Recoding of Patterns Between Hippocampus and Neocortex

David S. Touretzky

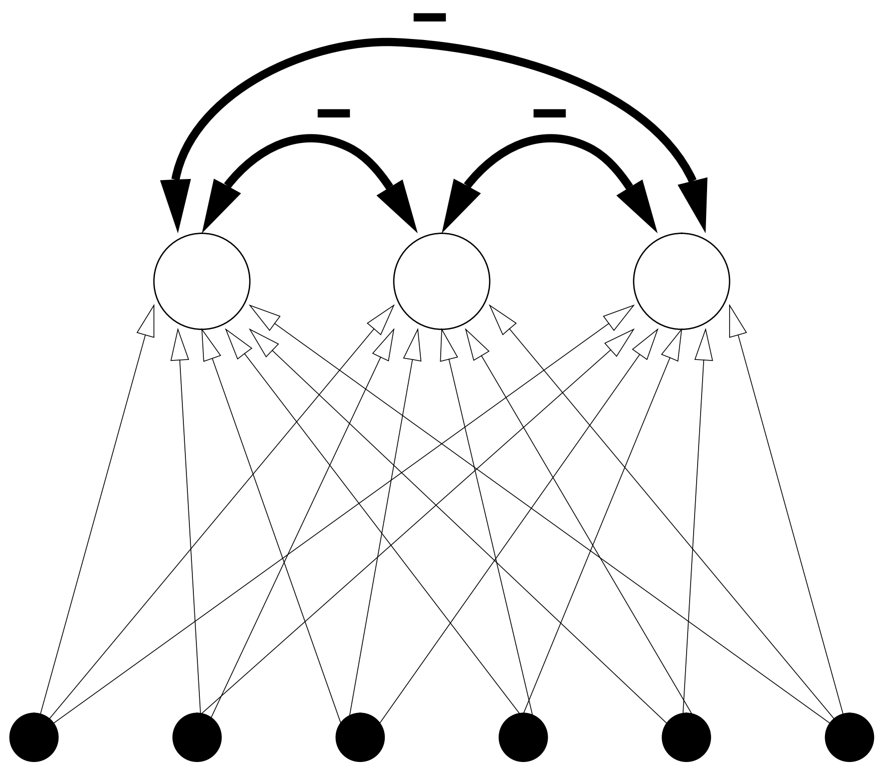
Rolls

- Pattern separation in CA3 due to low connectivity from DG:
  - DG fanout: only 14 mossy fiber synapses from each granule cell axon onto pyramidal cells in CA3
  - CA3 fanin: only 80 synapses from granule cells onto each pyramidal cell dendrite
- Training of perforant path inputs to CA3?
  - Roughly 2300 pp synapses per CA3 pyramidal dendrite
  - If mossy fiber synapse fires the cell, pp synapses can be modified by cooperative Hebbian learning
- Pattern separation in Dentate Gyrus:
  - Competitive learning in winner-take-all subnetworks.
  - Inhibitory cells in DG could produce WTA behavior.

First Rolls Simulation

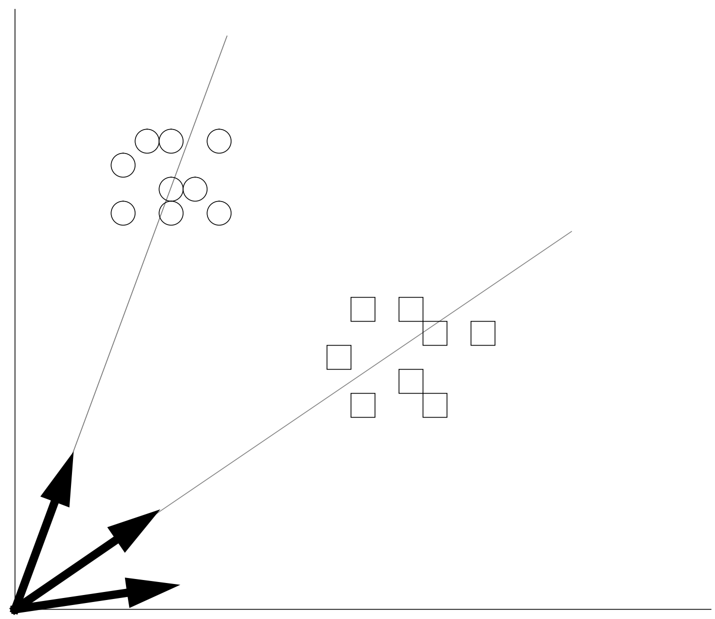
- Model of EC to DG projection.
- Looks like an associative matrix memory, but ...
- Not a hetero-associator, because no output patterns are specified.
- Not an auto-associator. No recurrent connections.
- It’s actually a competitive learning architecture.
- Lateral inhibition among the output units keeps the output vector sparse.
- Simulation results show that input patterns with high overlap produce output patterns with less overlap.
- Output patterns will be more orthogonal; better separated.

Competitive Learning



- Winner-Take-All dynamics through lateral inhibition.
- Hebbian learning on the input connections, with weights constrained so that weight vectors are all unit vectors.
- Result: input patterns are grouped into classes.

Competitive Learning



- Each unit's weight vector moves to the centers of the cluster of input patterns that unit captures.

- Some units might not capture any patterns.

Representations in CA3 and CA1

- Sparse connectivity from DG to CA3 means patterns will be even more orthogonal in CA3.
- Assume that an "episodic memory" is a collection of events.
- Its CA3 representation is a union of the activity patterns encoding the individual events.
- Recurrent connections in CA3 allow the entire episode to be recalled based on a small subset of the original events.
- Are these events simultaneously occurring in different sensory modalities: visual, auditory, olfactory, gustatory, tactile?
- Or can the events be temporally adjacent rather than simultaneous?
- CA1 uses competitive learning to form a unified, sparse representation of the entire episodic memory.
- This CA1 representation can be used to train the neocortex.

This is a highly speculative theory.

Inputs to Hippocampus

- Hippocampus receives cortical input from:
  Entorhinal cortex
  Parahippocampal gyrus

- These areas in turn receive input from:
  visual and auditory temporal association cortex
  parietal cortex (concerned with space)
  frontal cortex (goal-oriented behavior? consciousness?)
  amygdala (reinforcements; emotion)

- Subcortical inputs to hippocampus:
  cholinergic input from medial septum and diagonal band of Broca, via the fimbria/fornix

  other modulatory inputs

- Experiments on monkeys show HC cells sensitive to particular object-at-place stimulus combinations: a multimodal conjunction.

How Hippocampus Guides Cortex

- Unimodal sensory areas can't determine the significance of various input patterns.

- The hippocampal system learns multimodal patterns representing events.

- CA1 provides a unified representation for events, using competitive learning to separate dis-similar patterns and group similar ones together.

- This acts as a classification function for unimodal patterns: which ones are important? Which ones should be regarded as similar to which others?

- Feedback from CA1 to cortex can be used to fine-tune the unimodal representations to focus on their significant features.

- Example: food vs. toxic substances:
  dark brown: delicious
  med. brown with stripes: poisonous
  med. brown with speckles: edible, but not great
  light brown with swirls: pretty good

  any shade/texture of purple: inedible, harmless

Learning in Cortex

- Assume an association matrix memory in unmodal sensory association cortex similar to what was posited for DG.
- Competitive learning process mediated by cortical inhibitory interneurons.
- Backprojections from CA1 via subiculum and EC provide additional dendritic input to the excitatory cells, encriching the activity patterns.
- Result: if HC classifies two stimulus patterns the same way, the identical backprojections will make the stimuli look even more similar to cortex.
- If HC makes a distinction between the two patterns, e.g.

  med. brown/speckles + okay taste + pleasure ==> event1
  med. brown/striped + bitter taste + nausea ==> event2

  then the two patterns will generate different backprojections, and will tend to be coded differently in the cortex.

Sources of Guidance to Cortex

1. Cholinergic modulation from basal forebrain.
   Noradrenergic modulation from locus coeruleus.

   - These are sparse inputs: global reinforcement signals (“strobes”) that note behaviorally relevant events.

2. Backprojections from HC and amygdala. These are dense enough that they can transmit specific patterns rather than just a reinforcement signal.

   - Amygdala: associations between a neutral stimulus (vision) and a reinforcer (smell of food).

   - Hippocampus: associations between neutral stimuli, such as objects and spatial positions.

3. Backprojections from higher cortical areas. Teaches the lower areas how to build better filters, making them more useful to the higher areas.

   Note: individual backprojections cover a wider area of cortex (several mm^2) than forward projections.

Recall

- Rolls suggests that the backprojections alone could be enough to active a memory in a cortical area.
- Example: smell of food causes a backprojection to visual cortex, recalling the appearance of that food.
- Evidence from blood flow studies in humans suggests that early visual association areas are activated during recall of visual (but not auditory) information.

- Backprojections could also produce a sort of attentional modulation effect, guiding which aspects of the input are treated as significant.
- Semantic priming could be done this way as well. Example: show picture of a boat, ask subject for meaning of “deck”.
- Retrograde amnesia: if cortical learning requires many cycles of training between areas, then HC damage could produce graded amnesia: greater loss for events closer in time to the occurrent of damage.

Assessment of Rolls

- Recapitulates Marr's ideas about hippocampus training the neocortex.
- Cortical self-training via backprojections is related to various unsupervised learning ideas in the ANN literature.
- Supervised learning is also present, in the form of reinforcers (food, pain) influencing the amygdala and subcortical structures.
- The simulations reported here aren't very convincing. Too small, too simplistic.
- More recent work by other investigators further explores these ideas:

  pattern recoding

  one brain area training another

Gluck and Myers

Representational changes assumed in the hippocampal system:

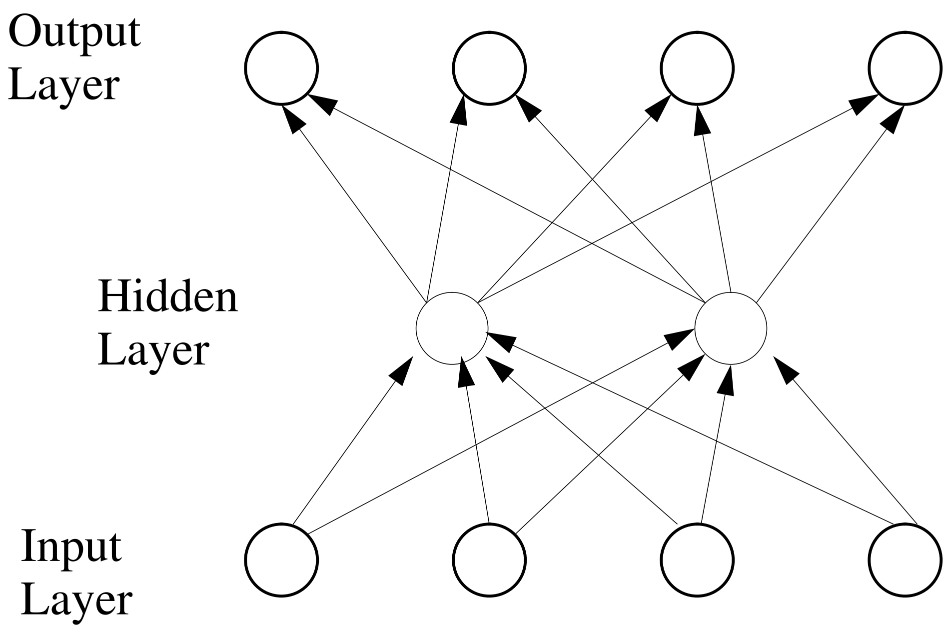
	Bias to Compress	Bias to Differentiate
Stimulus-stimulus regularity	IF stimuli reliably co-occur	IF stimuli reliably occur separately
Stimulus-outcome regularity	If stimuli reliably predict the same future event	If stimuli reliably predict different future events

Example: stimulus A occurs in context X. No reward.

- A predicts X, so the representations of stimuli A and X are made more similar: A/X. Stimulus-stimulus compression.
- A and X both predict no reward: stimulus-outcome compression.
- But X occurs both with and without A, so there is a tendency to separate A from X: stimulus-stimulus differentiation.

Which effect wins? Depends on input statistics.

Auto-Encoders



- Units compute a linear sum of their inputs and pass it through a non-linear squashing function (soft threshold.)
- Weights are trained by error back-propagation: a generalization of the Widrow-Hoff algorithm.
- With an auto-encoder: output pattern = input pattern.
- Hidden layer forms a compressed representation of the input.

Lesion Experiments

- Lesion H^{++}: remove or disable entire hippocampal system, including entorhinal and paraphippocampal cortices.

- Lesion H: remove just the hippocampal formation; leave entorhinal cortex intact.

- Lesion techniques in animals:

  aspiration (lots of damage to surrounding areas)

  electrolytic destruction

  chemical (ibotenic acid): destroys cell bodies but leaves fibers of passage intact

- H^{++} lesioning of the model: delete the auto-encoder. Eliminate any training of input-to-hidden connections in the feed-forward nework.

- H-lesioning of the model: replace auto-encoder with competitive learning network that does clustering but not pattern separation.

Effects of Stimulus Compression

- If stimuli A and B occur together, their representations will be compressed (made more similar).
- This is **stimulus-stimulus compression**.
- How can stimuli be made more similar?
- Competitive learning: clusters patterns together if the model tends to treat them similarly.
- Stimulus-stimulus compression can produce a variety of animal learning phenomena:

    sensory preconditioning

    compound preconditioning

    latent inhibition

    response decrement after context shift

Sensory Preconditioning

- Pattern for A and pattern for B will be more similar than before.
- If we then learn a response to stimulus A, it is more likely to transfer to B
- Presenting B is more likely to evoke the learned response than if AB had not been initially presented together.

Compound Preconditioning

- Initial training on A and B together; no reward: AB–.
- Then train on A+ and B–.
- Learning to distinguish A from B is harder as a result of the initial training.
- No lesion experiments have been done on compound preconditioning.
- Model predicts that compound preconditioning will not be affected by H lesion, but will be eliminated by H^{++} lesion.

Latent Inhibition

- Train on A in context X, with no reward. This is AX–.
- Then train on A with reward: AX+, X–.
- Learning is slower than if AX– had not been taught.
- This is **latent inhibition.**
- Causes: (1) stimulus-stimulus compression, (2) stimulus-outcome compression.
- Effect is eliminated by H^{++} lesion.
- Why? Stimulus compression does not take place, so the pattern for A is unchanged by exposure to AX–.
- Effect is preserved by H lesion (in rats and simulation), because stimulus-stimulus compression is thought to take place in EC.

Release from Latent Inhibition

- Train on AX–.
- Shift. to context Y.
- Learning to respond to A is dis-inhibited by the context shift.
- Can learn AY+ more quickly than AX+ after training on AX–.
- This is **release from latent inhibition.**

Response Decrement After Context Shift

- Train on A in context X: AX+, X–.
- Stimulus-stimulus compression causes patterns for A and X to become more similar.
- Shift to context Y.
- AY produces a smaller response than AX.
- This is the **response decrement**.
- Note: after substantial training on AX+, the response decrement goes away (in animals and the simulation.) Why?
- Stimulus-outcome differentiation has separated A from X, because AX+ has different outcome than X–.
- The positive outcome is now predicted by the component of the input most clearly associated with A, not A/X.
- Now the pattern for A in AY also predicts positive outcome.

Effects of Stimulus Differentiation

- If stimuli are made more different, learning of discrimination tasks should be enhanced.
- How can they be made more different?
- Orthogonalization separates patterns.
- Gluck & Myers suggest that pattern separation by is done by the hippocampus.
- Effects resulting from stimulus differentiation:

  disappearance of response decrement after context shift in overtrained animals

  transfer from easy to hard learning tasks

Easy-Hard Transfer

- Easy learning task: distinguish white from black stimulus.
- Hard task: distinguish light gray from dark gray stimulus.
- Training on easy task A+, B– speeds learning on harder task a+, b–.
- Performance improvement is greater than for an equal number of extra learning trials on the hard task alone.
- Performance improvement is visible even when the responses are reversed in the hard task: A+, B– followed by a–, b+.
- Stimulus-output differentiation causes the system to amplify differences among stimuli in the dimensions in which A and B differ.
- H^{++} lesion eliminates this effect in animals.
- Novel prediction of the theory: H lesion should also eliminate this transfer effect.

Assessment of Gluck & Myers

- Nicely captures a collection of animal learning effects.
- Makes testable predictions about lesion experiments.
- Abstract , high-level theory; pretty distant from the anatomy.
- Toy model; no attempt at realistic simulations.
- Entire model is feed-forward.
- Uses HC to modify cortical representations, but doesn't provide for changing cortical reps. to affect HC.