Modeling the cognitive spatio-temporal operations using associative memories and multiplicative contexts

Eduardo Mizraji

Group of Cognitive Systems Modeling
Sección Biofísica, Facultad de Ciencias,
Universidad de la República, Montevideo, Uruguay

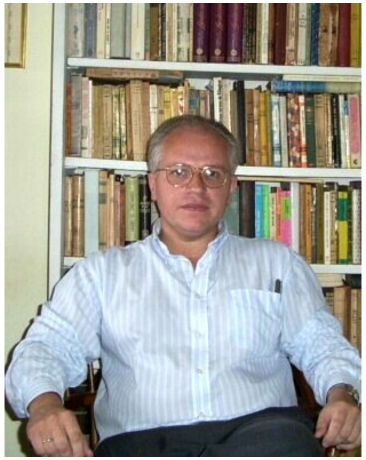
Tandem Workshop on Optimality in Language
and Geometric Approaches to Cognition
Berlin, December 11-13, 2010

Uruguay

Montevideo

Facultad de Ciencias, UdelaR

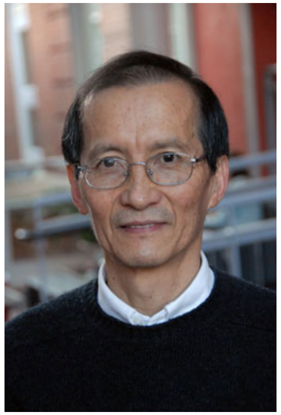
The other members of the team



Dr Andrés Pomi

Dr Juan C Valle-Lisboa

MSc Álvaro Cabana



Dr Juan Lin
Washington College
Chestertown, MD
USA

THE PROBLEM

How to build a minimal neural model capable of representing the coding of spatial and temporal relationships in the cognitive space created by the human mind?

THE PROBLEM

Important epistemological points

1. Nowadays, the neural modeling is (in a mathematical sense) an ill-defined objective because data is not enough to obtain unique solutions.

2. This fact provokes the existence of a family of acceptable coexistent neural models able to explain (provisionally) partial regions of neurobiological and cognitive realm.

3. In the present work, I explore only one member of this family.

THE "INSTRUMENT": Context-modulated matrix memories

Some antecedents

In the 1970's:

Teuvo Kohonen explored the non-linear processing of vector inputs

In the 1980's:

Ray Pike defined a matrix scalar product that allows context modulation of data

Paul Smolensky described a tensor product approach able to represent a variety of cognitive performances

Another approach by E. Mizraji was rooted in Ross Ashby's theory of adaptive control systems

Ashby's machine :

$$E \equiv \{a, b, c, d, e\}$$ State space

$$P \equiv \{\alpha, \beta, \gamma\}$$ Parameter space

$$\Gamma : P \times E \rightarrow E$$ Machine-with-input

$$\begin{array}{llllll} \Gamma: \downarrow & a & b & c & d & e \\ \alpha & b & c & d & e & a \\ \beta & a & a & b & d & d \\ \gamma & c & d & e & a & b. \end{array}$$

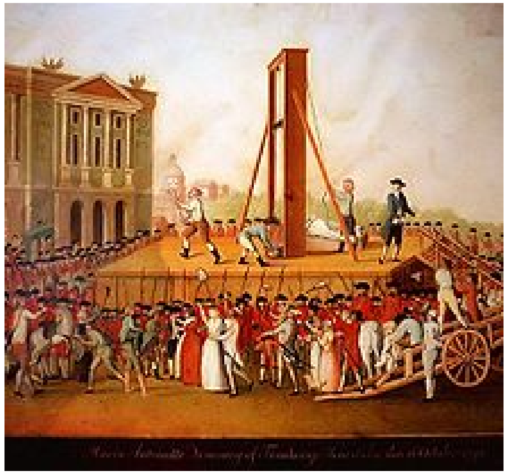
$$\Gamma(\alpha, b) = c$$

$$\Gamma(\gamma, b) = d$$

The parameters of a machine with input are 'gratuitous' contexts that allows evolutionary adaptation to changing and unpredictable environments.

What was the place and what was the right order?

Declaration of the Rights
of Man and of the Citizen



Execution of Marie-Antoinette

Storming the Bastille

To answer, we need :

(a) Information in our memories

(b) Computational abilities to deal with order relations

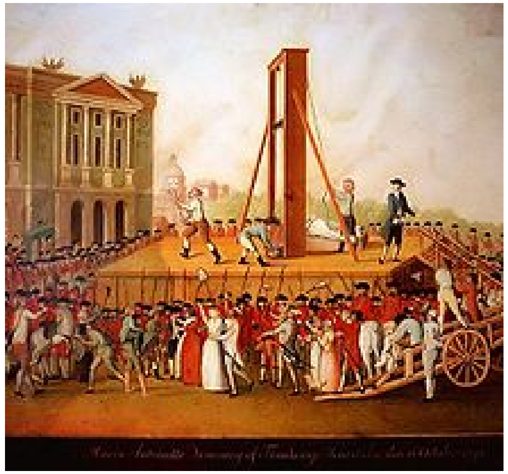
What was the place and what was the right order?

<u>The place</u> : France

<u>The right order</u> (from past to future) :

Storming the Bastille

Declaration of the Rights of Man and of the Citizen



Execution of Marie-Antoinette

Past → Future

The neural abilities to deal with order relations

A heuristic approach

Computing with words: logical words and prepositions

Miscellaneous examples

(a) "MOST cats are black" (from de Hoop, Hendriks and Blutner)

(b) "She is smart AND beautiful"

(c) "5 is NOT a negative number"

(d) "To live, it is NECESSARY to breathe"

(e) "To live, it is NOT POSSIBLE NOT to breathe"

(f) "The notebook is ON the table"

(g) "He is BEHIND you"

Note: Aristotle stated (and our brains usually confirm!) the equivalence of expressions (d) and (e)

Computing with words: logical words and prepositions

a) We postulate that these words give access to complex neural programs that compute the variety of logical or relational operations expressed by them.

b) Vectors are natural representations of concepts inside the neural system, and in what follows we are going to assume that different concepts map on orthonormal vectors.

An example: logical memories

Symbol-vector mapping :

$$1 \mapsto s \in E_q$$

$$0 \mapsto n \in E_q$$

NOT:
$N = sn^T + ns^T$

AND :
$C = s(s \otimes s)^T + n(s \otimes n)^T + n(n \otimes s)^T + n(n \otimes n)^T$

OR :
$D = s(s \otimes s)^T + s(s \otimes n)^T + s(n \otimes s)^T + n(n \otimes n)^T$

XOR :
$X = n(s \otimes s)^T + s(s \otimes n)^T + s(n \otimes s)^T + n(n \otimes n)^T$

IMPL :
$L = s(s \otimes s)^T + n(s \otimes n)^T + s(n \otimes s)^T + s(n \otimes n)^T$

A connection between logical memories and set operations (Mizraji 1992)

Characteristic function of a set S :

$$f(z, S) = \begin{cases} s & \text{iff } z \in S \\ n & \text{iff } z \notin S \end{cases}$$

$$f\left(z, \bar{A}\right) = N$$

$$f(z, A \cap B) = C$$

$$f(z, A \cup B) = D$$

Asymmetrical prepositions as words that compute spatial and time relationships

Some examples:

Before
After
On
Under
From
Towards

Which are the neural computations that underlie the understanding of these words?

Ziggurat metaphor for a hierarchical processing

High-level processing

Medium-level processing

Basal-level processing

Main Inspiration : It comes from the vector coding of order relations used in the modeling of hybrid neural representations of numbers by S. Dehaene and J.-P. Changeux, and by J.A. Anderson

High-level neural models for order relations (I)

PROVISIONAL ASSUMPTION

The asymmetric prepositions are installed as neural versions of anti-commutative functions :

$$f(x,y) = -f(y,x)$$

Let us use a hybrid representation based on the logical vectors s and n :

$$\boxed{\Psi(u,v) = N\Psi(v,u) \;\; , u,v \in E_R}$$

Strategy:

We assign specific coding vectors to the "previous-posterior" pairs

High-level neural models for order relations (II)

Some definitions

Logic truth values : $T = \{s, n\}$, $s, n \in E_q$

Positional values : $\Omega = \{b, i, a\}$, $b, i, a \in E_R$

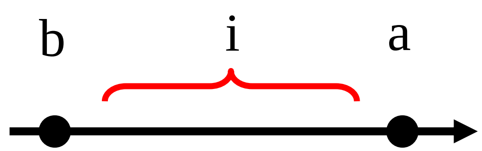
Important remark : all these basic vectors are orthonormal

Positional parameters of a coded event

$b \otimes \text{pattern} \rightarrow$ pattern positioned "before" ("under")

$i \otimes \text{pattern} \rightarrow$ pattern positioned "between"

$a \otimes \text{pattern} \rightarrow$ pattern positioned "after" ("on")



High-level neural models for order relations (III)

Monadic operators F and P $\quad \mathrm{Op1}: \Omega \rightarrow \mathrm{T}$

They are matrices that compute (similarly to the classic operators F and P of temporal logic) the answers corresponding to the following questions:

Matrix F : Will the event happen in the future?

Matrix P : Did the event happen in the past?

$$\boxed{F = nb^T + sa^T \quad , \quad P = sb^T + na^T}$$

Remark that $F = NP$

High-level neural models for order relations (IV)

Dyadic operators for asymmetric prepositions $\quad Op2 : \Omega \times \Omega \to T$

Let A be a matrix that codes abstractly the order relations and answer questions as: "is u in front of v?" or "is u on v?".
One of the possible formats of this matrix is

$$\boxed{A = s(a \otimes b)^T + n(b \otimes a)^T}$$

(matrix "after")

If now the questions are "is u behind v?" or "is u under v?", a possible operator is

$$\boxed{B = n(a \otimes b)^T + s(b \otimes a)^T}$$

(matrix "before")

Remark : $A(a \otimes b) = NA(b \otimes a)$ and $B = NA$

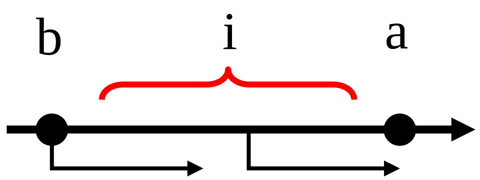
High-level neural models for order relations (V)

Dynamic order

The words “towards” and “from” describe dynamic order relations that code transitions. We can model the high-level processors using the following matrices.

Matrix “Towards” (events move towards a)

$$\boxed{\begin{aligned}\mathrm{Tow} = {} & n(b \otimes a)^T + n(b \otimes i)^T + n(i \otimes a)^T \\ & + s(a \otimes b)^T + s(a \otimes i)^T + s(i \otimes b)^T\end{aligned}}$$



Matrix “From” (events move from b)

$$\boxed{\mathrm{Fro} = N\ \mathrm{Tow}}$$

Notes :
(a) If the intermediate value I does not exist, these matrices degenerate in the previous matrices A and B
(b) The coding of intermediate positions using a single vector is similar to the strategy created by J. Lukasiewicz to define logical modal operators using a 3-valued logic.

Medium-level neural processing (I)

In the present model, the medium-level operations connect vectors that 'conceptualize' sizes (or positions, or temporal order), to the high-level vector set $\Omega = \{b, i, a\}$.

We use an additive composition of vectors with the purpose of modeling the emerge of transitivity. This composition is defined as follows :

Let $u = [u_1 \cdots u_R]^T$, $v = [v_1 \cdots v_R]^T$

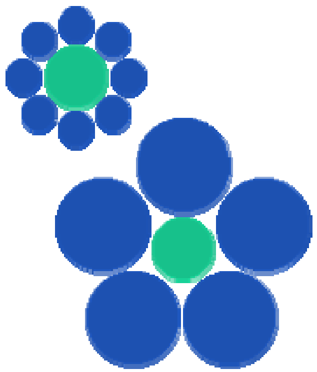
Then $u \cup v \equiv [u_1 \cdots u_R \; v_1 \cdots v_R]^T$

$u, v \in E_R$, $u \cup v \in E_{2R}$

<u>Consequence</u> : $\langle u \cup v, u' \cup v' \rangle = \langle u, u' \rangle + \langle v, v' \rangle$

Medium-level neural processing (II)

<u>Perceptual Titchener Effect</u>



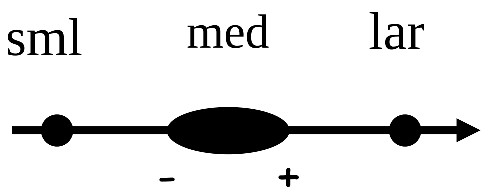
We define three basic 'size vectors' :

Small-size vector → sml

Medium-size vector → $\mathrm{med}[\pm]$

Large-size vector → lar

$$\mathrm{sml}, \mathrm{med}[\pm], \mathrm{lar} \in E_S$$



Let us define a "**<u>Conceptual Titchener Effect</u>**" that enhanced the contrast between the extreme sizes and the medium size, generating the following associated pairs :

$$\mathrm{sml} \cup \mathrm{med}[+] \quad , \quad \mathrm{med}[-] \cup \mathrm{lar}$$

Miniature 4-dim examples:

$$\mathrm{sml} = [1 \ 0 \ 0 \ 0]^T \ , \qquad \mathrm{med}[+] = [0 \ 0 \ 1 \ 0]^T$$

$$\mathrm{med}[-] = [0 \ 1 \ 0 \ 0]^T \ , \ \mathrm{lar} = [0 \ 0 \ 0 \ 1]^T$$

Medium-level neural processing (III)

LINKAGE MATRICES

(a) G is a matrix that connects adjacent pairs of size coding vectors, from sml to lar with the corresponding high-level order pair

$$G = (b \otimes a)\left[\left(\mathrm{sml} \cup \mathrm{med}[+] \right)^{T} + \left(\mathrm{med}[-] \cup \mathrm{lar} \right)^{T} \right]$$

(b) R is a matrix that connect adjacent pairs, from lar to sml, with the decreasing abstract coding vectors

$$R = (a \otimes b)\left[\left(\mathrm{lar} \cup \mathrm{med}[-] \right)^{T} + \left(\mathrm{med}[+] \cup \mathrm{sml} \right)^{T} \right]$$

(c) Global linkage matrix : $\boxed{\mathrm{GLM} = G + R}$

Medium-level neural processing (IV)

LINKAGE MATRICES : Some operations

Case 1 : Normal operation

$$\mathrm{GLM}\left(\mathrm{lar} \cup \mathrm{med}[-]\right) = 2(\mathrm{a} \otimes \mathrm{b})$$

Remark: The 'conceptual Titchener effect' could prevent interference (in this case with terms containing vector med in the second positions)

Case 2 : Medium level transitivity

$$\begin{aligned}\mathrm{GLM}(\mathrm{sml} \cup \mathrm{lar}) = & [\langle \mathrm{sml}, \mathrm{sml}\rangle + \langle \mathrm{med}[+], \mathrm{lar}\rangle + \langle \mathrm{med}[-], \mathrm{sml}\rangle + \langle \mathrm{lar}, \mathrm{lar}\rangle](\mathrm{b} \otimes \mathrm{a}) \\ & + [\langle \mathrm{lar}, \mathrm{sml}\rangle + \langle \mathrm{med}[-], \mathrm{lar}\rangle + \langle \mathrm{med}[+], \mathrm{sml}\rangle + \langle \mathrm{sml}, \mathrm{lar}\rangle](\mathrm{a} \otimes \mathrm{b})\end{aligned}$$

Hence

$$\begin{aligned}\mathrm{GLM}(\mathrm{sml} \cup \mathrm{lar}) &= (1+0+0+1)(\mathrm{b} \otimes \mathrm{a}) + (0+0+0+0)(\mathrm{a} \otimes \mathrm{b}) \\ &= 2(\mathrm{b} \otimes \mathrm{a})\end{aligned}$$

Basal-level neural processing of perceptual data (I)

Object-Size Pairs Associative Memories

Growing order associations

$$H_{sm} = \left(sml \cup med[+]\right)\sum_{i}\left(f_i \cup f_i^{'}[+]\right)^{T}$$

$$H_{ml} = \left(med[-] \cup lar\right)\sum_{j}\left(f_j^{'}[-] \cup f_j^{''}\right)^{T}$$

Decreasing order associations

$$K_{lm} = \left(lar \cup med[-]\right)\sum_{k}\left(f_k \cup f_k^{'}[-]\right)^{T}$$

$$K_{ms} = \left(med[+] \cup sml\right)\sum_{n}\left(f_n^{'}[+] \cup f_n^{''}\right)^{T}$$

Let

$$\boxed{M = H_{sm} + H_{ml} + K_{lm} + K_{ms}}$$

Basal-level neural processing of perceptual data (II)

An example

$$f_2 \rightarrow \text{elephant} \quad , \quad f_2'[-] \rightarrow \text{dog}$$

$$f_4'[+] \rightarrow \text{dog} \quad , \quad f_4'' \rightarrow \text{bacteria}$$

Consequences :

$$\text{(a)} \quad M\left(f_2 \cup f_4''\right) = \left(\text{lar} \cup \text{med}[-]\right) + \left(\text{med}[+] \cup \text{sml}\right)$$

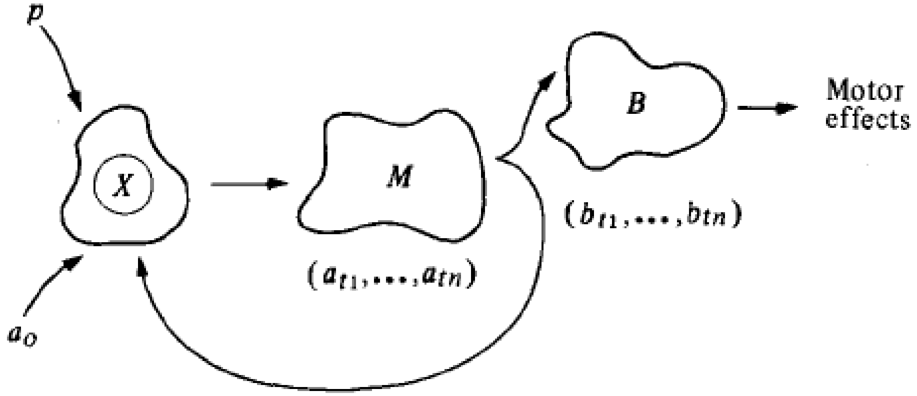
$$\text{(b)} \quad GLM\left[M\left(f_2 \cup f_4''\right)\right] = 4(a \otimes b)$$

Note : The model assumes that imperfections, errors, or inconsistencies are allowed by this kind of "empirical" matrix memories

Organizing episodes with contextual labels (I)

Two possible scales for the neural modeling of episodes

Scale 1: "Micro-episodes" as procedural associations without explicit time coding



(from Mizraji BMB, Vol 50, 1989)

In this case an associative sequence is stored inside a memory module, and the associative chain is acceded with a key initial pattern and a semantic context

Example: The phonetic production of a word gated by a conceptual pattern that acts as context.

Organizing episodes with contextual labels (II)

Two possible scales for the neural modeling of episodes

Scale 2: "Macro-episodes", as contingent associations where different memory modules integrated in a large network of networks, are connected with key context that explicitly specify time position and allows transitive computations.

Organizing episodes with contextual labels (III)

Macro-episodes: A theory for context-modulated searching trajectories
(Mizraji 2008, Mizraji, Pomi and Valle-Lisboa 2009)

Main idea : The selection of different associative pathways in a modular network can be guided by multiplicative contexts operating both at the input and at the output levels.

Each term of inside a memory module is

$$(\hat{p}'_i \otimes \hat{g}_{ij})\,(\hat{p}_i \otimes \hat{f}_{ij})^{\mathrm{T}} = (\chi_i p'_i p_i^{\mathrm{T}}) \otimes (\mu_{ij} g_{ij} f_{ij}^{\mathrm{T}})$$

being $\chi_i = \|\hat{p}'_i\| \cdot \|\hat{p}_i\|$ and $\mu_{ij} = \|\hat{g}_{ij}\| \cdot \|\hat{f}_{ij}\|$

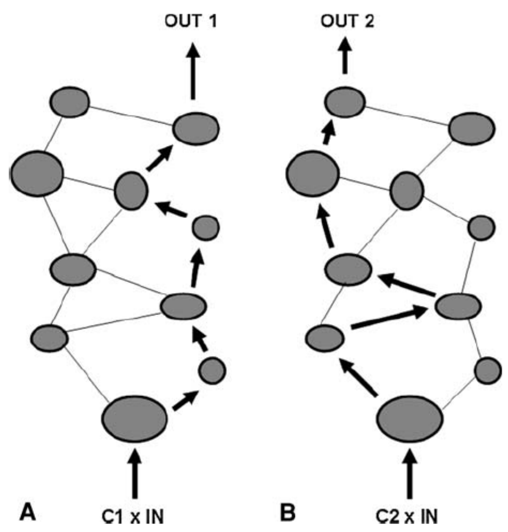
The structure of the whole memory module is

$$\mathrm{H} = \sum_i \left(\chi_i p'_i p_i^{\mathrm{T}} \otimes \sum_j \mu_{ij} g_{ij} f_{ij}^{\mathrm{T}} \right)$$

$$\mathrm{M}_i = \sum_j \mu_{ij} g_{ij} f_{ij}^{\mathrm{T}}$$

$$\mathrm{H} = \sum_i \left(\chi_i p'_i p_i^{\mathrm{T}} \otimes \mathrm{M}_i \right)$$

Different associative trajectories in the same network



Organizing episodes with contextual labels (IV)

Neural translation of an order-decision question

Imagine we ask: "is f smaller than f' ?"
We can transform this question in an ordered triple (p_b, f, f'').
p_b is a contextual parameter that codes for the question "is ---smaller ---?"

Similarly the question "is f larger than f" ?" can be represented by the triple (p_a, f, f'') with p_a representing the question "is ---larger ---?"

According to the previous theory, the memories that processed questions concerning orders, can be labeled by contexts as follows

High-level processing → $p_b^T \otimes B \quad , \quad p_a^T \otimes A$

Medium-level processing → $p_b p_b^T \otimes GLM + p_a p_a^T \otimes GLM$

Basal-level processing → $p_b p_b^T \otimes H + p_a p_a^T \otimes H$

Organizing episodes with contextual labels (V)

Using the theory to climb to the top of the ziggurat

At the top of the ziggurat the answer is NEGATIVE

(3) High-level processing

$$p_a^T \otimes A\left(p_a \otimes (b \otimes a)\right) \approx n$$

(2) Medium-level processing

$$\left(p_b p_b^T + p_a p_a^T\right) \otimes GLM\left(p_a \otimes \left(med[-] \cup lar\right)\right)$$
$$\approx p_a \otimes (b \otimes a)$$

(1) Basal-level processing

$$\left(p_b p_b^T + p_a p_a^T\right) \otimes H\left(p_a \otimes (f \cup f')\right)$$
$$\approx p_a \otimes \left(med[-] \cup lar\right)$$

Dog $\rightarrow f$
Elephant $\rightarrow f'$

<u>The basal question</u> : Is the **<u>dog</u>** bigger than the **<u>elephant</u>**? $\rightarrow (p_a, f, f')$

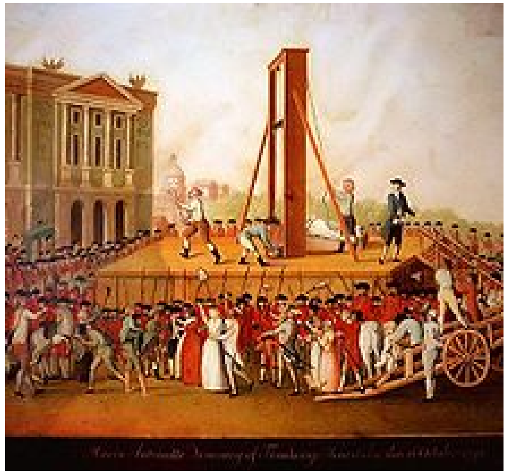
Conclusion 1

Episodes as orderly paths in the cognitive space

The cognitive establishment of an episode requires the neural ability to code order relations. Complex episodes plausibly involves many memory modules as well as contextual labels capable of linking one module to the next.

Storming the Bastille

Declaration of the Rights of Man and of the Citizen



Execution of Marie-Antoinette

Past → Future

Conclusion 2

<u>Order relations emerge from a process of abstraction</u>

a) This model assumes that a variety of perceptual processes converge to a more reduced repertoire of order concepts.

b) In turn, these order concepts are sent to a small set of neural computational modules, a kind of collective neuro-computational "final common pathway" that takes final decisions.

c) The hierarchical model we described here to compute order relations is capable of detecting transitive relations

Conclusion 3

Experimental counterpart

a) A prediction of the model is the convergence of order computations, towards neural modules that decrease their specificity and increase their level of abstraction

b) The experimental refutation of this model is possible

c) This refutation could result from the practical impossibility to prove the existence of hierarchical processing converging to a cognitive (and abstract) “final common pathway”

d) This refutation can be produced by a solid and reproducible series of experiments using the appropriate brain images techniques

THANK YOU VERY MUCH FOR YOU KINDNESS AND PATIENCE !

Appendix: Context-dependent matrix associative memories

A non-linear basis for linear neuronal models

The Nass-Cooper neurochemical version (1975) for the Knight's "integrate and fire" model

Basic theory :

$$\frac{dT_{ij}}{dt} = -\alpha T_{ij} + M'_{ij} d_j$$

$$V_{ij} = k_{ij} T_{ij} .$$

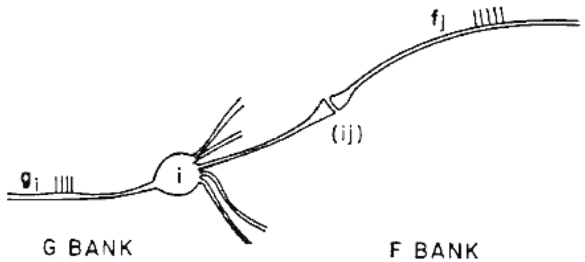
$$\frac{dV_{ij}}{dt} = -\alpha V_{ij} + k_{ij} M'_{ij} d_j .$$

$$V_i = \sum_{j=1}^{N} V_{ij}$$

$$\frac{dV_i}{dt} = -\alpha V_i + D_i$$

$$D_i \equiv \sum_{j=1}^{N} k_{ij} M'_{ij} d_j .$$

$$V_i(t) = e^{-\alpha t} \int_0^t e^{\alpha t'} D_i(t')\, dt' .$$

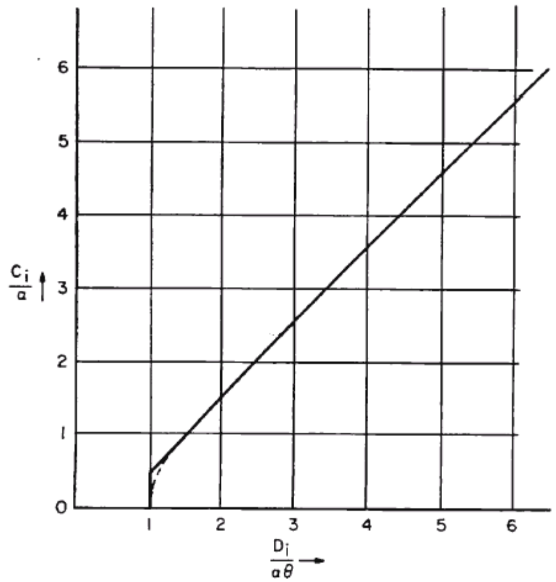


Main result ;

$$D_i(t') = D_i \qquad 0 < t < \tau$$
$$= 0 \qquad \tau < t .$$

$$\boxed{c_i = \frac{D_i}{\theta} = \Sigma M_{ij} d_j} \quad \left(\frac{D_i}{\theta} \gg 1\right).$$

$$M_{ij} \equiv \frac{1}{\theta} k_{ij} M'_{ij} .$$



$$c_i = \frac{-\alpha}{\ln\left(1 - \frac{\alpha\theta}{D_i}\right)} .$$

Matrix associative memories

(J.A.Anderson, T.Kohonen, S-I. Amari, L.N.Cooper, [decade of 1970])

Hetero-associative memory :

$$M^{(\mathrm{k})} = \sum_i \mu_i^{(\mathrm{k})} g_i f_i^{\mathrm{T}}.$$

$$M^{(\mathrm{k})} f_i = \mu_i^{(\mathrm{k})} g_i$$

Modeling the activity of an individual neuron including multiplicative terms

$$g(i, t+1) = \sum_{j} M_{ij}^{(1)} p(j, t) + \sum_{k} M_{ik}^{(2)} f(k, t) + \sum_{j,k} M_{i(jk)}^{(3)} p(j, t) f(k, t)$$

The two first terms correspond to a classical matrix pattern-associator. The third term involves coincidence detector synapses.

Multiplicative context-dependent associative memories (Mizraji, 1987)

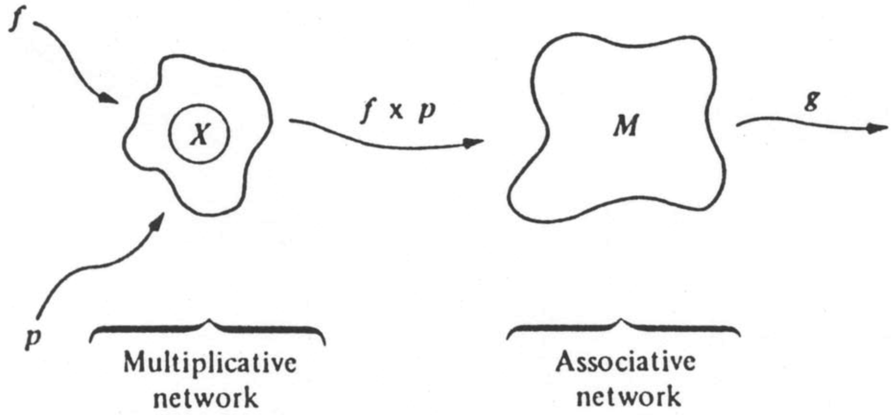
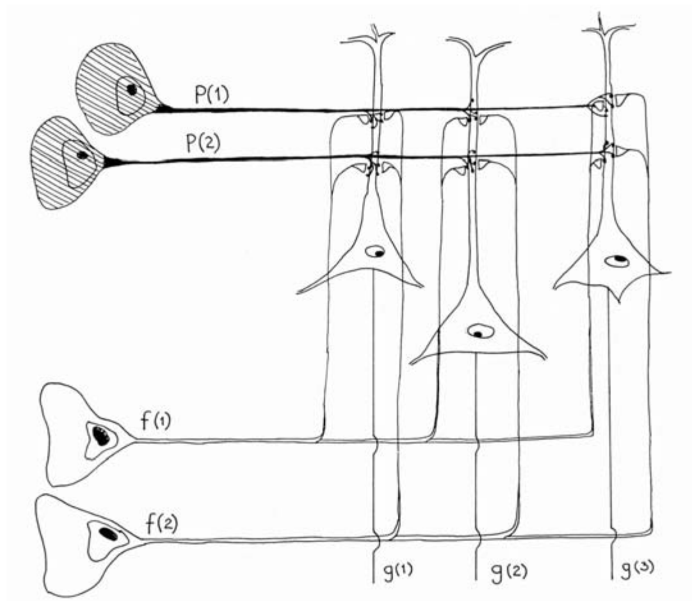


Figure 1. Context sensitive associative memory.



CONTEXT-DEPENDENT MEMORIES AS ASHBY'S NEURAL MACHINES

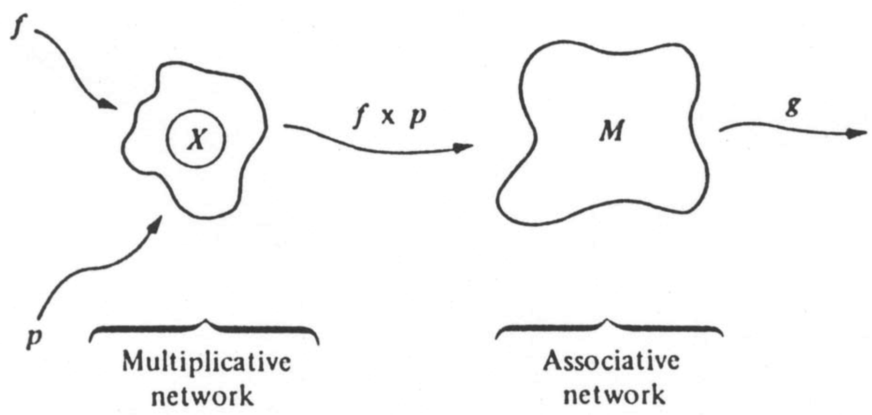


Figure 1. Context sensitive associative memory.

Associated matrix

These multiplicative context-dependent memories are represented using Kronecker (tensor) product based on the existence of coincidence detectors.

$$M = \sum_{i,j} g_{ij}(p_i \otimes f_j)^T$$

$$M(p_k \otimes f_h) = \sum_{i,j} g_{ij}\langle p_i, p_k\rangle\langle f_j, f_h\rangle$$

Reliability of the associations in the presence of incomplete Kronecker products