Machine Learning

Andrey V.Gavrilov

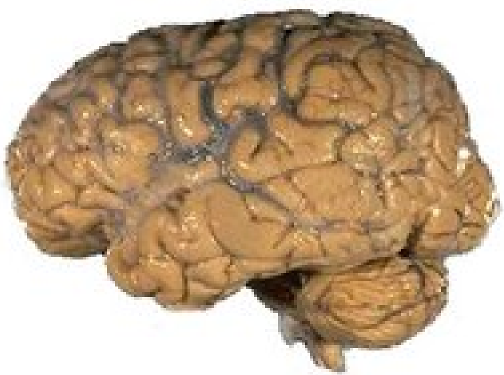
Colloquium 1

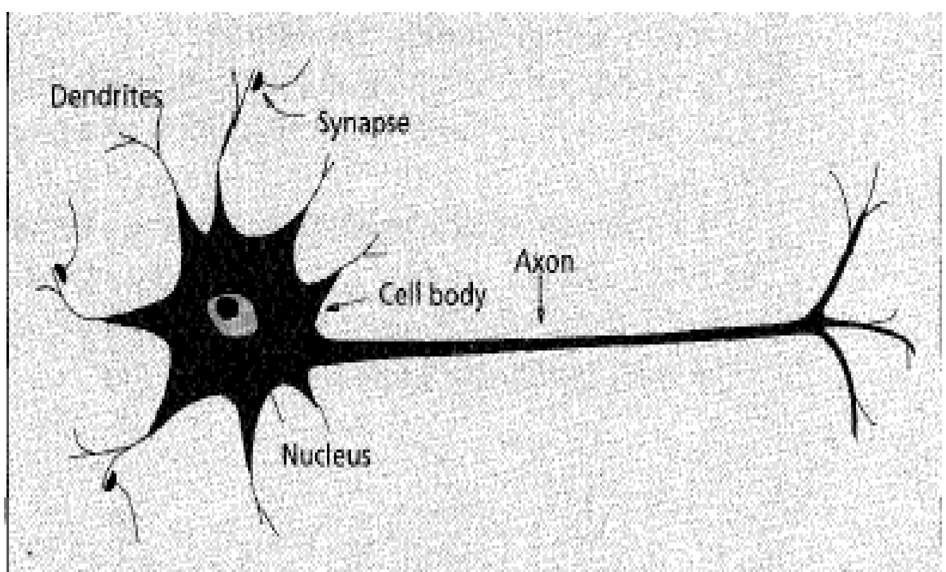
Features of construction and working of brain

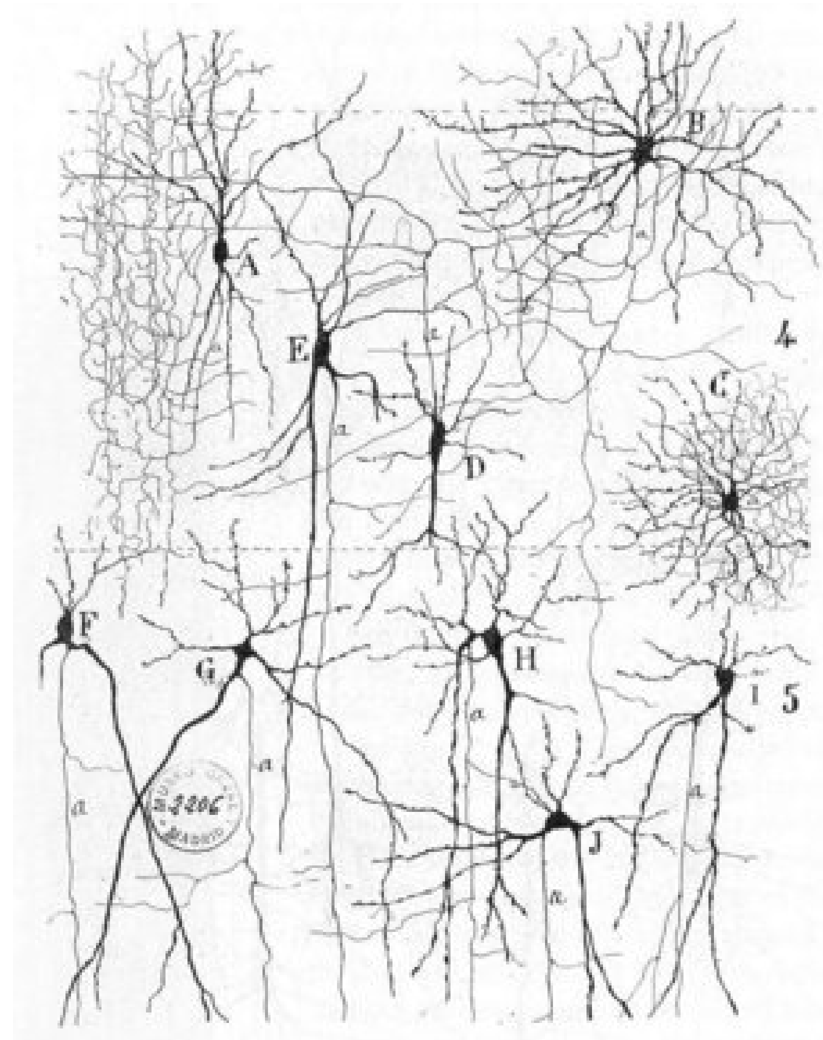
Kinds of learning algorithms in artificial neural networks

- Descend learning
  - EBP in MLP
- Competitive learning
  - SOM, ART
- Attractive learning, in particular, Hebbian rule
  - Hopfield model
- Stochastic learning
  - Boltzman machine
- Evolutionary learning
  - Genetic algorithms

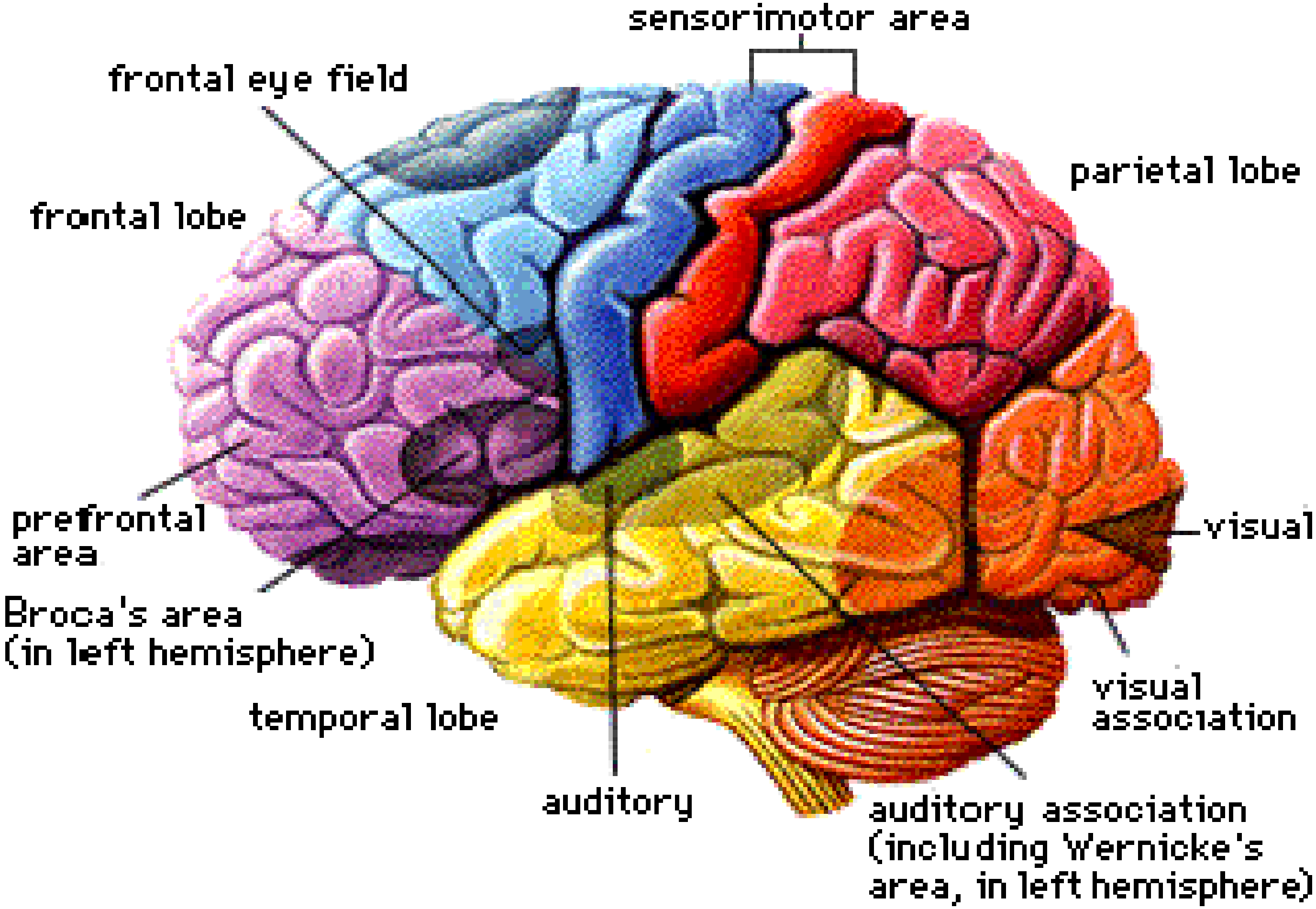
Human brain







The areas of brain



Criterions of Jeff Hawkins (in book "On Intelligence") for model of mind

- The inclusion of time in brain function
- The importance of feedback
- Any theory or model of the brain should account for the physical architecture of the brain

Senses

- We teach our children that humans have five senses: sight, hearing, touch, smell, and taste.
- We really have more.
- Vision is more like three senses— motion, color, and luminance (black-and-white contrast).
- Touch has pressure, temperature, pain, and vibration.
- We also have an entire system of sensors that tell us about our joint angles and bodily position. It is called the proprioceptive system (*proprio-* has the same Latin root as *proprietary* and *property*). You couldn't move without it.
- We also have the vestibular system in the inner ear, which gives us our sense of balance.
- Some of these senses are richer and more apparent to us than others, but they all enter our brain as streams of spatial patterns flowing through time on axons.

Neocortical memory

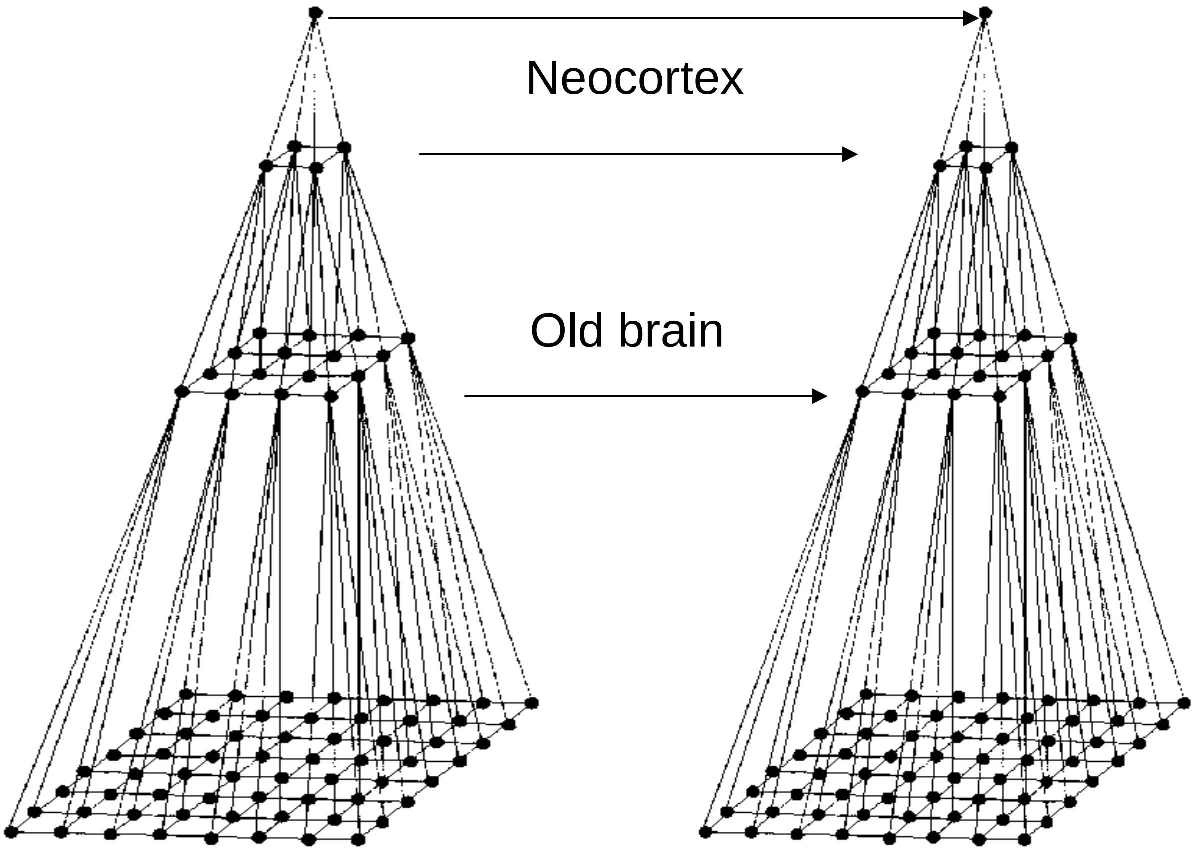
- The neocortex stores sequences of patterns
- The neocortex recalls patterns auto-associatively
- The neocortex stores patterns in an invariant form
- The neocortex stores patterns in a hierarchy
- Neocortex consists of 6 layers of neurons identical for different regions of one

Hierarchy in mind

Classification (rocognition) of situation (task)

Associative links

Forming of reaction on situation (solving)



Prediction

Basic behaviors

Essential features of action of brain

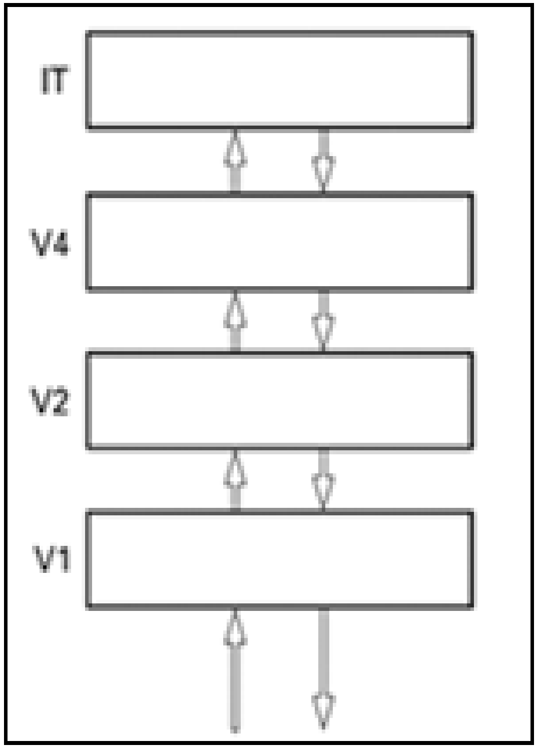
- Recognition of invariant patterns received from the world
  - Sufficient features of the environment
- Capability to predict
  - For avoidance of negative situations and for aspiration for positive ones

The understanding is prediction.

What we perceive is a combination of what we sense and of our brains' memory-derived predictions.

Instead of just making predictions based on the behavior of the old brain, the human neocortex directs behavior to satisfy its predictions.

The first four visual regions in the recognition of objects

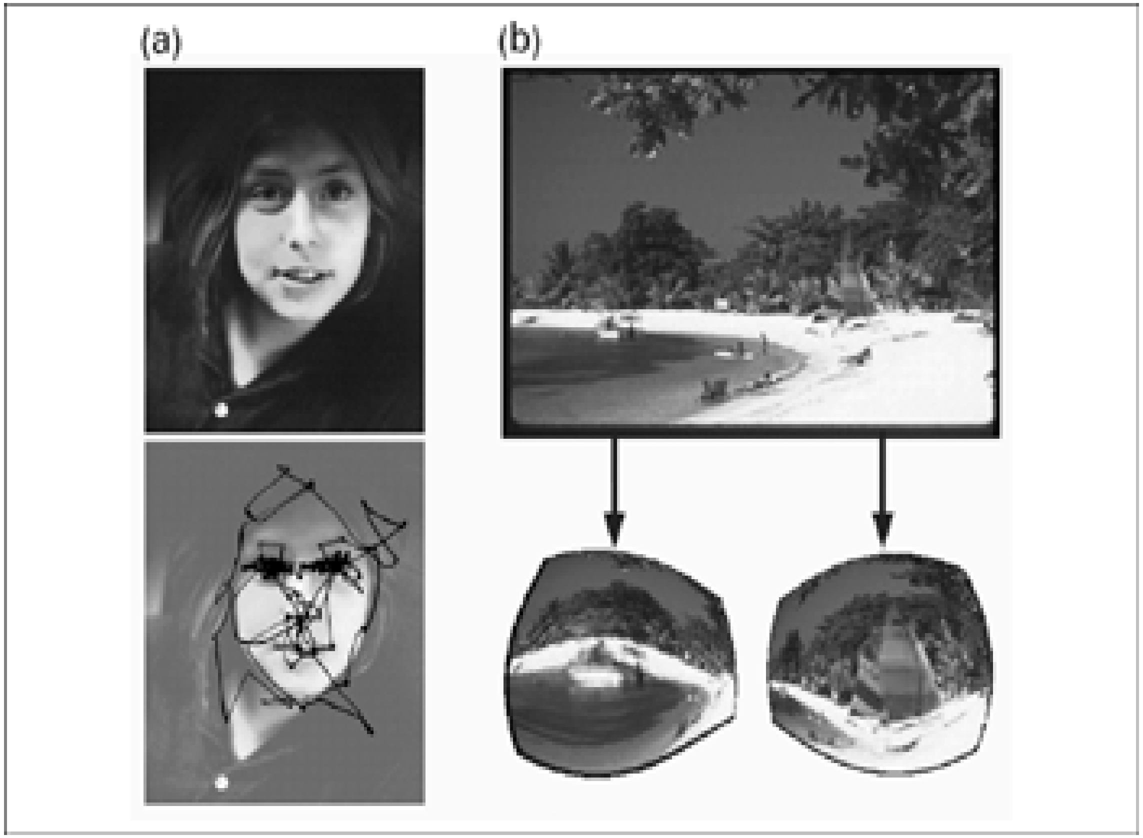


From retina

Vx – layers of visual region of neocortex

In the course of spanning four cortical stages from retina to IT, cells have changed from being rapidly changing, spatially specific, tiny-feature recognition cells, to being constantly firing, spatially nonspecific, object recognition cells. The IT cell tells us we are seeing a face somewhere in our field of view. This cell, commonly called a face cell, will fire no matter whether the face is tilted, rotated, or partially occluded.

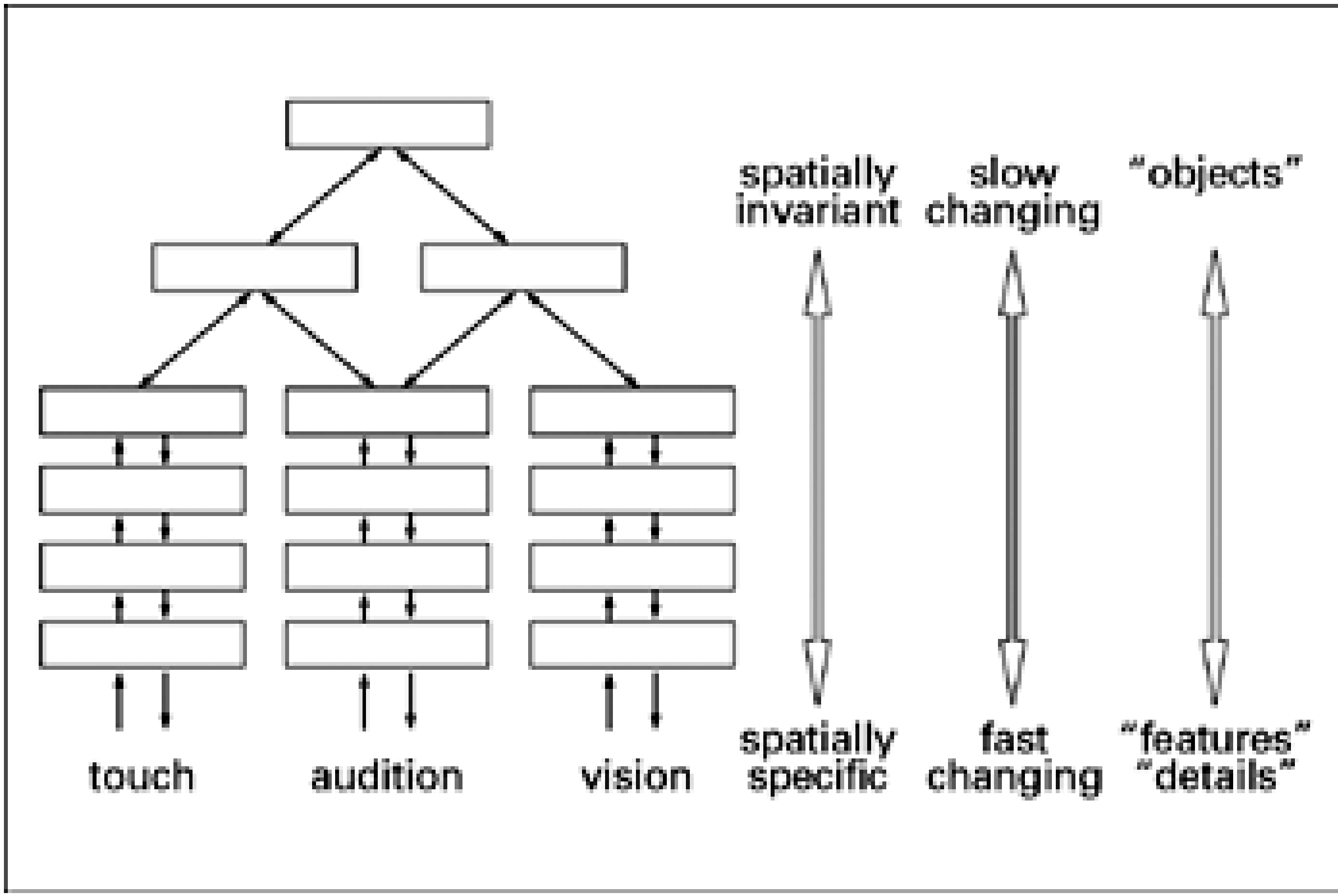
It is part of an invariant representation for "face”.



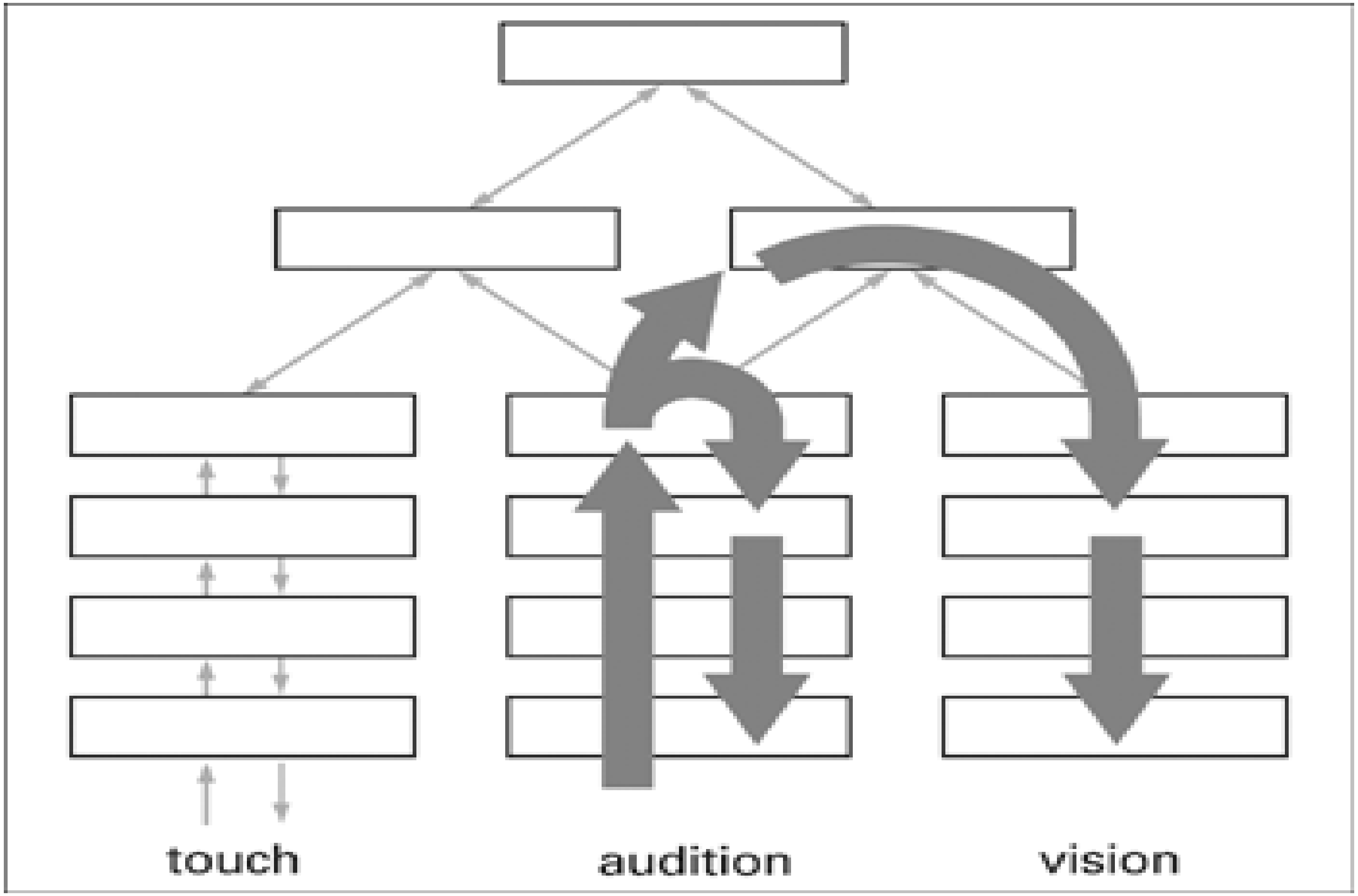
How the eye makes saccades across a human face

Distortion caused by the uneven distribution of receptors in the retina.

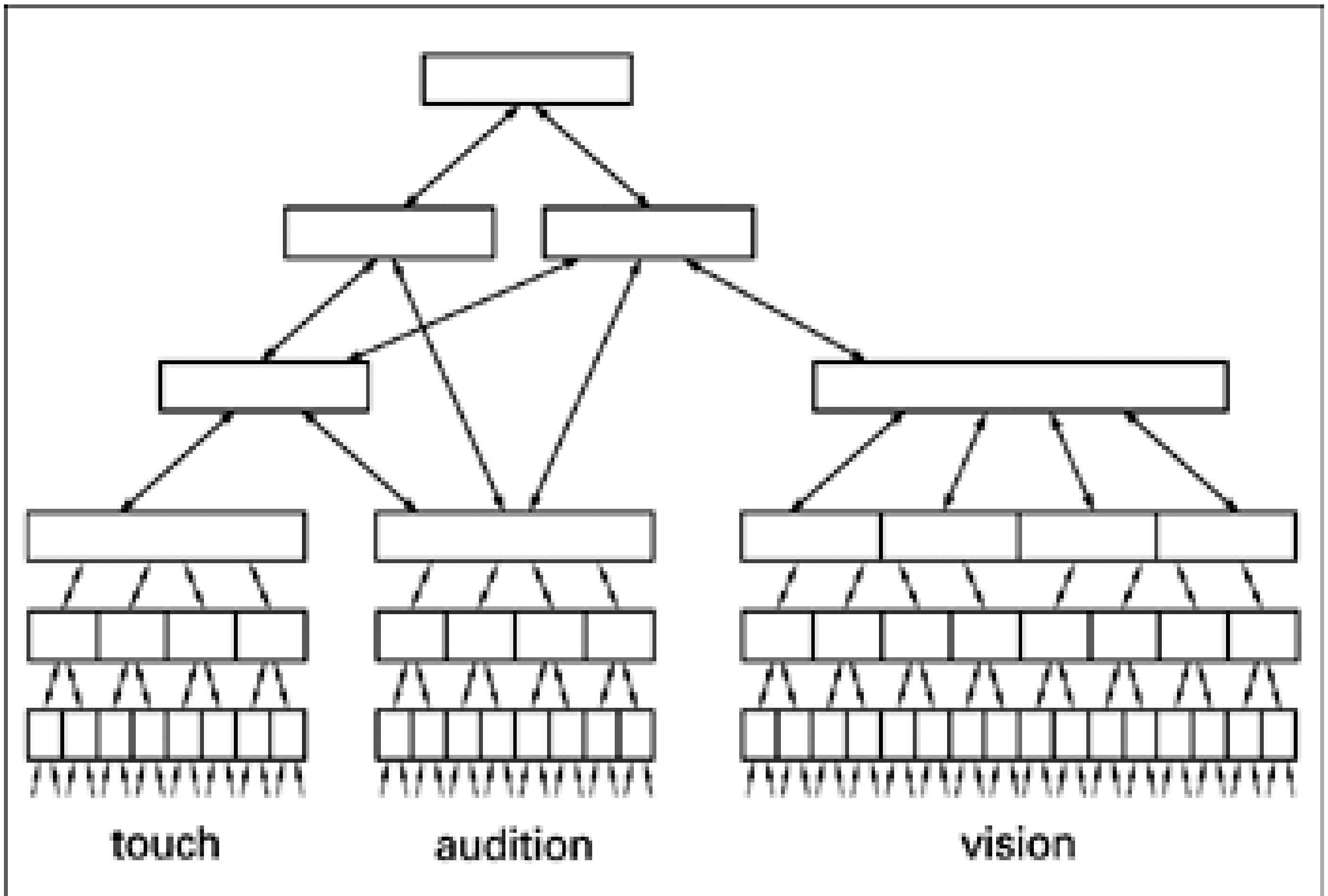
Forming invariant representations in hearing, vision, and touch



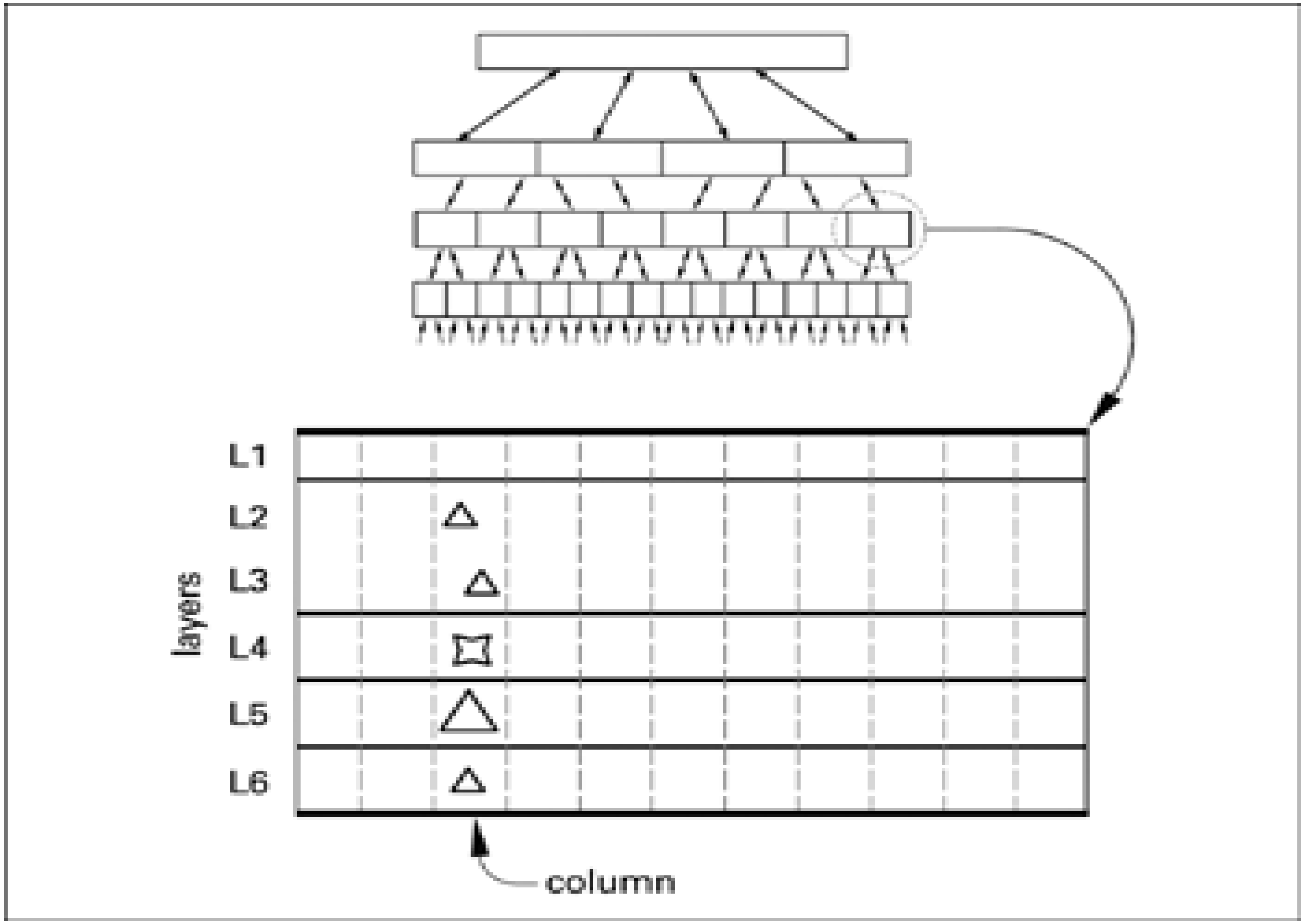
Information flows up and down sensory hierarchies to form predictions and create a unified sensory experience



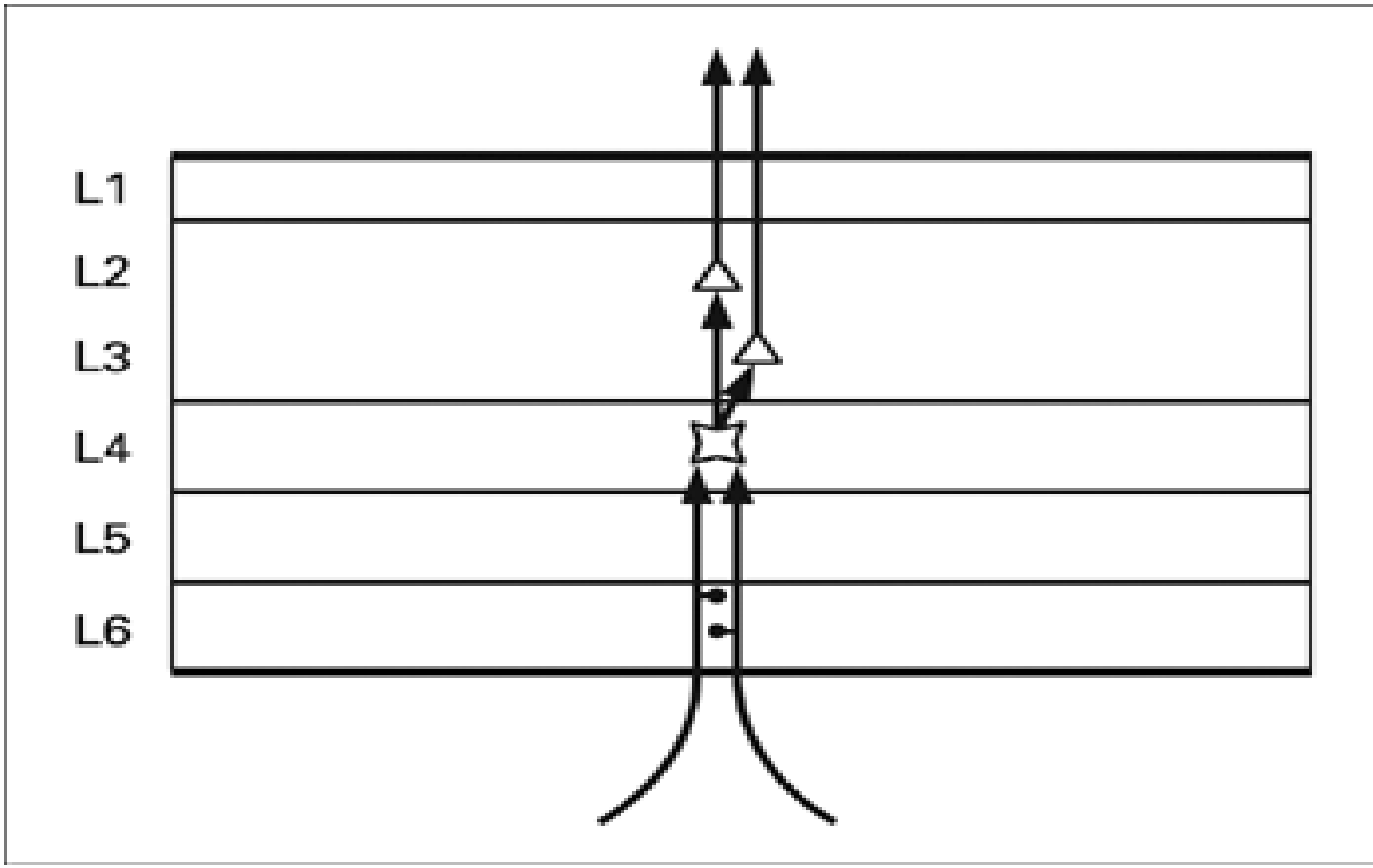
Alternate view of the cortical hierarchy taking into account left and right hemisphere



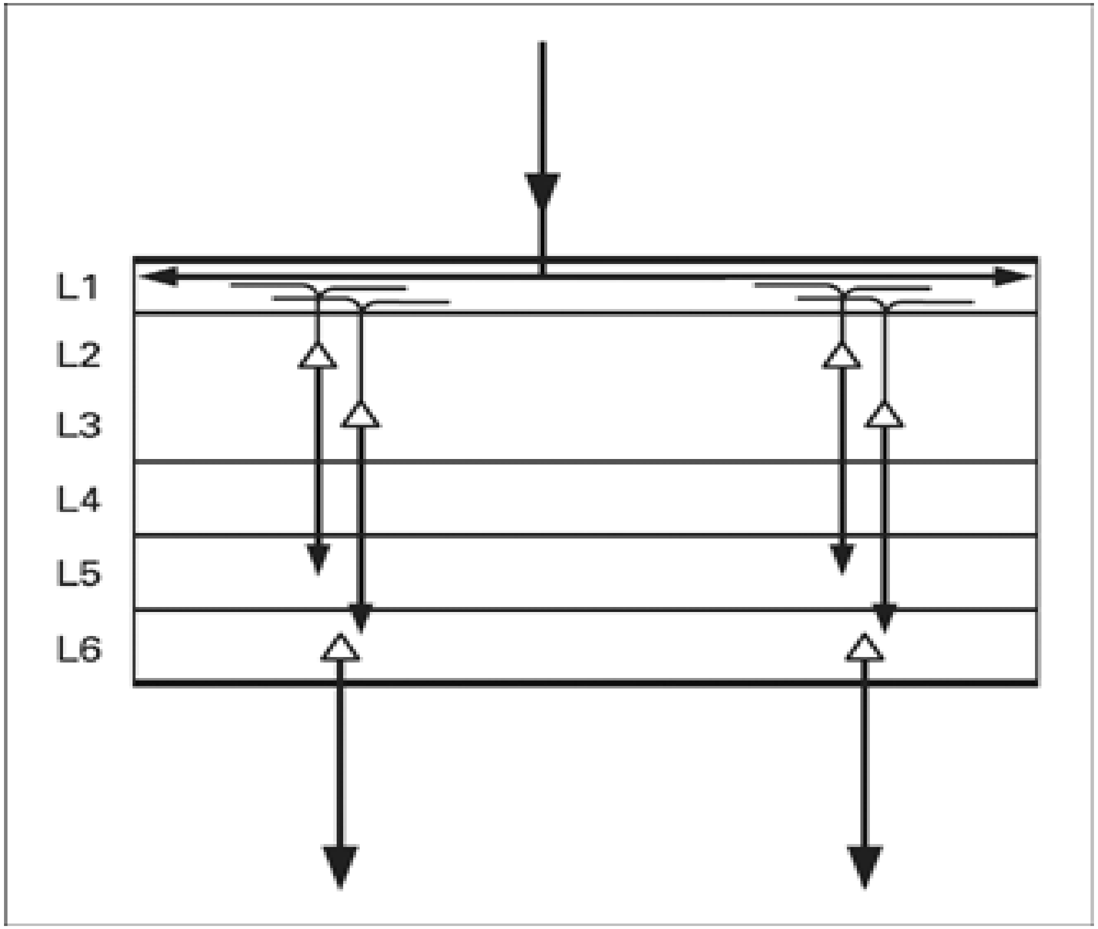
Layers and columns in a region of cortex



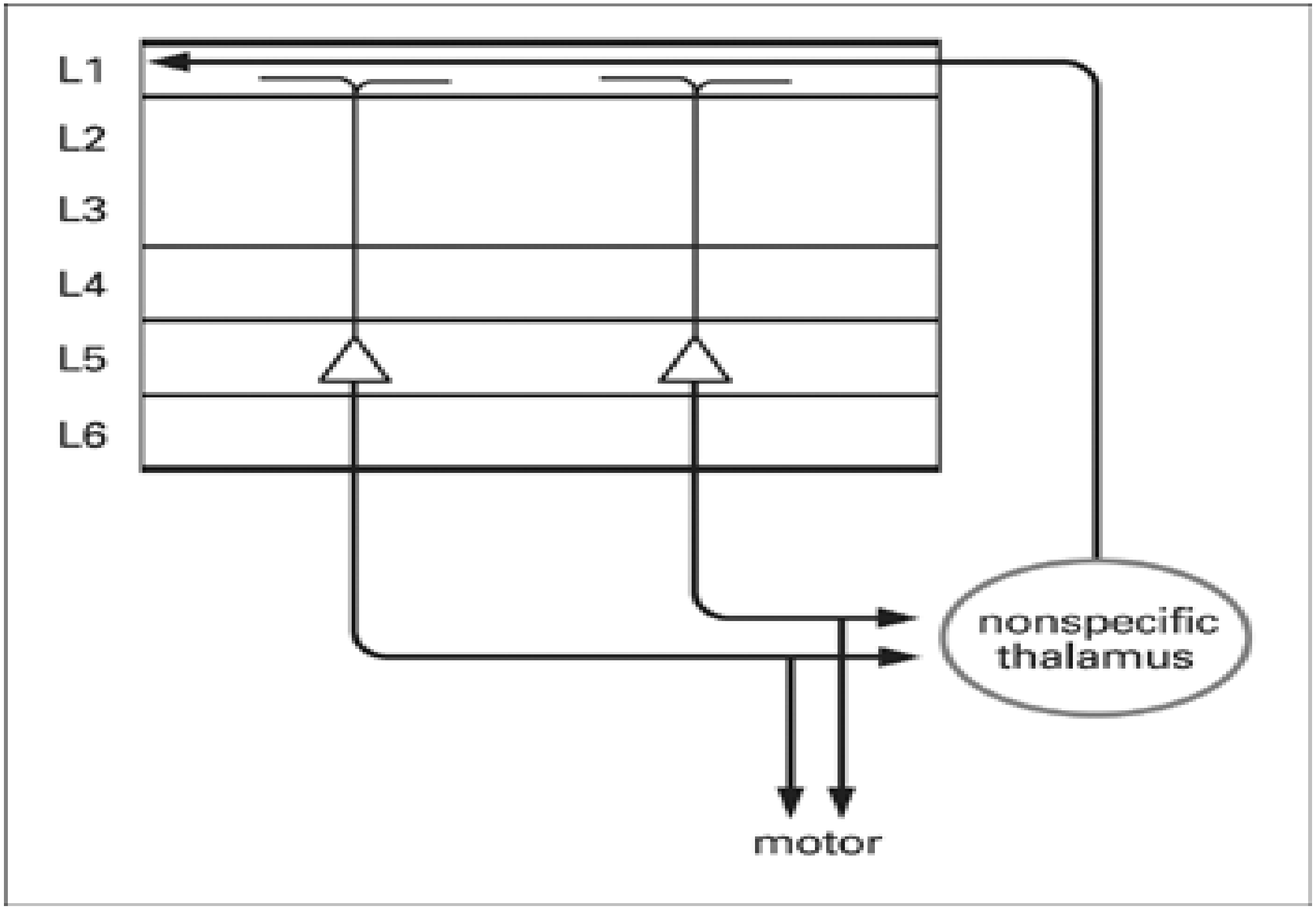
Upward flow of information through a region of cortex



Downward flow of information through a region of cortex



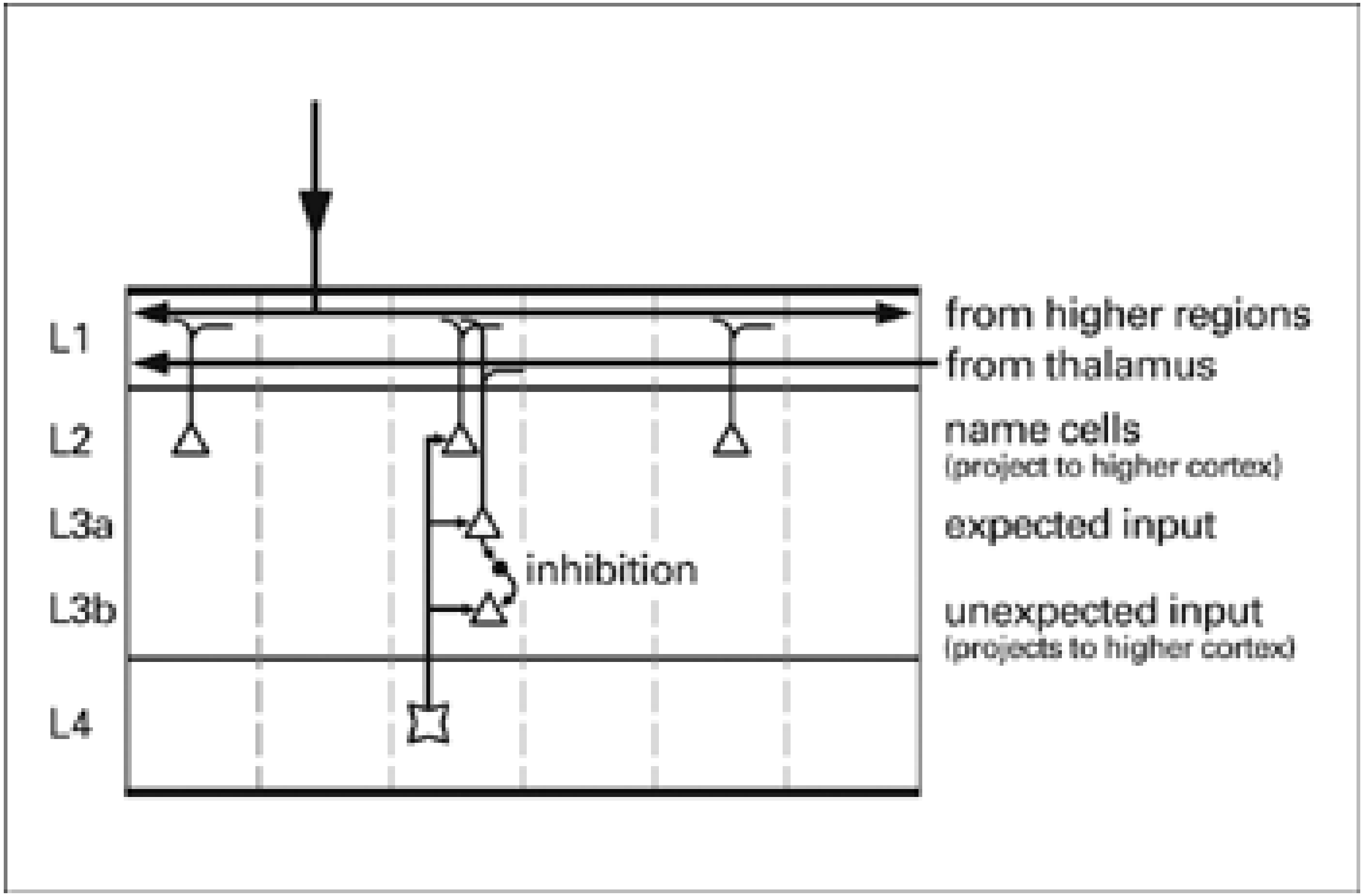
How current state and current motor behavior is communicated broadly via the thalamus



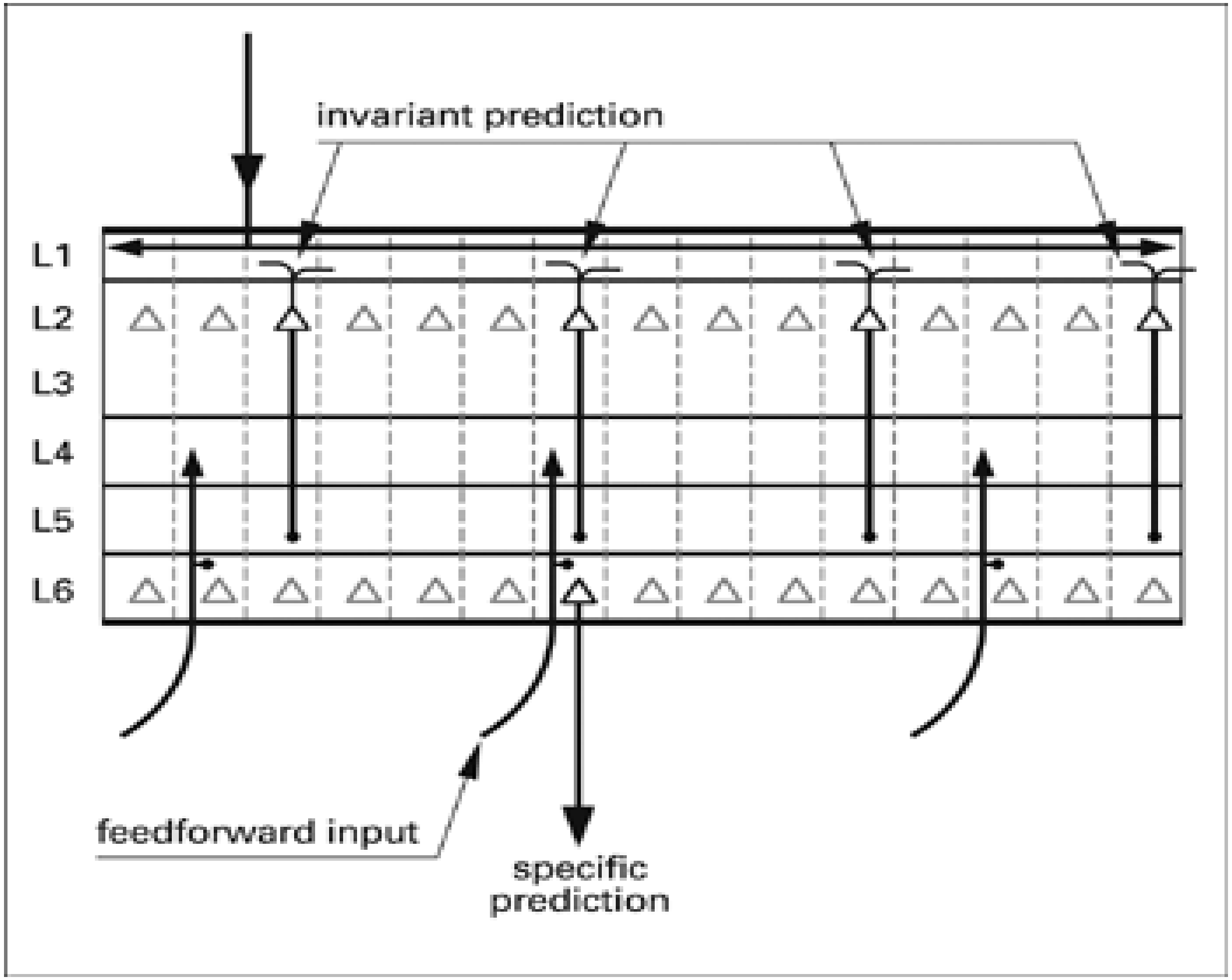
Three main circuits in mind

- Converging patterns going up the cortical hierarchy
- Diverging patterns going down the cortical hierarchy
- Delayed feedback through the thalamus

Forming a constant name for a learned sequence



How a region of cortex makes specific predictions from invariant memories



Hebbian learning

- When two neurons fire at the same time, the synapses between them get strengthened

Evolution of mind (or brain)

- Prediction of simple changes in environment (in protozoa, unicellular) and simple behavior based on it (tropism and avoidance)
- Prediction of simple changes in environment based on associative links between events (conditioned reflex) – appearance of sequences of signals
- Appearance of senses as fields of same sensors for more reliable recognition of events – appearance of associative memory of sequences of patterns (in reptiles)
- Invariant recognition of input patterns based on generalization – appearance of associative memory of sequences of preprocessed (generalized) patterns (prediction of generalized patterns), appearance of neocortex (in mammals)
- Increasing of role of neocortex on motor activity; Invariant recognition of sequences of generalized patterns – appearance of signs and manipulation of ones (in humans)

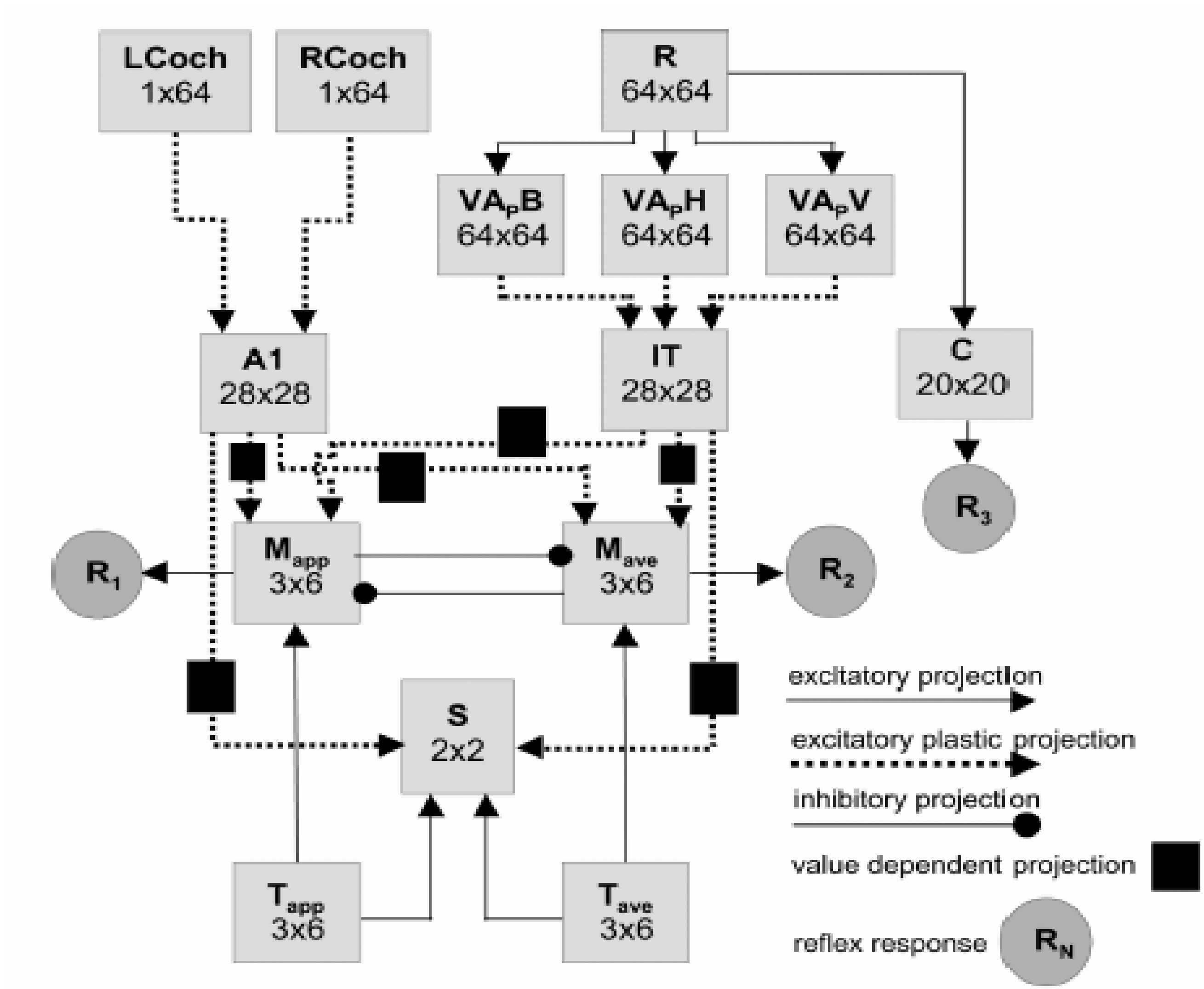
Investigations in Neuroscience Institute (California, San Diego)

Robot Darwin VII - mobile base equipped with several sensors and effectors and a neural simulation running on a remote computer workstation

Artificial environment

- Darwin VII's environment consisted of an enclosed area with black walls and a floor covered with opaque black plastic panels, on which we distributed stimulus blocks (6 cm metallic cubes) in various arrangements
- The top surfaces of the blocks were covered with two basic patterns: *blobs* (several white patches 2-3 cm in diameter) and *stripes* (width 0.6 cm, evenly spaced). *Stripes* on blocks in the gripper could be viewed in either horizontal or vertical orientations, yielding a total of three stimulus classes of visual patterns to be discriminated (*blob*, *horizontal* and *vertical*).
- A flashlight mounted on Darwin VII and aligned with its gripper caused the blocks, which contained a photodetector, to emit a beeping tone when Darwin VII was in the vicinity.
- The sides of the stimulus blocks were metallic and could be rendered either strongly conductive (appetitive or "good taste") or weakly conductive (aversive or .bad taste.). Gripping of stimulus blocks activated the appropriate taste neuronal units (either T_{app} or T_{ave} in Figure) to a level sufficient to drive the motor areas above a behavioral threshold. In the experiments, strongly conductive blocks with a striped pattern and a 3.9 kHz tone were chosen arbitrarily to be positive value exemplars, whereas weakly conductive blocks with a blob pattern and a 3.3 kHz tone represented negative value exemplars.

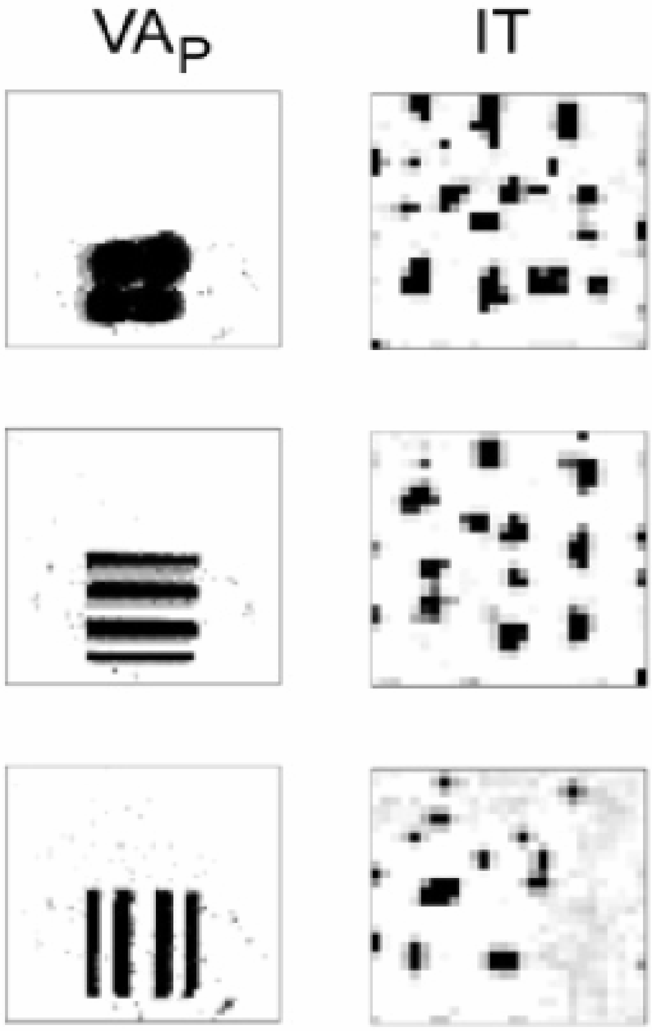
Schematic of the regional and functional neuroanatomy of Darwin VII. Each box denotes a neural area and the numbers signify the number of neuronal units in that area (height x width).



Comments to figure

Six major systems make up the simulated brain: an auditory system (*LCoch*, *RCoch*, *A1*) a visual system (*R*, VA_P, *IT*), a taste system (T_{app}, T_{ave}), sets of motor neurons capable of triggering behavior (M_{app}, M_{ave}), a visual tracking system (*C*), and a value system (*S*). The complete nervous system contained 18 neuronal areas, 19,556 neuronal units, and approximately 450,000 synaptic connections between neuronal units.

Visual system



Activity patterns in the visual receiving area R (left column) and in area IT (right column), after the development of visual perceptual categorization is completed.
The three stimulus classes (*blobs*, *horizontals* and *verticals*) are represented top to bottom, respectively.

Basic modes of behavior built into Darwin VII:

- IR sensor-dependent obstacle avoidance, visual exploration,
- visual approach and tracking,
- gripping and “tasting”,
- two main classes of innate behavioral reflex responses (appetitive and aversive)

With the exception of obstacle avoidance, selection among the above behaviors was under control of the simulated nervous system.

Some results of experiments

- Appetitive and aversive responses were triggered initially by taste, but after training or conditioning, these responses could be triggered by auditory or visual stimuli
- After conditioning, Darwin VII continued to grip and "taste" appetitive blocks, but learned to back away without picking up aversive blocks avoiding these blocks over 90% of the time
- After associating the initially neutral visual pattern with an innate valueloaded taste, the visual pattern was paired with a tone that emitted from the block. Darwin VII successfully learned this association, and avoided blocks with a tone predictive of bad taste and approached blocks with a tone predictive of good taste over 90% of the time

Features of nervous system of Darwin VII

- Connectivity from a topographically mapped primary area with transient activity to a non-topographically mapped higher area with more persistent activity (e.g. see V_{AP} ->*IT* in Figure)
- Sensory input that is continuous and temporally correlated with self-generated movement
- Activity-dependent learning in which competitive mechanisms categorize sensory information and select for appropriate behavioral repertoires.

Model of neuron

$$A_i(t) = \sum_{l=1}^{M} \sum_{j=1}^{Nl} c_{ij} s_j(t)$$

$$s_i(t+1) = \phi(\tanh(g_i(A_i(t) + \omega s_i(t))))$$

where $\phi(x) = 0$ for $x < \sigma_i^{fire}$, otherwise $\phi(x) = x$, g_i is a scale factor, and ω determines the persistence of unit activity

Changing of connections weights

$$\Delta c_{ij}(t+1) = \eta s_j(t) BCM(s_i(t))(V(t) - 0.1)$$

where η is a fixed learning rate ($\eta = 1.4$), $s_i(t)$ and $s_j(t)$ are the activities of the post- and pre-synaptic units respectively and $V(t)$ is the mean activity in area S. The term $(V(t) - 0.1)$ causes depression of plastic connections in the absence of value system activity.

The function $BCM()$ is based on the rule of Bienenstock *et al.* [2] and is implemented as follows ($\rho = 6$, $\theta_1 = \theta_2 = 0.1$, $k1 = k2 = 0.45$):

$$BCM(x) = \begin{cases} 0 & x < \theta_1 \\ k_1(\theta_1 - x) & \theta_1 \leq \frac{\theta_1+\theta_2}{2} \\ k_1(x - \theta_2) & \frac{\theta_1+\theta_2}{2} \leq x < \theta_2 \\ \frac{k_2 \tanh(\rho(x-\theta_2))}{\rho} & \text{otherwise} \end{cases} \qquad (11)$$

References

- *The Handbook of Brain Theory and Neural Networks.* – Eds. M.A. Arbib, MIT Press, 2003 (**Electronic version is available**)
- Hawkins J., Blakeslee S. *On Intelligence*. 2005. **(Electronic version is available)**