Learning, memory & how we come up with new ideas

Jaan Aru
jaan.aru@gmail.com
Learning objectives

How does learning unfold?
Which memory systems are there?
How do memory systems work together?
What is the role of sleep?
What is novelty?
How do we come up with novel ideas?
max hundreds of milliseconds
Mammalian memory systems

Long-term memory

Declarative memory
- Episodic memory
- Semantic memory

Nondeclarative memory
- Procedural memory: skills, habits
- Priming
- Simple classical conditioning
- Habituation, sensitization
- Reflex pathways

Neocortex
- Basal ganglia
- Medial temporal lobe, diencephalon
- Amygdala, cerebellum
Encoding requires attention

- Spikes per second vs. Time from stimulus onset (ms.)
- Graph showing preferred and poor stimuli with increased spikes during attend periods.
Attention is (usually) a prerequisite for learning & memorization
The most famous patient: H.M.
Patient H.M.

Hippocampus
Medial temporal lobe

HM

Normal Brain

- Errors vs. Days
- Trials vs. Errors
Types of memory

- Long-term memory
  - Procedural memories ("Knowing how")
  - Declarative memories ("Knowing that")
    - Semantic memories (General knowledge)
    - Episodic memories (Personal recollections)
Patient H.M. demonstrates that ... 

1) medial temporal lobe is important for memory
2) but it is not necessary for all memory!
3) there is a difference between short-term and long term memory
4) there is a difference between declarative and procedural memory
5) medial temporal lobe is necessary for long term declarative memory
Taxonomy of long-term memory

- **Declarative memory**
  - Episodic memory
  - Semantic memory

- **Nondeclarative memory**
  - Procedural memory: skills, habits
  - Priming
  - Simple classical conditioning
  - Habituation, sensitization
  - Reflex pathways

- **Medial temporal lobe, diencephalon**
- **Basal ganglia**
- **Neocortex**
- **Amygdala, cerebellum**
Time point of brain injury

Past

Preserved memories • Lost memories

Future

New information

Retrograde Amnesia

Anterograde Amnesia
complementary learning systems (CLS)

intelligent agents must possess two learning systems

neocortex and hippocampus.

NC **gradually** acquires structured knowledge (essentially deep learning)

HC quickly learns the specifics of **individual experiences**.
Cortical modules

Hippocampus

Time
Cell ensembles contributing to...

- Recently encoded neocortical part of a representation
- Recently encoded hippocampal part of a representation
- Associated pre-existing representation
- Unrelated pre-existing representation
**The Nobel Prize in Physiology or Medicine 2014**

**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.

![Fig. 1](image)

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](image)
Place cell activity marks the “state” (most likely not only location!)
A. Virtual reality spatial navigation task

B. Reward zone

C. Large scale readout
Two-photon calcium imaging

D. Functionally defined stimulation
Two-photon targeted optogenetics

E. Reward zone place cell stimulation increases licking in the stimulation zone
Cell ensembles contributing to...
- Recently encoded neocortical part of a representation
- Recently encoded hippocampal part of a representation
- Associated pre-existing representation
- Unrelated pre-existing representation
Replay of Hippocampal Memories
Replay of Hippocampal Memories

Circumventing the Statistics of the Environment
All the experiences don’t have the same priority!
replay is biased towards certain events
Novel, significant, emotional, rewarding
reweighting experiences shapes neocortical learning
For the cortex it is the same
We are novelty seekers

Novelty is intrinsically motivating
What kind of novelty matters?
The value (reward) of novel information depends on prior knowledge
I invented GANs back in the ‘90s
CHANGE MY MIND
LSTM

- forget gate
- cell state
- input gate
- output gate

sigmoid
tanh
pointwise multiplication
pointwise addition
vector concatenation
Size of the intrinsic reward = Amount of “compression progress, that is, the number of saved bits.”

Even reading this sentence will have different reward depending on a person!
“History of science is the history of compression progress.”
Jürgen Schmidhuber
Schmidhuber's Beauty Postulate (1994-2006): Among several patterns classified as "comparable" by some subjective observer, the subjectively most beautiful is the one with the simplest (shortest) description, given the observer's particular method for encoding and memorizing it.
The laws of physics must have mathematical beauty!
Minimize description length
Compression = comprehension
Finding a shorter description is rewarding ...
Hypothesis 1: new ideas (in science, innovation, arts) arise through compression
Hypothesis 2: this process can happen spontaneously (i.e. it is not dependent on external world!)
Henri Poincaré: “... we entered an omnibus to go some place or other. At the moment when I put my foot on the step, the idea came to me, without anything in my former thoughts seeming to have paved the way for it.”
Hypothesis 3: this process is highly rewarding