Some crazy ideas about temporal integration and object recognition that might be right

Dean Wyatte
April 15, 2013

Temporal integration

- World is full of temporal structure
- Brain probably leverages this important property to learn stable features from moment-to-moment
- Missing from most biological vision models, learn from discontinuous "snapshots"

Outline

Part I

- LeabraTI (Temporal Integration) framework
  - Generic framework applied to host of problems (perception, motor, WM, etc.)

Part II

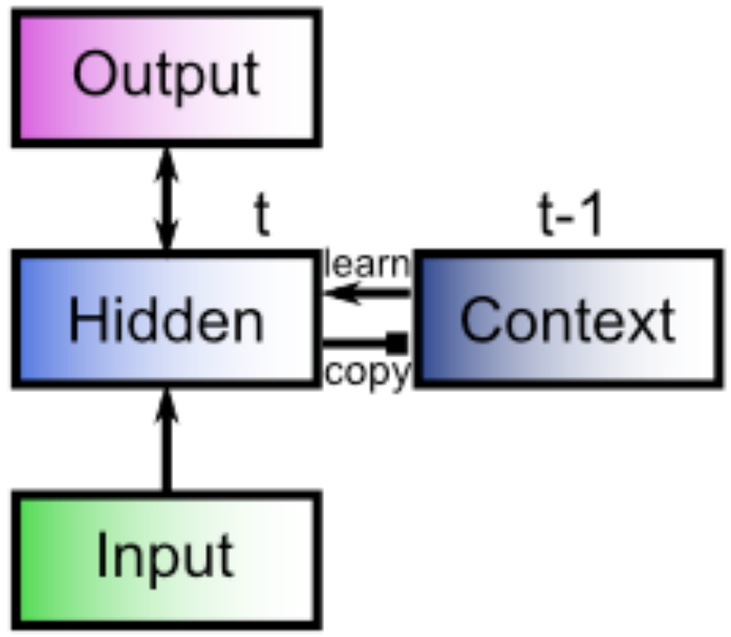
- Temporally integrated object learning
  - Previous research, predictions

General ideas: alpha rhythm, laminar differences, gating, phase resets

Computational requirements of TI

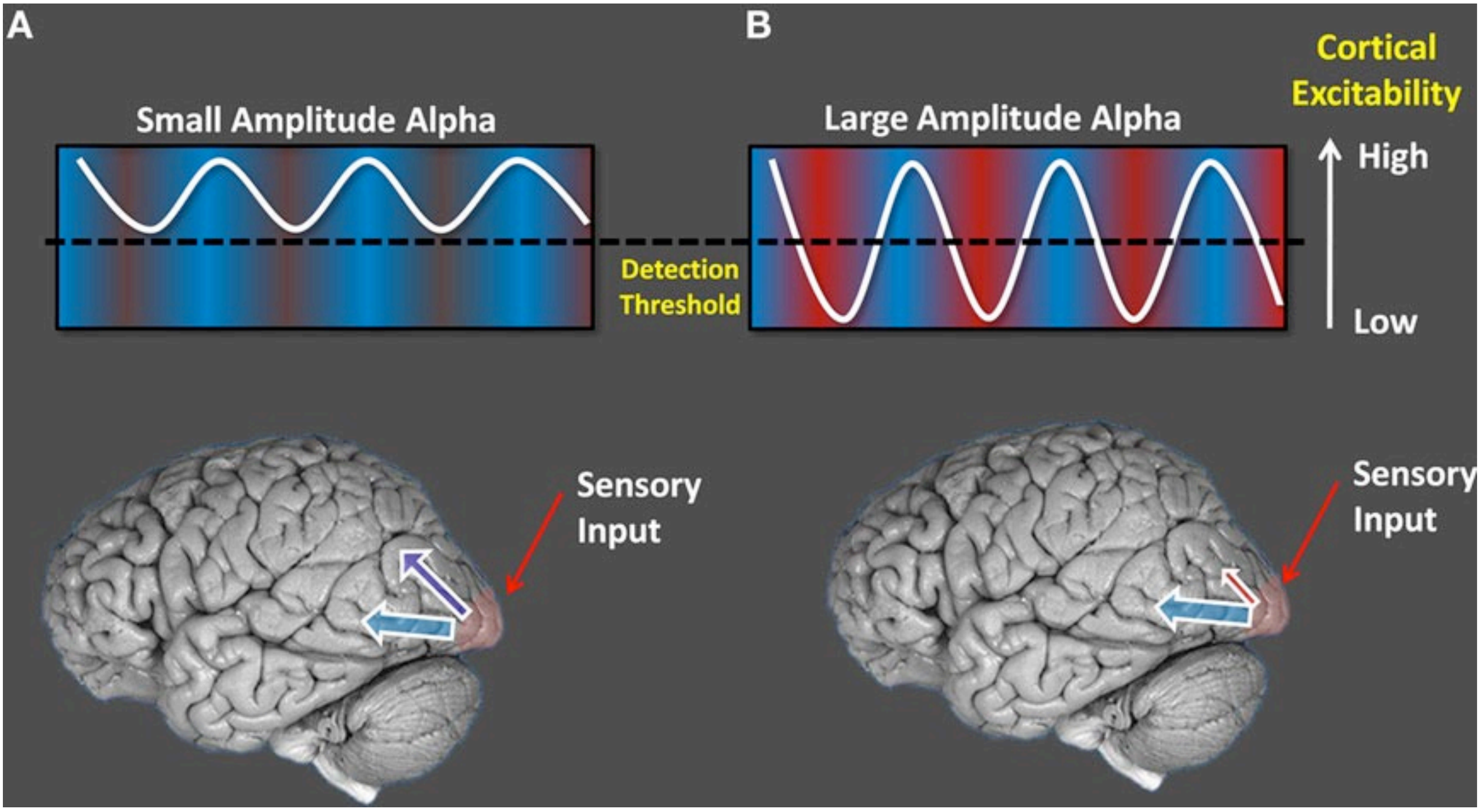
- Previous moment's representation needs to be "frozen" to integrate with new one
  - Simple recurrent network architecture is attractive computational model with t-1 static context
- How might this idea be realized in the brain, which is constantly changing state?
  - When to freeze and copy?
  - Frozen states highly unlikely (except for maybe WM maintenance)

Simple Recurrent Network



Key insights
Alpha rhythm

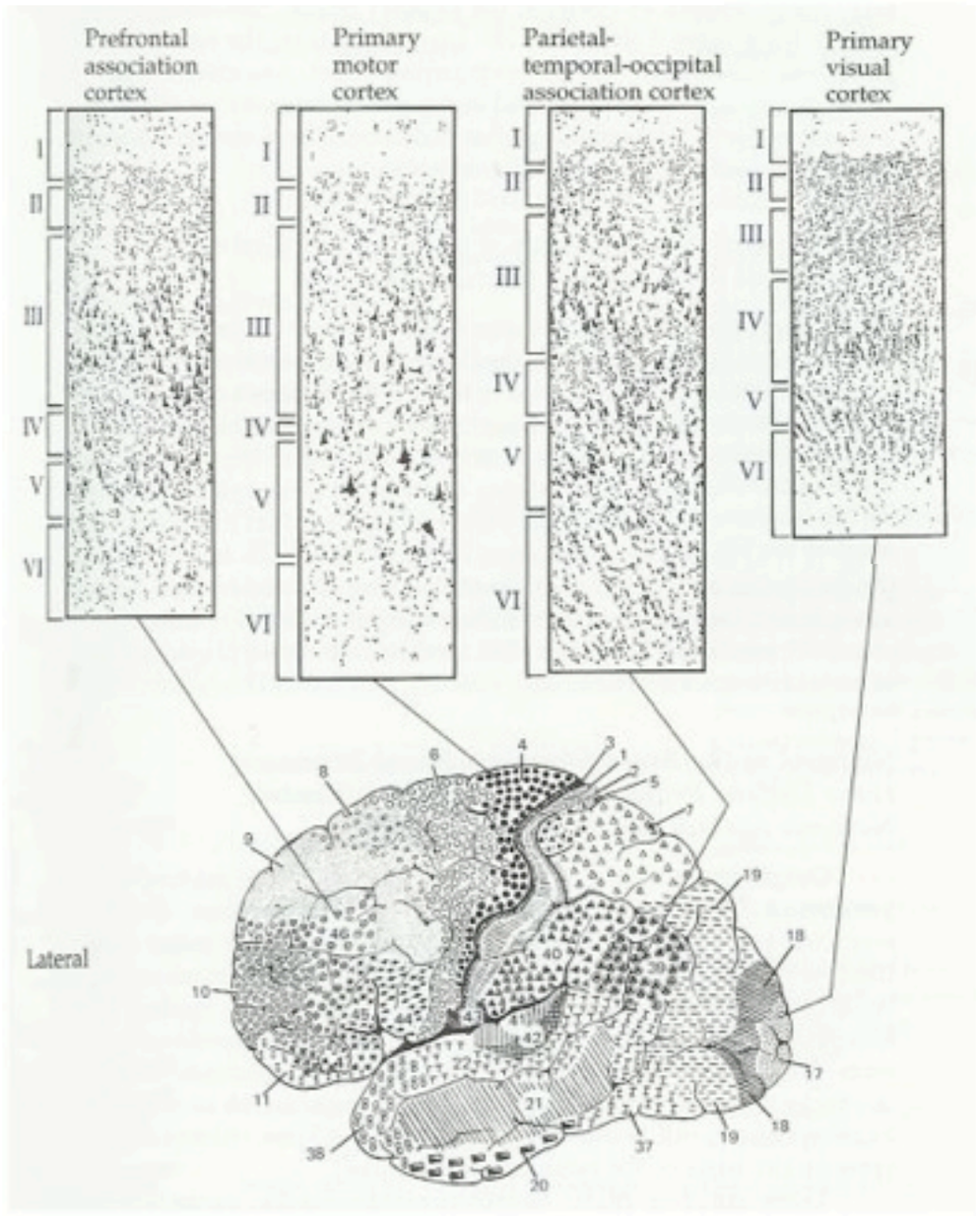
- Alpha rhythm (10 Hz) linked to cortical excitability
  - At-threshold stimuli less perceptible in certain parts of alpha cycle
- **Leading hypothesis:** Periodic high-gain sampling of perceptual stream
  - **Strong hypothesis:** Perception actually discrete
  - **More generally:** Active perception e.g., sniffing, whisking (also occur at 10 Hz...)



Mathewson et al. (2011), Frontiers in Psychology

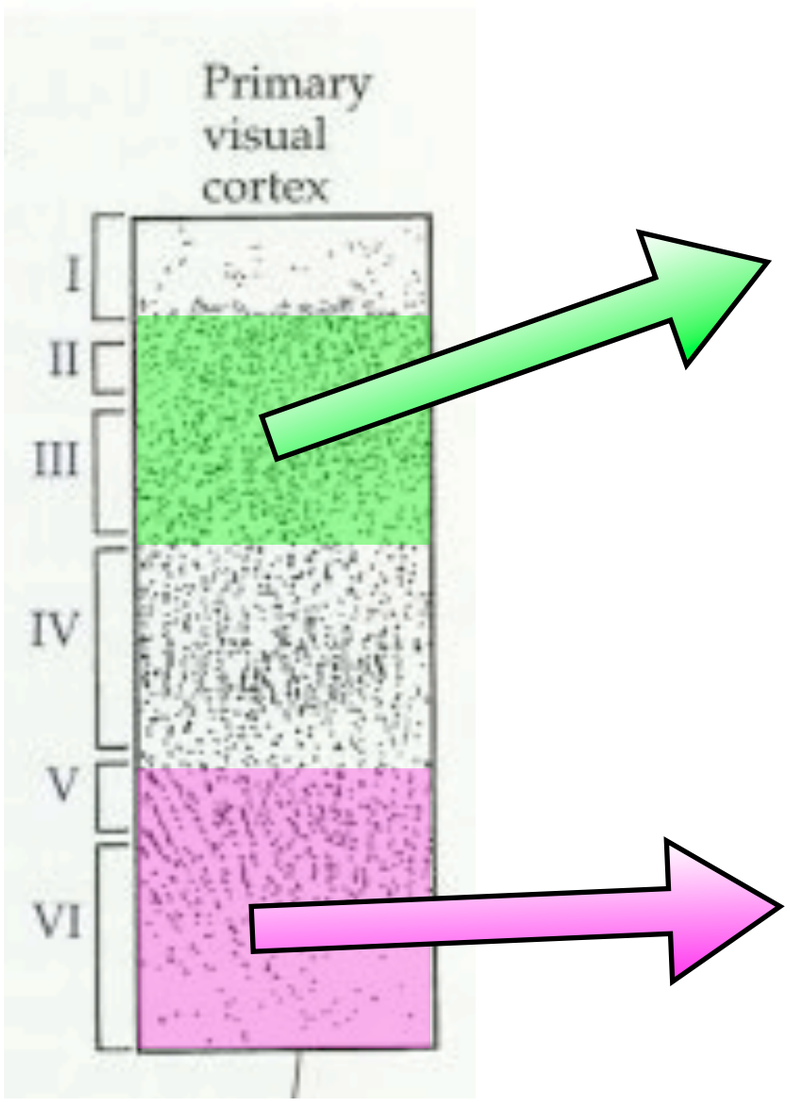
Key insights

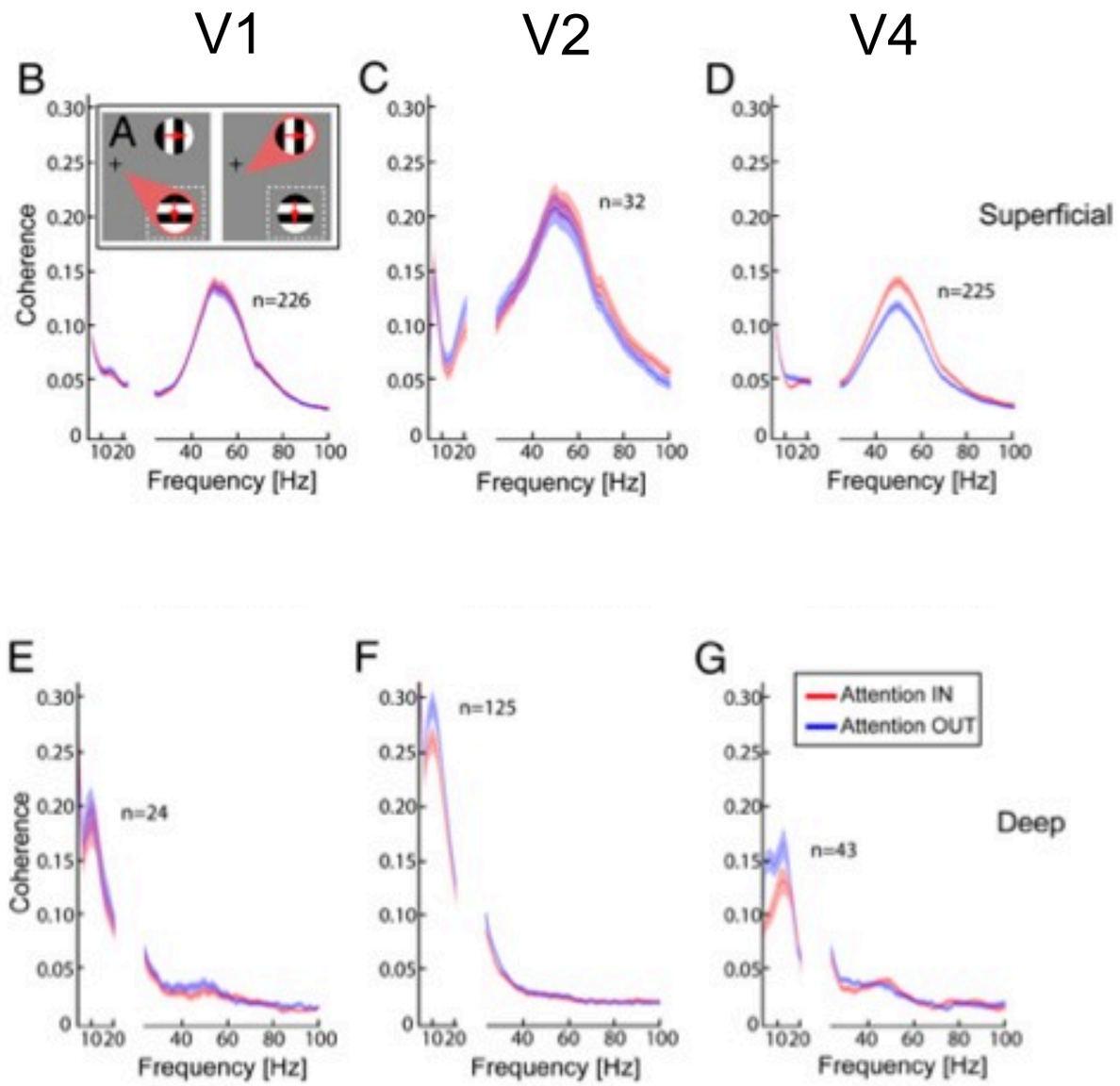
Laminar structure



Laminar structure

Spectral asymmetries



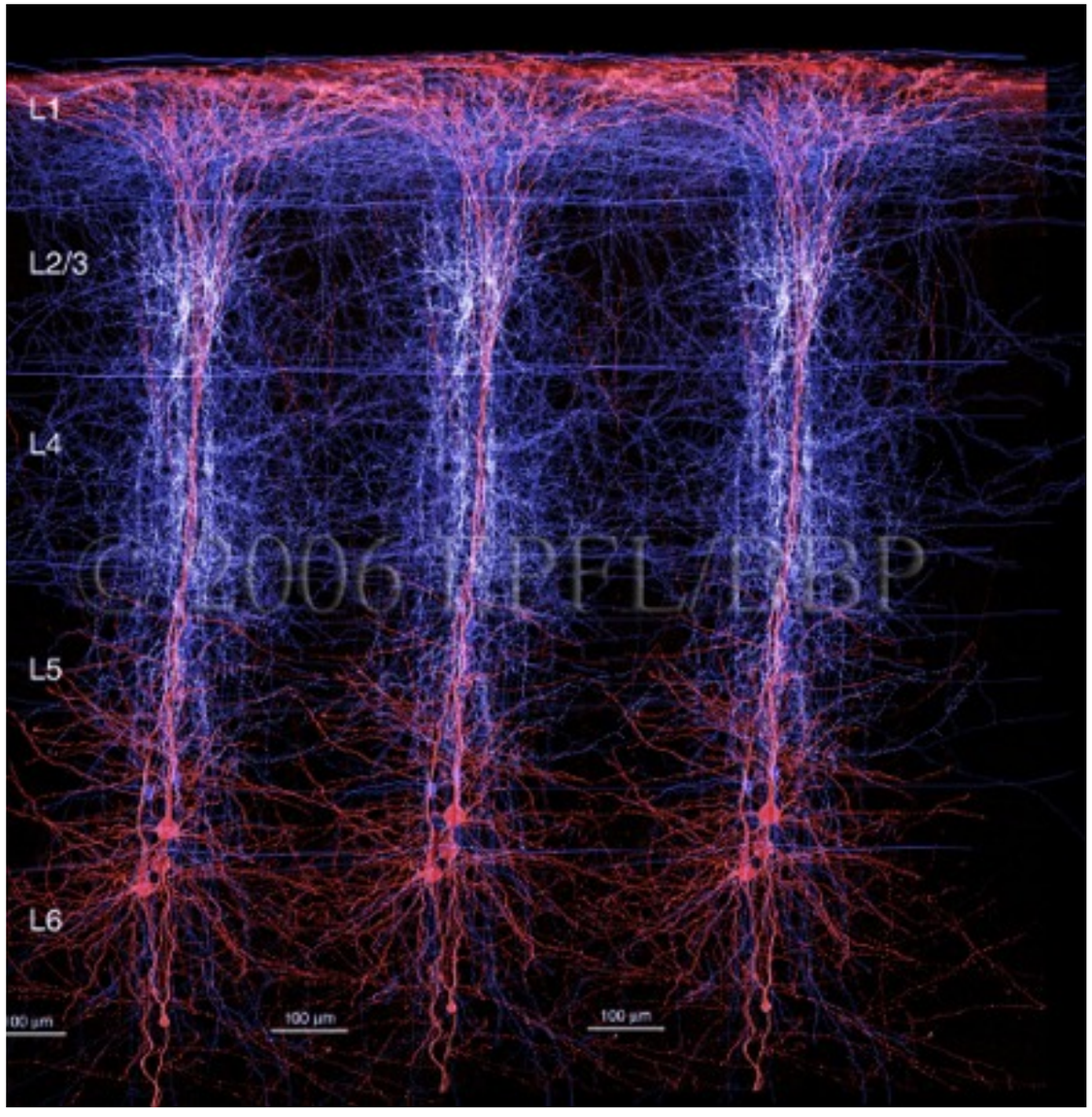


Buffalo et al. (2011), PNAS

Laminar structure

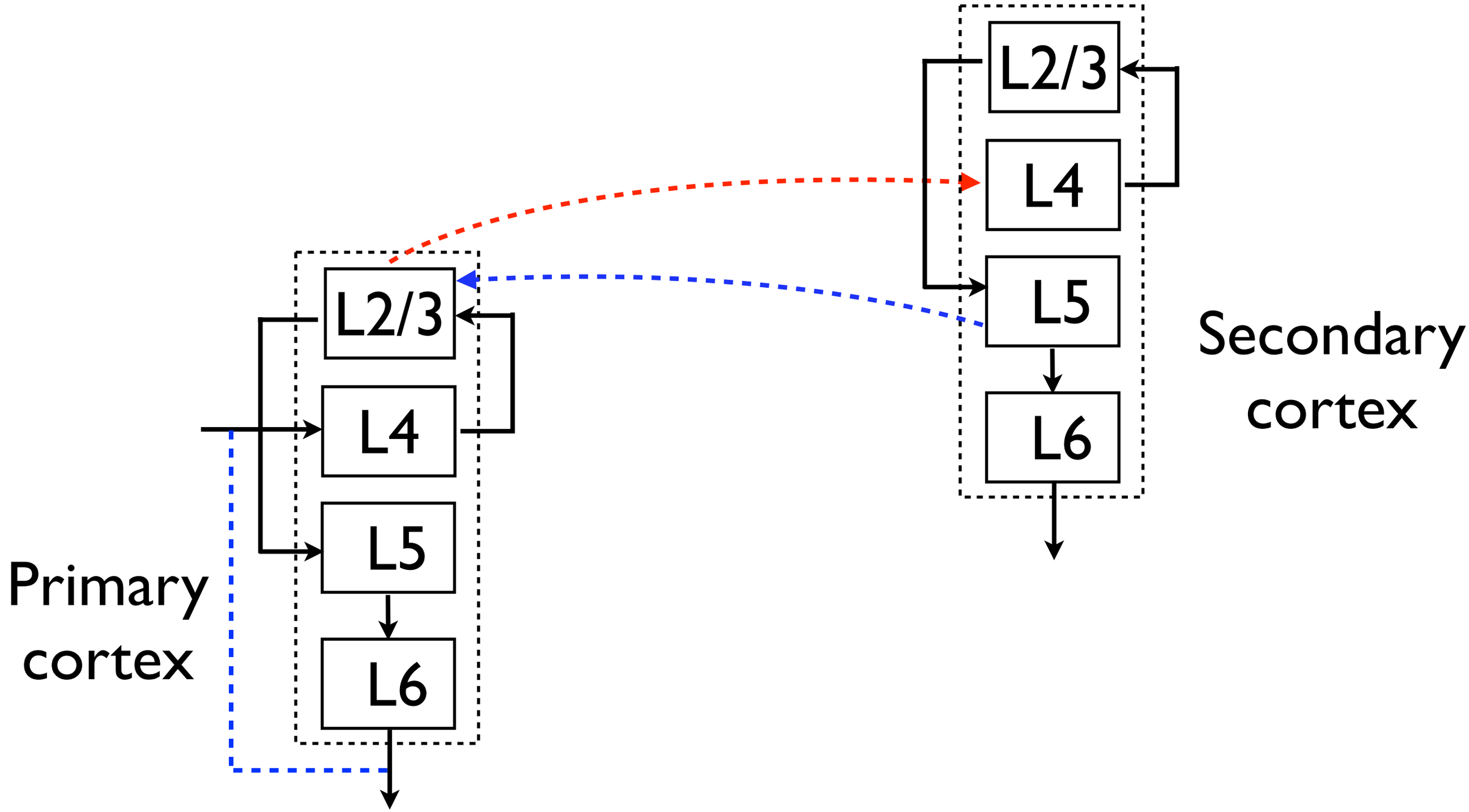
Columnar microcircuitry

- Microcolumn = 80-120 neurons
- Subtends 40-50 μm of cortex
- High degree of mapping within column, isocoding
- Replaces individual neuron as basic computational unit



Laminar structure

Columnar microcircuitry

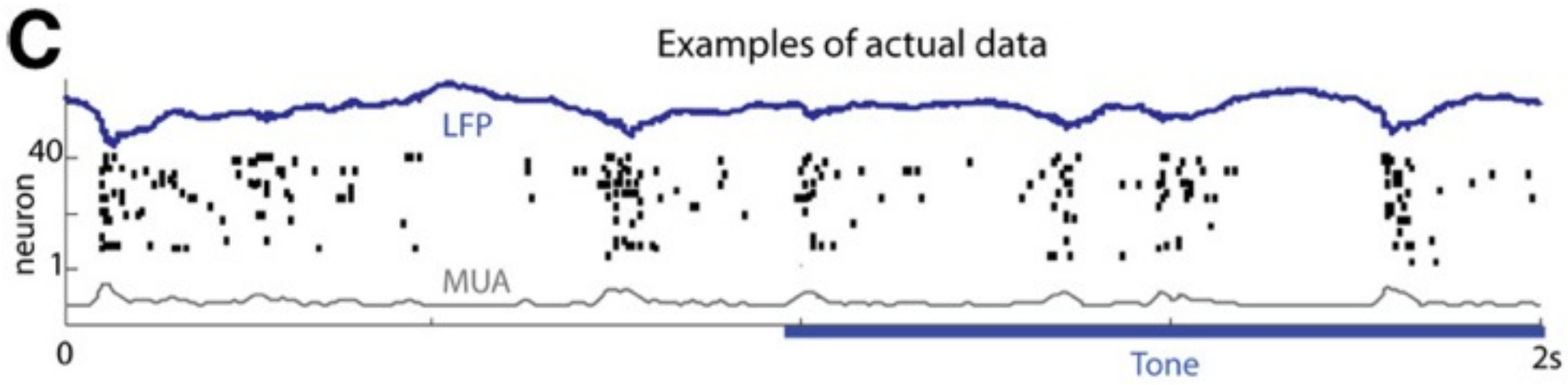


Laminar structure

Gating by activity

- Subset of L5 cells exhibit burst/quiescent response dynamics, even with sustained stimulation

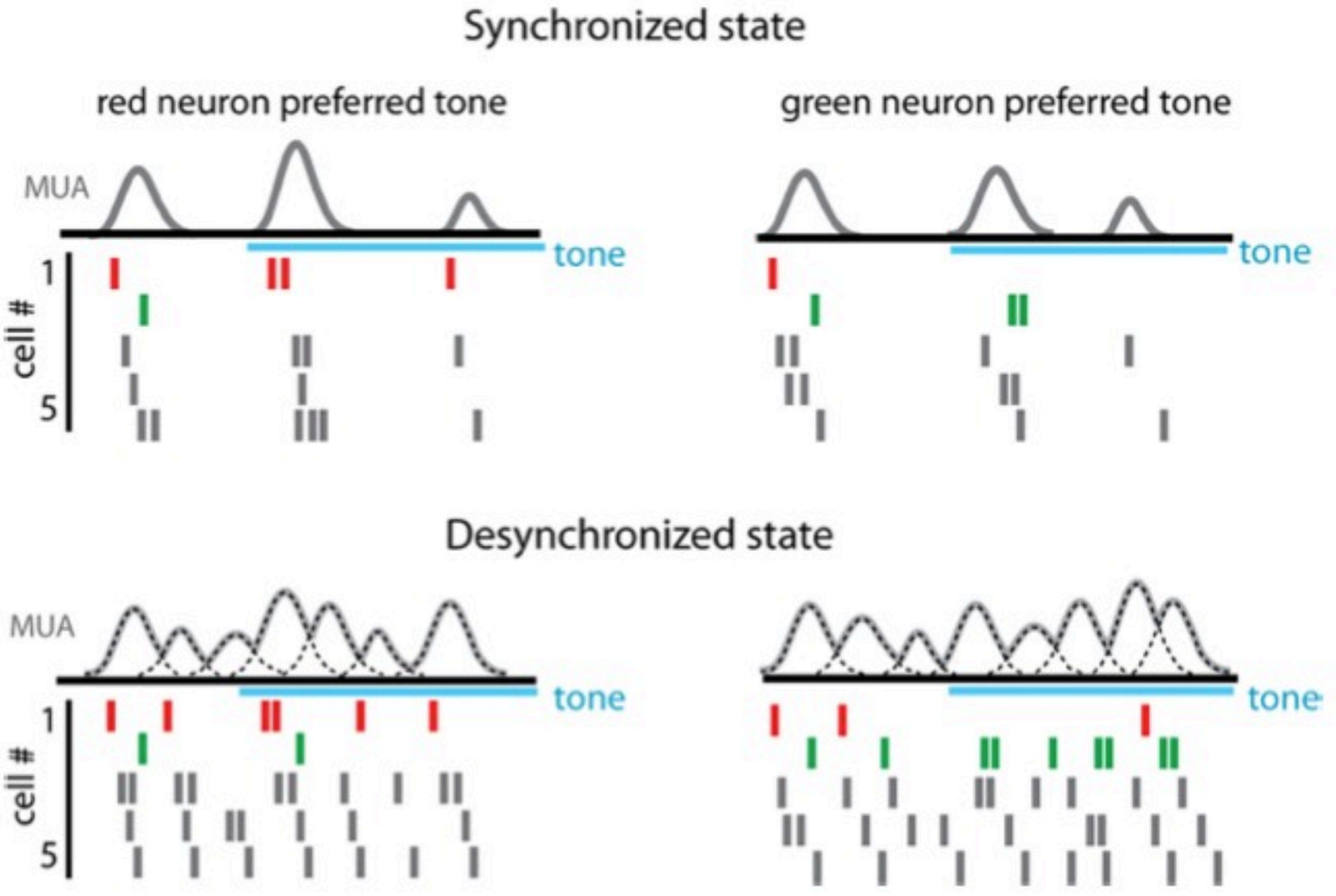
- Suggested to be due to gating by ongoing activity



Luczak et al. (2013), Journal of Neuroscience

Laminar structure

Gating by activity



Luczak et al. (2013), Journal of Neuroscience

Checking in...

Alpha rhythm

- Alpha peaks are a window of excitatory opportunity ("temporal pop-out")

Gating

- Increased excitation from alpha peak can cause spikes to penetrate deep layers

Open questions

- Environment *always* changing, what if important stuff happens during alpha troughs?
    - Alpha phase resets

Outline

Part I

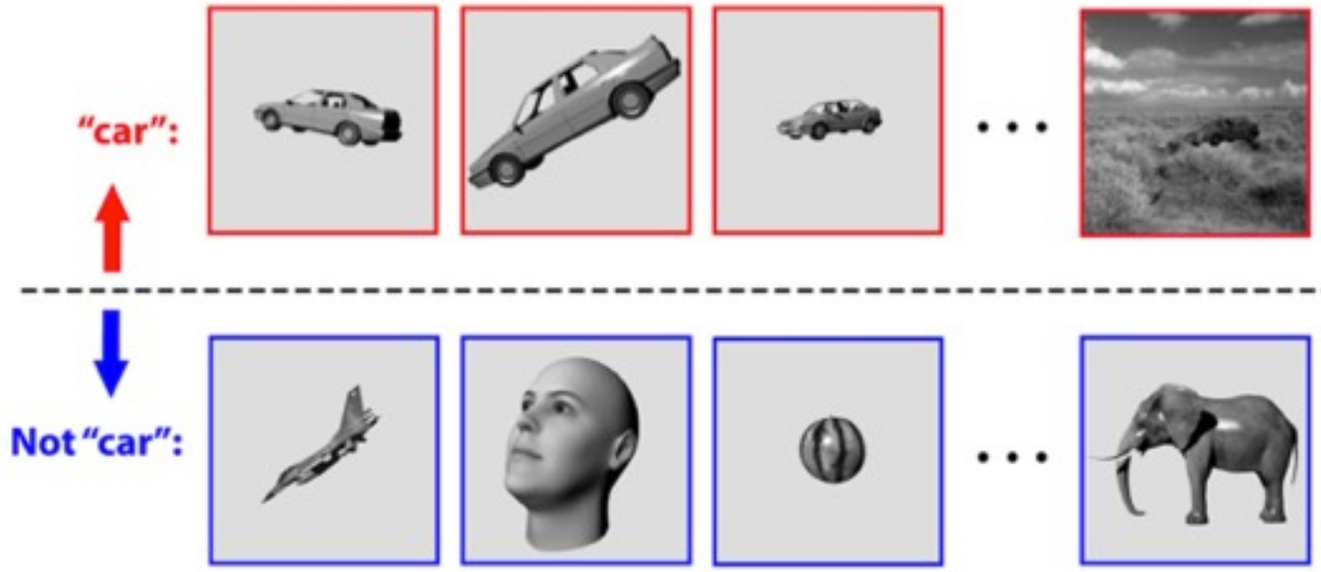
- LeabraTI (Temporal Integration) framework
  - Generic framework applied to host of problems (perception, motor, EF, etc.)

Part II

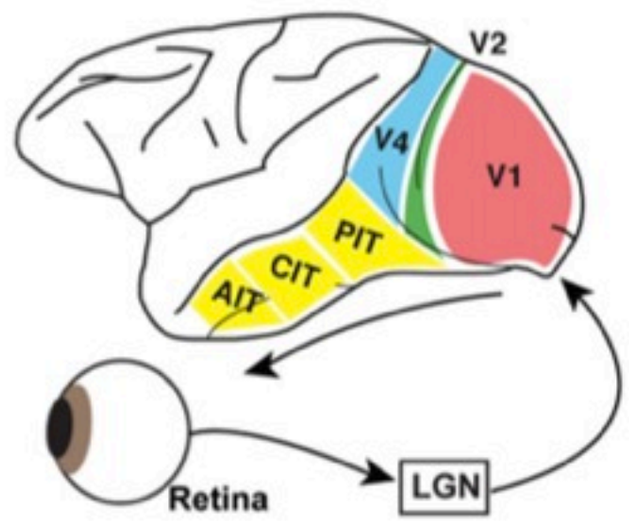
- Temporally integrated object learning
  - Previous research, predictions

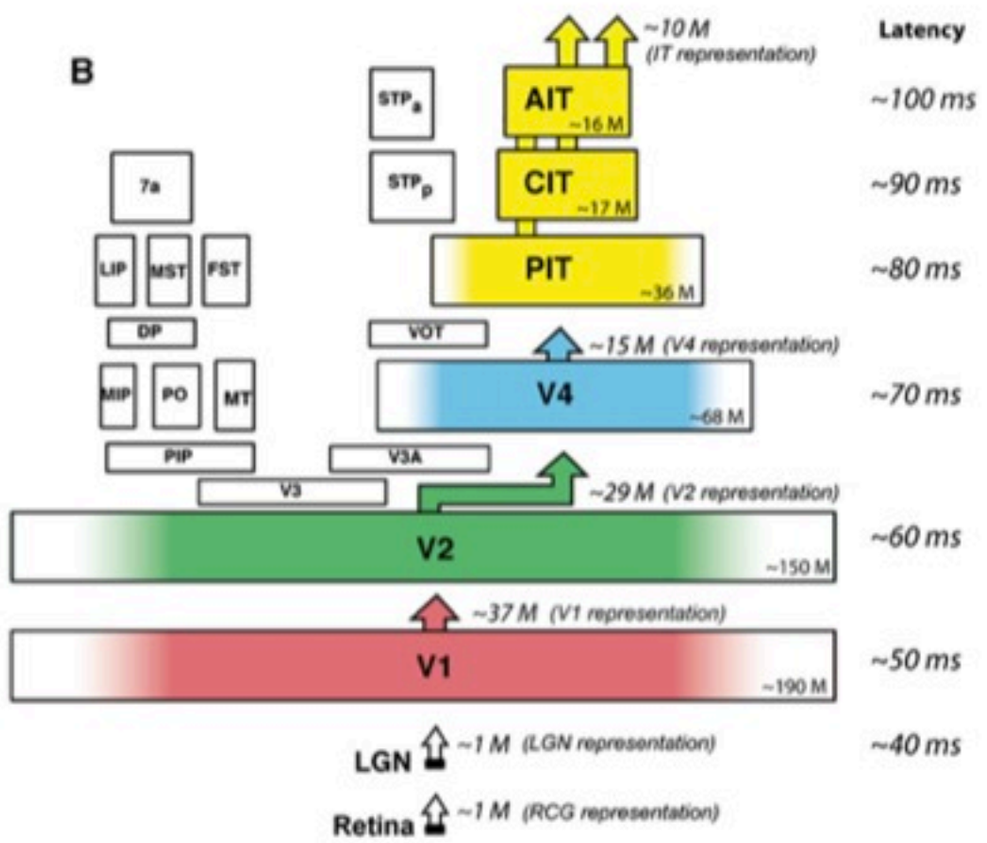
General ideas: alpha rhythm, laminar differences, gating, phase resets

Object recognition requires tolerance and selectivity

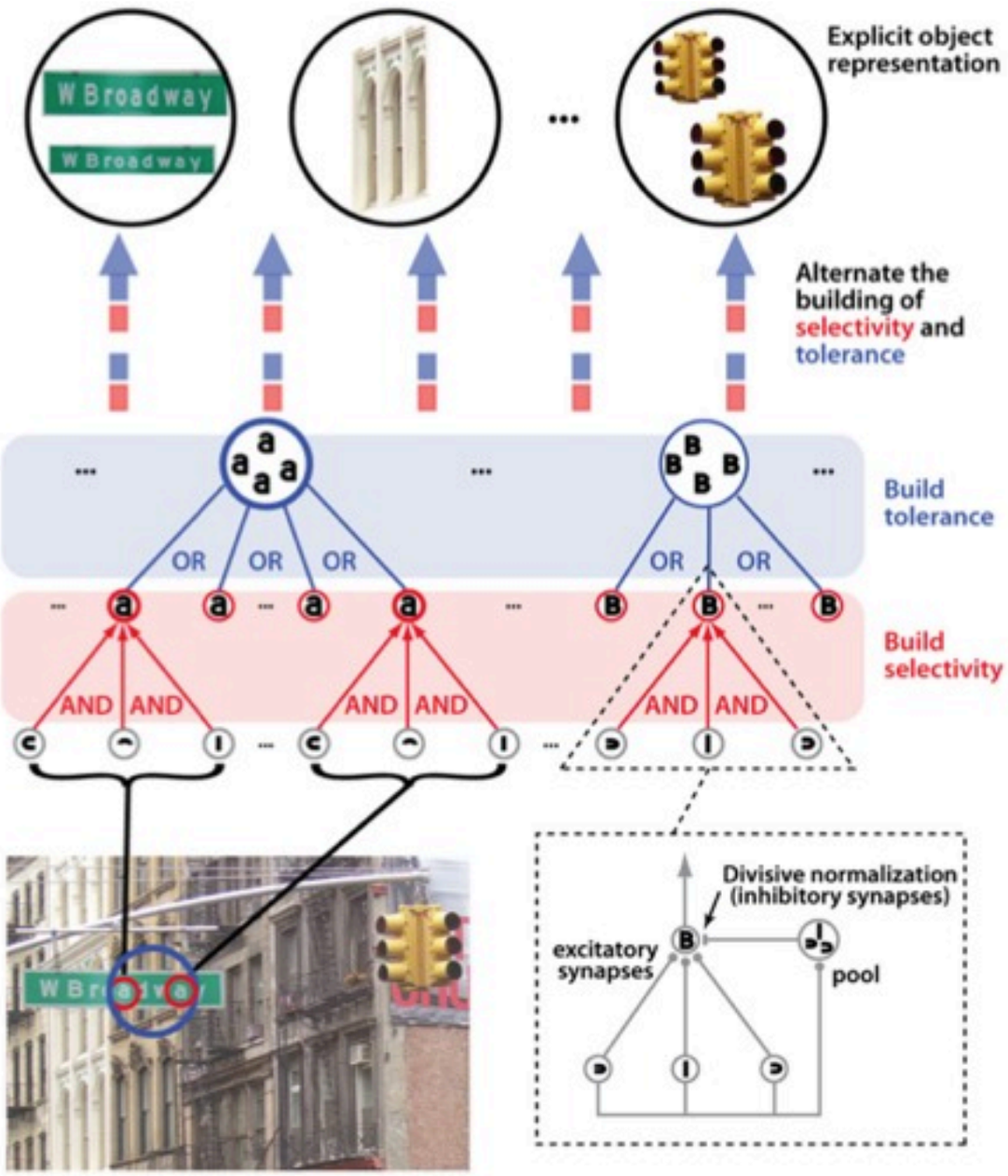


DiCarlo et al. (2012), Neuron





Object recognition requires tolerance and selectivity



DiCarlo et al. (2012), Neuron

Temporal integration and object recognition

- Wouldn't it be great if ventral stream neurons had a model of how objects in the environment change over time
  - Work from Jim Dicarlo's lab (MIT) suggests they do
- Temporal integration changes object tolerance
  - Plasticity persists into adulthood

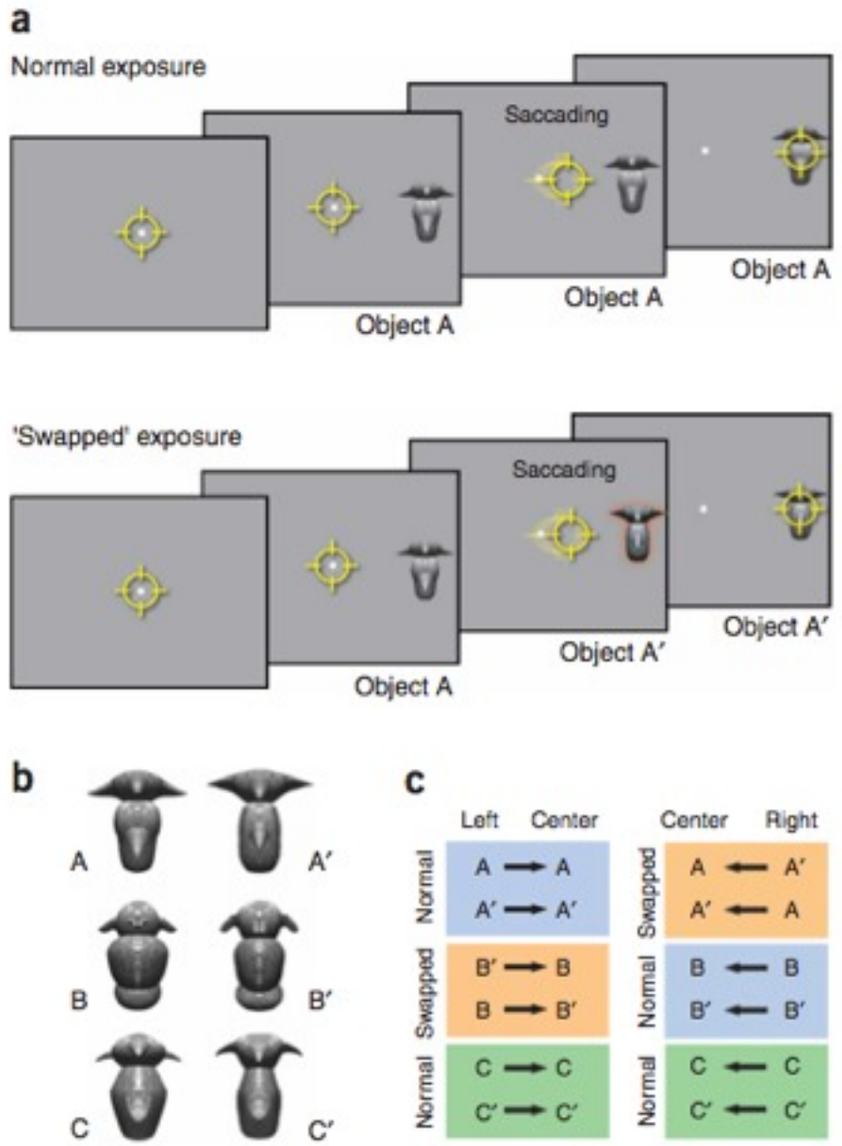
Three simple steps for tricking your ventral stream!

1. Object appears in periphery

2. Subject saccades to object

   *object identity swapped during saccade blindness

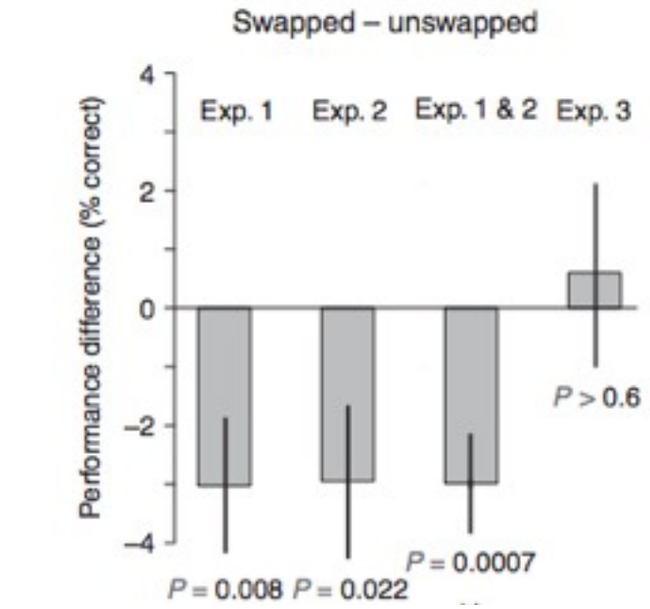
3. Subject foveates swapped object

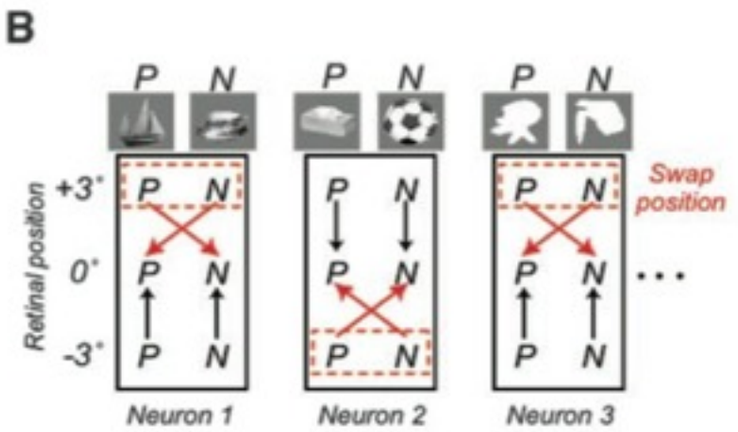


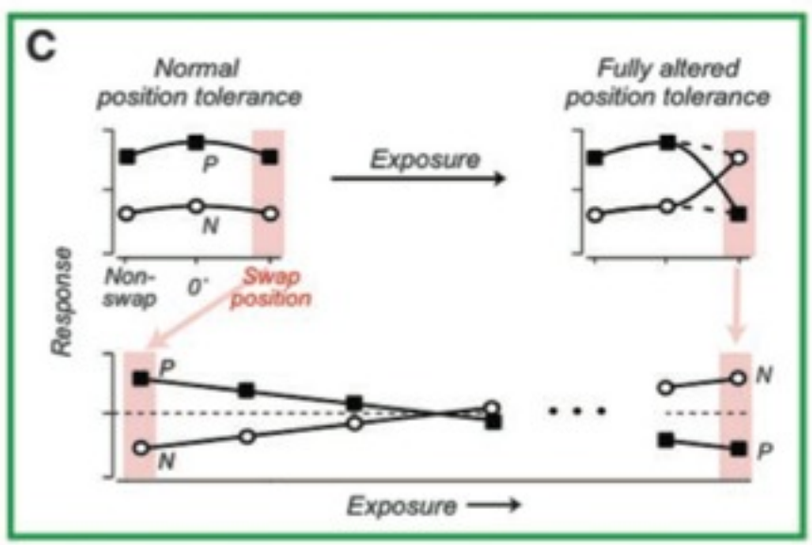
Cox et al. (2005), Nature Neuroscience

- Subjects perform worse for same object **at** swapped retina position compared to non-swapped position
    - Suggests that temporal learning mechanism associates image tolerance of t-1 peripheral input with t foveal input
- As little as ~1 hr of exposure training produces effect
- Somewhat mysteriously dependent on saccades in humans
    - No effect when subjects get "yoked" retina presentations

      (Although, subsequent studies with monkeys have obtained similar effect without eye movements for size tolerance)



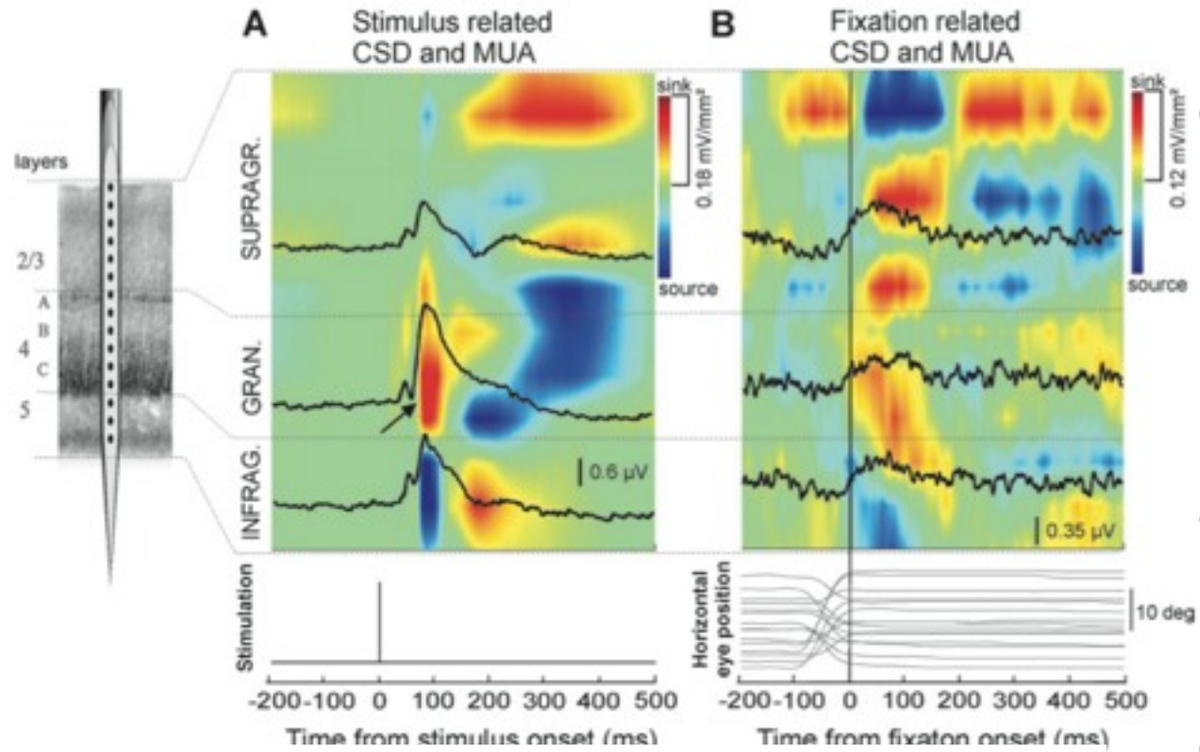


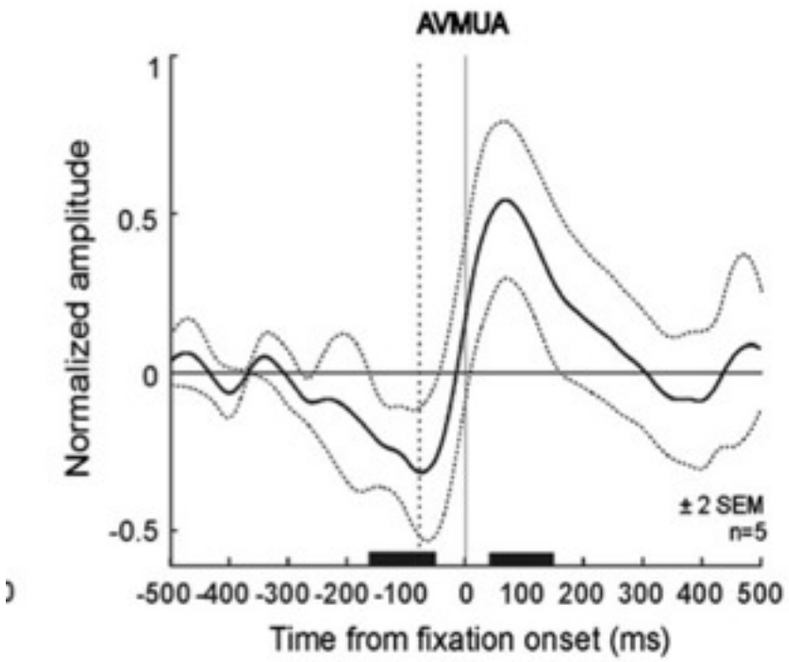


A potential explanation: Eye movements reset oscillatory phase

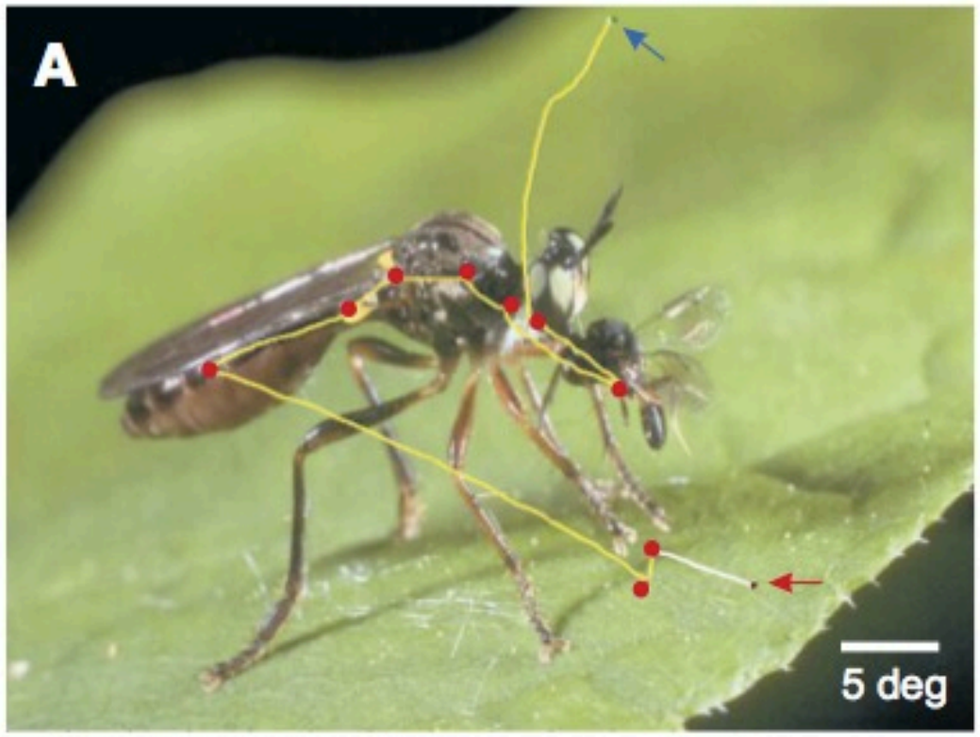
- Saccade causes modulation of neural activity for several hundred ms after fixation (and before)
  - Peak effect ~100 ms
- Suggested to be due to phase reset of 3-20 Hz oscillations, and then synchronization of stimulus-driven inputs with oscillatory peak
  - Preparatory state for brain
- Could strongly bias association of peripheral and foveated images

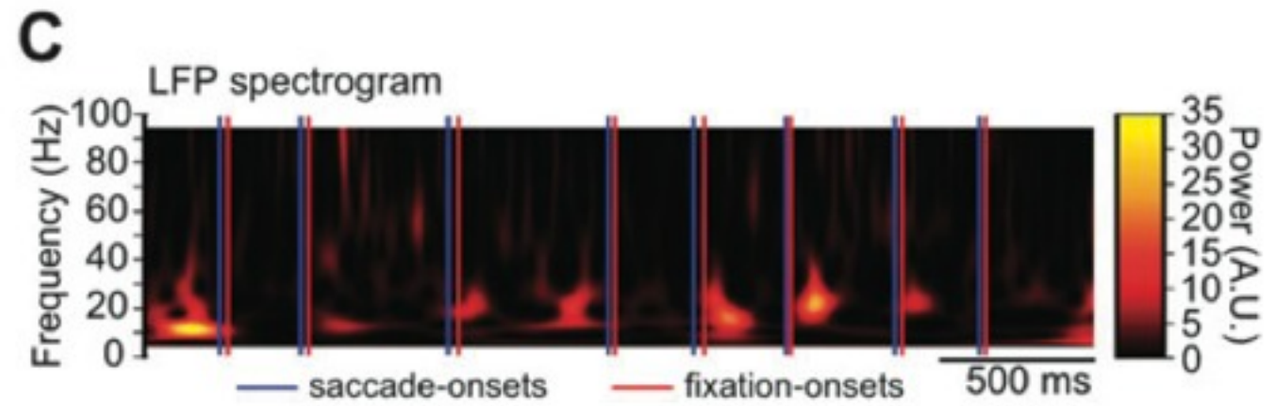
  (Exactly how still somewhat unclear to me)

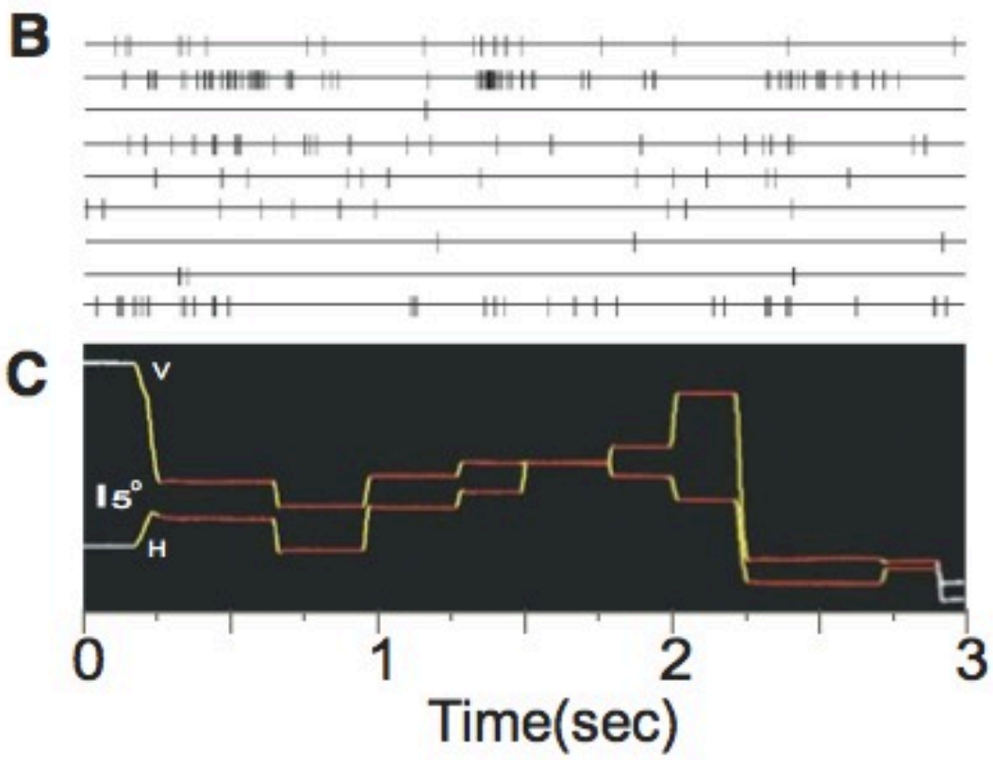


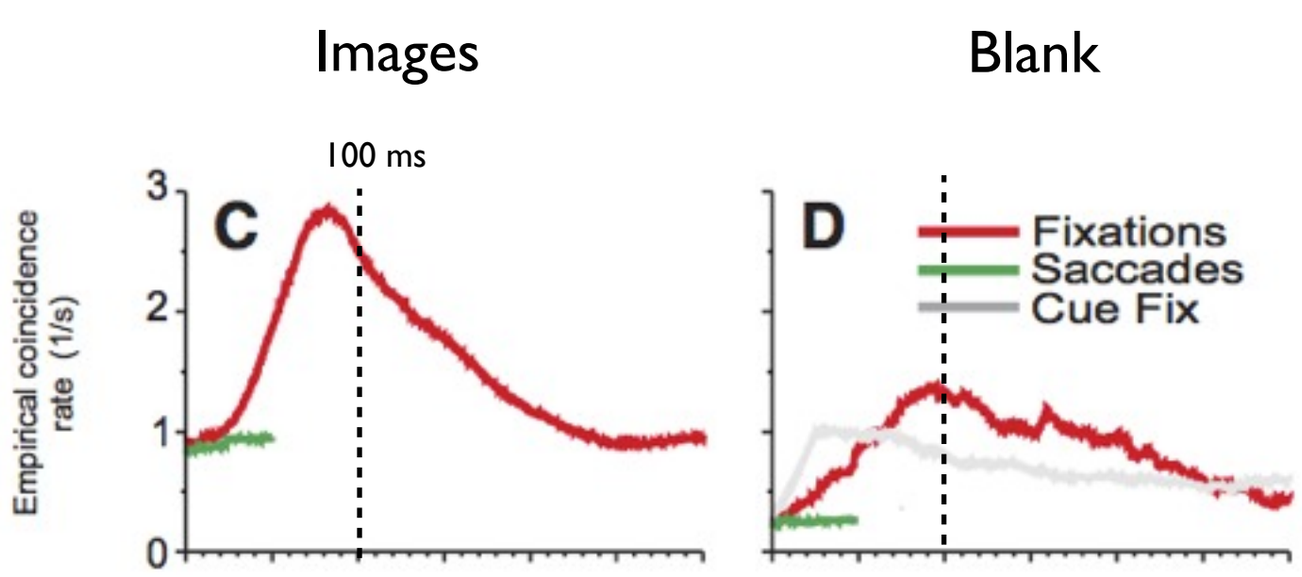


Rajkai et al. (2007), Cerebral Cortex


A
5 deg


C
LFP spectrogram
Frequency (Hz)
100
80
60
40
20
0
35
30
25
20
15
10
5
0
Power (A.U.)
saccade-onsets
fixation-onsets
500 ms


B
C
V
15°
H
0
1
2
3
Time(sec)

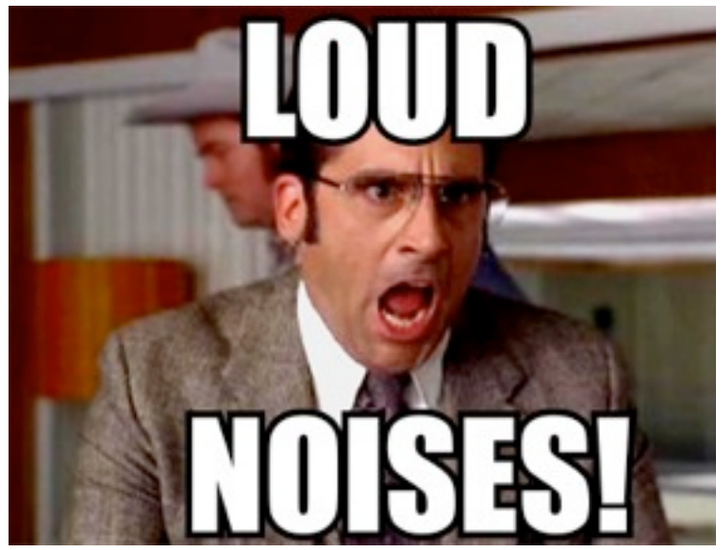

Images
Blank
100 ms
C
D
Empirical coincidence rate (1/s)
3
2
1
0
Fixations
Saccades
Cue Fix

Other things that reset oscillatory phase of visual cortex

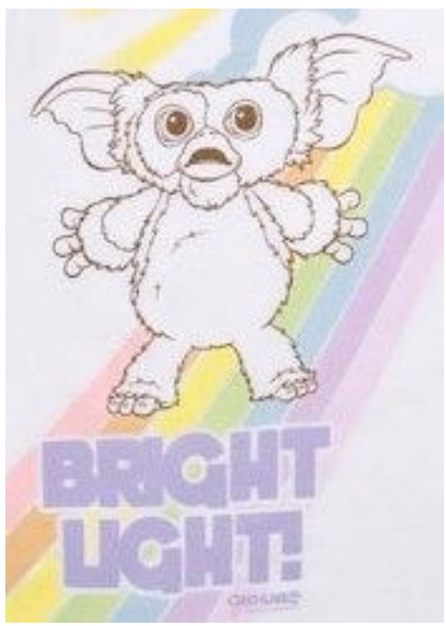
Eye movements

Salient events

- Noises



- Light flashes

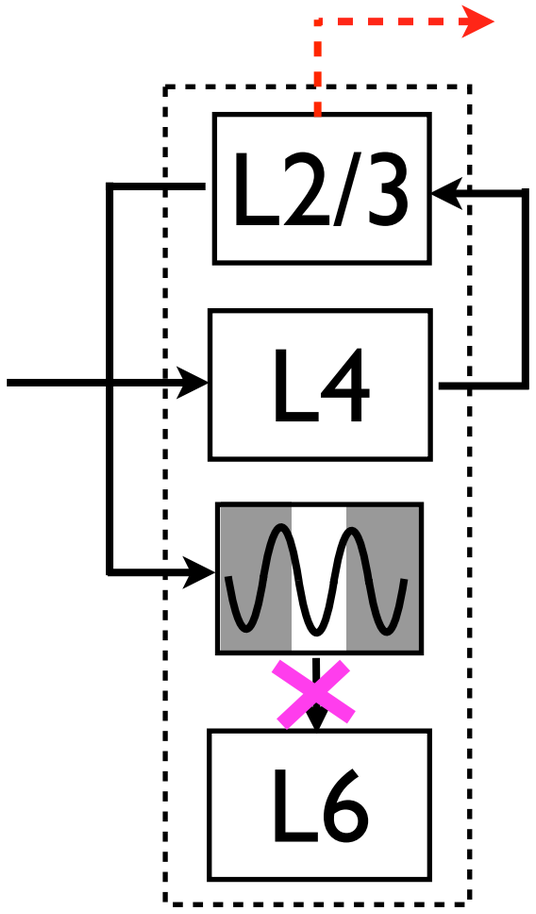


Multisensory inputs

- Auditory + somatosensory (monkeys)
- Visual + somatosensory (rats)

A general model

No alpha alignment, spikes generally only propagate through standard feedforward pathway (L2/3 onward)

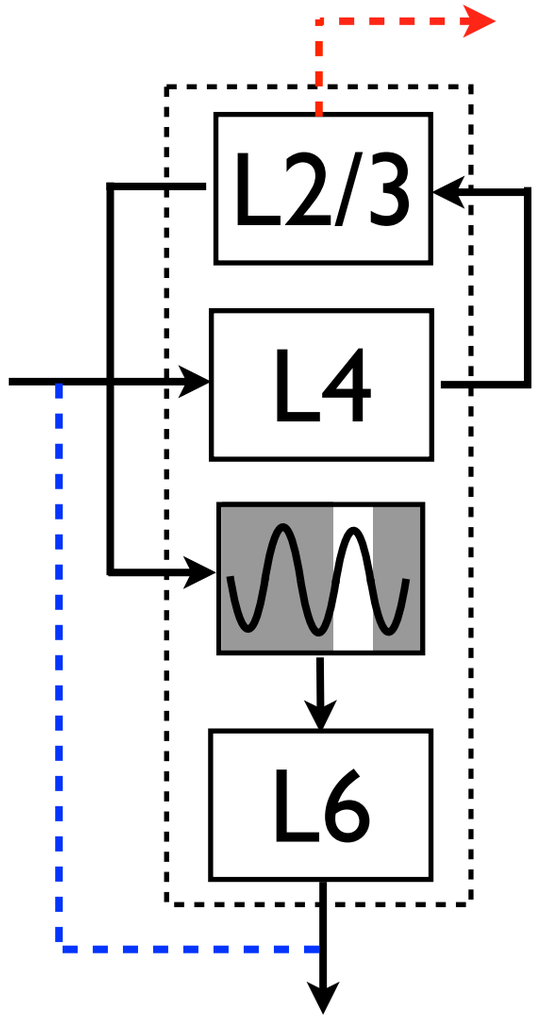


Alpha peak alignment (either from coincidence or hard reset), spikes also gated into deep layers and recirculated through microcolumn

A general model

No alpha alignment, spikes generally only propagate through standard feedforward pathway (L2/3 onward)

Alpha peak alignment (either from coincidence or hard reset), spikes also gated into deep layers and recirculated through microcolumn



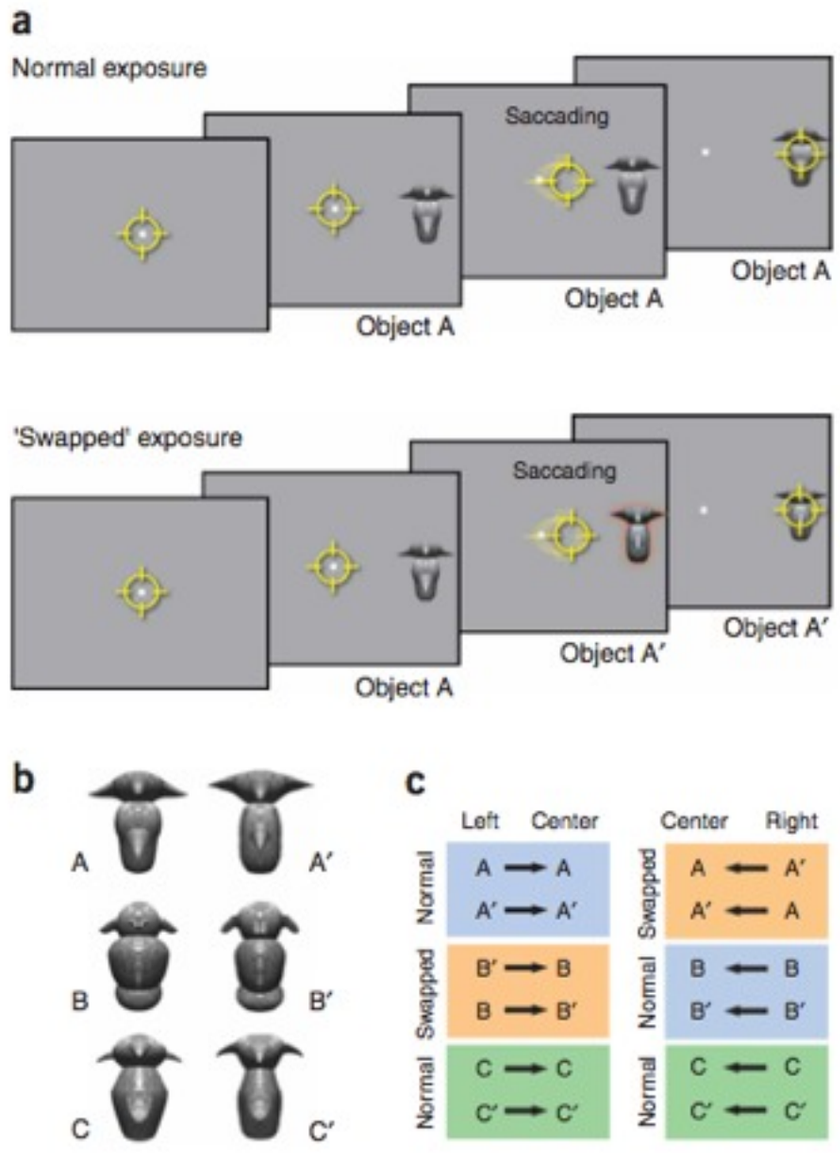
Summary

- Oscillations provide excitatory windows that dynamically routes information propagation, establishing "temporal frames"
  - Framing happens via intrinsic oscillations every 100 ms, but exogenous resets also possible to ensure synchronization
- This type of system seems ideal for temporal integration
  - Could provide novel methods for computational models of object recognition, scene description, information accumulation, etc.
- Theory integrates a huge range of empirical findings, each provide independent constraints

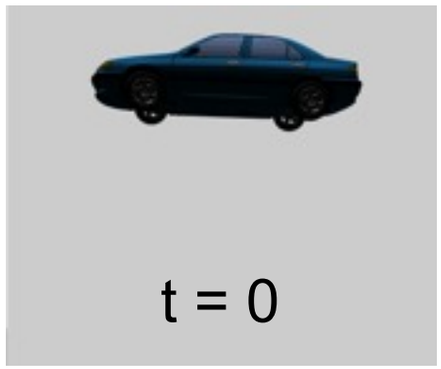
Almost too good to be true! Crackpot index high, must be overgeneralizing somewhere...

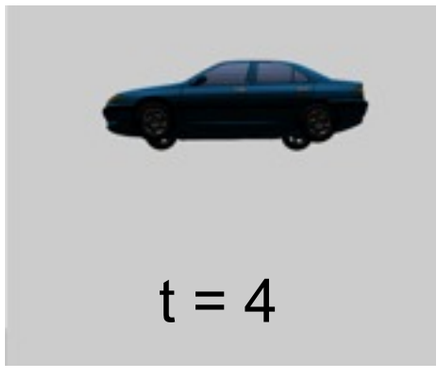
Actual neuroscience

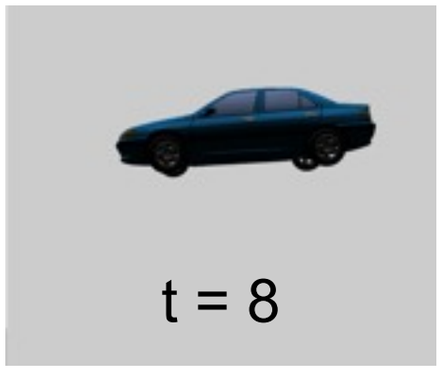
- Replicate results of DiCarlo lab and test idea that temporal object learning depends on alpha (and determine nature of interaction with saccades)
  - Interaction with other alpha effectors e.g., attention salient events, multisensory input, etc.
- Combine generic temporal integration framework with Leabra vision model

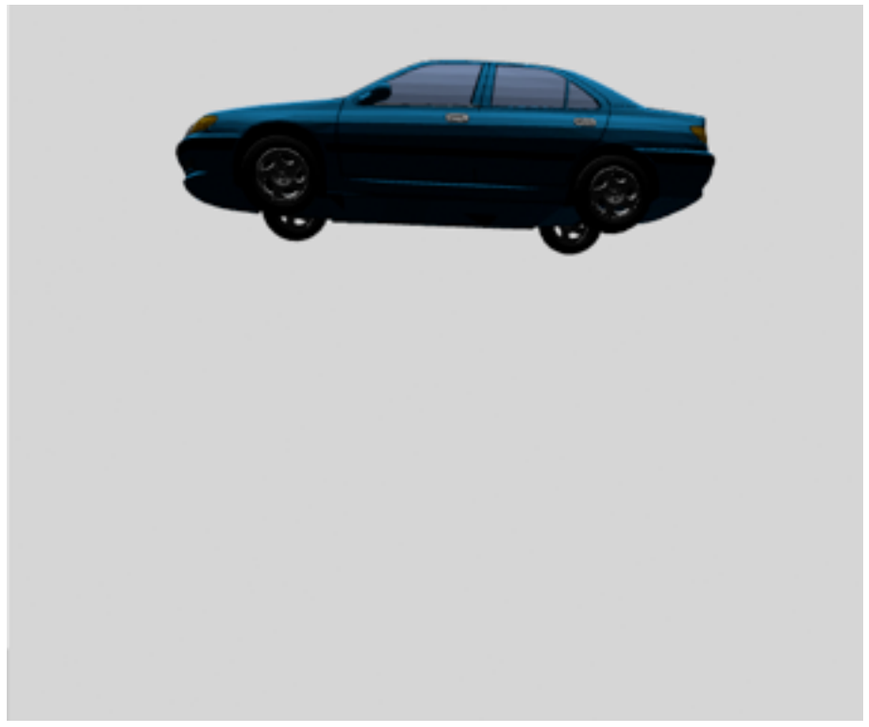


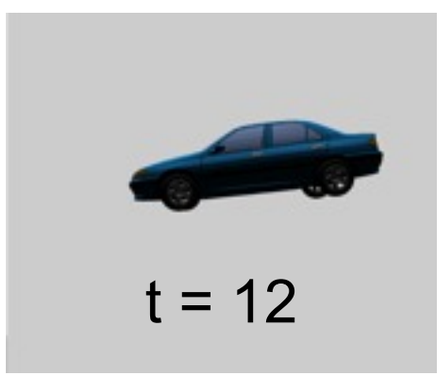
Initial modeling

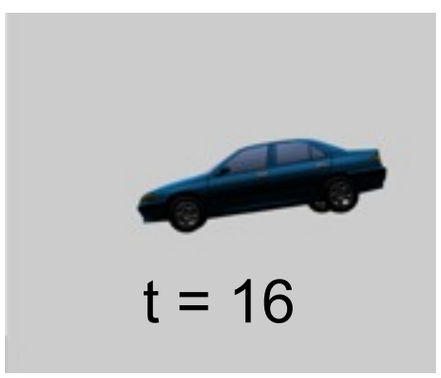


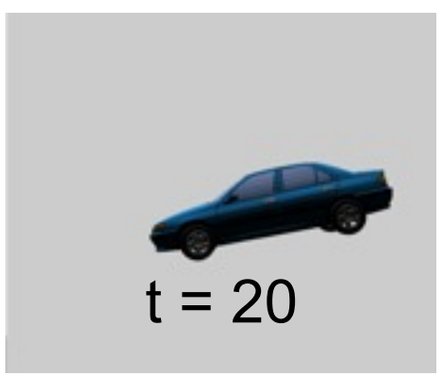






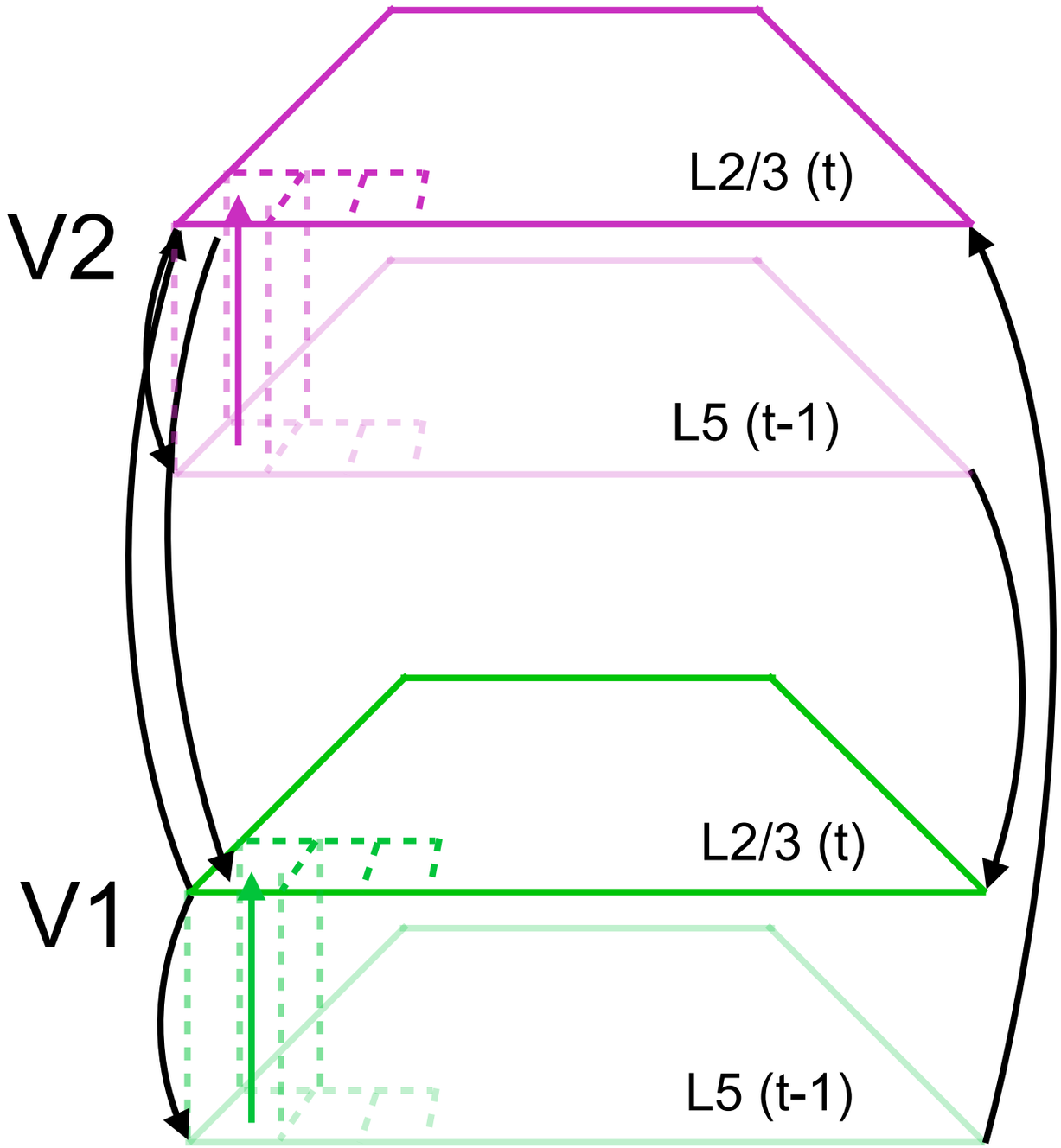




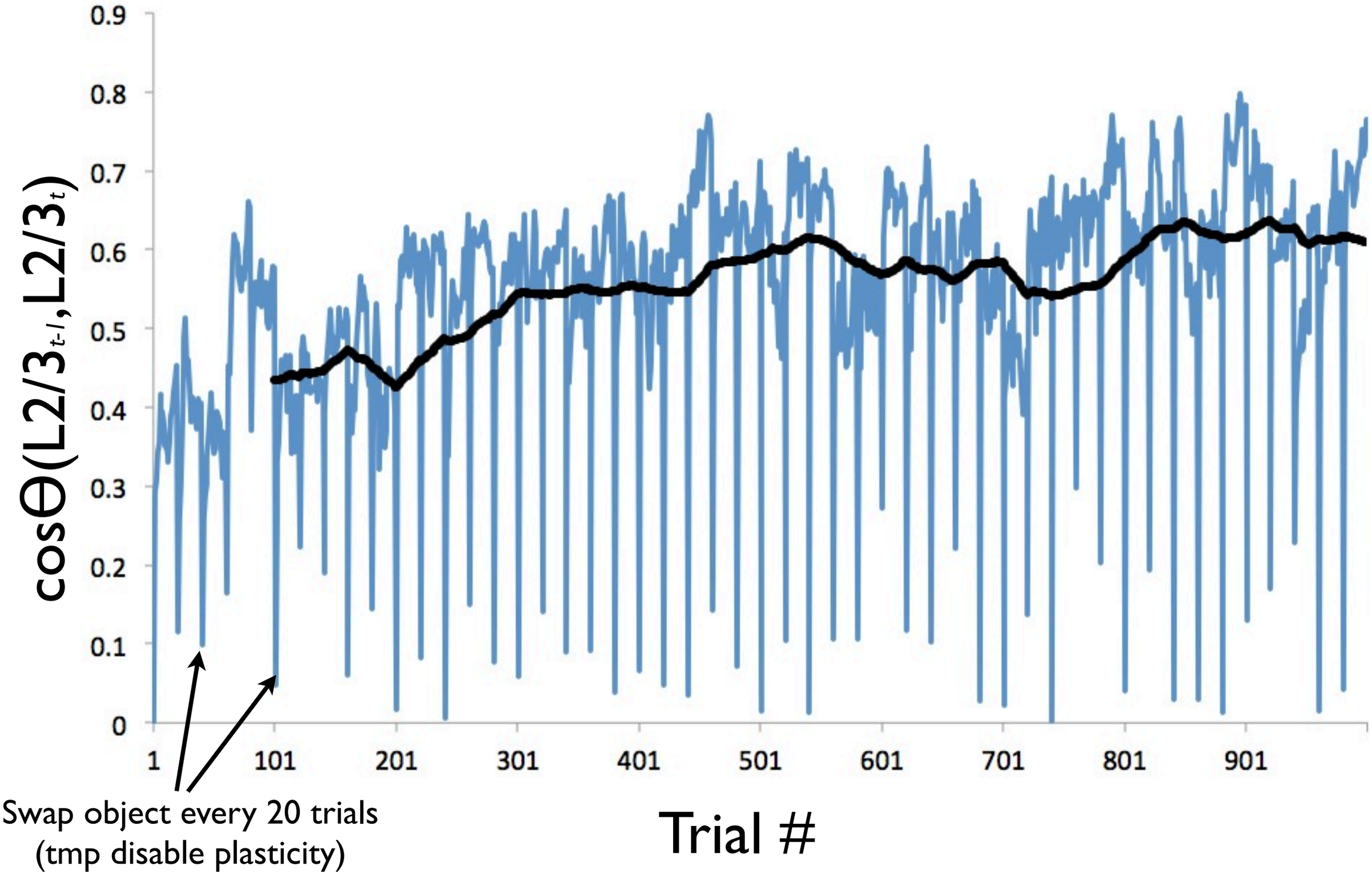


- Automatic deep layer gating every *t* (~100 ms)
- **Goal:** Learn what comes next

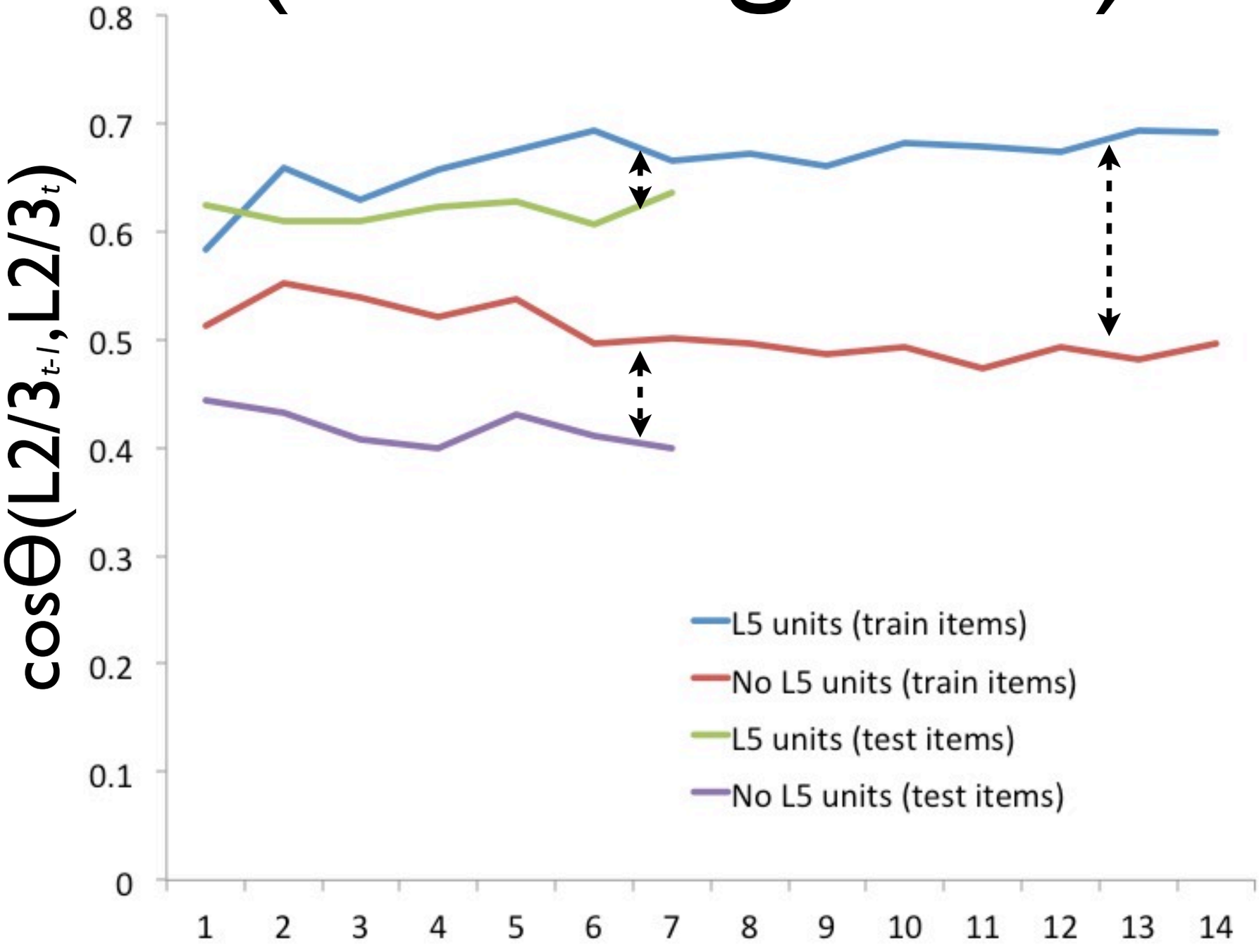
Preliminary modeling



Assessing learning



Long-term averages (30 categories)



Thanks

- **Randall O'Reilly**
- **Tom Hazy**
- Nick Ketz
- Jessica Mollick
- Scott Mackie
- Other numerous postdocs

The case for error-driven learning

- Learning "what comes next" in standard Leabra framework requires error-driven learning

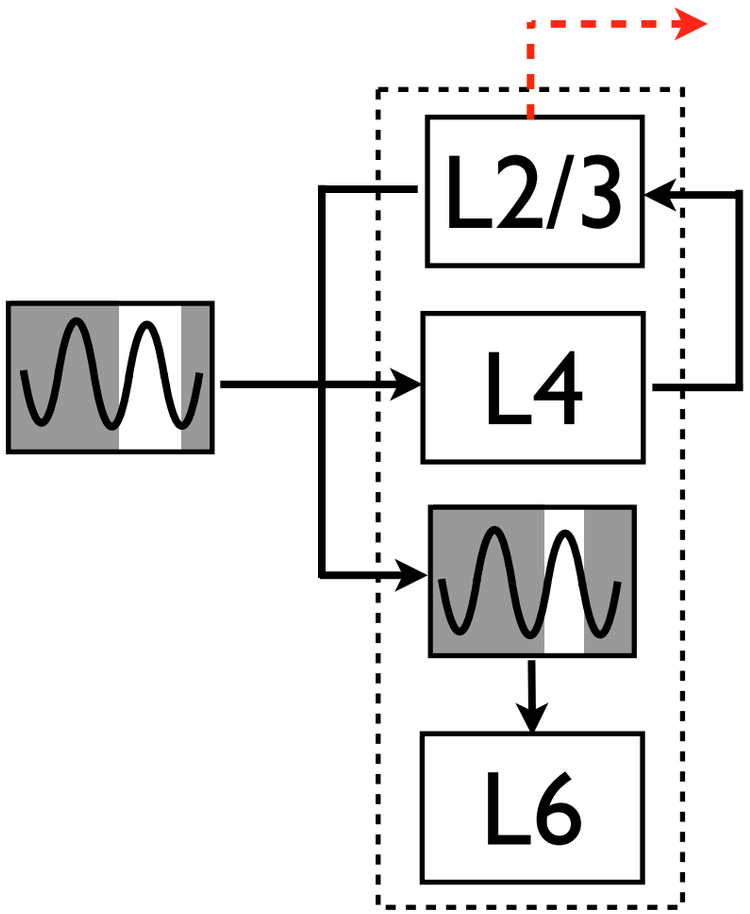
  t = Imagine what comes next

  t+1 = What actually comes next

  - Imagining happens on same neural substrate, driven by endogenous inputs
  - But retina always transmitting information
  - One solution: Thalamocortical modulation shifts transmission balance between exogenous/endogenous inputs
    - Portion of LGN neurons are bursting type (similar to L5)

Thalamocortical modulation

Thalamocortical alpha alignment, exogenous inputs pass through microcolumn and are gated into deep layers



No thalamocortical alpha alignment, microcolumn driven by endogenous deep inputs that do forward prediction