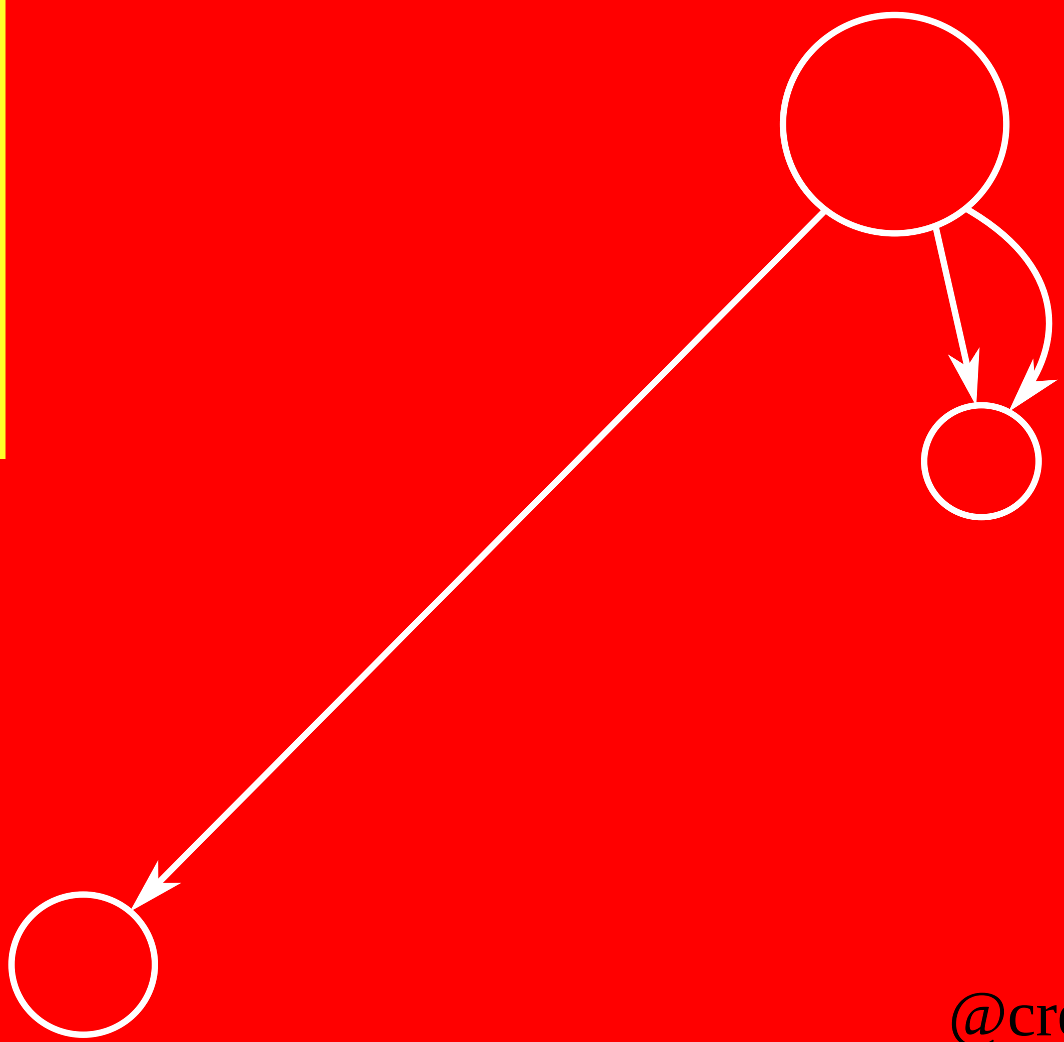
Towards
Artificial General Intelligence

A biologically plausible approach



@createai_org

The online version of this paper is available at www.createai.org , which has many interactive examples illustrating the concepts clearly.

It is recommended that the online version of this paper be viewed instead of this pdf version.

Contents

1. Introduction

1.1 Summary

A biologically plausible approach that has the potential to create *generic* Artificial Intelligence is described in this paper.
This paper describes how intelligent behaviour can be achieved by:

- incorporating timed delay in neural signalling
- encoding stimuli in the form of weighted neural connections
- creating neural circuits for specific functions
- enabling automatic creation of timed-associations between different types of stimuli
- enabling embodiment, motor-actions and autonomous subsystems in the creation of an Artificial Life Form.

A simulation of an 'Artificial Life Form' in a virtual environment is illustrated, wherein its behaviour is controlled purely by its 'brain' composed of interconnected neurons.

1.2 Approach required to create 'real' Artificial Intelligence

> *“I often refer to biology as setting a set of “constraints”. At first, the constraints make a problem more difficult to solve. Many machine learning and AI people avoid studying the anatomy of the brain for this very reason. But if you keep digging deeper into the biology eventually a solution emerges, and when a solution appears that satisfies the constraints you know you have the correct answer.”*
>
> *- Jeff Hawkins (Numenta) on building a biologically inspired AI*

It is widely accepted that contemporary techniques such as backpropagation will not lead to Artificial General Intelligence.[1] Any algorithm that attempts to create ‘generic’ artificial intelligence should be based on what we know about the human brain and should satisfy known biological constraints.
The following facts should be taken into account when formulating an algorithm for generic intelligence:

- Information is stored in the brain in the form of weighted connections between neurons.
- The connection strength between neurons may strengthen or weaken, enabling 'remembering' or 'forgetting'.
- Associations are formed between different types of stimuli by means of neural pathways connecting different brain regions.
- Any AGI algorithm should be able to explain empirically noted phenomena such as synaptic plasticity, sensory integration, false memories, deja vu, synesthesia, blindsight, phantom limb, etc.

2. The proposed Neuron model

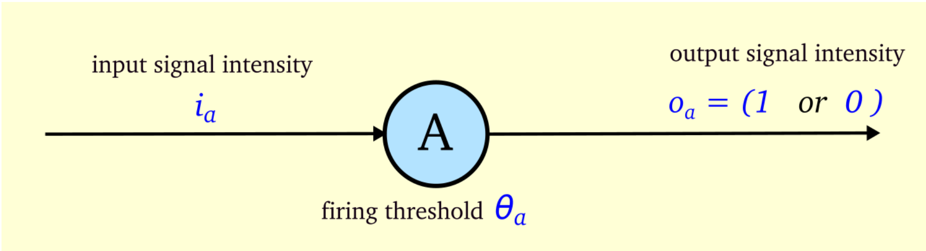
2.1 Description of the neuron model used

- Each neuron N has a threshold θ_n
- Each neuron receives input signal of intensity i_n
- When the input signal intensity i_n is greater than or equal to the neuron's threshold θ_n, then the neuron fires with output signal intensity O_n

2.2 Two types of neurons

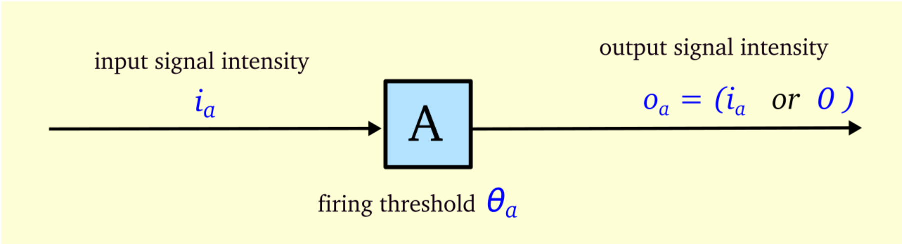
In this neuron model, there are two types of neurons :

1. **Binary neurons ,** which produce only binary output when activated;



If $i_a \geq = \theta_a$, then $O_a = 1$; else $O_a = 0$

2. **Non-binary neurons ,** which produce output signal intensity equal to input when activated;



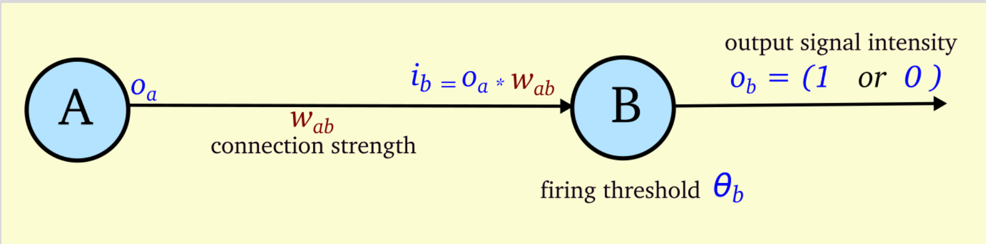
If $i_a \geq = \theta_a$, then $O_a = i_a$; else $O_a = 0$

Binary neurons are represented as circles while non-binary neurons are represented as squares.

2.3 Connection weights

- Connections may exist between two neurons and the *strength* of the connection is termed as *'connection weight'*
- If neuron A has a downstream connection to neuron B , the *'connection weight'* from A to B can be represented as $\boldsymbol{w}_{\mathbf{ab}}$
- When the neuron A sends an output signal $\mathbf{O_a}$ to neuron B, the signal is amplified by a factor $\boldsymbol{w}_{\mathbf{ab}}$
- When the input signal intensity $\mathbf{i_b}$ is greater than or equal to the neuron B's threshold $\boldsymbol{\theta}_{\mathbf{b}}$, then neuron B fires
    - ie., the input signal to neuron B is : $\mathbf{i_b} = \mathbf{O_a} * \boldsymbol{w}_{\mathbf{ab}}$

Connections in Binary Neurons

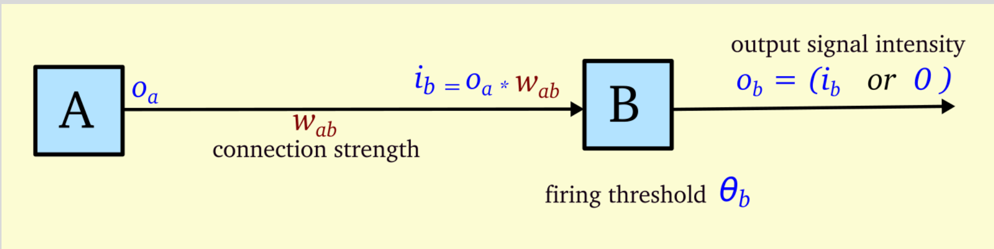


Input signal to Neuron B = output signal from A * (connection strength from A to B)

ie., $i_b = O_a * w_{ab}$

If $i_b \geq \theta_b$, then $O_b = 1$; else $O_b = 0$

Connections in Non-Binary Neurons



Input signal to Neuron B = output signal from A * (connection strength from A to B)

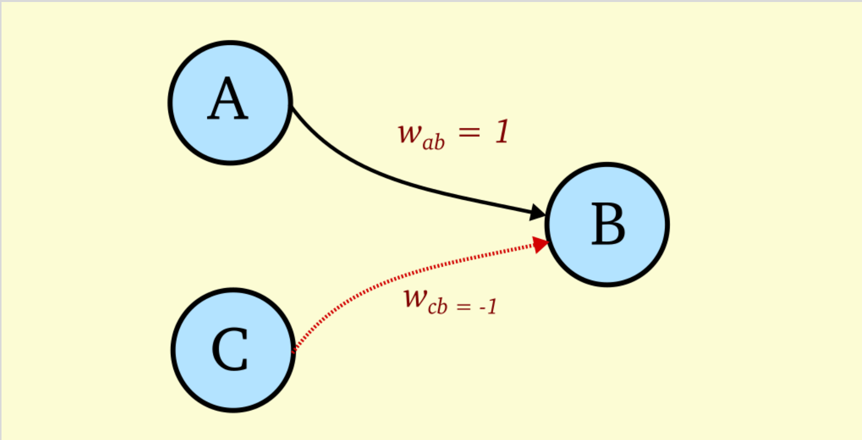
ie., $i_b = O_a * w_{ab}$

If $i_b \geq \theta_b$, then $O_b = i_b$; else $O_b = 0$

2.4 Excitatory and inhibitory connections

- The signal from one neuron to another could be either excitatory or inhibitory.
- An incoming inhibitory signal makes a neuron less likely to fire.
- Excitatory connections have positive connection weights, for example : $w_{ab} = 0.8$
- Inhibitory connections have negative connection weights, for example : $w_{ab} = -1$

Inhibitory connections:



Let Neuron B's firing threshold $\theta_b = 1$

An excitatory connection w_{ab} exists from neuron A to neuron B with connection weight 1

An inhibitory connection w_{cb} exists from neuron C to neuron B with connection weight -1

If neuron A and neuron C fire at the same time , with output intensity 1 each, then :

Input to neuron B : $i_b = (O_a * w_{ab}) + (O_c * w_{cb})$

$i_b = (1 * 1) + (1 * -1)$

$i_b = 0$

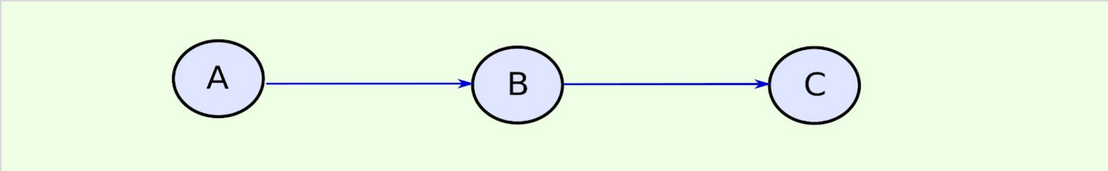
(The excitatory signal from A is cancelled out by the inhibitory signal from C)

Since $i_b < \theta_b$, neuron B doesn't fire.

2.5. Synaptic steps in a Firing sequence

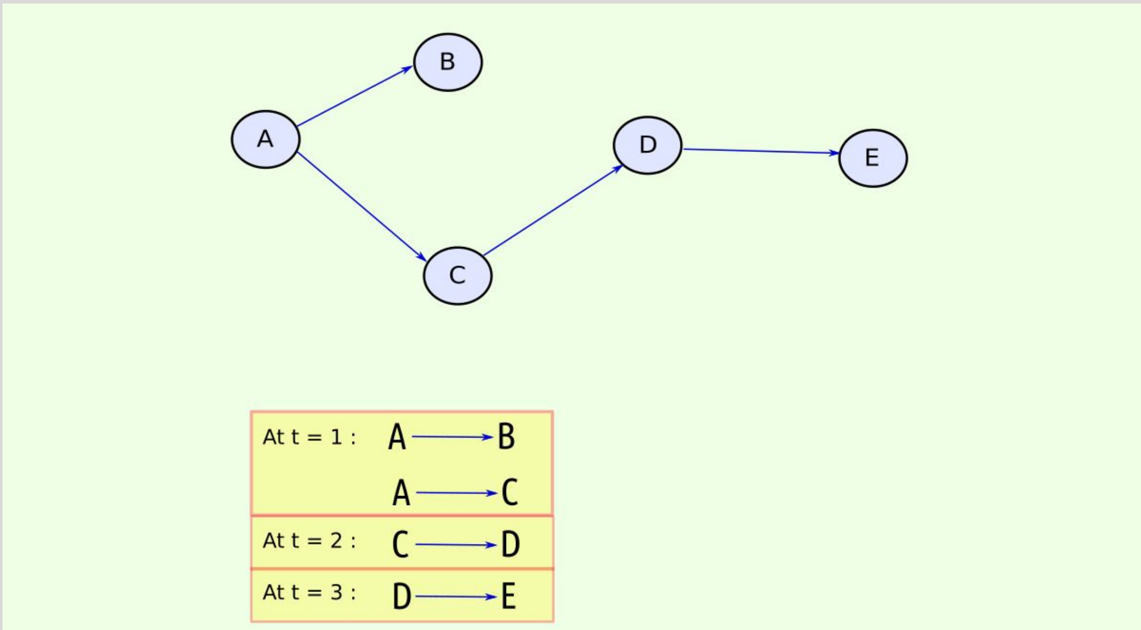
- In a neural circuit, all neurons do not fire at the same time.
- They fire in a sequential order, depending on the way they are connected to each other.
- The upstream neurons fire first, followed by the downstream neurons.

Firing sequence : Example 1



Sequence of firing :
Neuron A fires , then after a few milliseconds B fires , then after a few milliseconds C fires .
The signal from A reaches B in one "synaptic step"
The signal from A reaches C in two "synaptic steps"

Firing sequence : Example 2



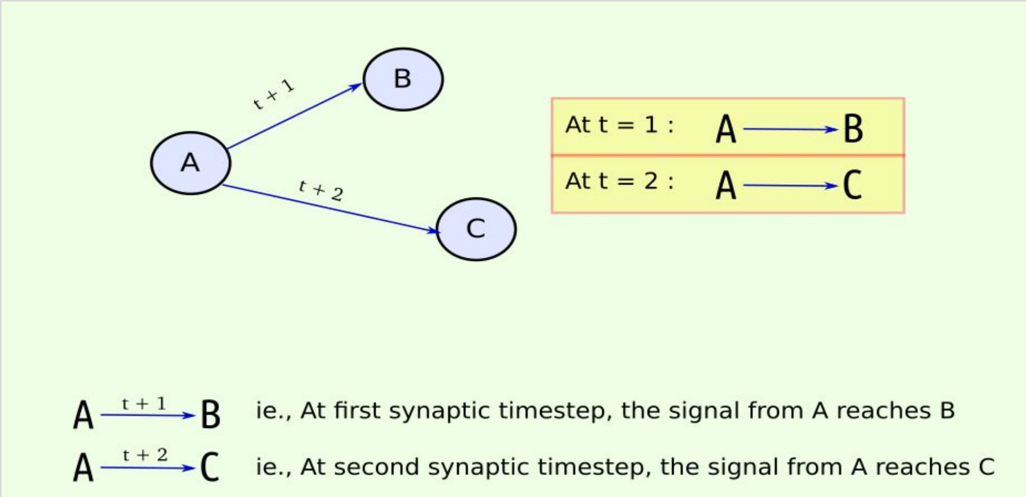
Sequence of firing :
Neuron A fires , then after a few milliseconds B & C fire at same time , then D fires, then E fires .
So it has taken 'three steps in time'(synaptic timesteps) for the signal from A to reach E
(assuming input to A,B,C,D,E cross their corresponding thresholds)

2.6. Temporal connections : Timed delay in neurotransmission

In this model, we introduce connections that incorporate a precisely timed delay in neural-signal transmission.

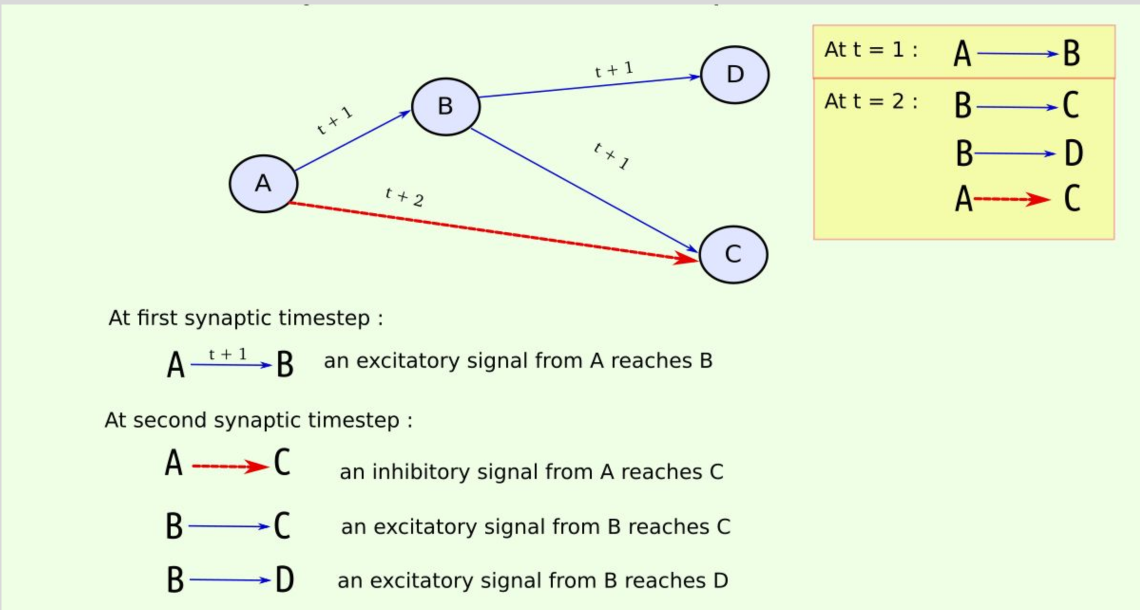
- A signal through a ***t+1*** connection reaches the target neuron in **one synaptic timestep**
- A signal through a ***t+2*** connection reaches the target neuron in **two synaptic timesteps**.
- A neural circuit can have upto ***t+H*** type connections, where **H** is the maximum delay allowed for signal transmission for that neural circuit.

Timed delay in neurotransmission : example 1



Sequence of firing : A , then B, then C.

Timed delay in neurotransmission : example 2

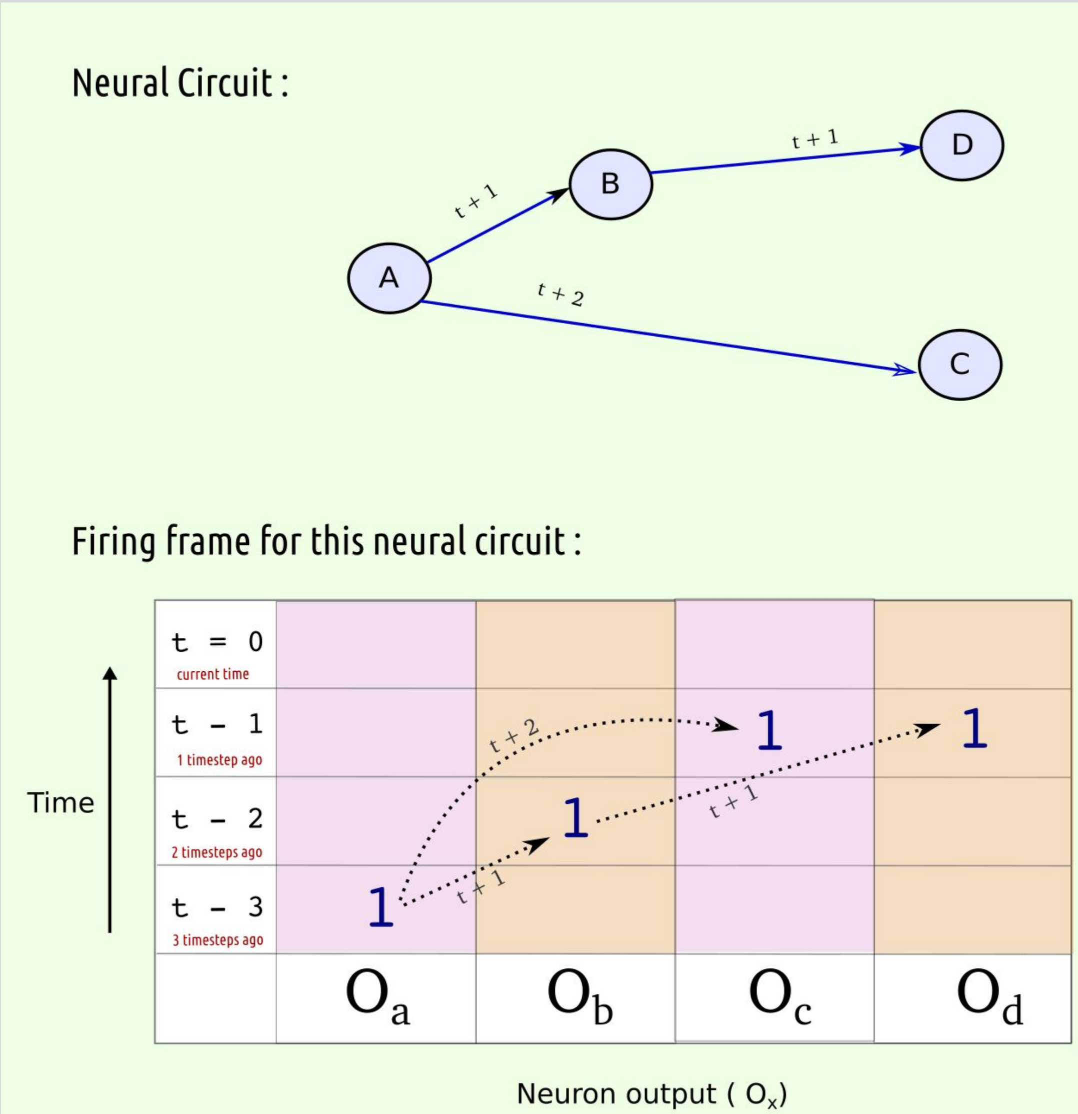


***Sequence of firing* :** Neuron A fires , then after a few milliseconds B fires , then after a few milliseconds C & D fire at the same time. Since there is a **t+2** connection from A to C, the signal from A reaches C at the sametime the signal from B reaches C

2.7. Firing frame : Representing neuron states over time

- To help with visualizing the firing states of neurons over time, we can tabulate the output signals of neurons over time.
- In this table, which we term as "Firing Frame", we have time in Y axis and neuron output intensities in X axis
- Each cell represents the output signal intensity of a neuron at specific time

Firing frame : example :



In this neural circuit, note that the connection from A to C is a t+2 connection, so it takes two synaptic steps for the signal from A to reach C ,upon which C fires.

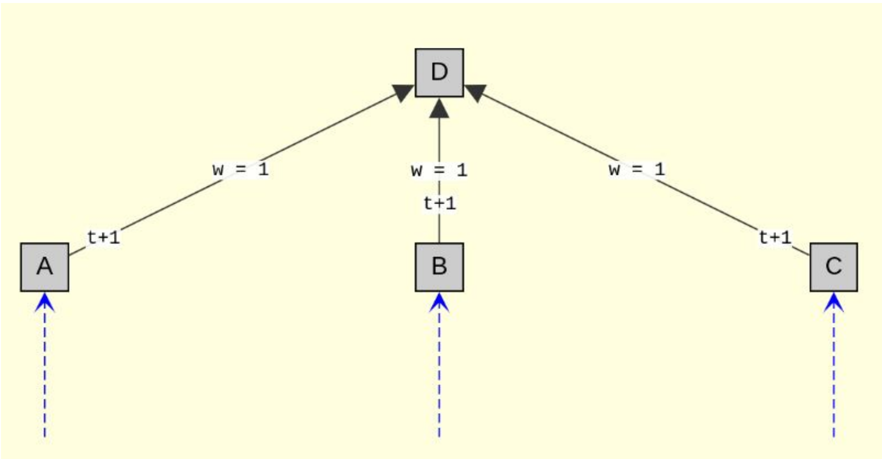
In this neural circuit composed of binary neurons, (assuming that firing thresholds are met) the sequence of firing is : A fires first, then B fires, then C and D fire at same time . The output of each of these binary neurons is 1 , which is entered in the cell corresponding to the time the neuron fired.

3. Neural circuits with specific functions

Using the concepts of time-delayed signalling, connection weights, thresholds and inhibitory connections, different types of neural circuits can be designed for specific goals, as illustrated below.

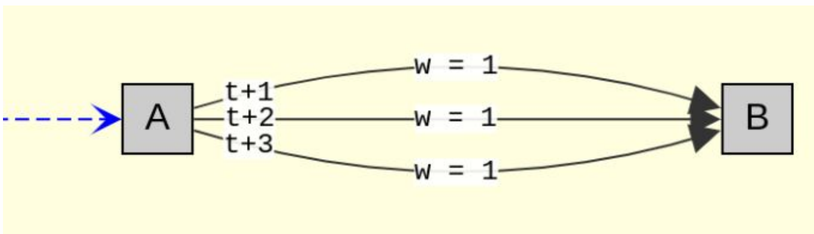
(Click on the images to go to the corresponding interactive demo)

3.1. Spatial summation



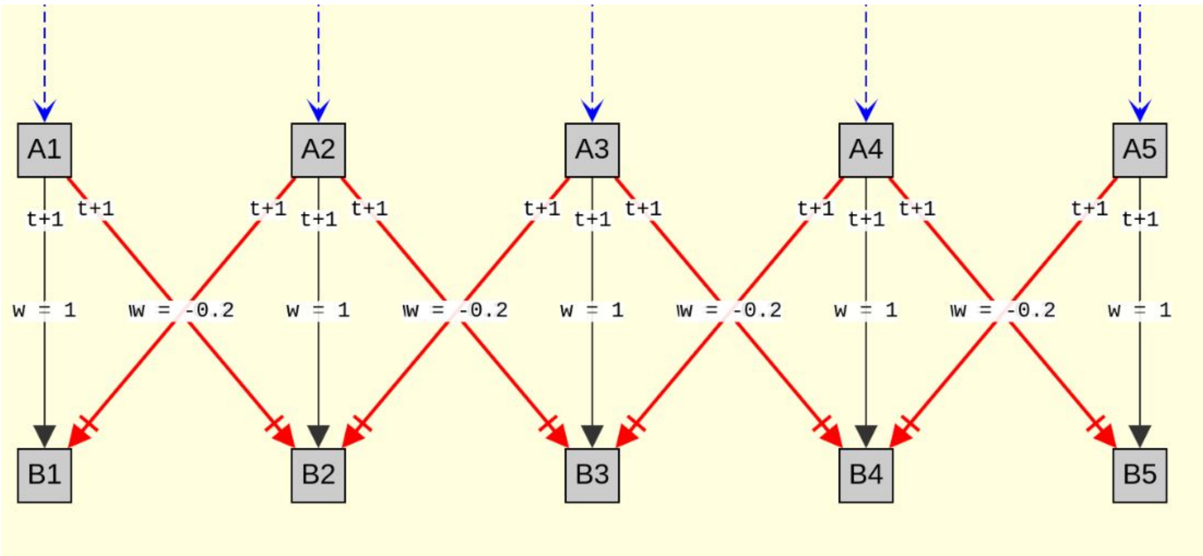
- Thresholds : $\theta_A = \theta_B = \theta_C = 100$; $\theta_D = 300$
- A, B and C have a (t+1) excitatory connection to D
- When the sum of output from A,B,C at (t-1) >=300, D fires at (t).

3.2. Temporal summation



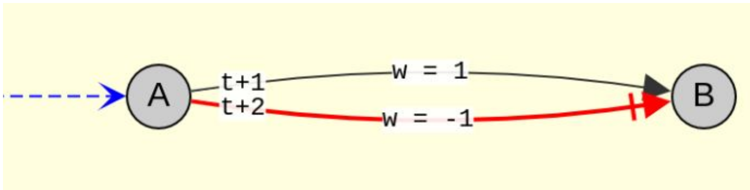
- Thresholds : $\theta_A = 1$; $\theta_B = 3$
- A has (t+1),(t+2),(t+3) excitatory connections to B , resulting in B accumulating output of A over three timesteps
- ie., $i_B = O_{A\,@\,t\text{-}1} + O_{A\,@\,t\text{-}2} + O_{A\,@\,t\text{-}1}$
- ie., When A fires consecutively thrice, B will have accumulated $i_B = 3$, So B reaches threshold and fires.

3.3. Lateral inhibition



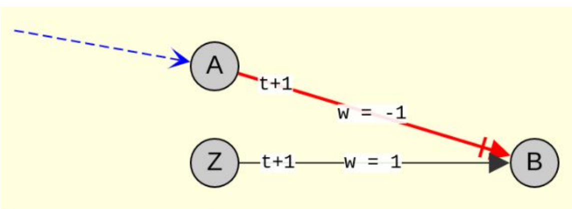
- Lateral inhibition occurs when an excited neuron reduces the inhibits its neighbouring neurons from firing.
- This phenomenon is seen in the human eye (see [Mach bands](#))
- This results sharpening of the signal over space, creating more contrast in the input signal.
- In this circuit, lateral inhibition is achieved by inhibiting the corresponding neighbour neurons in the next layer at t+1

3.4. Temporal sharpening



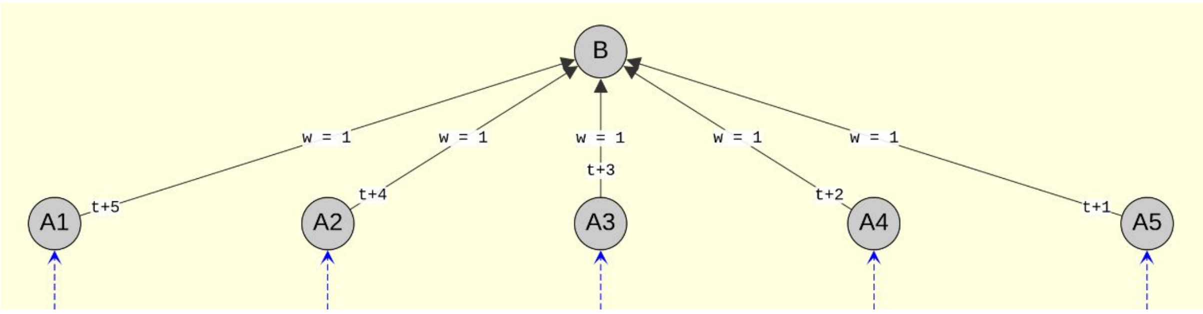
- Thresholds $\theta A = 1$; $\theta B = 1$
- A has a (t+1) excitatory and (t+2) inhibitory connection to B
- <u>This results in B firing only once, even if A fires consecutively many times.</u>
- Thus the signal from A is 'sharpened in time', providing information about the exact time of excitation
- This can be seen in touch sensory neural circuits in the human body, wherein a pinprick is detected and a reflex action initiated when B fires.

3.5. Inverse Neuron



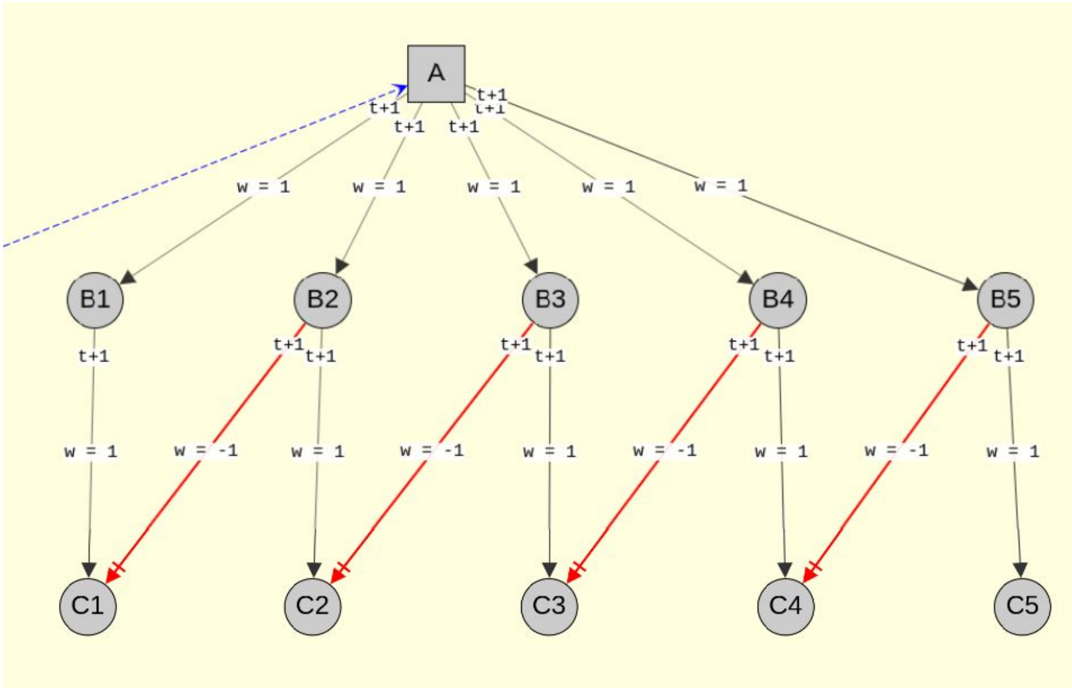
Goal : If A fires, B shouldn't fire, and vice-versa.
- To achieve this, A has a (t+1) inhibitory connection to B
- Z is a neuron that is always firing and has an excitatory connection to B
- When A fires at time=(t), B will NOT FIRE at (t+1)
- When A doesn't fire at time=(t), B will FIRE at (t+1)

3.6. Sequence detection



- Here, Neuron B will fire only if A1,A2,A3,A4,A5 fires one after the other.
- Any other sequence will not cause B to fire.
- This is achieved by creating the connections A1=(t+5)=>B , A2=(t+4)=>B , A3=(t+3)=>B , A4=(t+2)=>B , A5=(t+1)=>B .
- Thresholds $\theta_B = 5$; $\theta_{A1} = \theta_{A2} = \theta_{A3} = \theta_{A4} = \theta_{A5} = 1$;

3.7. Intensity estimation



- Let the input signal to A have an intensity value between 1 and 5
- Thresholds $\theta_{B1} = 1$; $\theta_{B2} = 2$; $\theta_{B3} = 3$; $\theta_{B4} = 4$; $\theta_{B5} = 5$; other thresholds = 1
- If input intensity to A i_A is greater than zero, one of the C_x neurons will fire based on its intensity i.e,
  - If $i_A = 1$: *C1* fires @ (t+2)
  - If $i_A = 2$: *C2* fires @ (t+2)
  - If $i_A = 3$: *C3* fires @ (t+2)
  - If $i_A = 4$: *C4* fires @ (t+2)
  - If $i_A >= 5$: *C5* fires @ (t+2)
- Eg: C4 fires at time (t) only if input value to A was 4 at (t-2)

4. Encoding stimuli to Neural code

4.1. Factors to consider in Sensory Stimuli Encoding

1. **Mapping :** Sensory neurons of the same type have a common pathway and terminate in the same region of the brain. (eg: Sound stimuli signals are sent to auditory cortex.) Neighbouring sensory neurons of the same type are mapped to corresponding neighbouring neurons in the brain region. In some cases, afferent neurons that carry signals to the brain will perform lateral inhibition/temporal sharpening to increase the contrast of the signal over space and time, prior to sending the signal to the brain.

2. **Intensity :** Each incoming sensory stimulus signal has a corresponding 'intensity' property.

Sensory type	Corresponding Intensity variable
Vision	Brightness of light
Hearing	Loudness of sound
Tactile	Pressure of touch
Smell / Taste	Concentration of the chemical

4.2. Stimuli encoding : Objectives

Sensory encoding involves activating specific neurons based on specific input stimuli. The same input stimuli should activate the same set of neurons everytime.
The process of encoding a sensory stimulus should facilitate the following :

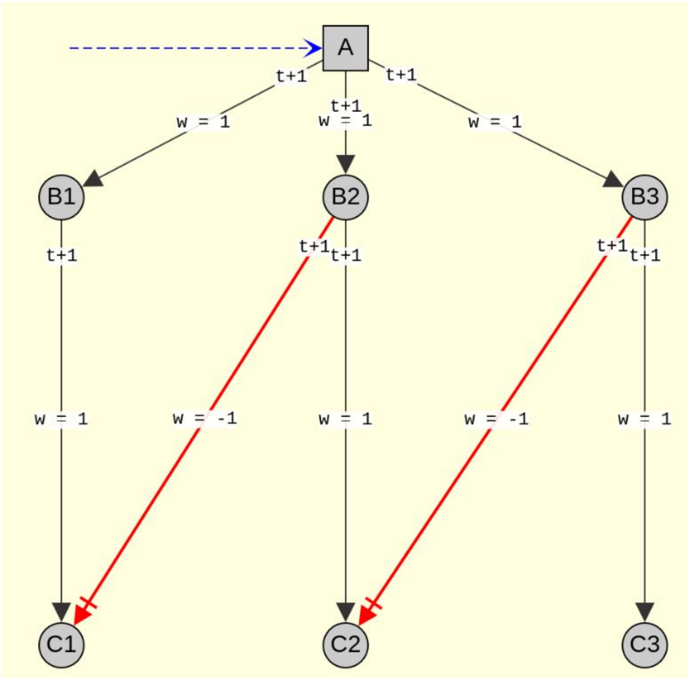
- Map stimuli of the same type to the corresponding neural region.
- Estimate the approximate intensity of the incoming stimulus within a set range
- Determine the specific sensory neuron's activation state (ON or OFF state)
- Detect changes in the sensory neuron's activation state
- Perform lateral inhibition and temporal sharpening where applicable
- Determine the duration of the current activation (to enable attention/habituation)
- Detect spatial/temporal patterns in incoming stimulus where applicable

4.3. Proposed encoding model

- In this encoding model, every input signal to the sensory neuron is just a scalar, ie., the intensity of the stimulus.
- Various downstream neural circuits are involved in encoding, so the Stimuli encoding objectives listed above are achieved

4.4. Encoding scalar value : Intensity estimation

A Neural circuit for signal Intensity estimation



The strength/intensity of the input signal can be estimated within two timesteps using this circuit.
Step 1: Determine the minimum and maximum possible values of the incoming signal's intensity
Step 2: Divide the range into a number of intervals ,depending on the precision required.
Step 3: In the second layer (B1,B2..) set thresholds that match the starting values of each interval.
Step 4: Inhibitory connections are added : B2 inhibits C1 , B3 inhibits C2.

Example :

- Let the input signal to A be a scalar with a value between 1 and 30
- Let the number of intervals chosen = 3 , so the range is split into three chunks (1..9), (10...19), (20...30)
- The thresholds of the second layer Bx are chosen to be the first number of each interval chunks.
- i.e., Thresholds $\theta_{B1} = 1$; $\theta_{B2} = 10$; $\theta_{B3} = 20$; other thresholds = 1
- If input intensity to A (i_A) is greater than zero, one of the C_x neurons will fire based on its intensity i.e,

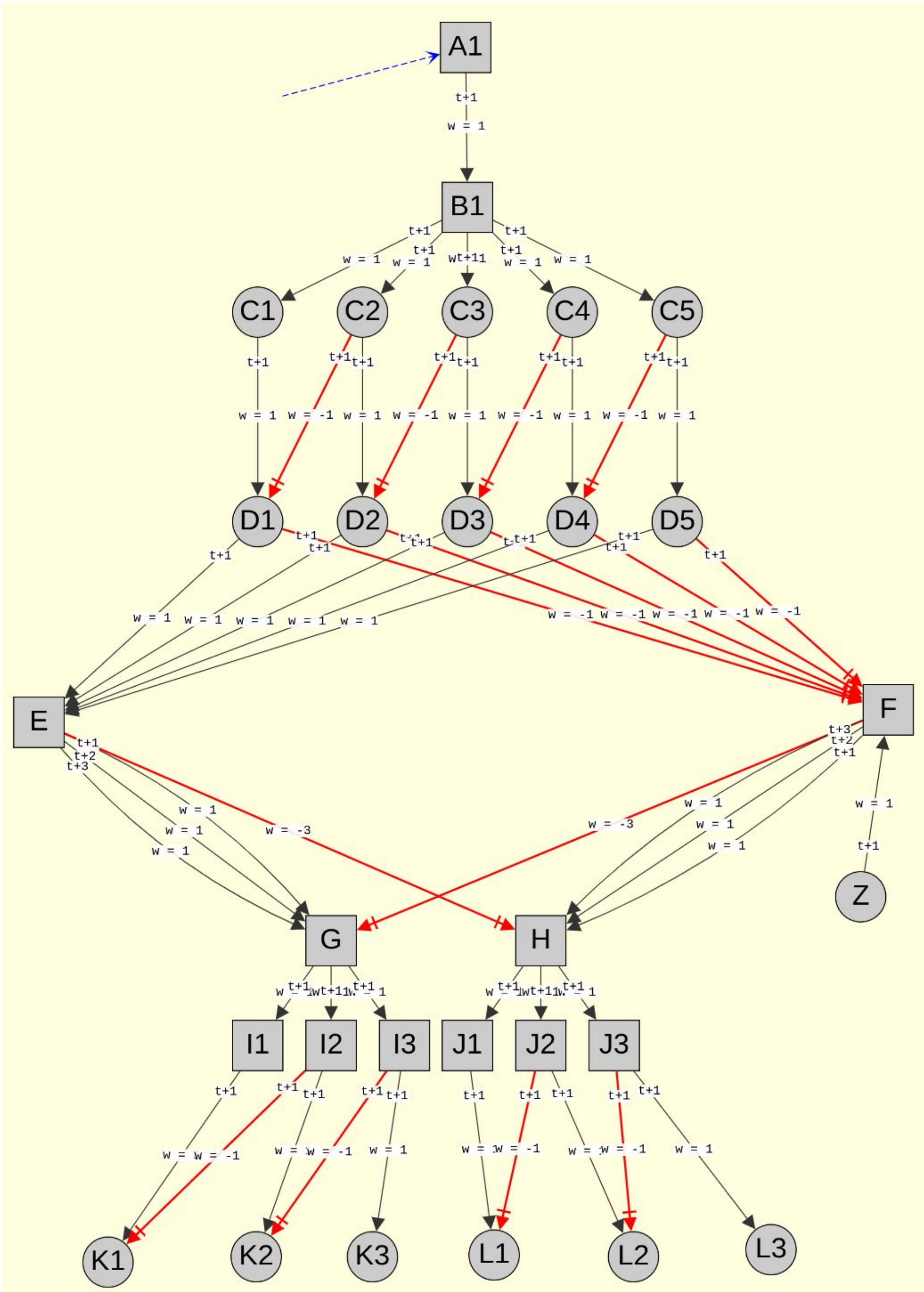
Input to A at time (t)	Result at time (t+2)
i_A = { 1,2,...,9}	*C1* fires @ (t+2)
i_A = {10,11,...19}	*C2* fires @ (t+2)
i_A >=20	*C3* fires @ (t+2)

4.5. Encoding a grayscale pixel input : (with change detection)

By building on the intensity detection circuit, it is possible to extract information about sensory stimuli's ON/OFF state and duration of current excitation/inhibition.

Demo: https://createai-org.github.io/paper/part4.html#section4.5

Consider this grayscale intensity detection circuit :

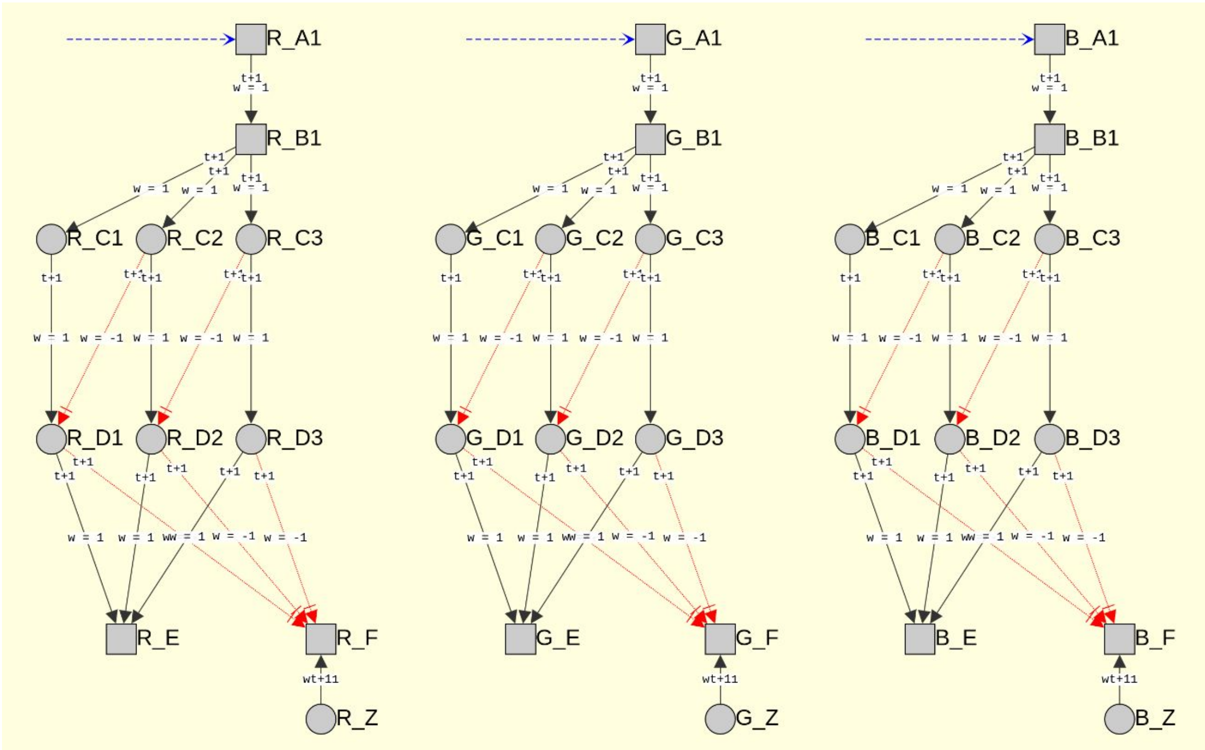


Description:

- The sensory neuron A1 gets a grayscale pixel's intensity value that is between 0 and 255 .
- The intensity of the input is grouped to be in one of the five ranges : (1 to 49), (50 to 99), (100 to 149), (150 to 200), (200 to 255)
- An intensity estimation circuit is created using B1, C* and D_* neurons by setting these thresholds : $\theta_{C1} = 1$; $\theta_{C2} = 50$; $\theta_{C3} = 100$; $\theta_{C4} = 150$; $\theta_{C5} = 200$;
- Neuron E is the 'ON state' detection neuron,which fires at time (t+4) when input > 0 at time t.
- Similarly, neuron F is the 'OFF state' detection neuron,which fires at time (t+4) when input = 0 at time t. This uses the "Inverse neuron" circuit shown in the earlier Section 3.5.
- Temporal summation connections (t+1,t+2,t+3) exist from E to G and F to H respectively, wherein G and H accumulate inputs over three timesteps.
- Further downstream, intensity estimation circuits exist for estimating the values in G and H

Scenario at time (t)	Result
Input to A1 = { 1,2,...,49}	D1 fires @ time (t+3)
Input to A1 = { 50,51,..,99}	D2 fires @ time (t+3)
Input to A1 = {100,101,,...,149}	D3 fires @ time (t+3)
Input to A1 = {150,151,...,199}	D4 fires @ time (t+3)
Input to A1 = { 200 and above}	D5 fires @ time (t+3)
Input to A1 > 0	E fires @ time (t+4)
Input to A1 is 0	F fires @ time (t+4)
Input to A1 changes from 0 to positive	K1 fires @ time (t+7)
Input to A1 is consecutively positive thrice or more	K3 fires @ time (t+7)
Input to A1 changes from positive to zero	L1 fires @ time (t+7)
Input to A1 is consecutively zero thrice or more	L3 fires @ time (t+7)

4.6. Encoding an RGB pixel input



Demo : https://createai-org.github.io/paper/part4.html#section4.6

Description:

• Each of the sensory neurons (R_A1 , G_A1, B_A1) get an input intensity value between 0 and 255 . s (R_A1 neuron for red, G_A1 neuron for green input, B_A1 neuron for blue input)

• The intensity of each input is grouped to be in one of the three ranges : (1 to 85), (86 to 170), (171 to 255)

• The intensity estimation circuit exists for each input, (using *_C* and *_D_* neurons); wherein the intensity of the input is determined to be in one of the three ranges.

• Eg: if the input red value to R_A1 >= 1, R_E will fire. (at t+4). Similar behaviour will be seen in the G_* (green) and B_* (blue) circuits.

Scenario at time (t)	Result
RED Input to R_A1 = { 1,2,...,85}	R_D1 fires @ time (t+3)
RED Input to R_A1 = {86,87,,...,170}	R_D2 fires @ time (t+3)
RED Input to R_A1 = {171 and above}	R_D3 fires @ time (t+3)
Colour input is white ie., (R_A1=255, G_A1=255, B_A1=255)	R_D3, G_D3, B_D3 fire @ time (t+3)
Colour input is black ie., (R_A1=0, G_A1=0, B_A1=0)	R_F, G_F, B_F fire @ time (t+4)

4.7. Encoding a one-dimensional array (with lateral inhibition)

• Consider a vision sensor with an one-dimensional strip of five light-sensors.

• Each sensor gets an input grayscale intensity value between 0 and 255 .

• In stimuli such as vision or touch, lateral inhibition is present , wherein neighbouring neurons inhibit each other's output to sharpen the contrast in the signal(see Mach Bands)

• In a comprehensive visual processing neural circuit, there will be a two-dimensional array of sensors instead of such an one-dimensional array, with corresponding lateral inhibitions to neighbouring neurons in two dimensions.

An example of lateral inhibition for an one-dimensional array of grayscale input is illustrated at https://createai-org.github.io/paper/part4.html#section4.7

4.8. Encoding sound

- Sound is composed of superposition of many frequencies, and each incoming frequency has an intensity factor(loudness).This concept is best visualized in a spectrogram.
- Sound can be thought of as a continuous stream of one-dimensional input, wherein each element is a sensor tuned to detect a specific frequency, and each element receives an input corresponding to the loudness in that particular frequency.
- Encoding of sound in such a way would also require Lateral inhibition to be enabled in the encoding neural circuit, similar to section 4.5. above.

An example of sound encoding for an input sound of just one frequency sensor (tuned to 6000Hz) is illustrated at https://createai-org.github.io/paper/part4.html#section4.8

5. Connectome

5.1. Neural connections

A *Connectome* is the map of the connections between neurons in the brain.
These inter-neuronal connections facilitate in encoding stimuli, deciphering stimuli and shaping the lifeform's overall behaviour. All incoming stimuli is encoded in the brain in the form of connections between neurons. When a new memory is formed, it is done so through modification of the weights of the inter-neuronal connections. Impulses to initiate and implement specific actions are encoded by these connections.

5.2. Two types of Connections

In this model, Connections are differentiated into two categories :

1. **hardwired connections ,** which are stable and don't strengthen/weaken
2. **dynamic connections ,** which are created based on neural activity and allowed to strengthen & weaken.

5.3. Innate (hardwired) connections for survival

Neuronal connections which enable food-seeking and danger avoidance aid the survival of the lifeform. When a biological organism has just been born, it has to survive in a specific environment, and innate connections related to food-seeking/danger-avoidance behaviour aid in its survival.

Such innate connections enable "Primitive reflexes" [2] , which is the execution of certain motor action sequences on encountering a particular stimuli. For example, in human infants, rooting reflex, stepping reflex, palmar grasp reflex,etc. are observed. After sometime after birth, such reflexes either come under voluntary control, or fade away.

Therefore any Artificial Lifeform should be preloaded with some innate connections to enable survival in the specific environment it is supposed to be 'placed'. Such *hardwired connections* are deemed stable and not strengthened or weakened.

5.4. Creation of new dynamic connections

Sensory organs receive the stimuli many times per second and transmit the signals to the brain simultaneously as a continuous datastream. The brain receives sensory datastreams of various types (vision,sound,etc.) simultaneously. Regions of the brain are specifically mapped to sensory type (eg: visual stimuli mapped to occipital lobe). To form an association between stimuli, connections need to be formed between the corresponding neurons that encode the different stimuli.

Let **H** be the **maximum delay allowed in forming connections** ie., connections with a delay of upto **t+H** synaptic timesteps are allowed in the brain.

Criteria for new connection formation :
In this model,criteria for forming new connections will be based on Hebbian learning, which can be loosely stated as "Neurons that fire (almost) together, wire together".
If neuron A and neuron B are activated within a short time (within **H** synaptic timesteps of each other), then a dynamic connection is created between neuron A and neuron B. (provided that neural pathway is permitted)
For example, if neuron A fires at time t and neuron B fires at **t+X** synaptic timestep, a **(t+X)** dynamic connection is created from neuron A to neuron B. (**where 1<=X<=H)**

5.5. Strengthening of existing dynamic connections

Criteria for strengthening an existing dynamic connection :

When neuron A fires at time t , and neuron B fires at timestep **t+X** , and a dynamic **t+X** connection already exists, then the connection strength of the **t+X** connection from A to B will be increased at each instance, upto a set limit. (where 1<=X<=H)

5.6. Weakening of existing dynamic connections

Criteria for weakening an existing dynamic connection :

Whenever a neuron fires, all its incoming dynamic connections are weakened except those dynamic connections which caused it to fire. This criteria is based on known research findings[3]

5.7. Explaining Classical conditioning using this model

Using the above criteria for dynamic connection creation/strengthening/weakening, it is possible to explain how (Pavolovian) classical conditioning can occur.

> Illustrations of how classical conditioning can be explained using this model is available at
> https://createai-org.github.io/paper/part5.html#section5.7

6. Designing an Artificial Life Form : Considerations

6.1. Sensory types

The lifeform should have sensors to detect stimuli types that are necessary for its survival in that specific environment. More the types of relevant stimuli, the more feedback it can get from environment/internal sensors, increasing the chances of its longer survival.

6.2. Motor actions

The lifeform should be able to interact with its environment and influence its environment through 'motor actions'. (eg: locomotion)

6.3. Embodiment

The lifeform should be 'embodied' , ie., should have a perceived physical form. Such a perceived physical form makes it possible to create a sense of 'self' ,by demarcating self-generated stimuli and environmental stimuli. The interaction of a lifeform with its environment through motor actions creates a change in sensory stimuli.(Eg: movement in a specific direction could create a change in visual input) The lifeform then learns the association between the executed 'motor action' and the resultant change in sensory stimuli from the environment. [4]
When an artificial lifeform starts its 'virtual life' in a new environment, it could execute all possible motor actions at random (*motor babbling*) to create the associations between the motor-actions and the resultant change in stimuli. Without embodiment, this association-creation is not possible. For example, in cases of lifeforms with limbs, the motor-action of moving a limb also generates proprioception sensory stimuli, which is internal stimuli that encodes relative position of a limb with respect to the body.

6.4. Reflex actions

Innate connections are required for incorporating '*primitive reflexes*' like food-seeking and danger-avoidance. Such primitive reflexes are required for initial survival until more complex behaviour are learnt using environmental/internal feedback. Some reflex actions can be made to fade over time (eg: fading of suckling , grasping reflexes in humans) after associations are formed to generate the equivalent voluntary behaviour.

6.5. Autonomous Subsystems

An Artificial Lifeform should be made to feel 'virtual pain' in order to shape its behaviour. To aid its survival, certain critical motor-actions should not be controlled directly by the Artificial LifeForm.(eg: actions equivalent to breathing,digestion) To achieve this, autonomous subsystems should be incorporated in the Artificial Lifeform's design.

Taking a Human analogy, the "human digestive system" is an autonomous subsystem. When stale food is ingested, the digestive system triggers sickness and causes pain, discouraging future ingestion of stale food. The human-digestive-system is autonomous in that it cannot be forced not to cause pain. It has its own independent sensors,and an independent ability to trigger pain neurons.

The autonomous subsystem should therefore independent of free will and powerful enough to trigger pain or satiation, thus shaping behaviour. Such autonomous subsystems enable implementation of motivation, positive reinforcement and negative reinforcement.

An Artificial Lifeform should therefore have several autonomous subsystems that trigger pain when undesirable stimuli is detected by the subsystem (eg: detecting proximity to fire) The pain should be inflicted on the Artificial Lifeform ONLY by an autonomous subsystem within the Artificial Lifeform. Inflicting and stopping pain should be possible only by the Artificial Lifeform's autonomous subsystem and not by the free-will of Artificial Lifeform. Similarly, triggering an "urge" and signalling "satiation" of the urge should be done only by the Artificial Lifeform's autonomous subsystem and not by the free-will of Artificial Lifeform.

Such autonomous subsystems would help to :

- Trigger quick reflex actions on certain stimuli , bypassing free-will (eg: analogous to human spinal reflex of pulling back arm on touching a flame)
- Ensure continuous activation of highly critical motor sequences : It is important that some motor-actions be initiated involuntarily . For example, it is dangerous if life-critical actions such as breathing have to be initiated voluntarily. (Eg: Forgetting to breathe .) Other bodily functions such as digestion, involves a sequence of actions (acidity increase in stomach, periodic contraction of stomach muscles, etc.) which should not be directly controllable by the lifeform's will.
- When an Artificial Lifeform performs an undesirable motor action, the undesirable action will be detected by one of its autonomous subsystems , and pain is delivered to the Artificial Lifeform, thus dissuading similar undesirable behaviour in the future.
- Various types of 'virtual pain' could be created, corresponding to different types of negative stimuli (hunger, thirst, physical pressure). Various levels of pain could be created, because an incremental scale of pain could stop undesirable behaviour at earlier stages.

6.6. Inflicting pain on an Artificial Lifeform

'Virtual pain' can be designed to cause the following changes to the Artificial Lifeform : [5]

- Pain could inhibit the motor-action neurons, reducing its responsiveness
- Pain could prevent or intermittently block the flow of sensory data from the environment
- Pain could trigger specific reflex actions, preventing other intended motor actions(eg: pain neurons trigger loss of balance maintenance, forcing robotic AI to bend/fall)
- Pain could impair involuntary actions of its subsystems (eg: impair the equivalent of digestion)
- Higher levels of pain could cause shutdown of all sensory information flow from environment
- Higher levels of pain could cause shutdown of all motor action capabilities,effectively putting it in a "machine coma"

The most important factor to note is that the Artificial Lifeform should associate the pain with its recently executed undesirable motor action. This association of an undesirable action with 'virtual' pain will discourage the Artificial Lifeform from performing the undesirable action in the future.

6.7. Creating a motivation system

Research has shown that non-satiation of an *urge* (eg:hunger) encourages more activity in pursuit to satisfy the urge.[6]
An Artificial Lifeform can be designed with neural circuits that execute various combinations of possible motor-actions until it finds the means to satiate its urge. A need (food) creates an unsatiated feeling(hunger), which triggers discomfort(pain),which in-turn triggers some sort of action.

Therefore, when the Artificial Lifeform performs a desired set of motor action sequences (eg: eats virtual food), its own autonomous subsystem (eg: digestive system equivalent) detects the virtual food, inhibits the hunger/pain neurons, thus providing satiation.

By introducing different types of needs (food,water) and corresponding feelings(hunger,thirst), the Artificial Lifeform can :

- be motivated to explore its environment
- obtain positive feedback/reward when it performs a desirable action
- obtain negative reinforcement(pain) when it performs an undesirable action.

Using such a motivation system and feedback system from its autonomous subsystems, its behaviour can be shaped.

A plausible motivation circuit : with Positive reinforcement is illustrated at https://createai-org.github.io/paper/part6.html#section6.7

A plausible motivation circuit : with Negative reinforcement is illustrated at https://createai-org.github.io/paper/part6.html#section6.7

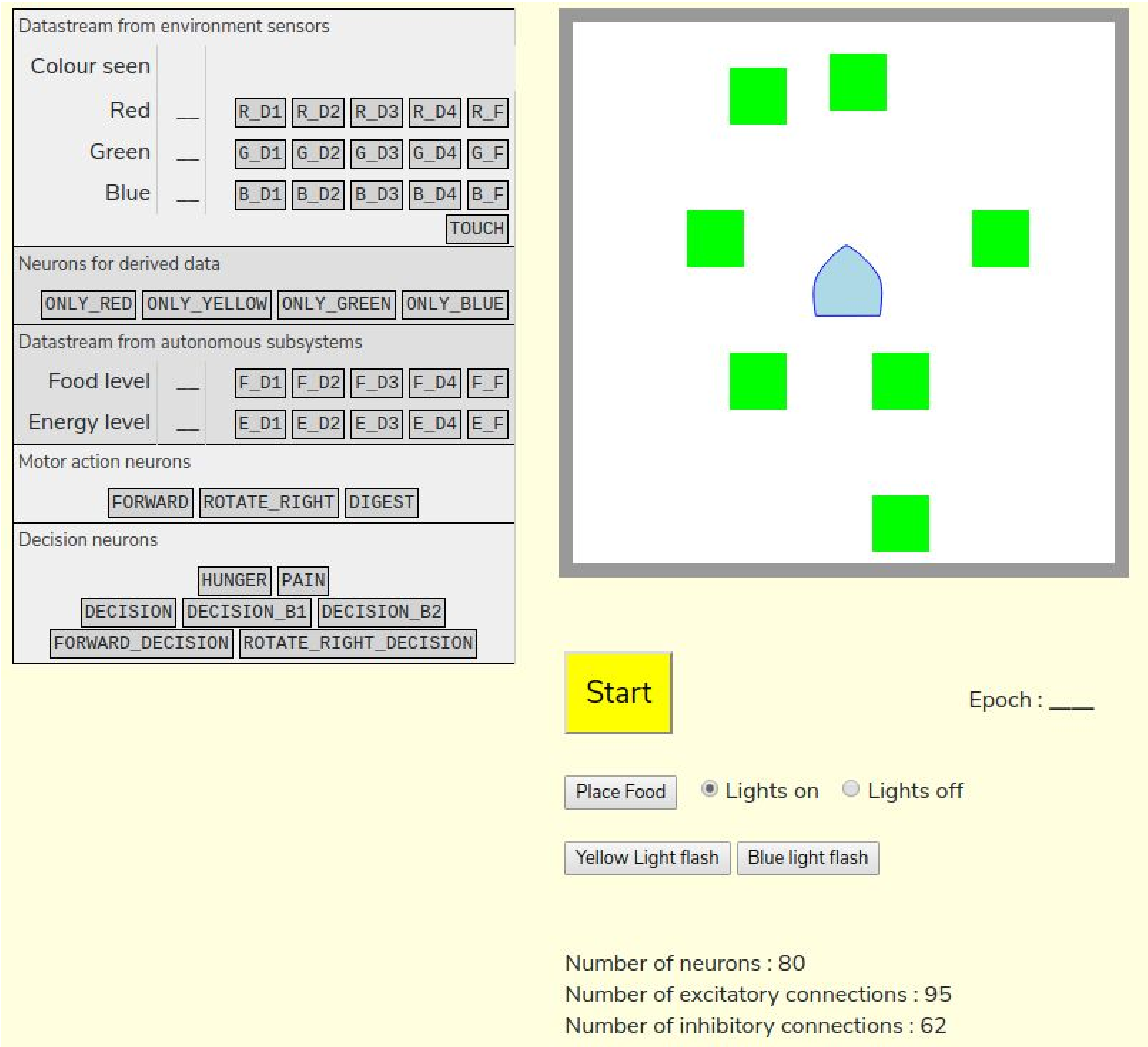
7. Simulation of an Artificial Life Form in a 2d environment

7.1. Simulation

The interactive simulation of the *Artificial Life Form* is available at
https://createai-org.github.io/paper/part7.html

Environment description : The blue coloured object is the Artificial Life Form , situated in a 2d environment with gray coloured walls on each side. The green boxes represent Food.In this simulation, to simulate energy expenditure of the Artificial Life Form, energy-level is designed to decrease by 10 units for every 20 epochs.When food is eaten, food-level increases by 20 units.The DIGEST motor action decreases food-level by 10 units and increases energy-level by 10 units.When energy level reaches 0, Alfie dies.

A screenshot of the simulation is shown below :



7.2. Detailed description of the simulation

7.2.1. Virtual Environment

1. To illustrate the concepts in a simple manner, a simple *two-dimensional* virtual environment is used as the location of the "Artificial Life Form".
2. The *Artificial Life Form*, which we'll call **Alfie** , is represented by the pointed blue coloured shape.
3. The two-dimensional environment is the white-coloured area and has gray coloured walls on each side.
4. Food placed in this environment is represented by green coloured squares.

7.2.2. Sensory Information

The 'body' of **Alfie** is simulated such that it has a 1-pixel colour sensor "eye" at its pointed end. Based on the alignment of Alfie's body, the 1-pixel colour sensor receives light corresponding to what the Alfie "sees". The colour actually seen by Alfie's eye is shown in the top left of the simulation UI in the box corresponding to *"Colour seen"*. The 'body' of Alfie is simulated such that it also has a "touch sensor" at its pointed end.

The "brain" of the Artificial Life Form receives the values of the following six sensory stimuli every epoch :

1. **Red intensity level :** This stimuli represents the red intensity level of the incoming light pixel . (Intensity range : 0 to 255)
2. **Green intensity level :** This stimuli represents the green intensity level of the incoming light pixel . (Intensity range : 0 to 255)
3. **Blue intensity level :** This stimuli represents the blue intensity level of the incoming light pixel . (Intensity range : 0 to 255)
4. **Touch :** When Alfie collides with the wall, a **TOUCH** sensor in its pointed end is activated. (Intensity range : { 0,1})
5. **Food level :** This internal sensory stimuli indicates the level of food in Alfie's stomach. (Intensity range : { 0,1,..100})
6. **Energy level :** This internal sensory stimuli indicates Alfie's energy level. The **DIGEST** motor action converts food to energy when energy level is low and food is available. Energy level decreases by 10 units every 20 epochs. When energy level reaches 0, Alfie dies. (Intensity range : { 0,1,..100})

Note:

- The Red, Blue, Green sensory signals and the **TOUCH** sensory signal are *external stimuli* arriving to Alfie's brain, while the remaining stimuli (Food level,Energy level) are *internal stimuli*.
- The intensity of the above 6 variables are available to Alfie in a continuous datastream.
- A snapshot of the datastream is taken every epoch(300millisec) and Alfie's brain receives that snapshot of intensity values every epoch.
- In this simulation one epoch is approximately 300 milliseconds.

7.2.3.Motor actions

Alfie is provided with these motor-actions :

- **`FORWARD`** , which makes it move forward by 10 units
- **`ROTATE_RIGHT`** , which makes it rotate right from its current orientation by 90 degrees.
- **`DIGEST`** motor-action is also present, but this action cannot be initiated voluntarily by Alfie, but only by its autonomous-subsystem.

7.2.4. Autonomous subsystems

An autonomous subsystem is designed for 'virtual digestion' of the food eaten by Alfie. The autonomous subsystem is required to generate and manage internal stimuli such as food level,energy level,etc. (The need for autonomous subsystem is described in section 6.5)

This autonomous subsystem will perform the following functions :

- activate **HUNGER** when food level is low
- trigger **DIGEST** motor action when **EnergyLevel** is less and food is available
- inhibit **HUNGER** when **FOOD** level is maximum

7.2.5. List of Neurons

The following **80** neurons are added to Alfie's virtual brain :

- Neurons for incoming stimuli with non-binary values (65) :
    - **Red, Green & Blue** sensory input stimuli have an input intensity range of (0 to 255).
    - **Food-level & Energy-level** input stimuli have an input intensity range of (0 to 100).
    - For each of these stimuli, we need to detect their approximate intensity using an intensity detection circuit.
    - An intensity detection circuit that groups incoming values into **four** "range buckets" will have **13** neurons.
    - So, for **five** input stimuli, the total number of neurons = 5 * 13 = 65 neurons.
- Neurons for incoming stimuli with binary values as input (1)
    - **TOUCH**
- Neurons for derived data (4)
    - **ONLY_RED**
    - **ONLY_BLUE**
    - **ONLY_GREEN**
    - **ONLY_YELLOW**
- Motor action neurons (3)
    - **FORWARD**
    - **ROTATE_RIGHT**
    - **DIGEST**

- Motivation circuit neurons (7)
    - **PAIN**
    - **HUNGER**
    - **DECISION**
    - **DECISION_B1**
    - **DECISION_B2**
    - **ROTATE_RIGHT_DECISION**
    - **FORWARD_DECISION**

TOTAL NEURONS = (65 + 1 + 4 + 3 + 7) = 80 neurons.

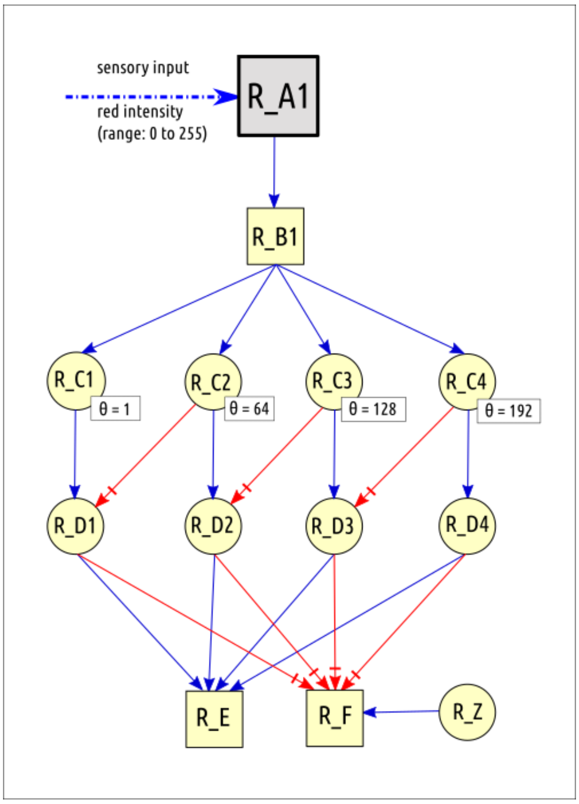
7.2.6. Connecting neurons to create desired behaviour

The following "neural circuits" (inter-neuronal connections) are created to impart the desired behaviour in the Artificial Life Form.

> ***Note:*** *In the neural circuit diagrams below, unless specified otherwise, all connections are* ***t+1*** *connections ; inhibitory connection weights are* ***-1.0*** *; excitatory connection weights are* ***1.0*** *; neuron thresholds are* ***1.0***

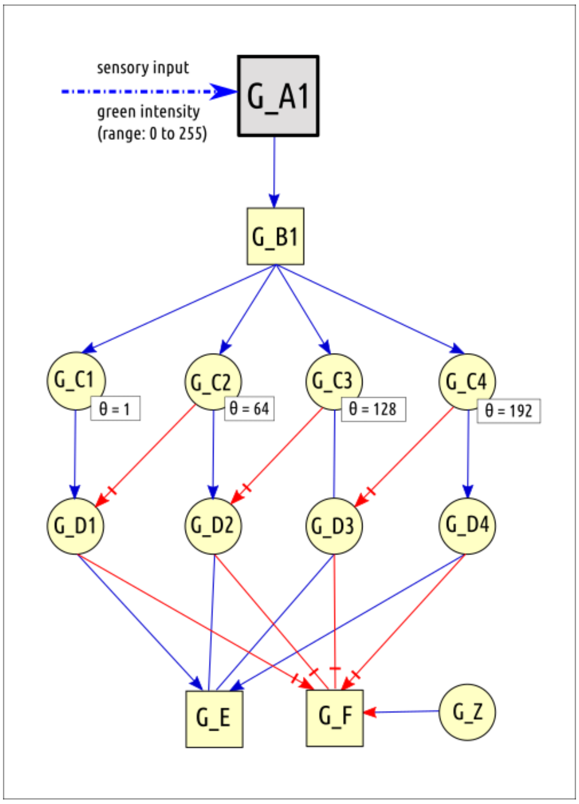
7.2.6.1. Intensity detection circuits for detecting intensity level of red/green/blue in light pixel

This intensity detection circuit is similar to the example seen in section 4.4



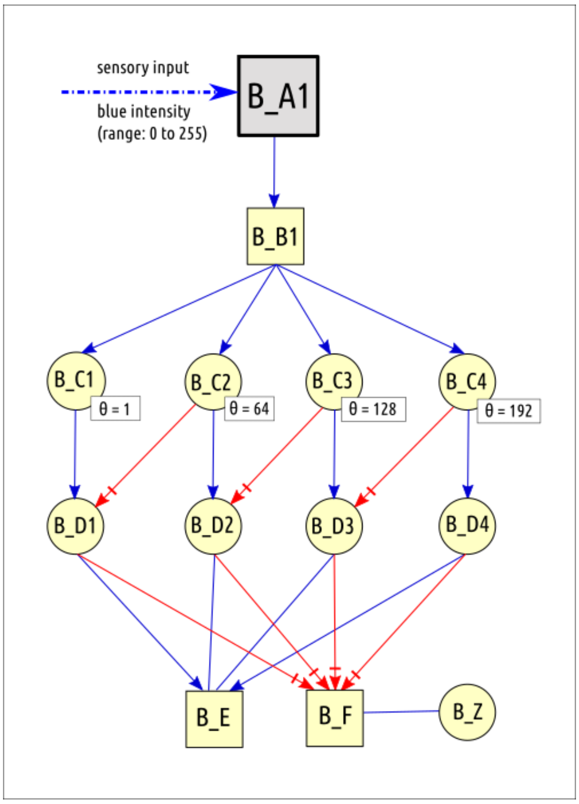
The consequence of this circuit is as follows:

Input to R_A1 at time (t)	Result
Red level = { 1,2,...,63}	*R_D1* fires @ (t+3)
Red level = {64,65,..127}	*R_D2* fires @ (t+3)
Red level = {128,129,..191}	*R_D3* fires @ (t+3)
Red level >= 192	*R_D4* fires @ (t+3)
Red level = 0	*R_F* fires @ (t+4)
Red level > 0	*R_E* fires @ (t+4)



The consequence of this circuit is as follows:

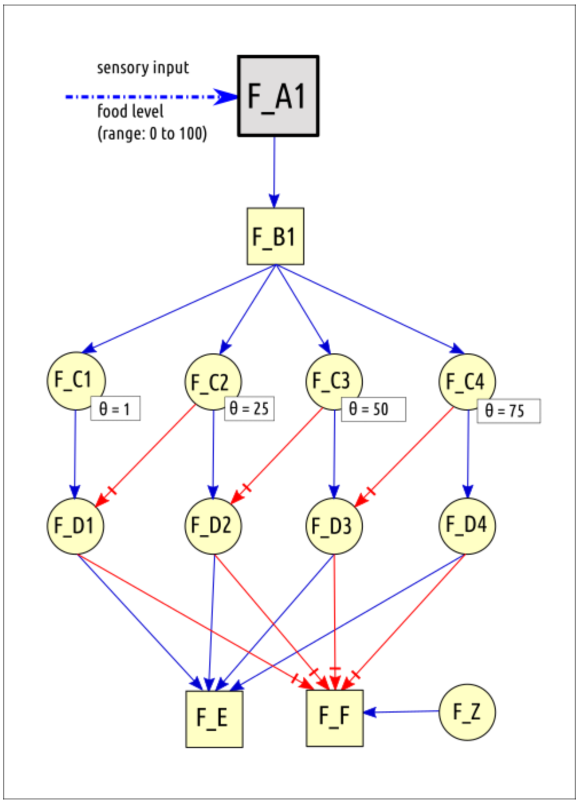
Input to G_A1 at time (t)	Result
Green level = { 1,2,...,63}	*G_D1* fires @ (t+3)
Green level = {64,65,..127}	*G_D2* fires @ (t+3)
Green level = {128,129,..191}	*G_D3* fires @ (t+3)
Green level >= 192	*G_D4* fires @ (t+3)
Green level = 0	*G_F* fires @ (t+4)
Green level > 0	*G_E* fires @ (t+4)



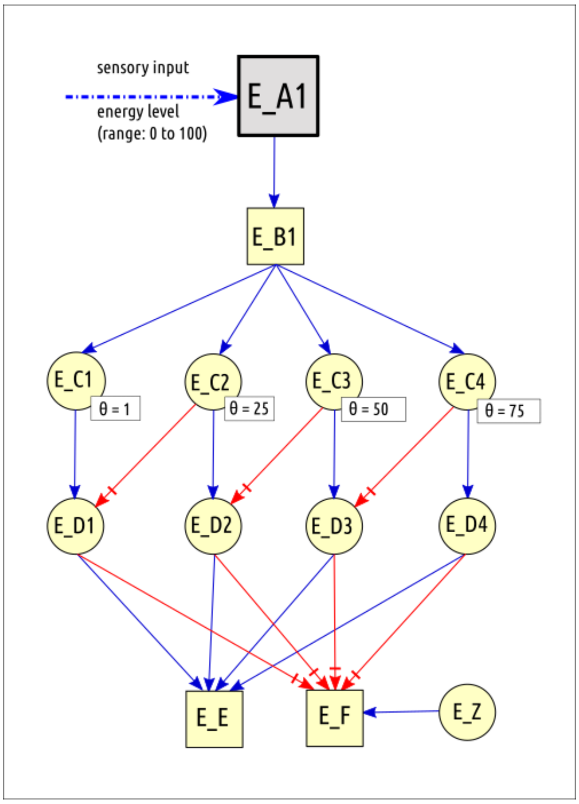
The consequence of this circuit is as follows:

Input to B_A1 at time (t)	Result
Blue level = { 1,2,...,63}	*B_D1* fires @ (t+3)
Blue level = {64,65,..127}	*B_D2* fires @ (t+3)
Blue level = {128,129,..191}	*B_D3* fires @ (t+3)
Blue level >= 192	*B_D4* fires @ (t+3)
Blue level = 0	*B_F* fires @ (t+4)
Blue level > 0	*B_E* fires @ (t+4)

7.2.6.2 Intensity detection circuits for detecting Food-level and Energy-level

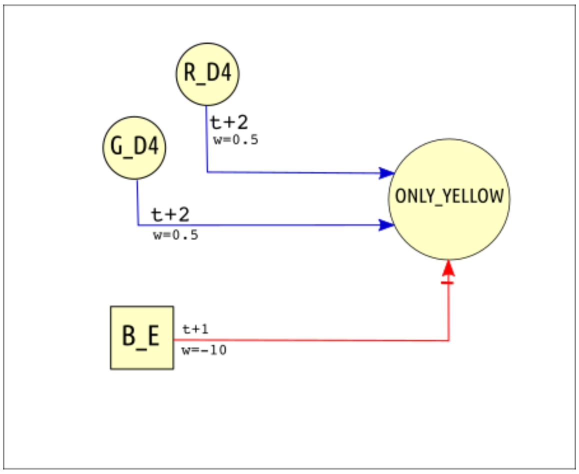


Input to F_A1 at time (t)	Result
Food level = { 1,2,...,24}	*F_D1* fires @ (t+3)
Food level = {24,25,..49}	*F_D2* fires @ (t+3)
Food level = {50,51,..74}	*F_D3* fires @ (t+3)
Food level >= 75	*F_D4* fires @ (t+3)
Food level = 0	*F_F* fires @ (t+4)
Food level > 0	*F_E* fires @ (t+4)

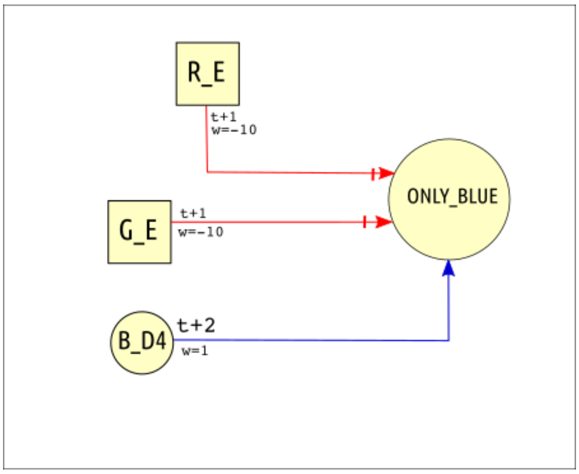


Input to E_A1 at time (t)	Result
Energy level = { 1,2,...,24}	*E_D1* fires @ (t+3)
Energy level = {24,25,..49}	*E_D2* fires @ (t+3)
Energy level = {50,51,..74}	*E_D3* fires @ (t+3)
Energy level >= 75	*E_D4* fires @ (t+3)
Energy level = 0	*E_F* fires @ (t+4)
Energy level > 0	*E_E* fires @ (t+4)

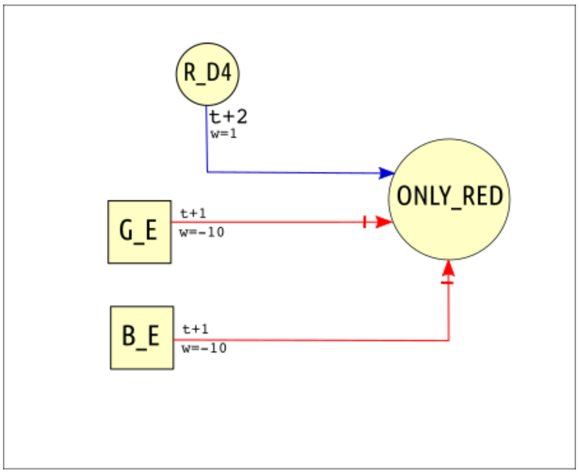
7.2.6.3. Colour detection circuits to detect strong levels of RED,BLUE,GREEN and YELLOW colours



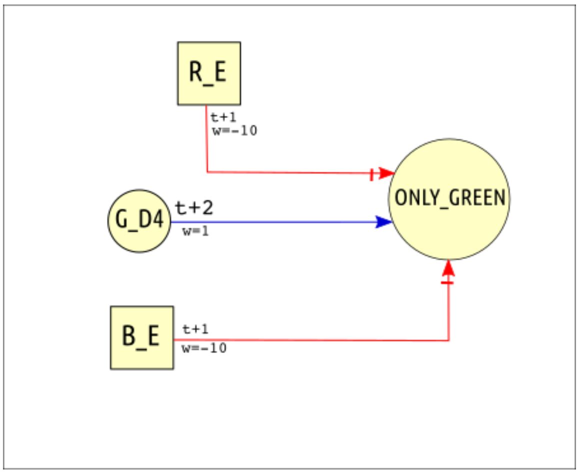
When Red and Green are at maximum levels(**R_D4,G_D4**), and Blue is zero, then **ONLY_YELLOW** is activated



When Blue is at maximum level(**B_D4**), Red is zero and Green are zero, then **ONLY_BLUE** is activated

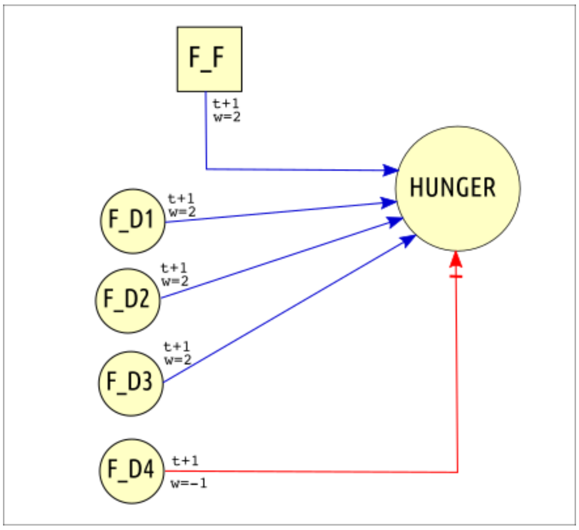


When Red is at maximum level(**R_D4**), Green is zero and Blue are zero, then **ONLY_RED** is activated

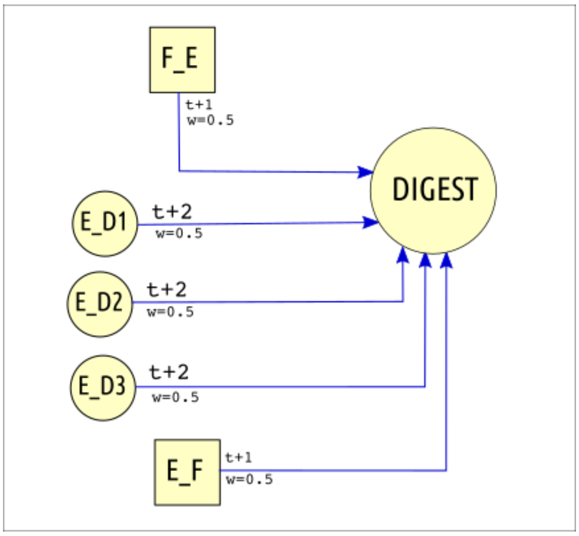


When Green is at maximum level (**G_D4**) , Red is zero and Blue are zero, then **ONLY_GREEN** is activated

7.2.6.4. Hunger and digestion circuits

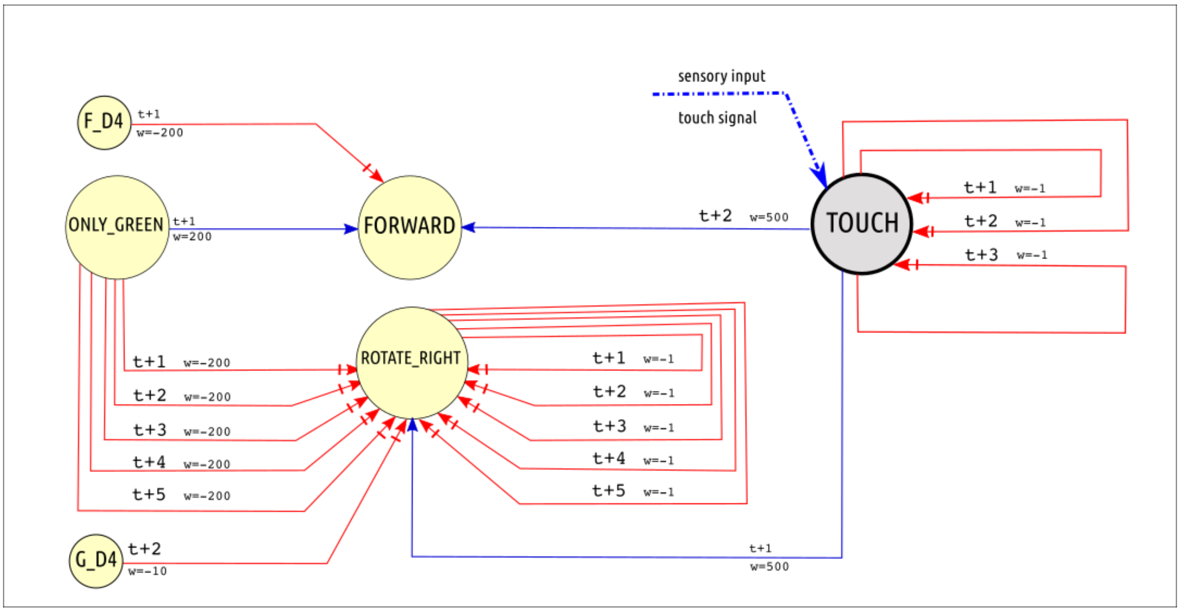


If Food is zero(**F_F**) or Food is low(**F_D1**,**F_D2**,**F_D3**) , then activate HUNGER neuron.When Food is at maximum(**F_D4**), then **HUNGER** is inhibited.



If Energy is zero(**E_F**) or Energy is low(**F_D1**,**F_D2**,**F_D3**) AND Food >0 (**F_E**), then activate **DIGEST** neuron.

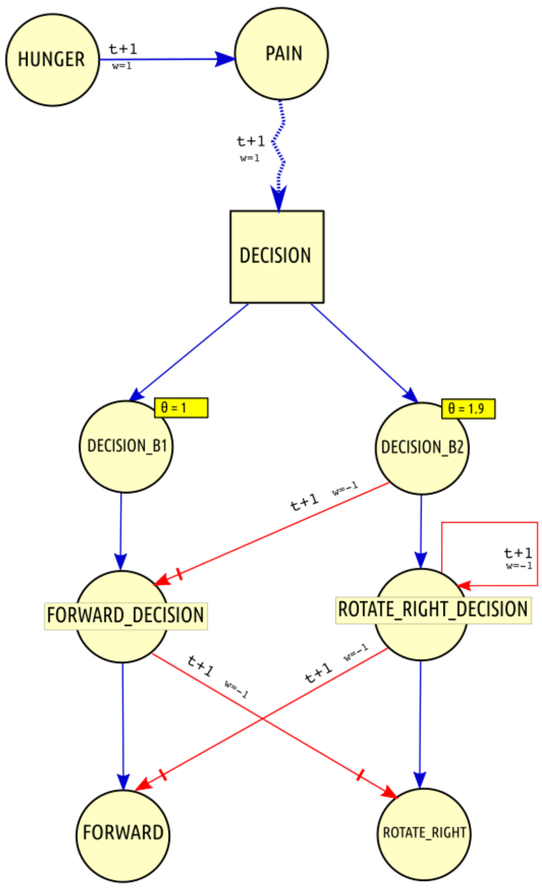
7.2.6.5. Food-seeking innate connection and Touch reflex



Behaviour created by this neural circuit :

- Temporally sharpen **TOUCH** input signal, ie., avoid continuous excitation, just fire once and inhibit immediate self-excitation
- If **TOUCH** detected, then rotate-right only once and then go-forward once.
- Avoid rotating repeatedly, limit consecutive rotations
- If colour seen is pure **GREEN** (food) , then send a very strong signal to move forward and inhibit rotate_right strongly for the next few timesteps.
- Prevent rotate right at same time as **ONLY_GREEN** fires
- When food is at maximum level (**F_D4**) , inhibit **FORWARD** motion.

7.2.6.6. Motivation and Decision circuit



Behaviour created by this neural circuit :

- **HUNGER** neuron activates **PAIN** neuron
- **PAIN** neuron then activates the **DECISION** neuron with a fluctuating random weight, resulting in an input to **DECISION** neuron of between **1.0** to **2.0**
- The output from **DECISION** neuron is sent to **DECISION_B1** and **DECISION_B2** neurons. **DECISION_B1** fires if threshold **1.0** is crossed,and **DECISION_B2** fires only if a higher threshold of **1.9** is crossed.
- This creates the probability of **DECISION_B1** firing 90% of the time and **DECISION_B2** firing 10% of the time. (In this particular environment, moving forward 90% of the time provides a higher survival advantage compared to moving forward 50% of the time, that is the reasoning behind assigning threshold 1.9 for **DECISION_B2** neuron.)
- Since **DECISION_B1** triggers **FORWARD** motor action and **DECISION_B2** triggers **ROTATE_RIGHT** motor action, **FORWARD** motor action is 90% likely to be activated, and **ROTATE_RIGHT** motor action is 10% likely to be activated.
- So finally, depending on whether **FORWARD** or **ROTATE_RIGHT** is activated, Alfie moves or rotates accordingly.

7.2.7. Observed behaviour in this simulation

Scenario 1 : When Alfie sees green coloured food , the input (R_A1=0, G_A1=255, B_A1=0) is sent as visual input. The **ONLY_GREEN** neuron is activated after a few time steps, which in turn strongly triggers **FORWARD** motor action and strongly inhibits **ROTATE_RIGHT** motor action Alfie keeps moving until it collides with the food and consumes it.

Scenario 2 : Click "Yellow light flash" button in the UI: The visual sensor receives yellow light ie., (R_A1=255, G_A1 = 255, B_A1=0). This triggers the yellow-colour detection circuit and causes **ONLY_YELLOW** neuron to be activated after a few epochs and the **ONLY_YELLOW** neuron flashes in the UI

Scenario 3 : Click "Blue light flash" button in the UI: The visual sensor receives blue light ie., (R_A1=0, G_A1 = 0, B_A1=255). This triggers the blue-colour detection circuit and causes **ONLY_BLUE** neuron to be activated after a few epochs and the **ONLY_BLUE** neuron flashes in the UI

Scenario 4 : Click "Place food" button in the UI repeatedly to place green coloured food in a random place in the environment. When FoodLevel is high, it inhibits the **HUNGER** neuron.
This in turn inhibits **PAIN**, which inturn inhibits the **DECISION** neuron from firing.
So Alfie just stays where it is, unmotivated and idle ; Until food level drops over the next dozen epochs.

Scenario 5 : Click "Lights off" radio button in the UI : Alfie explores environment even if it is dark. But less likely to survive in the dark due to inability to see **ONLY_GREEN** coloured food and move forward towards it.

Scenario 6 : Alfie hits wall : **TOUCH** neuron is activated (just) once. The **TOUCH** neuron activates **ROTATE_RIGHT** motor in the next timestep and **FORWARD** motor action in the second timestep.

7.2.8. Other notes on this simulation

This simulation demonstrates how the fusion between perception and action can be achieved. In this simulation, only food-seeking behaviour is demonstrated. However, it is indeed possible to demonstrate more complex behaviours using the same model, like generalization,causality,reinforcement-learning,etc., by adding more connections, motor-actions and permitting dynamic-connection creation , as seen earlier in section 6.7.

8. Conclusion

The concept of time-delayed neural signalling is biologically plausible [7]. By constructing neural circuits based on time-delayed neural signalling, any type of stimulus can be encoded as a scalar to a "standard" circuit, in such a way that it is possible to deduce intensity, detect change and detect patterns in the incoming stimulus. Some innate connections are incorporated for initial survival. Further learning-by-association is enabled by setting criteria for formation, strengthening and weakening of neural connections. The approach described in this paper has the potential to explain other empirical phenomena related to the human brain.

In a neural circuit based system, both data and action-initiation could be encoded in the connectome of the neural network. The neural circuits and other methods described in this paper could be leveraged for use in software applications related to encryption, cybersecurity, and various other use-cases. Software designed based on such neural circuits will be able to determine context, determine the importance of the current ongoing action in the environment and trigger motor actions accordingly. For example, if virtual “pain” is associated with an undesirable action, (eg: detection of unauthorized edit to a database) , a motor action (eg: email alert) can be triggered.
By providing positive or negative reinforcements , more behaviours can be taught to such a system purely by interacting with it. It should be noted that such systems based on Neural circuits are not good at “brute force” remembering (eg: like recalling a long text sequence); however such neural circuits would excel at integrating sensory information, taking prompt action based on context and then adjusting behaviour based on feedback.

Further work is needed on the following topics:

- Designing comprehensive neural circuits for visual stimuli processing(eg: processing colour information, feature extraction,etc) and speech recognition.
- Enhancing/refining the criteria for managing connectome (connection creation/strengthening/weakening)
- Illustrating how abstract concepts are coded in neural connections and associated with relevant sensory stimuli,and how visual feedback pathways enable invocation of "mind's eye"
- Incorporating dynamically activated neural pathways in an Artificial Life Form by setting neuron thresholds as a function of a variable (eg:virtual hormone level), which activates dormant neural pathways and simulates biological phenomena such as circadian rhythm, phobias to certain stimuli,etc.

References

1. Dispense with Backpropagation : Artificial intelligence pioneer says we need to start over : https://www.axios.com/artificial-intelligence-pioneer-says-we-need-to-start-over-1513305524-f619efbd-9db0-4947-a9b2-7a4c310a28fe.html

2. Reflexes in newborns: Neonatal reflexes : https://www.khanacademy.org/test-prep/mcat/behavior/human-development/v/neonatal-reflexes

3. The law of the synapses: punishing the weak to maintain strong synapses strong : https://www.u-tokyo.ac.jp/focus/en/articles/a_00038.html

4. How the body shapes the way we think: Rolf Pfeifer at TEDxZurich : https://www.youtube.com/watch?v=mhWwoaoxIyc

5. How will an AI be made to feel pain? : https://www.quora.com/How-will-an-AI-be-made-to-feel-pain/answer/Kesavan-A-8

6. Motivation and Reward in Learning : Yale University : https://archive.org/details/Motivati1948

7. Neurons can learn temporal patterns : https://www.sciencedaily.com/releases/2017/05/170529090526.htm

Bibliography

1. Robert Sapolsky : Behave: The Biology of Humans at Our Best and Worst
2. Aubrey Manning, Marian Stamp Dawkins : An Introduction to Animal Behaviour