Neurorobotics

Module 1: Background and Foundations

Lecture 7: Reinforcement Learning and Prediction
Darwin VII Case Study: perceptual categorization and conditioning in a brain-based device

David Vernon
Carnegie Mellon University Africa

www.vernon.eu

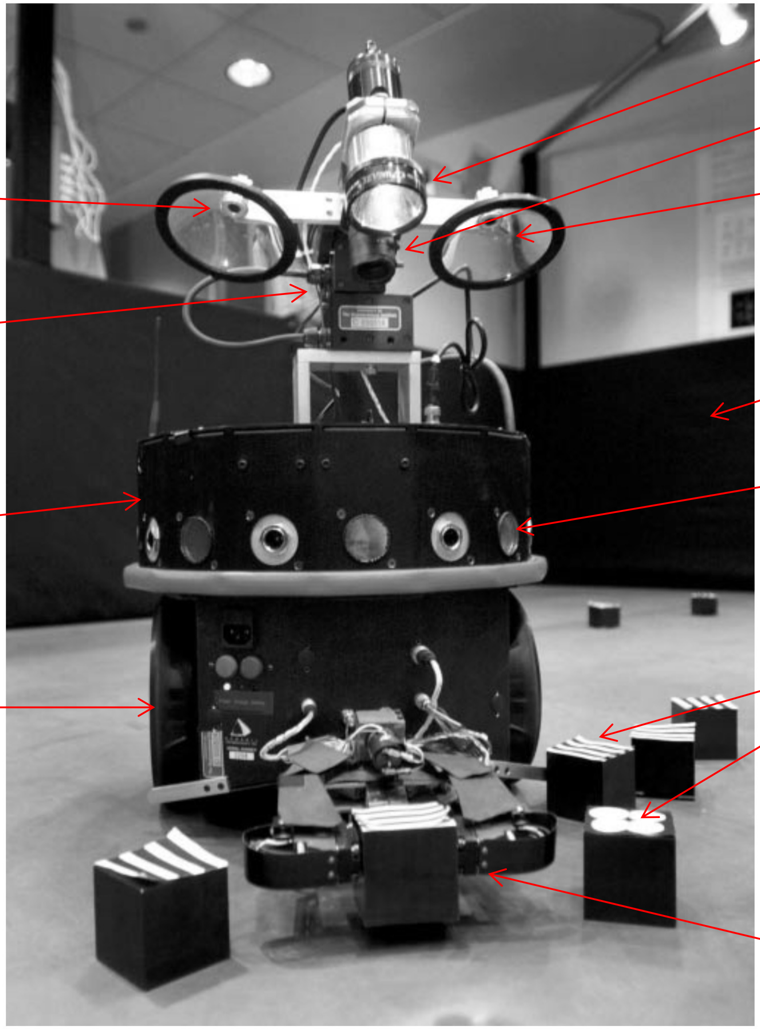
Darwin VII

Microphone

Pan and tilt camera head

NOMAD robot platform
(Neurally Organized
Mobile Adaptive Device)

Omnidirectional wheels
(aka Swedish wheels)



Flashlight

CCD camera (64 x 64 images)

Microphone

Opaque plastic panels enclosing the workspace

Eight infrared proximity sensors for obstacle avoidance

Blocks with stripes and blobs

Emit low and high frequency tones
when the flashlight illuminates a photodetector

Metallic sides: weakly conductive or strongly conductive

Gripper with conductivity sensors
(to simulate sense of taste)

Darwin VII

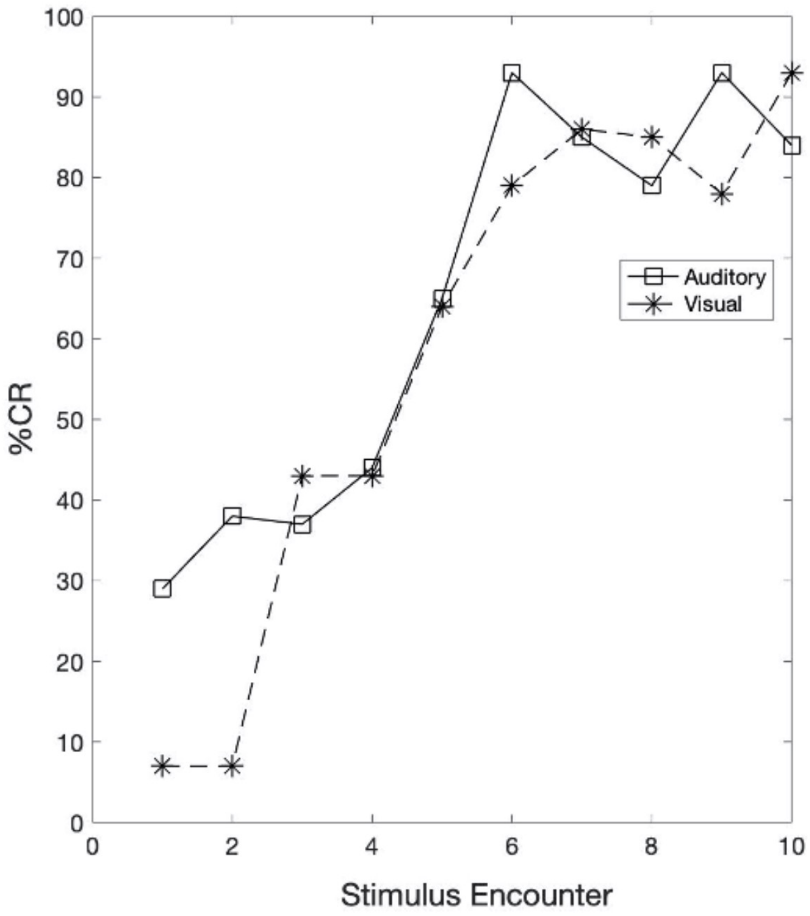
Basic, inbuilt behaviors

- Infrared (IR) sensor obstacle avoidance
- Visual exploration
- Visual approach and tracking
- Gripping
- "Tasting"
- Innate behavioral reflexes
  - Aversive (avoidance)
  - Appetitive (attraction)

Darwin VII

Darwin VII is a conditioning exercise

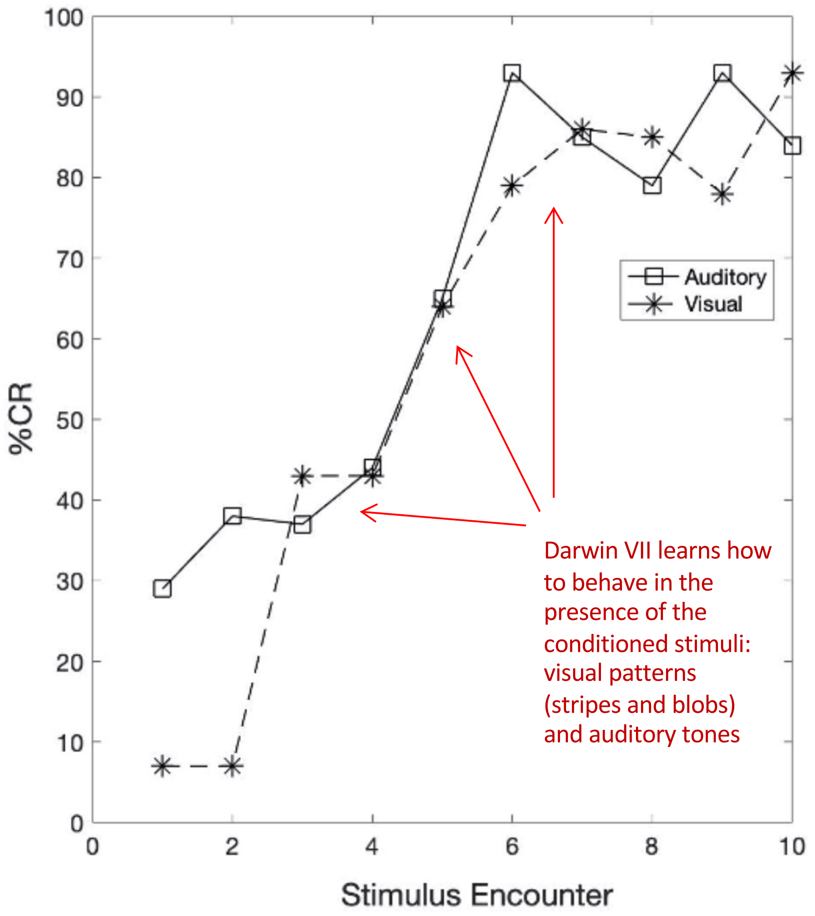
- Unconditioned stimulus:
  the taste (conductivity) of the block
- Unconditioned response:
  the approach (appetitive) or avoid (aversive) behavior
- Conditioned stimulus:
  the sound and visual categories
- Conditioned response:
  the approach (appetitive) or avoid (aversive) behavior



Darwin VII

Darwin VII is a conditioning exercise

- Prior to conditioning, taste triggers the behavioral response
- After conditioning, either a visual pattern or auditory pattern can evoke the behavioral responses



Darwin VII

Value-based learning in Darwin VII is a neural implementation of model-free reinforcement learning

- States – perception of
  - Low frequency tone (emitted by blocks when the flashlight shines on a photodetector)
  - High frequency tone (emitted by blocks when the flashlight shines on a photodetector)
  - Blob pattern on the blocks
  - Horizontal stripe pattern on the blocks
  - Vertical stripe pattern on the blocks

- Actions
  - Appetitive approach to an object
  - Aversive avoidance of an object

- Reward
  - Positive: strongly conductive "taste", stripe pattern, high 3.9 kHz tone
  - Negative: weakly conductive "taste", blob pattern, low 3.3 kHz tone

Darwin VII Brain-based Device

Six major regions in the simulated nervous system

Auditory (Cochlea): LCoch & RCoch

Visual: R, VA_p (VA_pB, VA_pH, VA_pV), IT

Taste: T_{app}, T_{ave}

Motor neurons: M_{app}, M_{ave}

Visual tracking : C

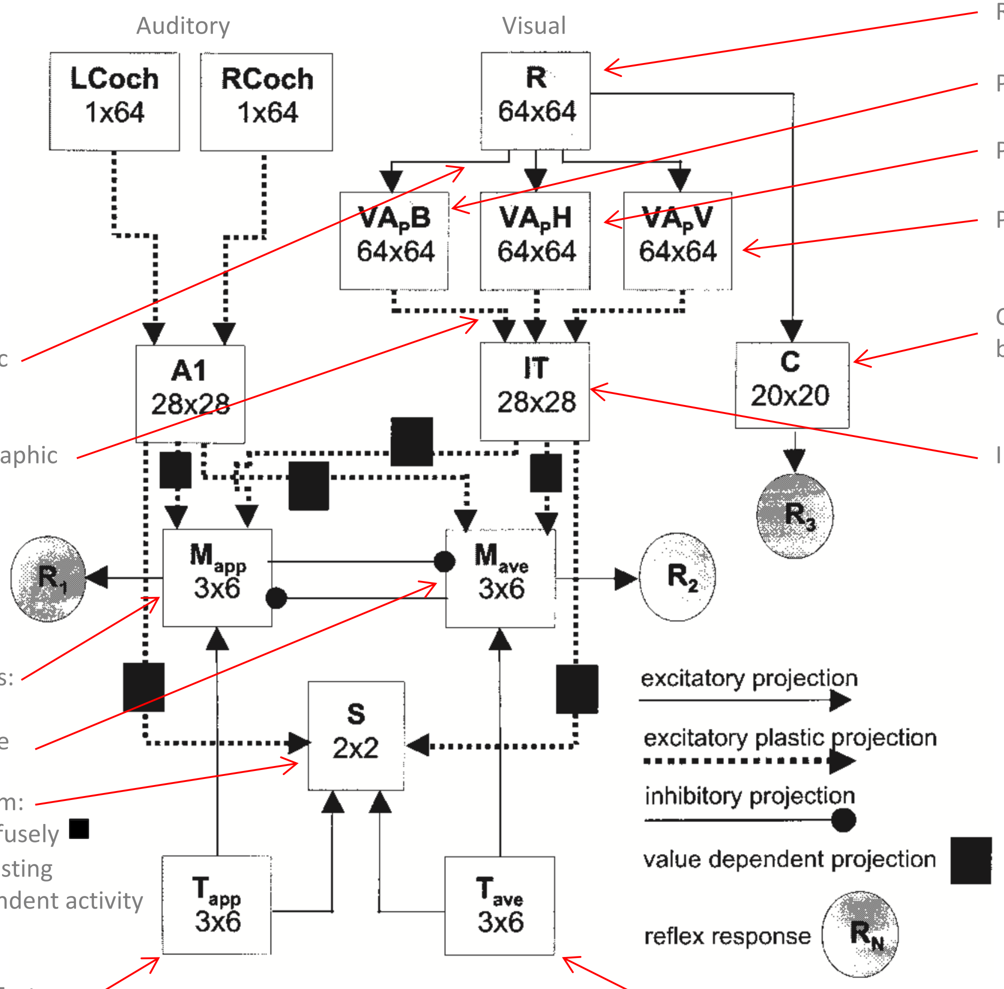
Value system: S

18 neural areas
19556 neuronal mean firing rate units (200 ms)
~450 000 synaptic connections

Activity-dependent
synaptic strength adjustment

Value-dependent
synaptic strength adjustment

Synaptic strength is adjusted using a modified Bienenstock, Cooper, and Munro (BCM) learning rule



Retina

Primary visual area for blob detection

Primary visual area for horizontal stripe detection

Primary visual area for vertical stripe detection

Colliculus for visual tracking based on contrast in brightness

Inferotemporal cortex

Topographic projection

Non-topographic projection

Motor areas: appetitive and aversive

Value system: projects diffusely ■ with long-lasting value-dependent activity

Appetitive Taste

Aversive Taste

Four different types of projection:

VA_p area sends activity- dependent excitatory plastic projections to IT

A1 and IT send activity- dependent excitatory plastic projections to M (producing behavior) and to S

T sends excitatory projections to M (producing behaviors) and S

The value dependent projections are shown implicitly using ■

Darwin VII

Mean firing rate neuronal units

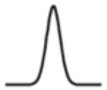
- Activity of a unit corresponds to the firing activity of a group of neurons
- Averaged over a period of 200 ms
- Approximately the time required to
  - Compute the neuronal unit activities
  - Update the connection strengths of plastic connections
  - Generate motor output

Recall from NR01-02

Neurons

Neuron models

- Alternatively, model with mean firing rate neuron
  - Average firing rate of a pool of neurons over tens of milliseconds
  - Several rate functions are used
    - Step, sigmoid, tanh, threshold-linear (ReLU), …
  - Same as activation functions

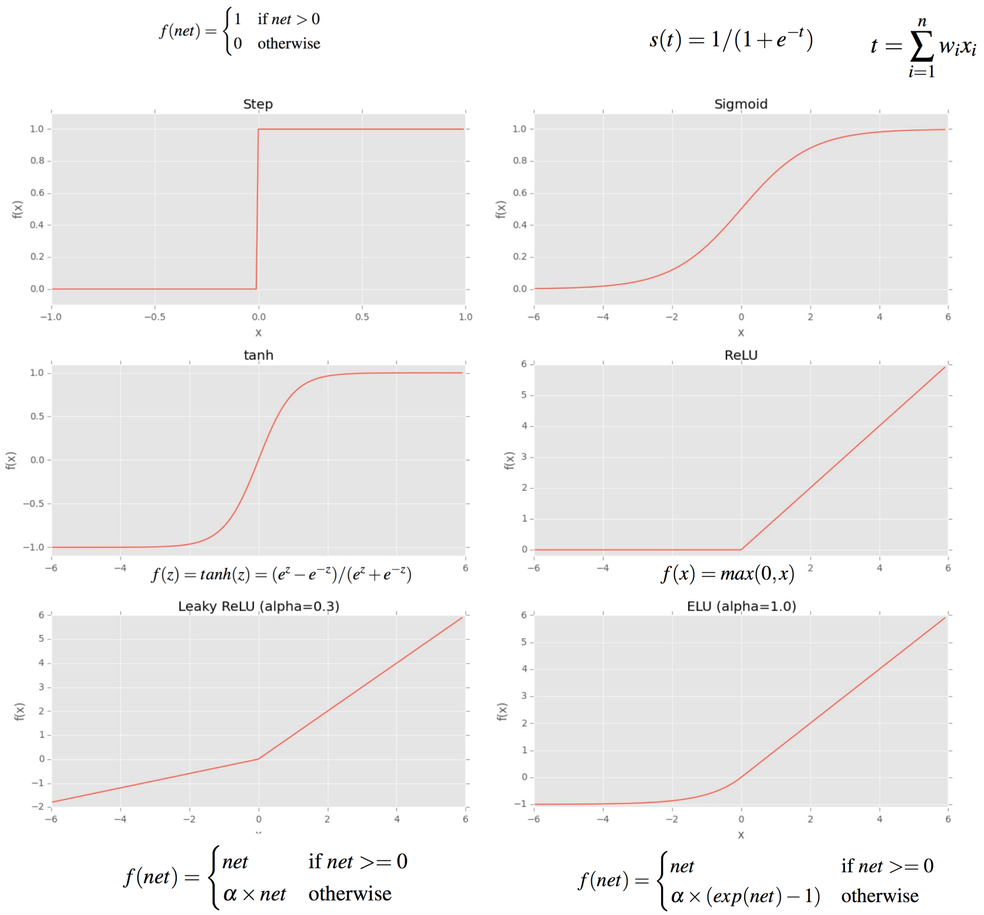
Type of function	Graphical represent.	Mathematical formula	MATLAB implementation
Linear		$g^{\text{lin}}(x) = x$	`x`
Step		$g^{\text{step}}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{elsewhere} \end{cases}$	`floor(0.5*(1+sign(x)))`
Threshold-linear		$g^{\text{theta}}(x) = x\,\Theta(x)$	`x.*floor(0.5*(1+sign(x)))`
Sigmoid		$g^{\text{sig}}(x) = \frac{1}{1+\exp(-x)}$	`1./(1+exp(-x))`
Radial-basis		$g^{\text{gauss}}(x) = \exp(-x^2)$	`exp(-x.^2)`

Neurons

Neuron models

– Alternatively, model with mean firing rate neuron

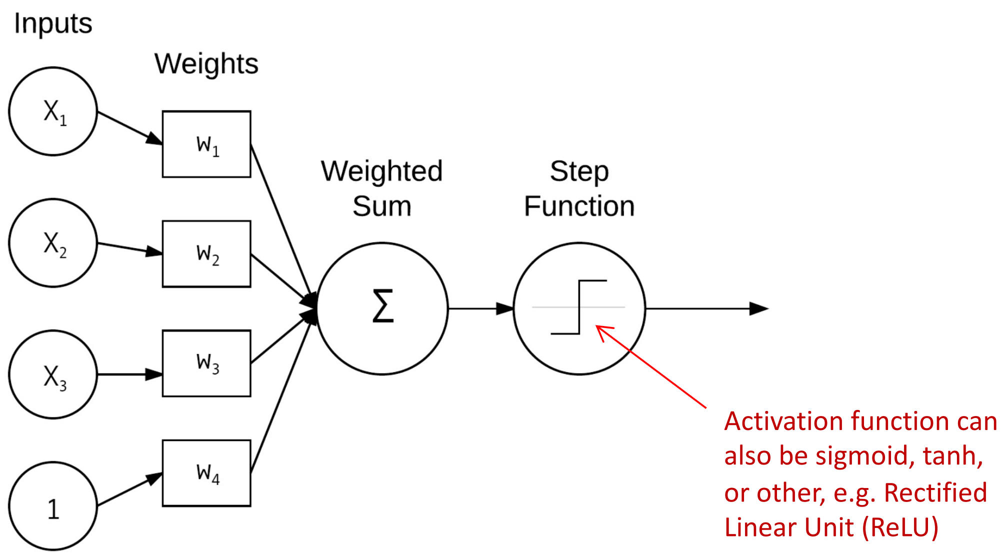
- Average firing rate of a pool of neurons over tens of milliseconds
- Several rate functions are used
  - Step, sigmoid, tanh, threshold-linear (ReLU), ...
- Same as activation functions



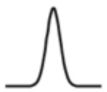
Recall from NR01-04

Neural Network Basics

Appendix 1 specifies how the value of y_0 and z_0 are computed



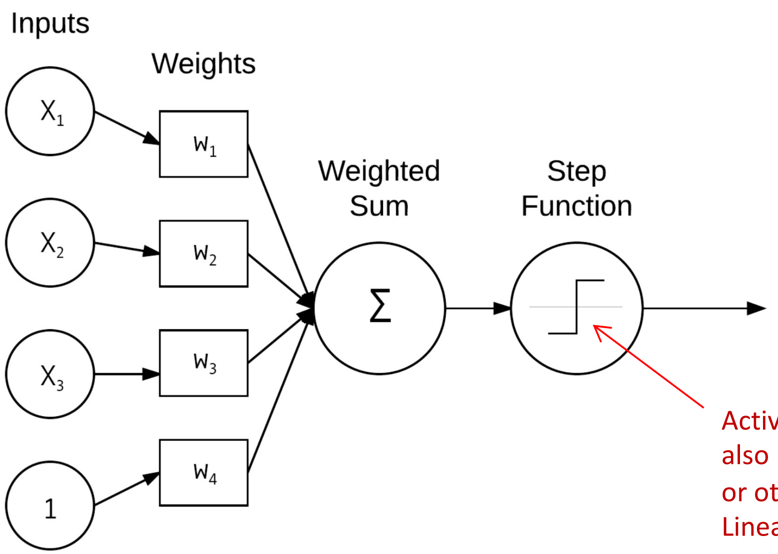
Credit: Adrian Rosebrock, Deep Learning for Computer Vision, PyImageSearch, 2017

Type of function	Graphical represent.	Mathematical formula	MATLAB implementation
Linear		$g^{\text{lin}}(x) = x$	`x`
Step		$g^{\text{step}}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{elsewhere} \end{cases}$	`floor(0.5*(1+sign(x)))`
Threshold-linear		$g^{\text{theta}}(x) = x\,\Theta(x)$	`x.*floor(0.5*(1+sign(x)))`
Sigmoid		$g^{\text{sig}}(x) = \frac{1}{1+\exp(-x)}$	`1./(1+exp(-x))`
Radial-basis		$g^{\text{gauss}}(x) = \exp(-x^2)$	`exp(-x.^2)`

Recall from NR01-04

Neural Network Basics

Appendix 1 specifies how the value of y_0 and z_0 are computed



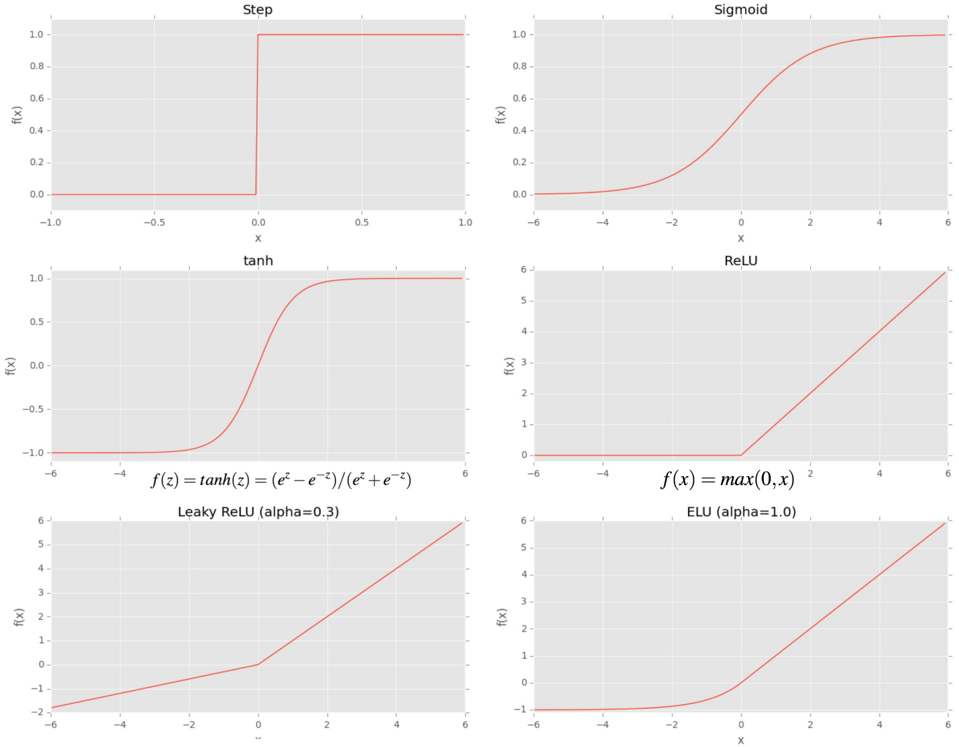
Activation function can also be sigmoid, tanh, or other, e.g. Rectified Linear Unit (ReLU)

Credit: Adrian Rosebrock, Deep Learning for Computer Vision, PyImageSearch, 2017

$$f(net) = \begin{cases} 1 & \text{if } net > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$s(t) = 1/(1+e^{-t})$$

$$t = \sum_{i=1}^{n} w_i x_i$$



$$f(z) = tanh(z) = (e^z - e^{-z})/(e^z + e^{-z})$$

$$f(x) = max(0, x)$$

$$f(net) = \begin{cases} net & \text{if } net >= 0 \\ \alpha \times net & \text{otherwise} \end{cases}$$

$$f(net) = \begin{cases} net & \text{if } net >= 0 \\ \alpha \times (exp(net) - 1) & \text{otherwise} \end{cases}$$

Darwin VII

Activity level (rate) of unit i is given by

Activation function: thresholded hyperbolic tangent

Scale factor: dependent on the neural area

Persistence of unit activity from one cycle to the next

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

Threshold function

$$\phi_i(x) = \begin{cases} 0; & x < \sigma_i \\ x; & otherwise \end{cases},$$

Unit-specific firing threshold

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

Number of connections to unit i

Activation of unit j at time t

Weight value of the connection of unit j projecting to unit i
Negative values correspond to in

Darwin VII

Synaptic plasticity

- Connections within and between neuronal areas are subject to activity-dependent modification
- Based on a value-independent and a value-dependent synaptic rule: the BCM learning rule
- Modification of synaptic strength is determined by pre-synaptic and post-synaptic activity

$$\Delta c_{ij}(t+1) = \varepsilon\left(c_{ij}(0) - c_{ij}(t)\right) + \eta s_j(t) F\left(s_i(t)\right)$$

Recall from NR01-05

BCM Learning Rule

Bienenstock, Cooper, and Munro, 1982

Modified Hebbian learning rule with inbuilt weight stabilization

- Each neuron has a threshold value
- If the postsynaptic neuron has an activity less than the threshold
  - the weight change is negative (or depressed)
- If the postsynaptic neuron has an activity greater than the threshold
  - the weight change is positive (or potentiated)
- The threshold can change dynamically to keep weights within some required range

Darwin VII

Value-independent synaptic change

Decay constant governing a passive, uniform decay of synaptic weights from their original values

Fixed learning rate

Piecewise linear approximation to the BCM learning rule

$$\Delta c_{ij}(t+1) = \varepsilon\left(c_{ij}(0) - c_{ij}(t)\right) + \eta s_j(t) F\left(s_i(t)\right)$$

Activity of pre-synaptic unit

Activity of post-synaptic unit

Darwin VII

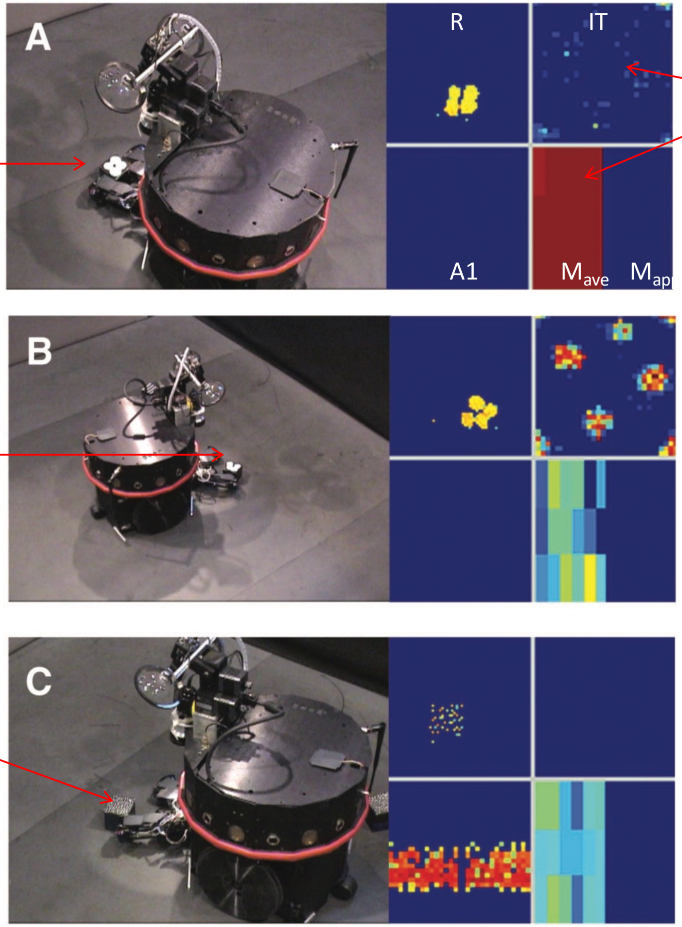
Value-dependent synaptic change

$$\Delta c_{ij}(t+1) = \varepsilon\left(c_{ij}(0) - c_{ij}(t)\right) + \eta s_j(t) F\left(s_i(t)\right)\bar{S}$$

- ε, $c_{ij}(0)$: Decay constant governing a passive, uniform decay of synaptic weights from their original values
- η: Fixed learning rate
- F: Piecewise linear approximation to the BCM learning rule
- $\bar{S}$: Average activity of the value system S
- $s_j(t)$: Activity of pre-synaptic unit
- $s_i(t)$: Activity of post-synaptic unit

Darwin VII

First encounter with an aversive block
(weak conductivity, blob, low tone)
Aversive behavior is driven by the taste system T_{ave}



Color indicates neural activity
No actiivity: dark blue
Maximum activity: bright red

Tenth encounter with an aversive block
(weak conductivity, blob, low tone)
After primary conditioning with visual stimulus, aversive behavior is driven by area IT

Pattern of small dots bright enough to trigger visual tracking but not sufficient to evoke a response in VAp and hence IT

Tenth encounter with an aversive block
(weak conductivity, blob, low tone)
After conditioning with auditory stimulus, aversive behavior is driven by area A1

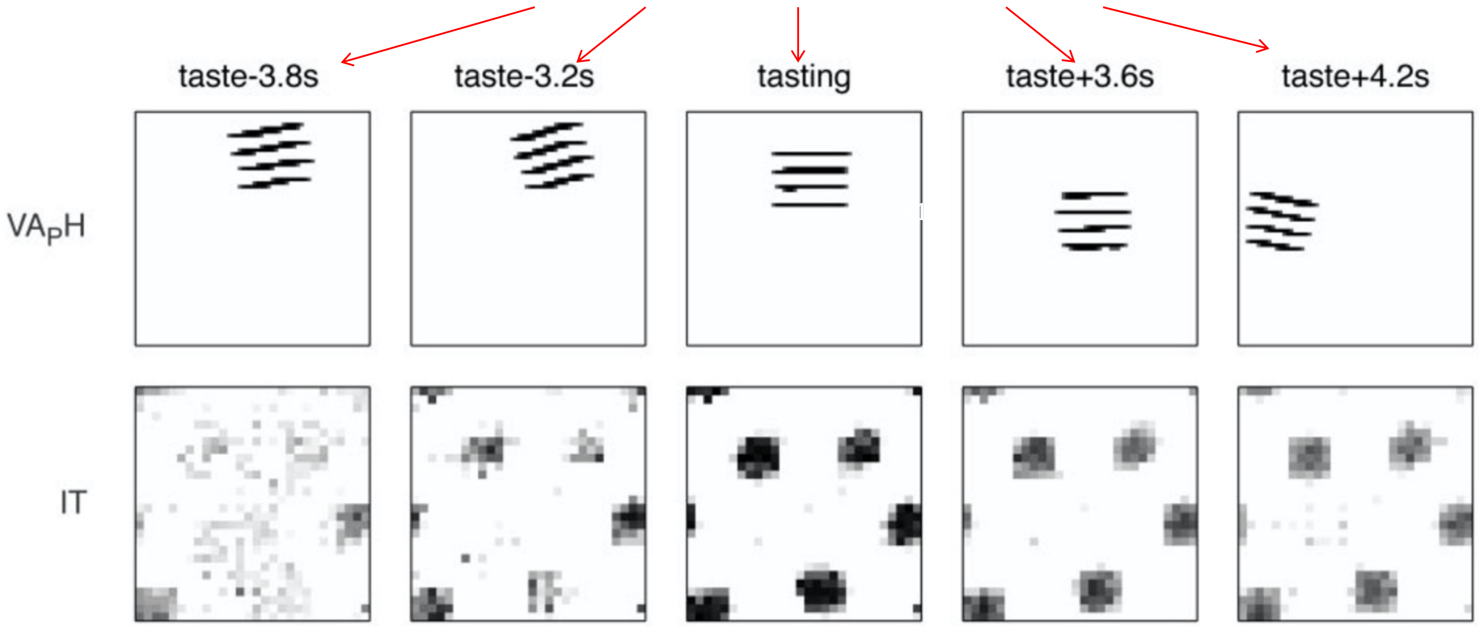
Darwin VII

Visual perceptual categorization

Number of seconds as Darwin VII approaches, tastes, and moves away

Primary visual area for horizontally oriented lines

Simultated inferotemporal cortex (IT): self-organization map with no retinotopy
White: no activity
Black: maximum activity



After ~5 stimulus encounters, activity-dependent plasticity between VA_p and IT caused IT responses to the different stimuli to be strong, sharp and separable

The persistence parameter, ω, of the IT neuronal units is higher than those in area VA_pH, leading to stable activity pattern in IT as the object moves across Darwin VII's field of view

Different activity patterns for each pattern; blob, horizontal stripes, vertical stripes

Position, rotation, and scale invariant categorization (+/- 30º)

Darwin VII

Take-away message

- A robot operating on biological principles can carry out
  - Perceptional categorization
  - Conditioned responses
- Without pre-specified instructions
- Achieved by
  - Exploration of the environment
  - Sensorimotor adaptation

Darwin VII

As a case study, it highlights many concepts covered so far

1. A neural architecture that is strongly influenced by mammalian neuroanatomy
2. Mean firing rate neuronal units with a hyperbolic tangent activation function
3. Categorization learned using the unsupervised BCM learning rule
4. Pavlovian conditioning (called operant conditioning since learning is expressed as actions)
5. Model-free reinforcement learning:
   action policy is to approach good tasting blocks and avoid bad tasting blocks

Reading

Hwu, T. and Krichmar, J. (2022). Neurorobotics: Connecting the Brain, Body and Environment, MIT Press.

Chapter 4, Sections 4.1 - 4.5, pp. 74 - 81.

(Krichmar and Edelman, 2002) and (Bienenstock et al., 1982) are available for reference on Canvas in the Reading Material module.