Cognition
and the Prefrontal Cortex

Zi-Seok Lee

2023-09-07

Contents

Cognitive Functions

- **Working Memory**: How we keep useful information active
- **Attention**: What really matters
- **Cognitive Control**: Detecting and resolving conflicting stimuli

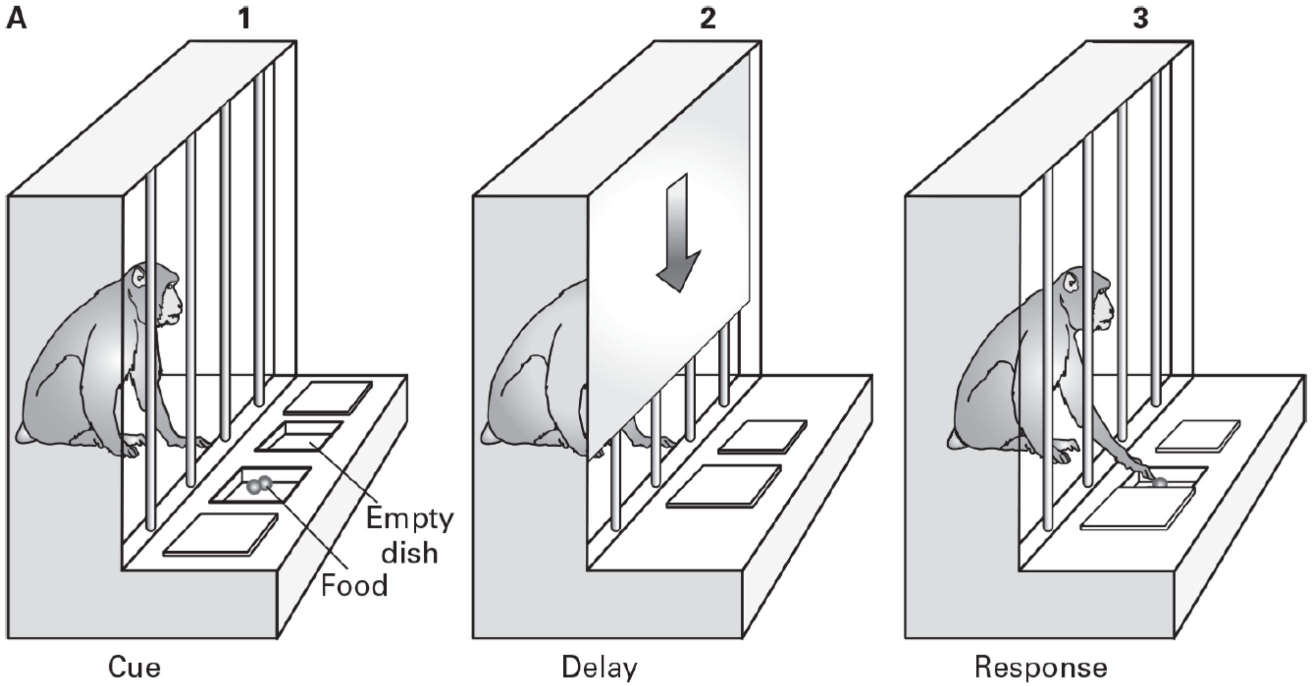
- History of Neuroscience:
  - Phrenology; feeling the skull (1810s~1840s) → Localization (Paul Broca, 1861)
  - Broca proposed that the frontal lobe was important for judgement, reflection, abstraction.
  - Broca, 1987: “great cerebral cortical system” – Broca believed that the brain of primates was different because of frontal dominance and the devolution of the limbic lobe.

Measuring Electrical Activity from Cortical Neurons

- Refinement in electrophysiological recording technique in the late 1950s accelerated the discovery of functions of brain parts.
  - Parts of the cortex that respond to <u>sensory stimulus</u>
  - E.g., Visual system: V1 responds to orientation of a bar, V2, V3, V4 responds with different sensitivity to object shape, color, and motion.
  - “Jennifer Aniston cell” – a cell with a preference for Jennifer Aniston pictures (Likely reflects memory association of the epilepsy patient)
- The focus of research shifted from stimulus properties to <u>cognitive variables</u>: How does an animal remember, decide to pay attention to, or switch between stimuli?

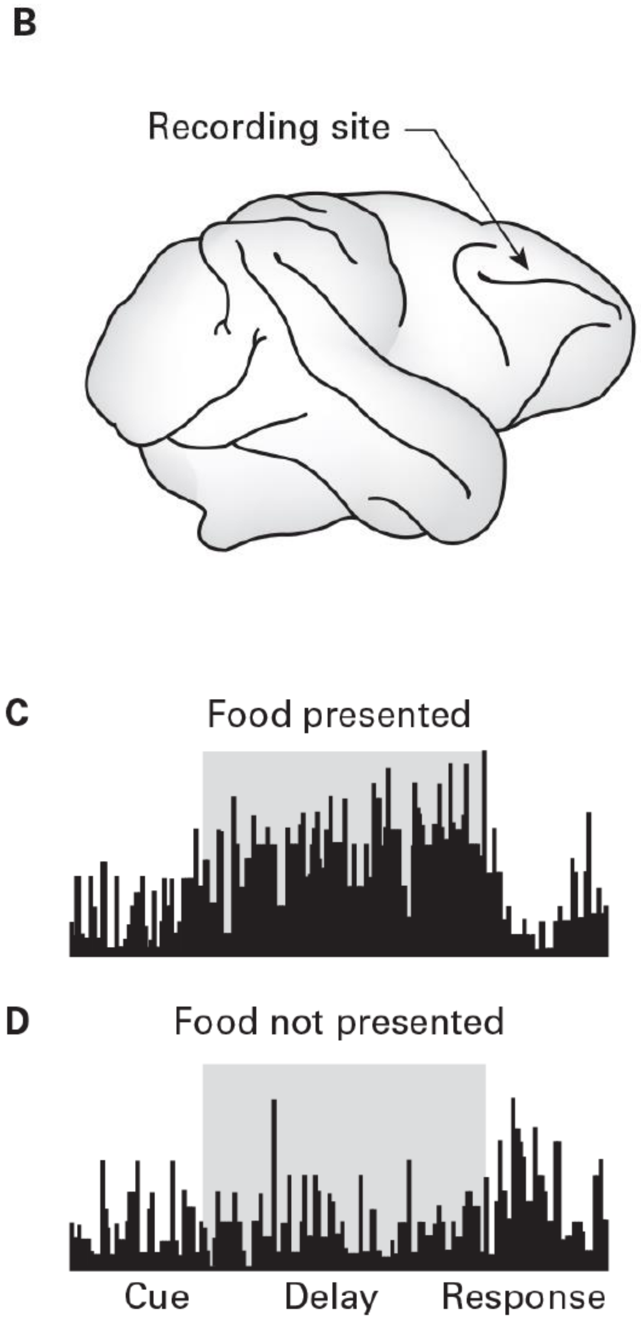
Holding Information in Mind

- Monkey experiments in the 1930s
  - Lesions of the dorsolateral part of the prefrontal cortex substantially impair working memory.
- Discovery of ***sustained firing*** in 1971
  - Visual Stimuli → Delay → Test Stimuli
  - What type of neural signal could bridge this gap? Probably *sustained firing*!
  - Prolonged activity = neural correlate of remembering during the delay



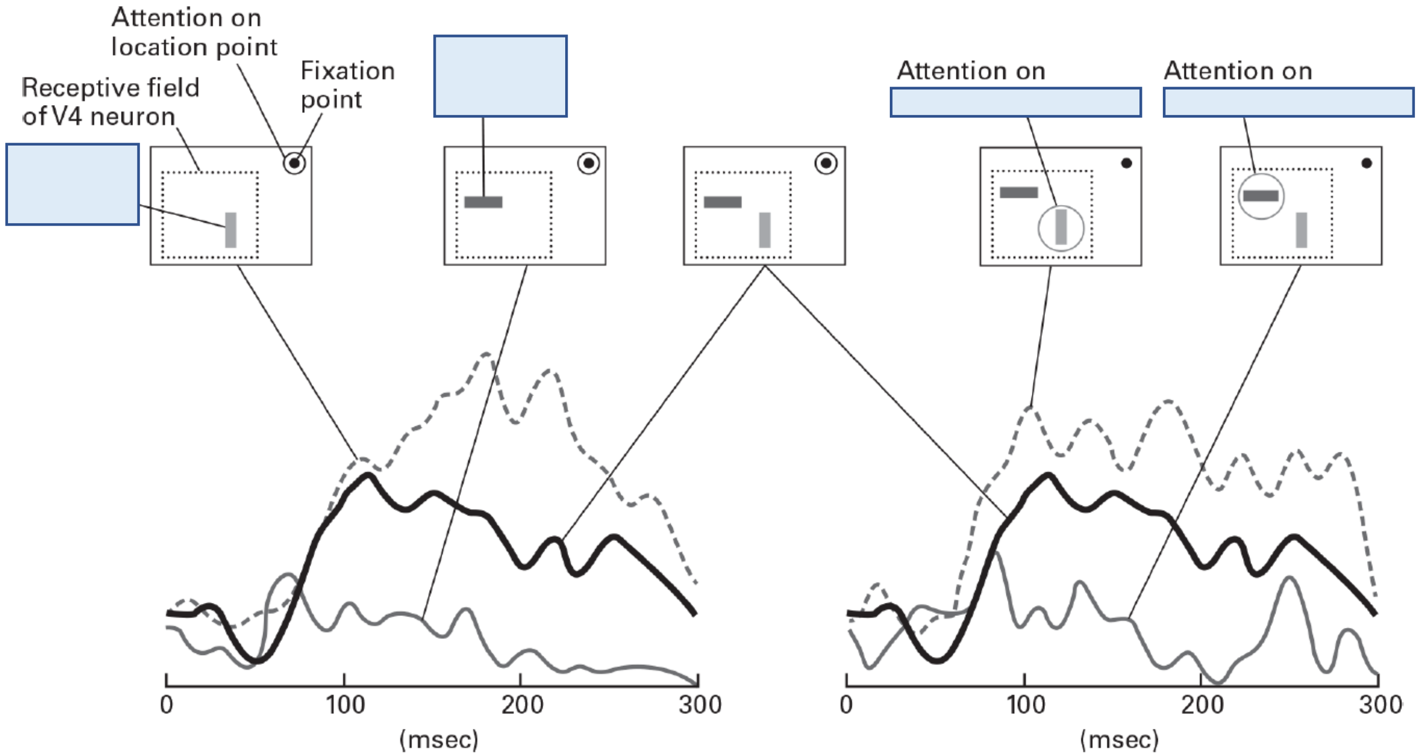
Holding Information in Mind

- What does *sustained firing* imply?
  - Short-term remembering?
  - Motor preparation?
- Cells in the prefrontal cortex prefer certain types of stimuli
  - Sustained responses are tuned to the **same attributes** of the stimulus
  - Shape, color, roundness (e.g., Constantinidis *et al.* 2001)
  - Then which attributes should the monkey pay attention to?



Should I Pay Attention to You?

- How do animals *select* some information while ignoring other information sources?
- The process of attention appears to “filter out” the unattended stimulus (mid 1980s)



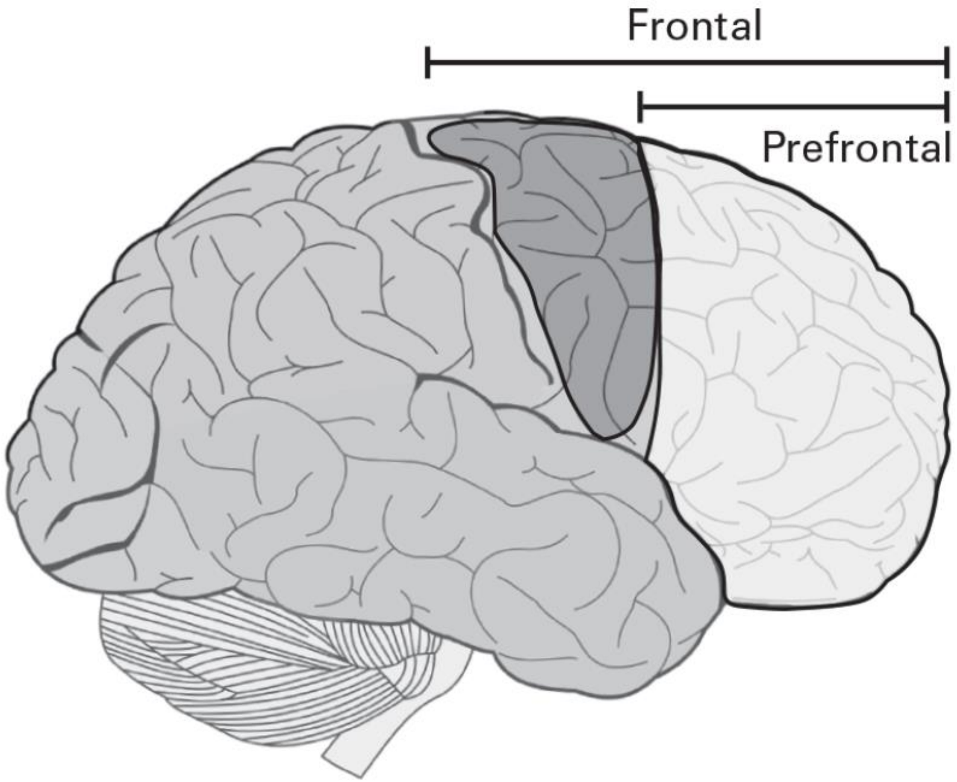
- This cell has an *effective* stimulus and an *ineffective* stimulus

Fill in the blanks!

- Attention is a collection of selection processes; a *competition mechanism*
- NLP: RNN transformer networks, attention blocks

The Final Frontier: Frontal Cortex

- **Central sulcus** separates the parietal and frontal cortices
- **Primary motor cortex** found on the first gyrus on the frontal side, linked to action
- **“Premotor” areas** reflect motor intentions and planning ➔ Prefrontal Cortex

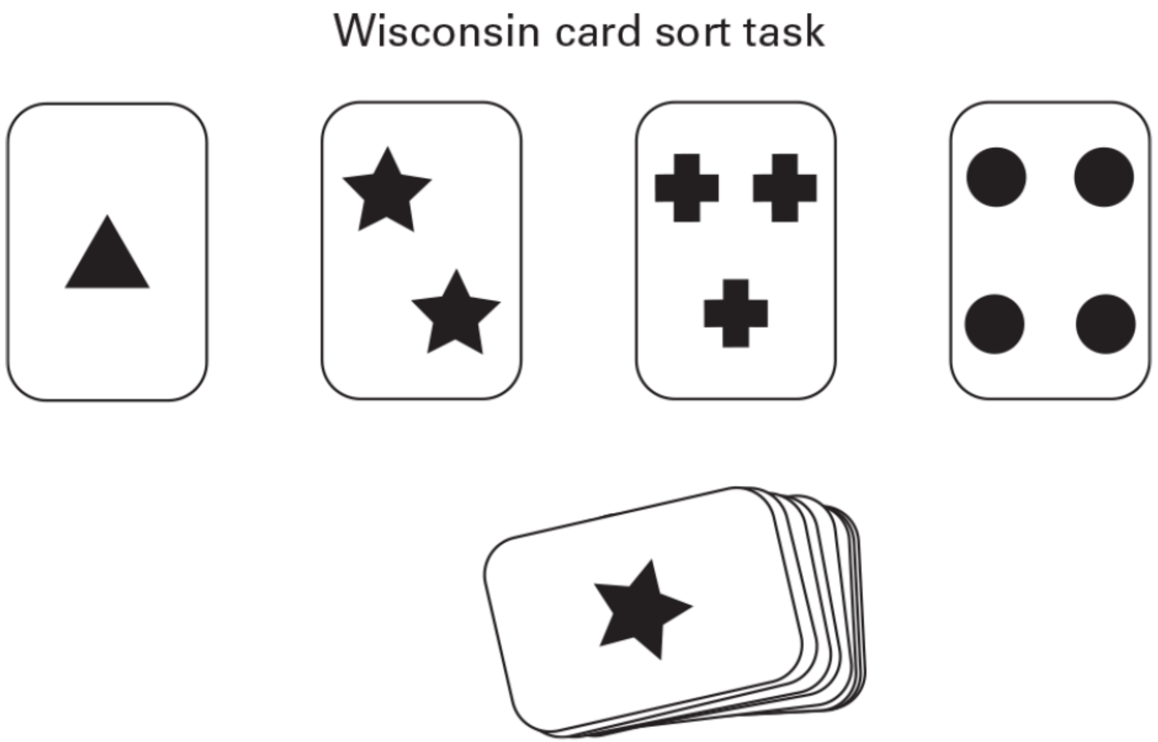


Prefrontal Cortex

- Frontal lesions do disrupt function, but it’s difficult to detect with conventional tests of intelligence like the standard IQ test.
- Prefrontal Cortex: ***Cognitive Control***
    - Lhermitte 1983; experiments with a hammer, a nail, a picture / a hypodermic needle demonstrate a condition of only being able to focus on the immediate present absent social knowledge.
- Wisconsin card sort task
    - Milner discovered that patients with frontal lesions performed worse that patients with lesions in other regions. They tended to stick to one rule even as they received negative feedback

Wisconsin Card Sort Task

- You can try it!



Stroop Task

Red

Blue

- You can try this, too! What does the Stroop Task tell us?

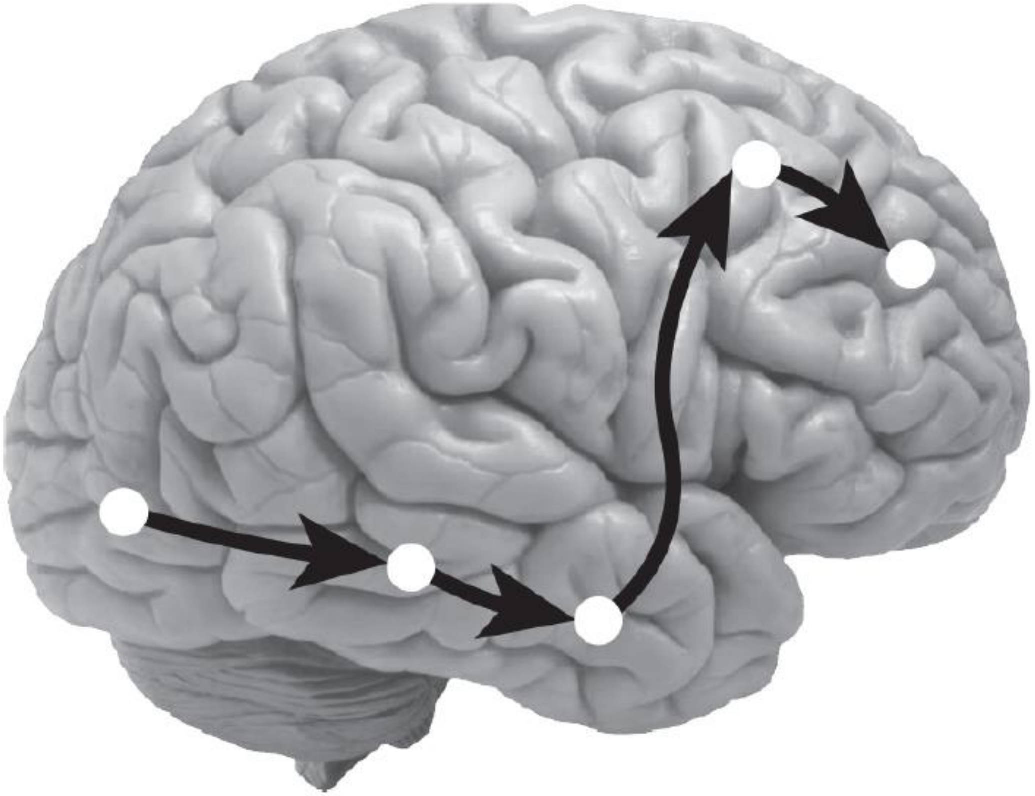
a) *Conflict Task*: Both detecting the presence of conflict (color vs. meaning) and resolving it are important cognitive functions

b) *Cognitive Control* comes to play when a *habitual action* is uncalled for, and should be recalibrated, modified, or completely inhibited.

c) Cognitive control in *goal-directed behavior* is the ability to select a weaker but task-relevant response in the face of competition from an otherwise stronger but task-irrelevant one. (It happens a lot when you try to solve math problems)

From Sensation to Cognition, One step at a Time

- *Sequential* information processing in the brain
  - A scheme favored until the late 1980s
- Cortical *hierarchy*
  - sensory processing → abstract processing
- Example: Vision & motor action



An Alternative View, the Predictive Brain

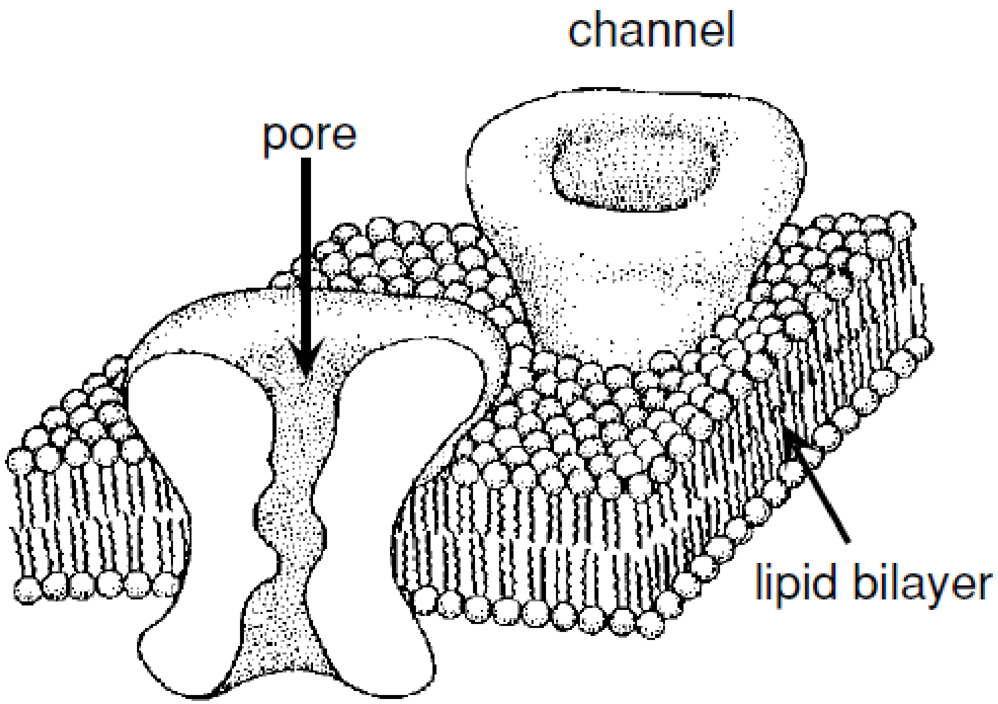
- ***Perception*** is directed by ***action*** so that ***effective behaviors*** can be generated.
- Vision is active and guided by endogenous computations that try to anticipate the ***most valuable future information*** for the animal
- Feedforward and feedback pathways: the **bidirectional architecture** has profound implications for our understanding of how the brain works
    - SGD, CBO
- Brains are not passive. A stimulus is registered against a host of expectations
- The brain *constructs* a probable version of the external world.
    - Just like humans construct theories and models of the world.

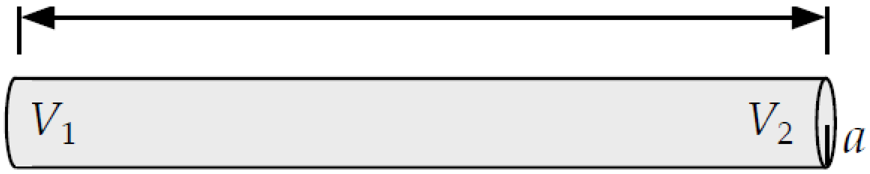
Coda

- Distinct dimensions of the same stimulus can be relevant based on the context at hand, and the system needs to be updated on the fly.
- Abundant *bidirectional* connectivity fosters a view that processing is as much about exogenous as about endogenous signals, leading to an active, **predictive** system.
- Furthermore, parallel pathways are capable of conveying information in a ***distributed manner***, creating an elaborate anatomical infrastructure that can support nonsequential and decentralized mechanisms.

So, what is a network model?

Neuroelectronics





$I_L = (V_2 - V_1)/R_L$

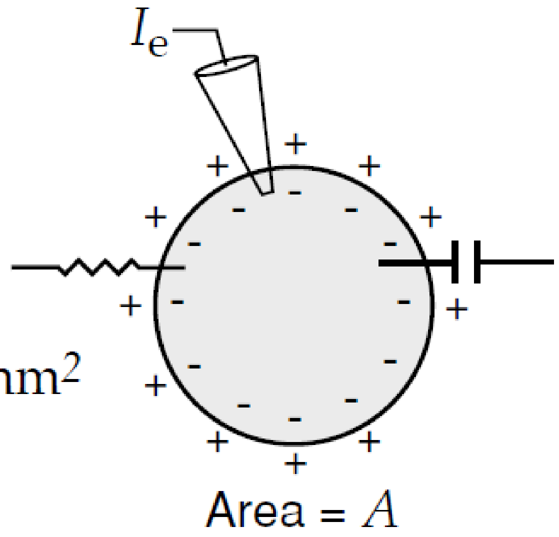
$R_L = r_L L/\pi a^2 \quad r_L \approx 1\ \mathrm{k\Omega\ mm}$

- Membrane Capacitance

$\Delta V = I_e R_m$

$R_m = r_m / A$

$r_m \approx 1\ \mathrm{M\Omega\ mm^2}$



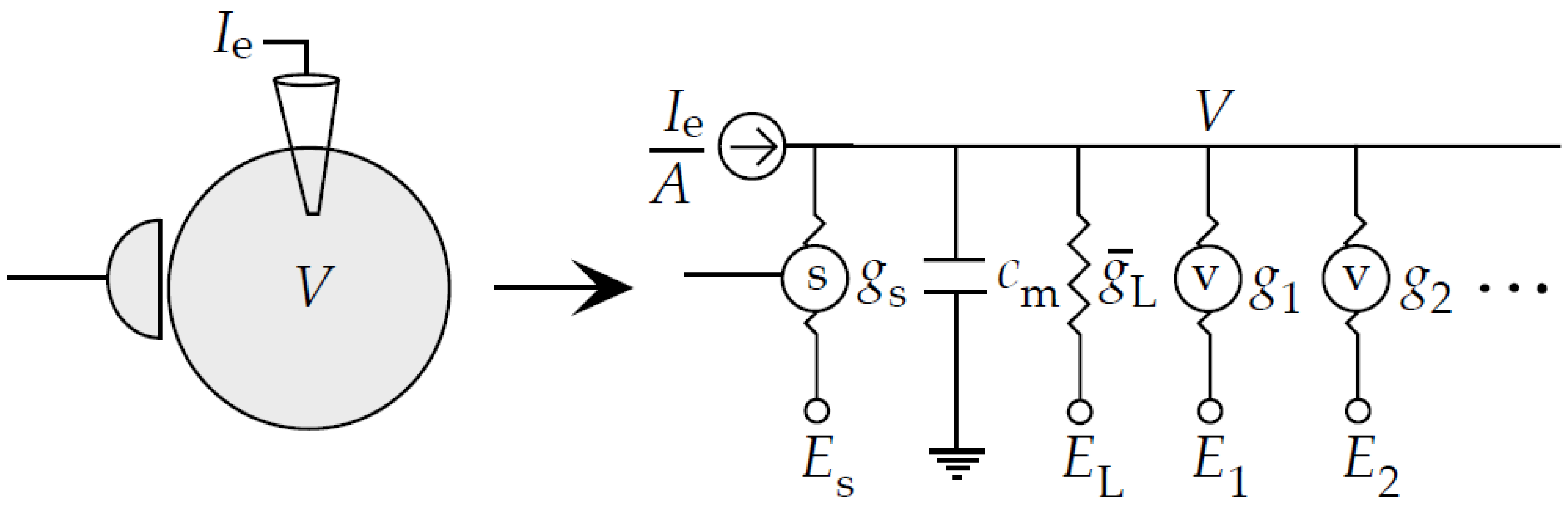
$Q = C_m V$

$C_m = c_m A$

$c_m \approx 10\ \mathrm{nF/mm^2}$

$$C_m \frac{dV}{dt} = \frac{dQ}{dt}$$

One-Compartment Neuron Model



$$c_m \frac{dV}{dt} = -i_m + \frac{I_e}{A}, \qquad i_m = \overline{g_L}(V - E_L) \implies \tau_m \frac{dV}{dt} = E_L - V + R_m I_e$$

i_m is the membrane current per unit area of cell membrane. The total current is $i_m A$.

Each g_i represent the conductance per unit area due to ion channels → lump into $\bar{g}$

The Math/Computer Science/Physics Perspective

- Firing rate models: **Feedforward** Network Model

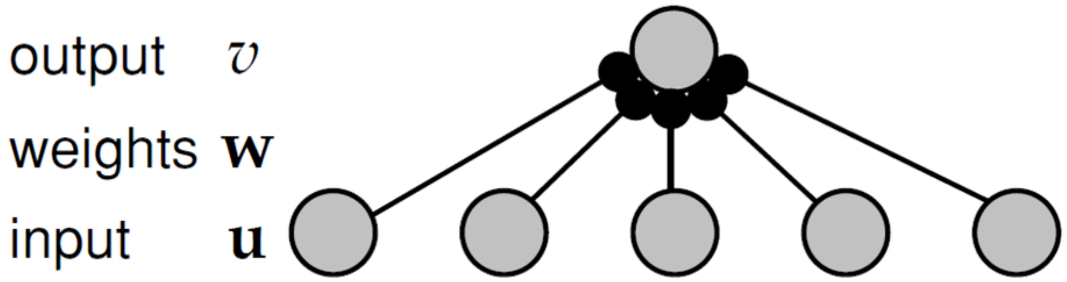


Figure 7.1 Feedforward inputs to a single neuron. Input rates $\mathbf{u}$ drive a neuron at an output rate v through synaptic weights given by the vector $\mathbf{w}$.

$$\tau_s \frac{dI_s}{dt} = -I_s + \mathbf{w} \cdot \mathbf{u}, \qquad v = F(I_s)$$

$$\tau_r \frac{dv}{dt} = -v + F(I_s) \approx -v + F(\mathbf{w} \cdot \mathbf{u}), \qquad (\tau_r \gg \tau_s)$$

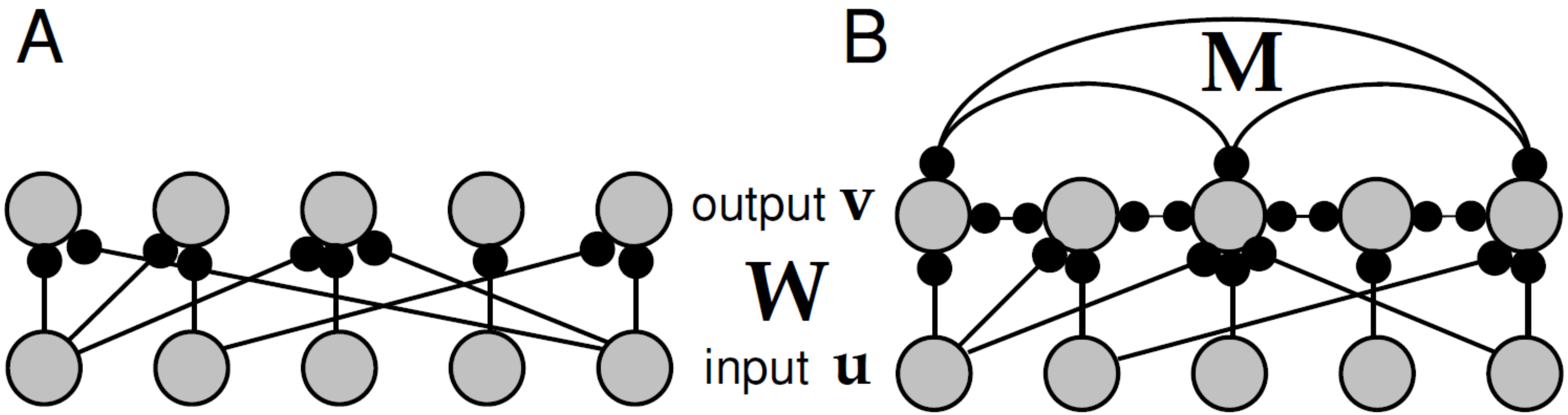


Figure 7.3 Feedforward and recurrent networks. (A) A feedforward network with input rates **u**, output rates **v**, and a feedforward synaptic weight matrix **W**. (B) A recurrent network with input rates **u**, output rates **v**, a feedforward synaptic weight matrix **W**, and a recurrent synaptic weight matrix **M**. Although we have drawn the connections between the output neurons as bidirectional, this does not necessarily imply connections of equal strength in both directions.

$$\tau_r \frac{d\mathbf{v}}{dt} = -\mathbf{v} + F(\mathbf{W}\cdot\mathbf{u} + \mathbf{M}\cdot\mathbf{v}), \qquad \mathbf{h} \coloneqq \mathbf{W}\cdot\mathbf{u}$$

Note that M is a symmetric matrix, so its eigenvalues are real and the corresponding eigenvectors are orthogonal.

Recurrent Networks

- Under linear approximation,

$$\tau_r \frac{d\mathbf{v}}{dt} = -\mathbf{v} + \mathbf{h} + \mathbf{M} \cdot \mathbf{v}$$

- Eigenvector expansion of **v**

$$\mathbf{v} = \sum_{i=1}^{N_v} c_i \mathbf{e_i} \quad \text{where } \mathbf{M} \cdot \mathbf{e_i} = \lambda_i \mathbf{e_i}$$

- It is easier to solve the differential equation for c_i than for **v**.

Recurrent Networks

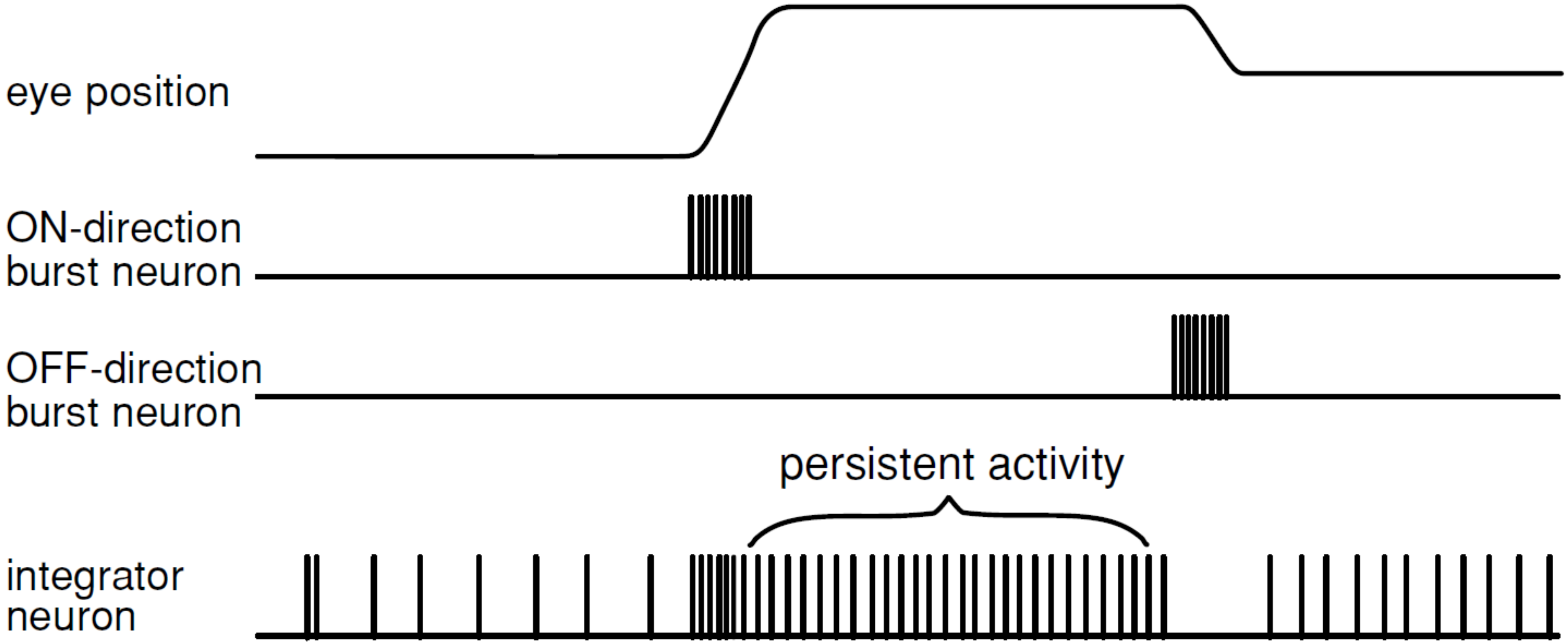
- Network Integration : when one of the eigenvalues is equal to 1, all others λ_v<1

$$\tau_r \frac{dc_\nu}{dt} = -(1-\lambda_\nu)c_\nu(t) + \mathbf{e}_\nu \cdot \mathbf{h}\,. \qquad \tau_r \frac{dc_1}{dt} = \mathbf{e}_1 \cdot \mathbf{h}\,,$$

- Then c_1 grows linearly with t while all other $c_v \rightarrow 0 \quad \forall v \neq 1$

$$\mathbf{v}(t) \approx \frac{\mathbf{e}_1}{\tau_r} \int_0^t dt'\, \mathbf{e}_1 \cdot \mathbf{h}(t')\,.$$

- The network activity is **sustained** in the absence of input, providing a memory of the integral of prior input.



The steady-state firing rate of the integrator neuron provides a memory trace of the maintained eye position. (Adapted from Seung et al., 2000.)

Learning Rules

- The Hebb Rule

$$\tau_w \frac{d\mathbf{w}}{dt} = v\mathbf{u}$$

- Neural Network Training
  - Objective: Minimize a loss function $f(\theta)$ where $\theta \in \mathbb{R}^n$
  - Algorithm: Gradient Descent (learning rate $\alpha_k > 0$)

$$\theta^{k+1} = \theta^k - \alpha_k \nabla f\left(\theta^k\right)$$