences Vol.7 No.9 September 2003

Persistent activity in the prefrontal cortex during working memory

Clayton E. Curtis[1] and Mark D'Esposito[2]

[1]New York University, Department of Psychology, 6 Washington Place, Room 859, New York, NY 10003, USA
[2]Helen Wills Neuroscience Institute and Department of Psychology, Henry H. Wheeler Jr Brain Imaging Center, University of California, 3210 Tolman Hall, Berkeley, CA 94720-1650, USA



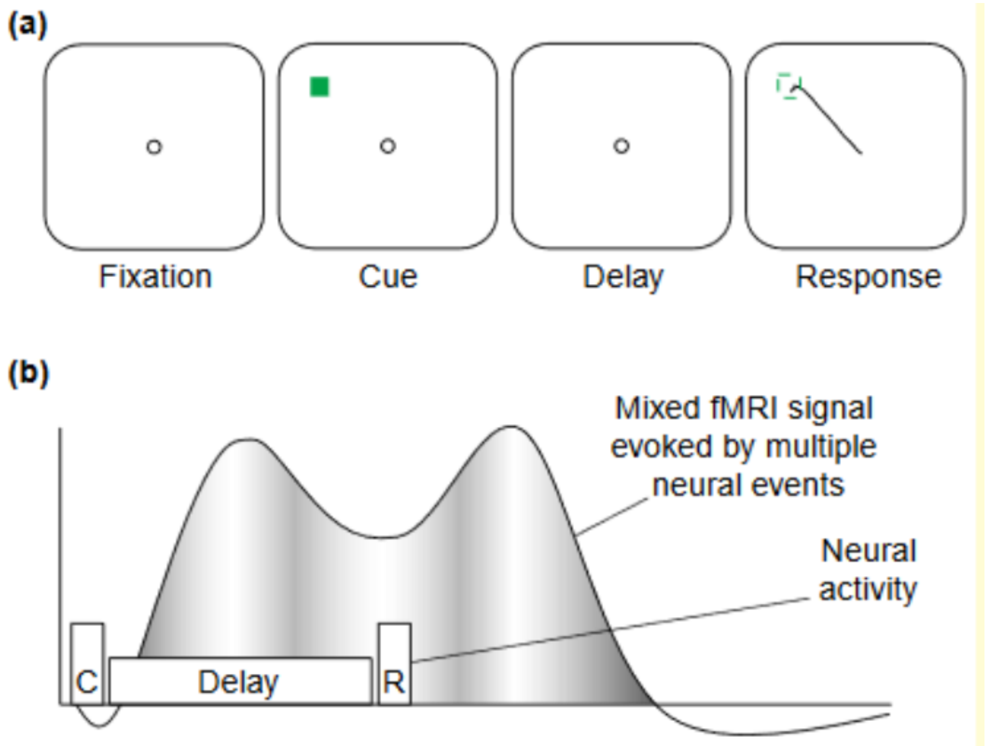
Working Memory

Review TRENDS in Cognitive Sciences Vol.7 No.9 September 2003

Persistent activity in the prefrontal cortex during working memory

Clayton E. Curtis[1] and Mark D'Esposito[2]

[1]New York University, Department of Psychology, 6 Washington Place, Room 859, New York, NY 10003, USA
[2]Helen Wills Neuroscience Institute and Department of Psychology, Henry H. Wheeler Jr Brain Imaging Center, University of California, 3210 Tolman Hall, Berkeley, CA 94720-1650, USA



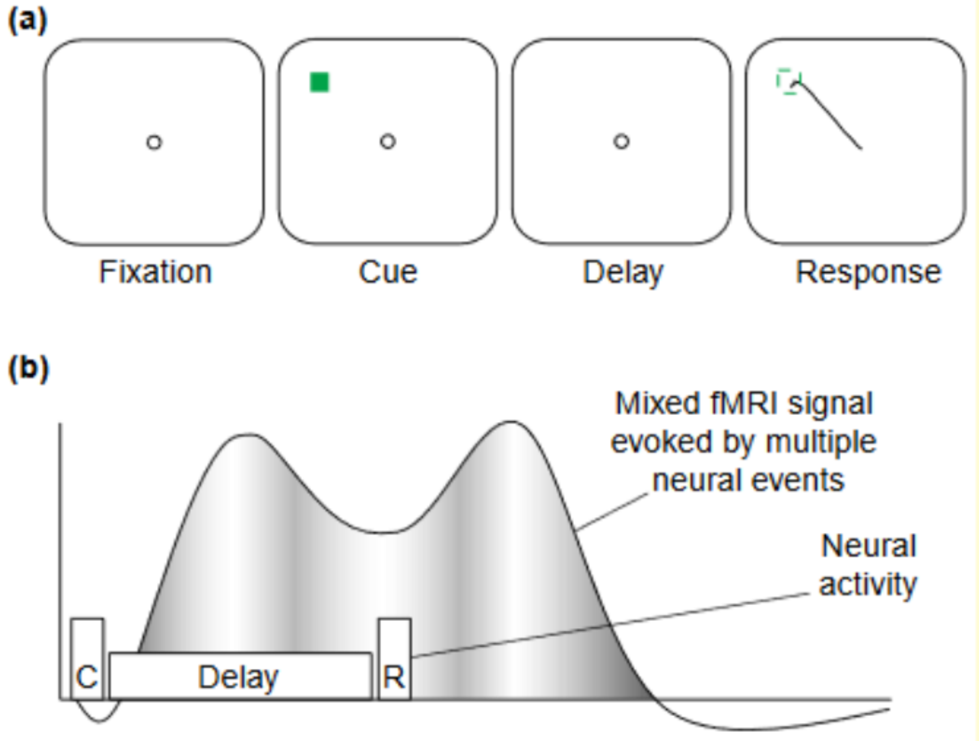
Working Memory

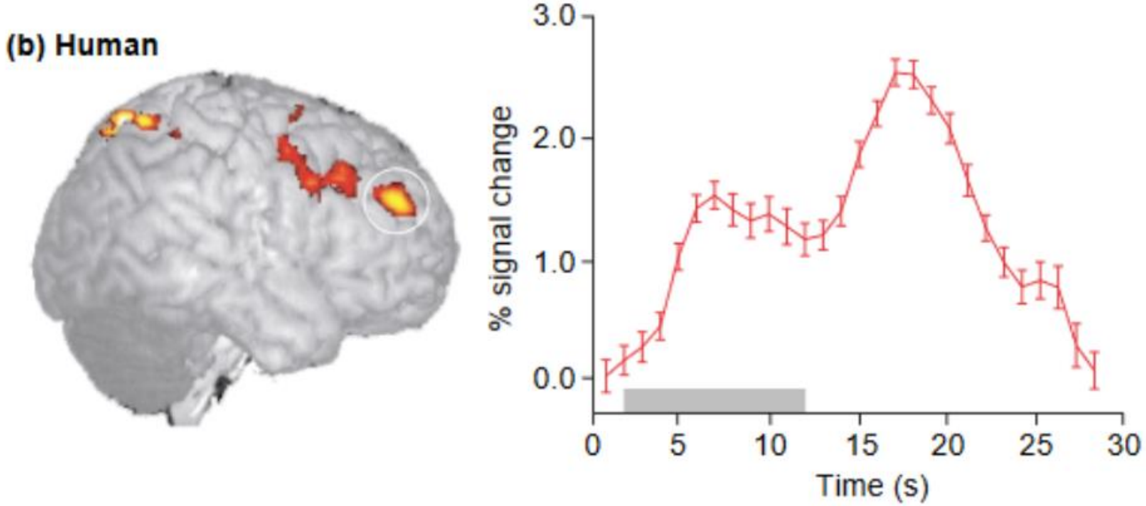
Review TRENDS in Cognitive Sciences Vol.7 No.9 September 2003

Persistent activity in the prefrontal cortex during working memory

Clayton E. Curtis[1] and Mark D'Esposito[2]

[1]New York University, Department of Psychology, 6 Washington Place, Room 859, New York, NY 10003, USA
[2]Helen Wills Neuroscience Institute and Department of Psychology, Henry H. Wheeler Jr Brain Imaging Center, University of California, 3210 Tolman Hall, Berkeley, CA 94720-1650, USA





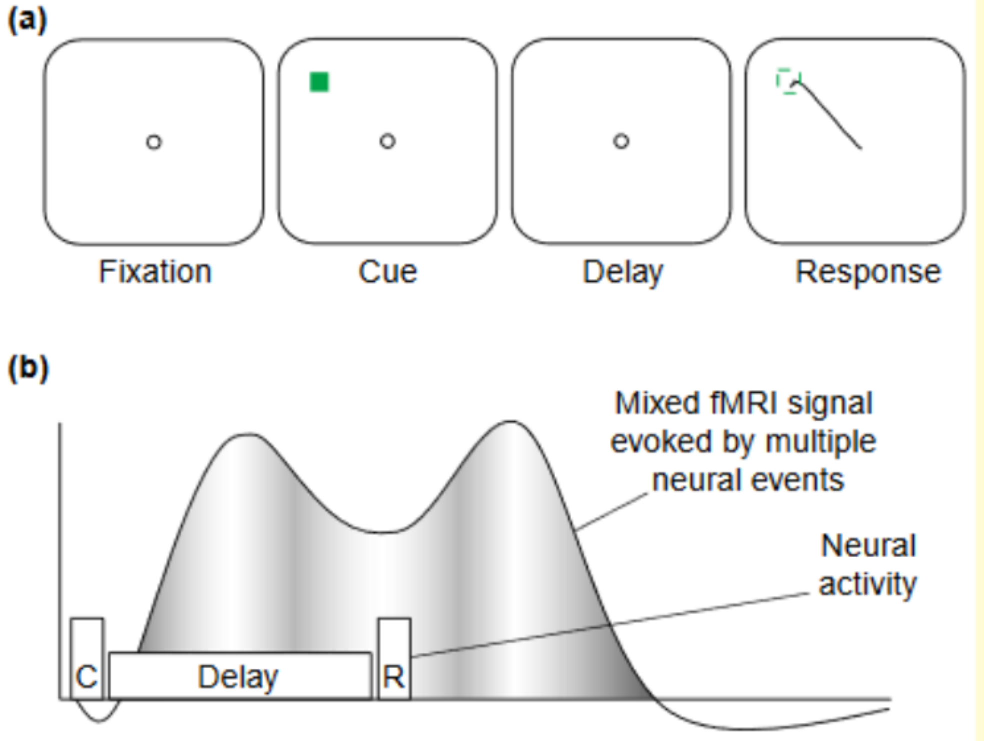
Working Memory

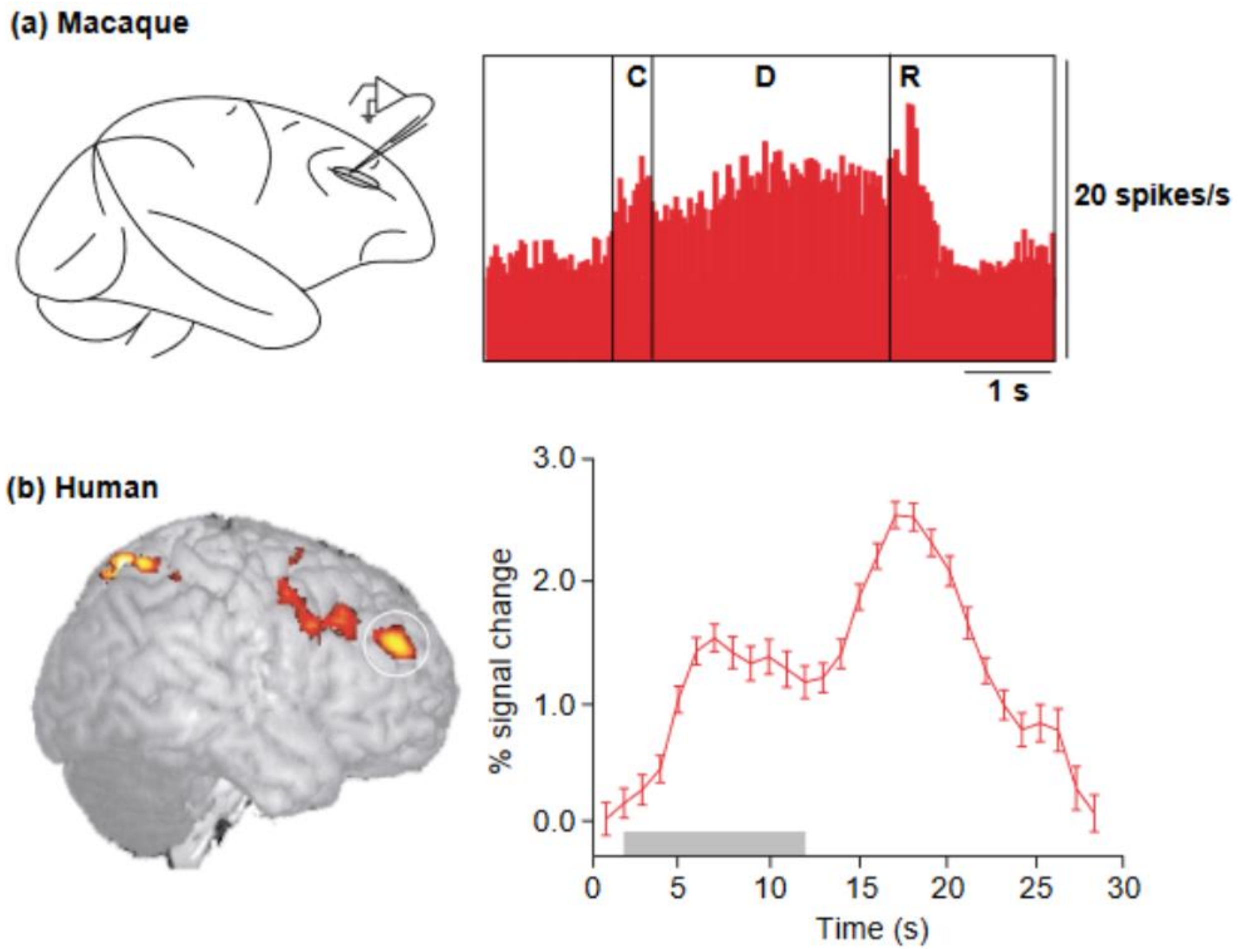
Review
TRENDS in Cognitive Sciences Vol.7 No.9 September 2003

Persistent activity in the prefrontal cortex during working memory

Clayton E. Curtis[1] and Mark D'Esposito[2]

[1]New York University, Department of Psychology, 6 Washington Place, Room 859, New York, NY 10003, USA
[2]Helen Wills Neuroscience Institute and Department of Psychology, Henry H. Wheeler Jr Brain Imaging Center, University of California, 3210 Tolman Hall, Berkeley, CA 94720-1650, USA

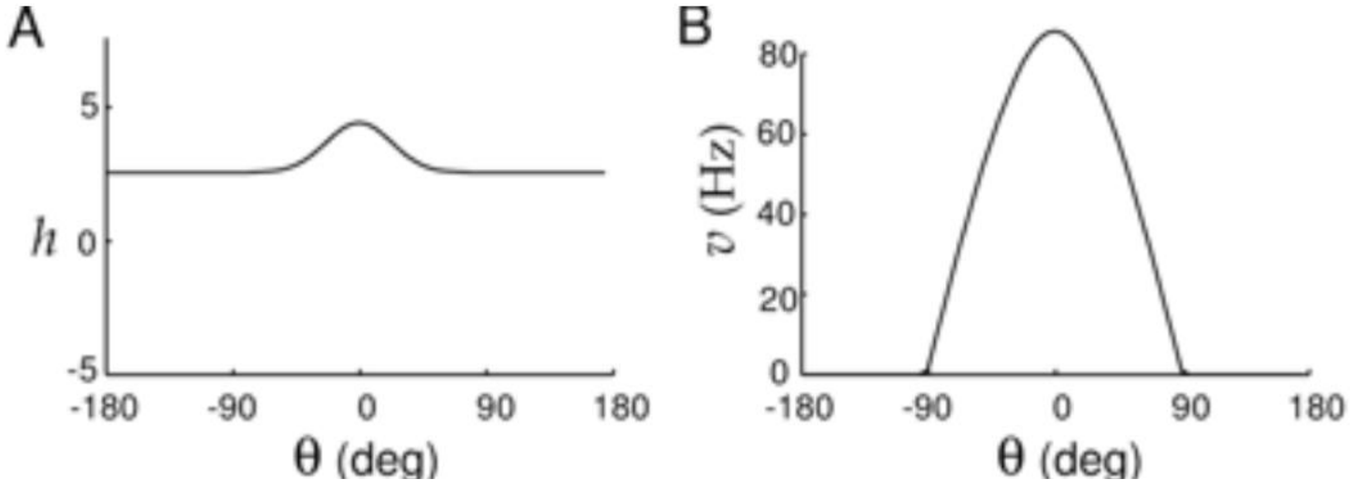




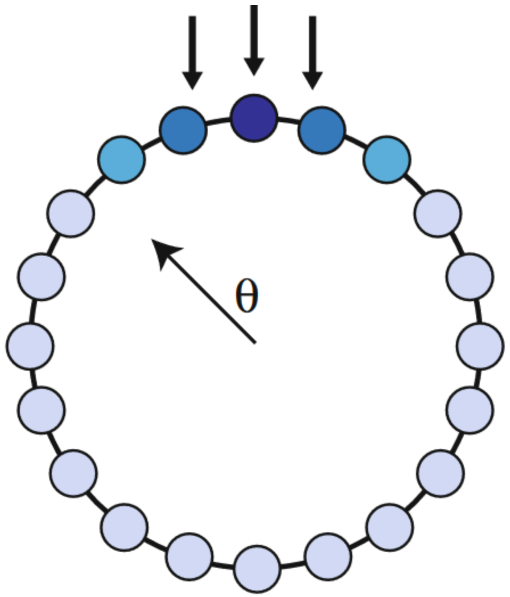
Working Memory – Ben-Yishai Ring Model

- Recall the Ben-Yishai model for orientation tuning in V1

Cue

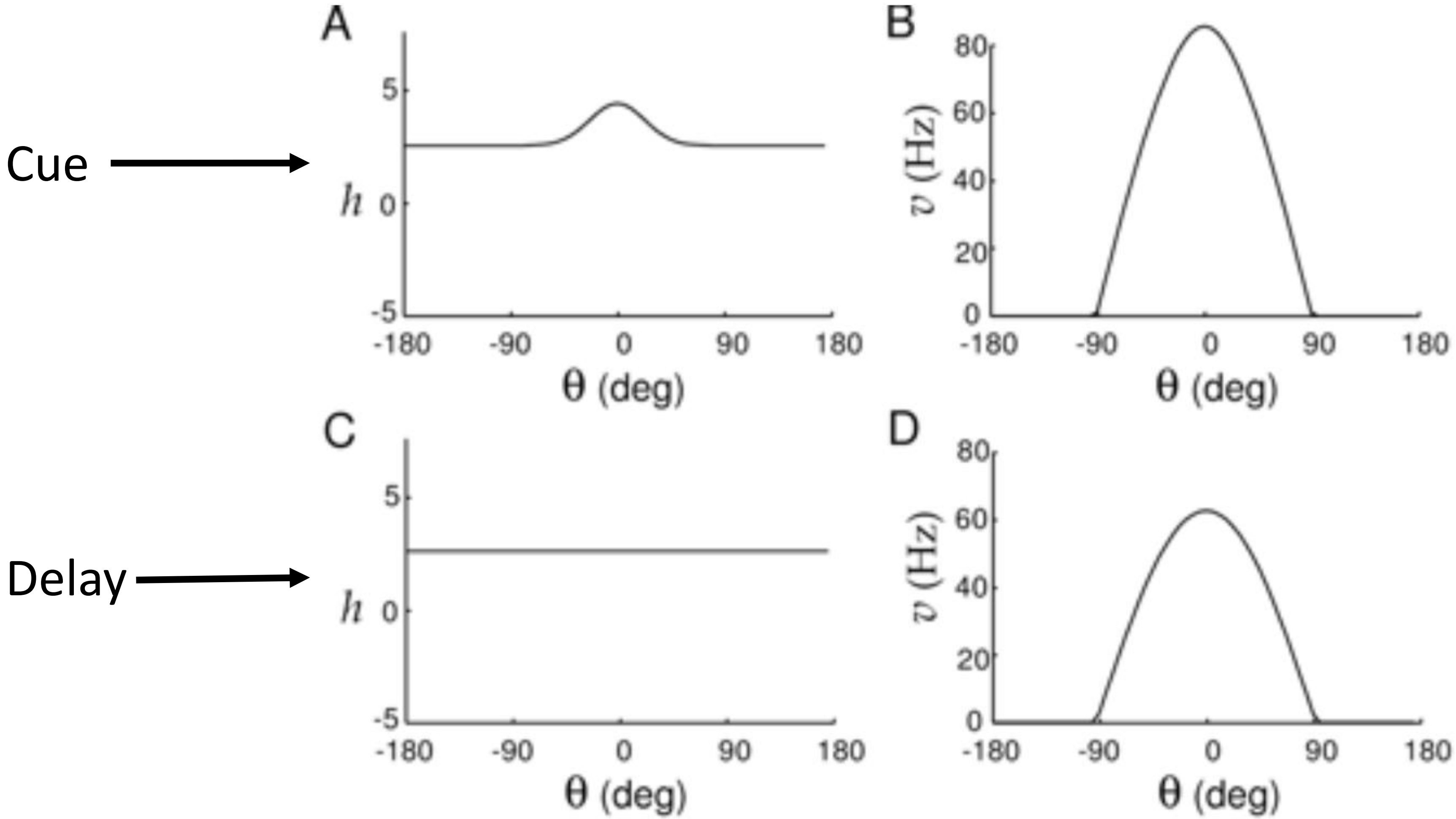


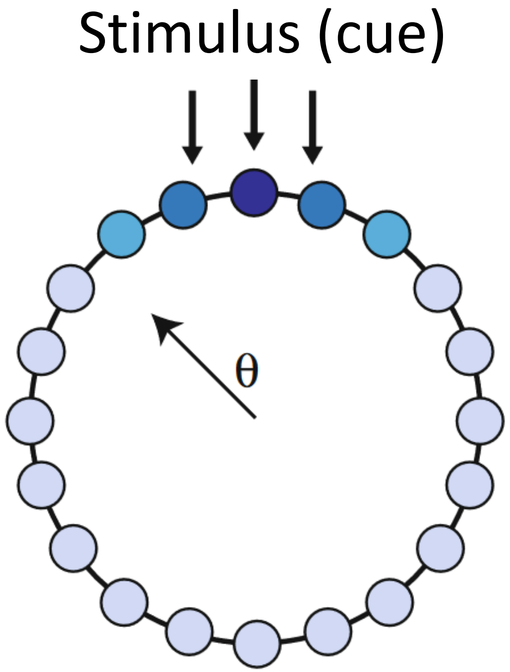
Stimulus (cue)



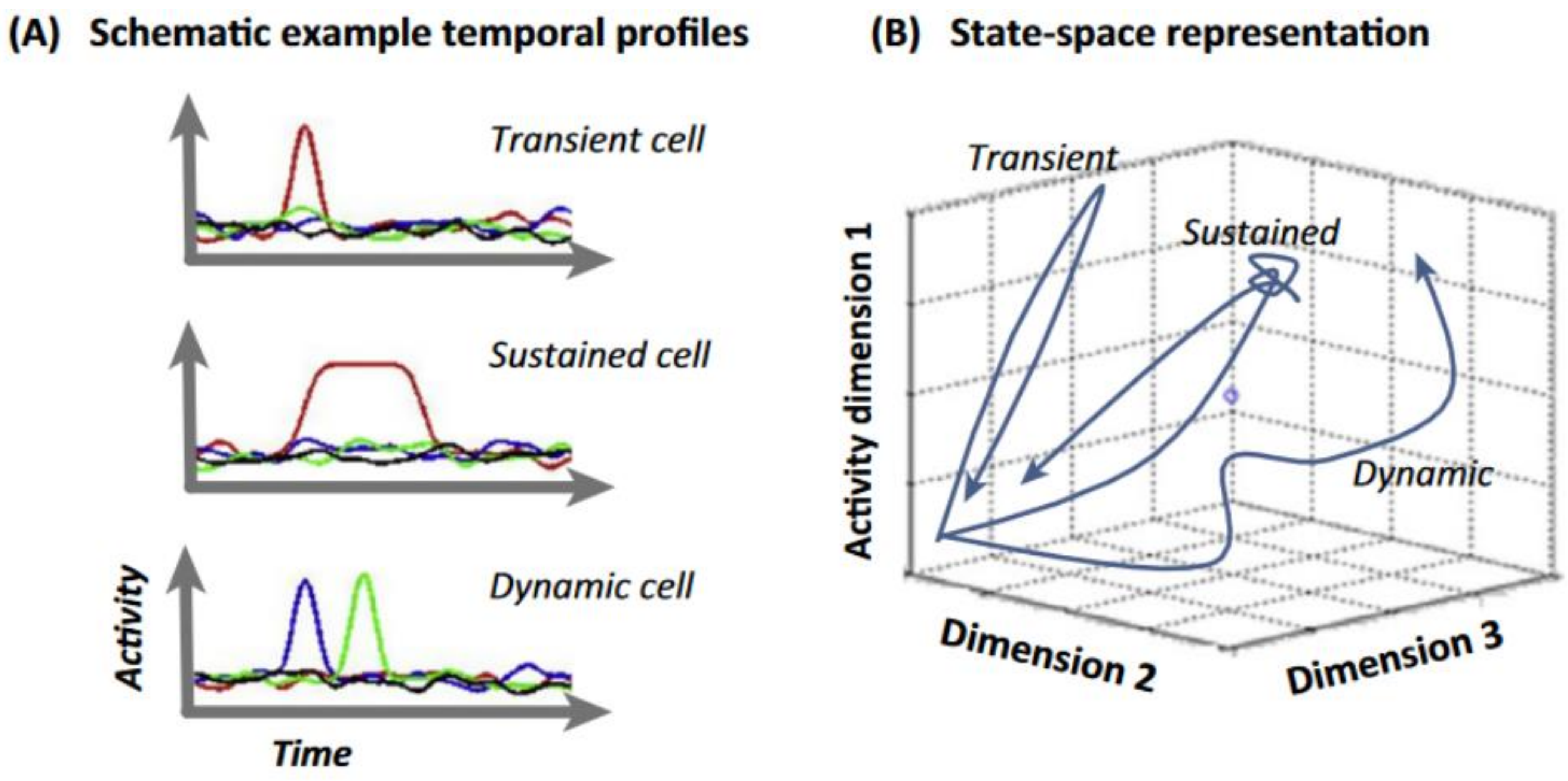
Working Memory – Ben-Yishai Ring Model

- Recall the Ben-Yishai model for orientation tuning in V1
- When the stimulus is removed, the network maintains the activity through recurrent dynamics (an attractor)

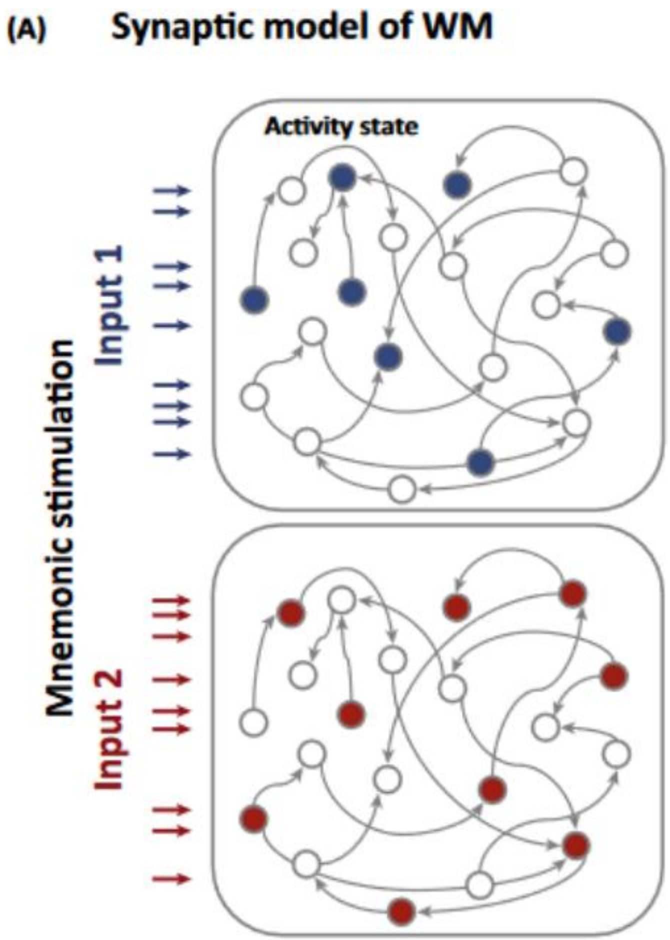




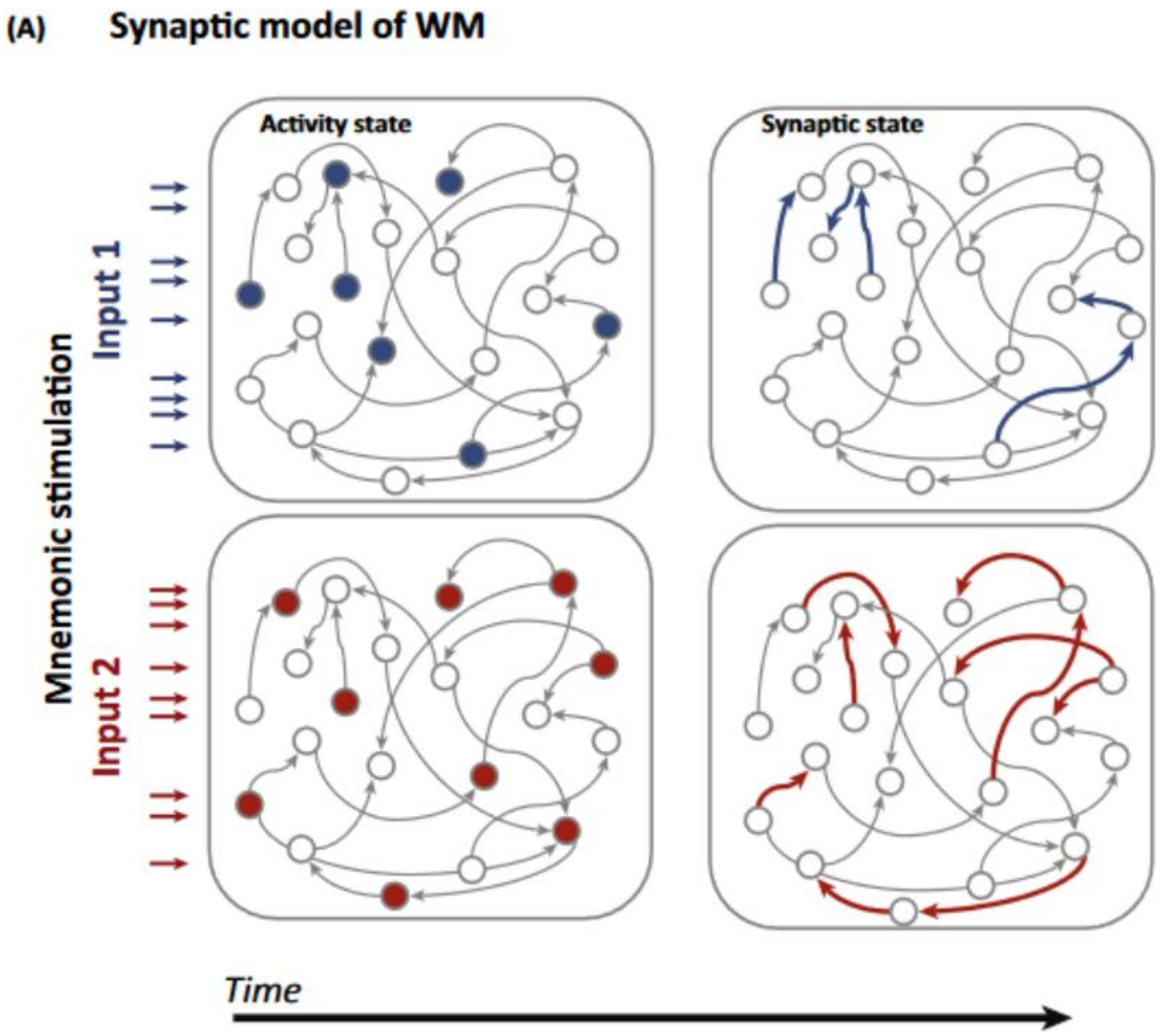
Working Memory – Dynamic Coding Models



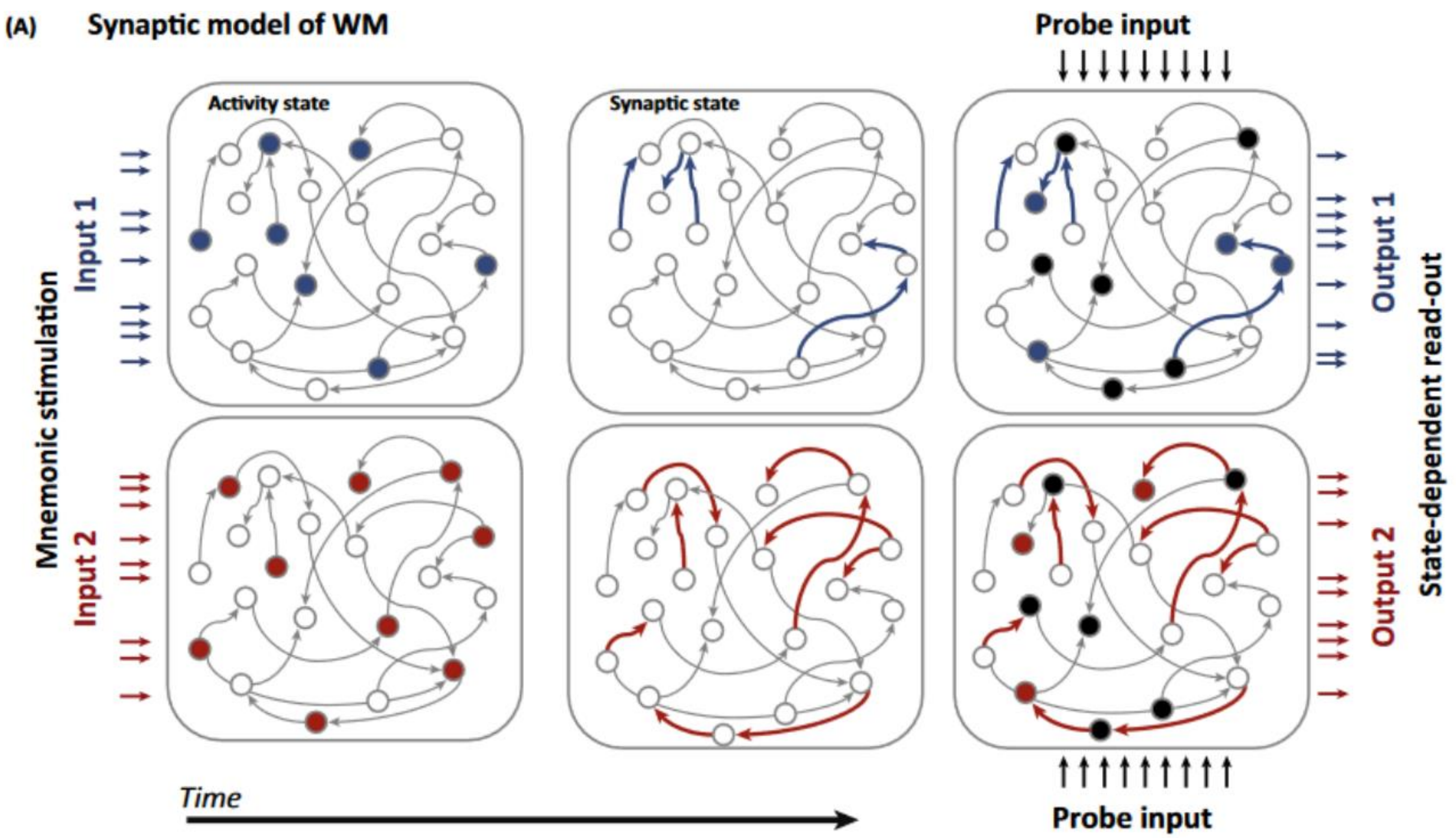
Working Memory – Activity-Silent Models



Working Memory – Activity-Silent Models



Working Memory – Activity-Silent Models



Working Memory Summary

1) Involves persistent activity in neural circuits during delay period

2) Various brain regions involved, including prefrontal cortex

3) Three major hypotheses: attractor models, dynamic coding models, activity-silent models

4) Probably involves a combination of all three mechanisms

Summary of Lecture

- Various kinds of memory – short vs long, implicit vs explicit etc.
- Memory is linked to persistent activity (network dynamics) and synaptic plasticity
- Long term memory involves temporary storage in hippocampus and then transfer to cortex (standard model)
- Spatial memory and navigation are well-studied as a model system (in hippocampus and related brain systems)
- Working memory involves persistent activity in recurrent networks

Further Reading (optional)

- Dayan and Abbott (Chapter 8)
- Gerstner (Chapter 17,19)
- Eichenbaum – The Cognitive Neuroscience of Memory