Manuscript draft, November 20, 2014

Thoughts on the Brain,

A NEURO-LOGICO-MATHEMATICAL ESSAY

E.Engeler

Abstract

The mathematical model introduced in this paper attempts to explain how complex scripts of behavior and conceptual contents can reside in, combine and interact on large networks of interconnected basic actors.

The approach is exemplified by modeling the neural structure and dynamics of the connectome of a brain. The neurological hypothesis attributes functions of the brain to sets of firing neurons, dynamically to sets of cascades of such firings, typically visualized by imaging technologies. Such sets are represented as the elements of what we call a neural algebra, and their interaction as its basic operation. For convenience, we name specific represented objects, and particularly kinds of objects, using the vocabulary of mental functions for its richness and suggestiveness, using names such as ”thoughts”, ”concepts”, ”memories”, ”scripts”, and use logical terminology for naming some of the combinations of these objects.

The main thrust of this paper develops from the fact that characteristic properties of these suggestive notions can be cast in the form of equations of the neural algebra. Analyzing the solutions leads to a complete description of the necessary structure of their neural correlates. In particular we analyze the representation of perception in the form of ”concepts” and of control in its various forms, distributed, hierarchical and especially reflexive control, the latter modelling a conception of ”consciousness”.

1 Introduction: About Thinking

Neuroscience has demonstrated that mental objects such as individual concepts and memories are locatable in the brain as specific assemblies of neurons (and their connections). Encoded in living matter, they are not static, but participate in interacting processes as part of ”thinking”. So, even if we know to identify some selected individual concepts as structures in the brain, the challenge is to understand them dynamically in their interaction.

Thinking means applying thoughts to thought. If the thought A is applied to the thought B, then the result $A \cdot B$ is again a thought: A theory of thinking inherently has this algebraic aspect. But the tradition of Logic reduces the operational aspect to linguistic categories, namely to operations on propositions such as ”and”, ”or”, etc, or to modularities such as ”necessarily”. While this leads, since Aristotle, to very rich and fruitful logical theories it also impoverishes logic as a basis for a theory of the mind. In contrast, we hold that an algebra of thoughts should be based directly on an analysis of what it means to apply a thought to a thought.

Such an approach is detailed in the following section. It results, as we shall see, in an algebraic system for formally representing thought objects and their mutual interactions. This algebraic framework allows to distinguish types of thoughts by their algebraic properties, equations as it were. For example, conceptual thoughts, concepts, are characterized as retraction operations. More generally, perceiving, acting on thoughts, thinking about other peoples thoughts are captured by sets of equations allowing to discuss various hypotheses on the functioning of the brain and its connectional structure. Problem solving, that major aspect of nature generally, is thus captured as solving equations for unknown brain mechanisms.

This implied relation between our formal model and biological reality is to be understood suggestively. The the names we choose for our objects are simply names and may be substituted by names chosen from completely different contexts that deal with intercommunicating actors. But we do attempt, in the Discussion at the end, to relate these development to (our appreciation of) current neuroscience. The discussion is referred to by section-numbers; references to the literature are listed there

,

2 Patterns of Thought

THE BRAIN MODEL $\mathcal{A}$

The conceptually simplest model of a brain represents its connectivity, the *connectome* A, as a directed graph whose nodes, called neurons, fire at discrete time instances $t \in \mathbb{Z}$. The global activity of the brain, the firing history of these neurons, is represented by the *firing function* $f(a, t)$ which takes the value 1 if the neuron a fires at time t and 0 otherwise. Modelling a brain is accomplished by imposing restrictions on the functions f by specific a *firing law* inherited from abstracting neurological findings. A firing law specifies the condition under which the firing of neurons $a_1, \ldots, a_k$ at times $t_1, \ldots, t_k$ causes the firing of a neuron a_{k+1} at some later time t_{k+1} , assuming the former are connected to it by directed edges.

For example: In artificial neural nets a rudimentary firing law is based on assigning weights to the individual directed edges of the graph A: If sum of weights the incoming edges (synapses) exceeds a given threshold, then the firing of the corresponding source neurons at time t causes the firing of the target neuron at time $t + 1$. Positive weights correspond to excitatory, negative weights to inhibitory synapses.

Remark: To view living neurons as purely reacting entities is too restrictive in my opinion. As the result of a very long line of descent from unicellular ancestors, it seems reasonable to suspect that they retain some mechanisms of memory, optimization, goal functions, etc. These could conceivably be modeled in the fashion of our brain models, indeed subsumed in them.

The directed graph A together with the firing function f and the firing law constitute our brain model $\mathcal{A}$; it describes a full history of the modeled brain.

Taking a causal point of view of the sequences of individual firings, we are able to distinguish cascades of firings: Starting with some arbitrarily selected firings at some time instances, a *cascade* is a branched sequence of firings of neurons which causally follow from these original activations during a finite time interval. A *firing pattern* is simply a set of such cascades.

Our aim is to view firing patterns as functions, brain functions. In analogy to the usual set-theoretic definition of functions, the individual cascades need to be understood as tuples of input / output cascades. This is accomplished by choosing the firing of a specific neuron as the key point of causality: the parts of the cascade that are its temporal antecedents are understood as inputs; a cascade that follows it is understood as output. – This analysis of cascades is formalized below by representing such cascades as *track expressions*. Note that the input- and output-cascades should reasonably also be represented as track expressions, thus structuring the whole cascade. This leads to the formal recursive definition of track expressions in the next section.

Firing tracks and firing patterns

Let $\mathcal{A}$ be a given brain model. Track expressions are defined recursively as follows:

The basic track expression $a(t)$ denotes the activation of a single neuron a at an integer time instance t.

Composite track expressions are based on paths of directed edges in A and the firing function f, i.e. on cascades of firings. They all have the form $x_c(t)$ for some neuron c, the *key neuron* of $x_c(t)$, and time instance t. In particular, the key neuron of $a(t)$ is a, thus this expression may also be written $x_a(t)$.

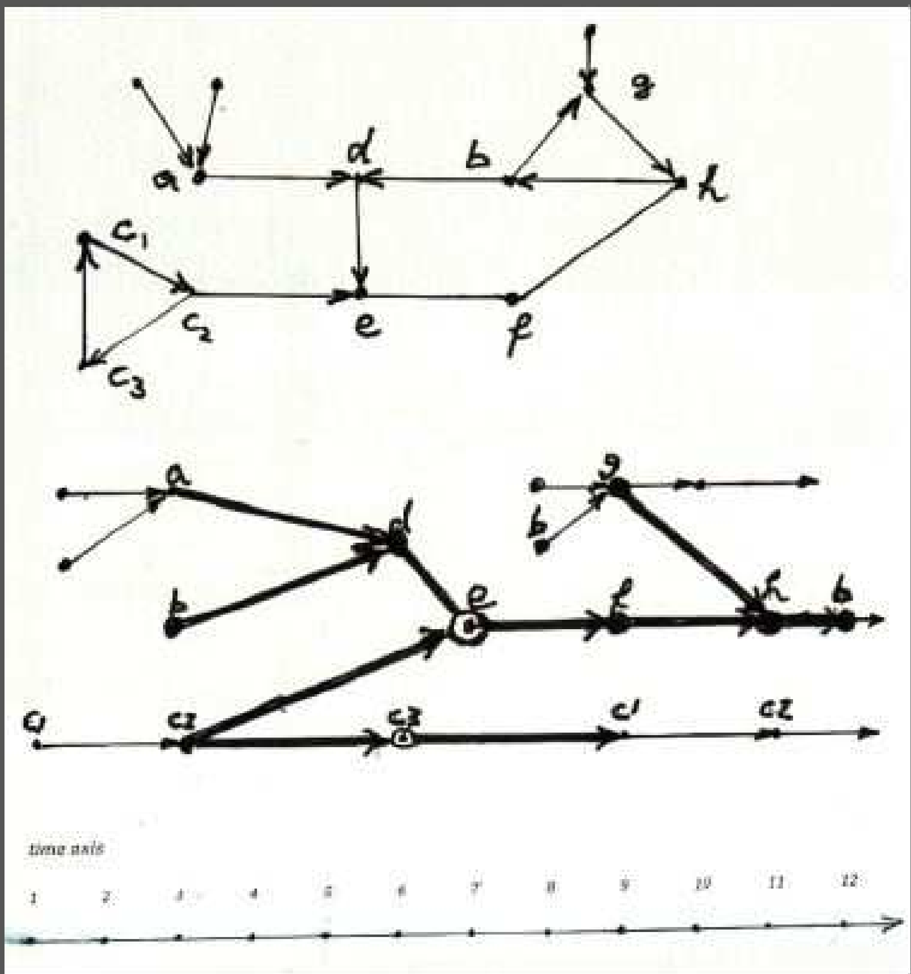
The *antecedent neurons* $a_1 \ldots a_n$ are connect to neuron b along paths of one or more edges in A, which in turn connects by such a path to neuron a_{n+1}. By defining

$$x_b(t) = \{x_{a_1}(t_1), \ldots x_{a_n}(t_n)\} \xrightarrow[b]{t} x_{a_{n+1}}(t_{n+1}),$$

where

$$t, t_1, \ldots t_{n+1} \in \mathbb{Z}, t_1, \ldots t_n < t < t_{n+1},$$

we recursively compose the track expressions $x_{a_1}, \ldots x_{a_{n+1}}$. The neuron b is called the *key neuron* of $x_b(t)$, and $a_1, \ldots, a_{n+1}$ are the key neurons of the track expressions $x_{a_1}, \ldots, x_{a_{n+1}}$. Each such track expression, by timing of the key neurons, describes a firing of the neurons occurring in it. This leads to:



$$\{\{a(3), b(3)\} \xrightarrow[d]{6} e(7), c_2(3)\} \xrightarrow[e]{7} (\{e(7)\} \xrightarrow[f]{9} (\{f(9), g(9)\} \xrightarrow[h]{11} b(12))),$$
$$\{c_2(3)\} \xrightarrow[c_2]{6} c_1(9)$$

Figure 1: A neural net, a cascade and two causal tracks expressions.

Definition 1 (Causal Track Expressions) *The track expression* $x_a(t) = a(t)$ *is* causal, *if* $f(a,t) = 1$*;*
The composite track expression

$$x_b(t) = \{x_{a_1}(t_1), \ldots x_{a_n}(t_n)\} \xrightarrow[b]{t} x_{a_{n+1}}(t_{n+1})$$

is causal, if $f(a_i, t_i) = 1, i = 1, \ldots n+1$ *and* $f(b,t) = 1$*, and the firing of* $a_1, \ldots a_n$ *suffice according to the firing law for the activation of* b *as well as for all neurons on the paths from* $a_1, \ldots a_n$ *to* b *and from* b *to* a_{n+1} *at the times given in the expression.*

Note that the same neuron may occur repeatedly in a track expression, reflecting the fact that activations may be cyclic.
Fig.1 shows a tiny example of a neural net in which we may observe the cascade figured below it. In this cascade two causal tracks can be extracted, one of them the cycle (c_1, c_2, c_3). The corresponding track expressions are supplied, using the time indices from the given scale. Neuron e is the key neuron of the first expression. – One should realize that nets, cascades (and track expressions) of size larger by several orders of magnitudes should be envisioned.
As special cases we admit initial and terminal track expressions with empty antecedents, $\emptyset \xrightarrow[b]{t} x_{a_{n+1}}(t')$, respectively empty consequents $\{\{x_{a_1}(t_1)\}, \ldots \{x_{a_n}(t_n)\}\} \xrightarrow[b]{t} \perp$, where "$\perp$" stands for the missing consequent.

Each composite causal track expression is divided by its key neuron into argument track expressions and a value track expression, representing the causality. (What is different from the classical function case is that these arguments and values could again be, necessarily causal, track expressions; this is a central aspect of the model.)

Definition 2 (Firing Patterns) *Any set of track expressions which are causal with respect to the firing law is called a firing pattern.*

3 Neural Algebras

Firing patters are the basic objects of our theory. They are eventually meant to embody brain functions.

The challenge is to identify those firing patterns on which a reasonable *theory of the mind* may be based. These brain functions, firing patterns, are quite complex infinite sets, a fact to which we have become quite oblivious in the case of analysis. The development of analysis has singled out its realm by adding additional structure: We can add, multiply, differentiate and integrate functions. With this in mind, our goal is to develop a corresponding operability with firing patters.

COMPOSITION

Firing patterns are related by acting on each other as determined by the structure of the net and the firing function. We untangle these interactions by basing them on the concept of *applying* a firing pattern to another. Recall that in each causal track expression the part to the left of the main arrow represents the cascades that prompt the key neuron to fire. The cascade denoted by the expression on the right denotes what new firings this firing produces. The same is true for sets of causal track expressions, i.e. for firing patterns.

This observation motivates the following definition of composition of such sets:

A firing pattern M *composed* with a firing pattern N applies the causation, represented by causal track expressions in M, on N as follows:

$$\begin{aligned} M \cdot N = \{\; & x_{n+1}(t_{n+1}) : && \textit{there exists}\; \{x_1(t_1), \ldots, x_n(t_n)\} \overset{t}{\underset{b}{\rightarrow}} x(t_{n+1}) \\ & \textit{in } M && \textit{such that}\; \{x_1(t_1), \ldots, x_n(t_n)\} \subseteq N\}\,. \end{aligned}$$

Definition 3 (Neural Algebras) *Given $\mathcal{A} = (A, f)$, a firing law and the operation of composition, the set* B *of subsets of the set of causal firing tracks, closed under this operation, defines an algebraic structure, the* neural algebra $\mathcal{N}_{\mathcal{A}} = \langle \mathbf{B}, \cdot \rangle$.

Neural algebras $\mathcal{N}_{\mathcal{A}}$ are our candidates for an algebra of thoughts as envisioned in the introduction. The elements of a neural algebra are sets of track expressions describing activities of neural assemblies, ”thoughts” among them. The algebraic operations correspond to basic acts of thinking and equations serve to describe the interoperation of thoughts, concepts, memories and scripts for actions.
Remark: $\mathcal{N}_{\mathcal{A}}$ is essentially a submodel of a Plotkin-Scott-Engeler graph model of the untyped λ-calculus, (see [13, 14, 15]).

4 Episodes of Mental Activities

The neurological hypothesis posits that all mental activities are embodied in the brain as patterns of firing neurons. Such patterns typically involve a great number of neurons, linked over considerable distances and active for considerable time relative to the time scale of the individual neuron. Indeed, any mental concept and activity is episodic in character, in particular in the way in which it is activated and used.

It may be argued that in reality the brain does not work on a time scale from minus to plus infinity, that is $\mathbb{Z}$, but during a finite lifetime. In the same vein, a set of track expressions R makes only sense as a mental activity if its firing is *sustained* for a time interval $[t_0, t_1]$. Such a sustention is called appreciable if $t_1 - t_0 > \nu$ for some arbitrarily fixed number ν, say 10^5. Given a time interval $[t_0, t_1]$ and track expressions $x_{a_1}(t), \dots x_{a_n}(t)$ we denote by $R = \{x_{a_1}(t), \dots x_{a_n}(t)\}_{t_0}^{t_1}$ firing pattern of their sustained firing during that interval. This means that the set of firing times of the key neurons of R cover the given time interval, the *sustaining interval* of R. The composition of sustained firing patterns may not be sustained.

Notation: If $x(t)$ is a track expression, then $x(t')$ is the result of substituting t' for t everywhere in $x(t)$, including of course all instances of the dependent firing times, modified according to their place in the track expression.

Two sustained firing patterns X and Y are *approximate* if their sustaining intervals overlap for an appreciable subinterval. We write $X \approx Y$ in this case.

There are two ways by which we are able to realize sustention of a firing pattern in our brain brain model: We may assume a source of signals at one or more neurons which continue to activate the input neurons of R during a given time interval, conceivably by biochemical messaging. Or we may have an autonomous sustention in form of one or more causal cycles in the connectome. Such a cycle could be the base of a cyclically repeated firing track as in some of the connectomes that we introduce later. In fact, external sustention may be mimicked by cycling a source track expression upon itself.

Familiar mental activities are typically based on sustained firing patterns, having an *episodic character*, and can be described as thoughts, scripts or as memories: *Thoughts* are general mental activities, conscious or not, *Scripts* act situationally and are templates for procedures, projects, processes, etc., *Memories* are invoked by triggers and store auditory and visual perceptions, thoughts, emotions, etc.

Any sustained firing pattern is called an episode. Instead of ”episode” we may use some other suggestive term, e.g. ”reaction pattern”, ”perception”, or more generally ”distinct brain function”, depending on the kind of objects we wish to identify in applications of the model.

Even for small brains the set of firing patterns is enormous and seems to defy structuring. Neuroscience is concerned with identifying the embodiments of specified functions of the brain. Such functions are distinguished by giving them well accepted names if they have proved to be stable and express the gist of the matter.

Fleeing upon being threatened, (Fig.2), may serve here as a simple example of a script; it is based on the embodiment of an instinctive reaction pattern, consisting of the mental episodes of perception T of threat, the perception D of danger, L of the lack of cover, and F the reaction of flight. For example:

$$T = \{\{u_1(t'), v(t'')\} \xrightarrow[a_1]{t} s(t'''), \{u_2(t')\} \xrightarrow[a_2]{t} s(t''')\}_{t_0}^{t_1}$$

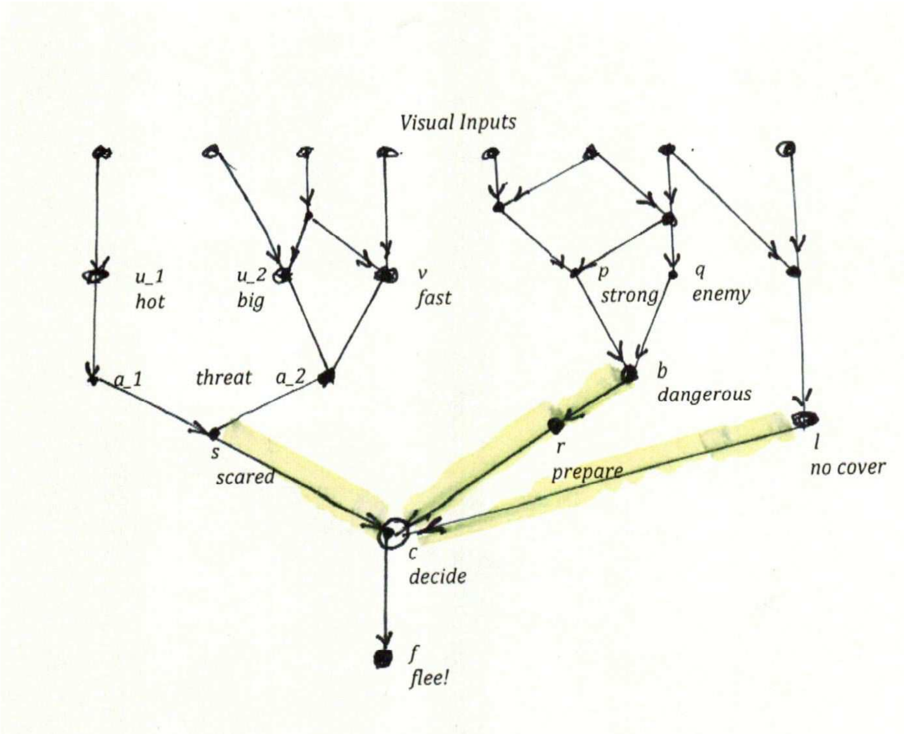


Figure 2: Fleeing on a Threat.

It is composed of the track expressions u_1 ”it is hot”, u_2 ”it is big”, v ”it moves fast towards me”, with the key neurons a_1, a_2 naming the threat, and s ”I’m scared”. Correspondingly for D and L with key neurons s, resp.l. The script of this instinct, is simply $F = \{r(t'), s(t''), l(t''')\} \underset{c}{\overset{t}{\rightarrow}} f(t'''')\}_{t_0}^{t_1}$. Figure 2 highlights the cascade which correspond to the track expression of F, the reaction upon a threat.

5 Perception, Concepts and Control

The above example of flight-upon-threat is, of course, more conceptional than realistic. There are two considerations:

First, it is not plausible that a real brain has individual neurons that completely and exclusively correspond to terms such as ”big”, ”fast approaching” and their combinations into ”threat” and ”danger”. There would simply have to be too many neurons in a realistic brain to embody all the concepts necessary for its functioning in the world. Indeed, it is reasonable to embody these concepts by larger functional objects in the brain, by connectomes that support the brain function corresponding to the perception of these individual concepts. Then we have am exponentially larger choice of representations; concepts are represented by elements of $\mathcal{N}_{\mathcal{A}}$.– Instead of ”fleshing out” the above example, we shall turn to the problem of what it means abstractly for an element of $\mathcal{N}_{\mathcal{A}}$ to perceive a concept. Observe that perceptions are commonly expressed as predicates, something being called ”big” or ”threatening”, etc. Thus the problem comes down to characterize perceptions and concepts as types of objects in the neural algebra. A solution to this problem is proposed in the following section.

Second, *control*. In the example, the control of the reflexive behaviour of ”flight” is embodied in a single neuron. This also is far from being neurologically plausible. Again, we do not propose here to ”flesh out” the neural control mechanism for this example but concentrate on an analysis on how the abstract notion of control can be represented in a neural algebra by one or more objects of $\mathcal{N}_{\mathcal{A}}$, followed by an attempt of some structuring of the notion of neurological control.

To understand how an element A of $\mathcal{N}_{\mathcal{A}}$ could be said to have control over firing patterns, consider the connectome of Fig.3 on which the controlling object A is assumed to be based. The key neurons of A, the ”controlling neurons”, are emphasized in black. Control is effected by applying A to an input B resulting in the controlled output C.

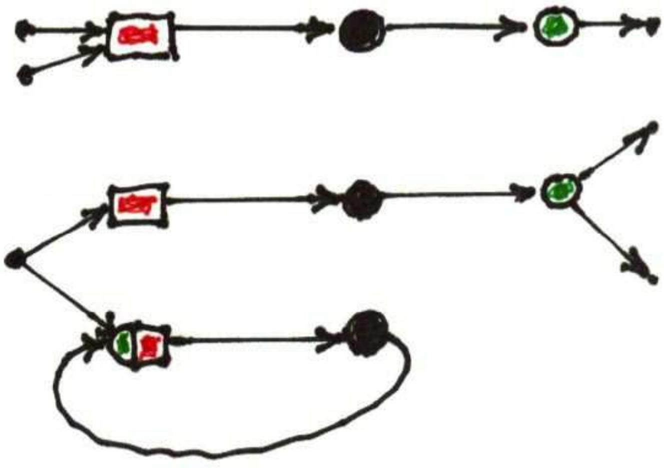


Figure 3: Connectome of Feedback Control.

$$A \cdot B = C$$

Looking at the connectome, it is clear that both B and C are based on connectomes that are parts of the connectome of A. The key neurons of B and C are colored red and green respectively. Observe that some of these key neurons are both red and green: control may include feedback of course.

In a sense, control with more than one controlling neuron may be considered *distributed control*. This could be expressed by splitting A into $A = A' \cup A'' \cup A''$ according to the different controlling neurons.

coupled control arises if two controlling objects A_1 and A_2 act in a coupled manner, e.g.

$$A_1 \cdot B = C, A_2 \cdot C = B;$$

a situation encountered in neurology (e.g. hand-eye movement) just as in mechanics (e.g. coupled pendulum).

The two above forms of distributed control are represented by first-order equations. Higher-order equations represent *hierarchical control*: the controlling object A_1 is itself controlled by a separate controlling object A_2. For example

$$A_1 = A_2 \cdot B, A_1 \cdot C = D$$

is an example of second-order control. Again, neurological examples are easily imagined.

On a still higher conceptual level of control is what could be called *reflexive control* to which we return in a later section.

Altogether, this shows the richness of forms of control and, conversely, the challenge to recognize, localize and describe natural control mechanisms in the brain, concretely in the brain model $\mathcal{N}_{\mathcal{A}}$. The result may perhaps be of help in structuring the exploration and representation of structure/function relations in neuroscience.

Control of activation in a neural net may be seen as residing in specific neurons, control neurons. These may be distinguished and obtained by a mathematical analysis of the connectome, (e.g. [30]), an approach that may nicely complement the present one.

6 Predicates, Concepts and their Structure

To arrive at the notion of concept, we look at a special type of firing patterns, corresponding to the mental action of predication and go from there to look for the kind of patterns that may be identified as concepts and thus have chance to be relevant for a theory of thoughts in the brain.

PREDICATION

Predicating about the snow that it is white means to apply the thought [whiteness] of being white to the thought of the snow. In $\mathcal{N}_{\mathcal{A}}$ this is represented as [whiteness] · [snow]. Generally, the composed thought $R \cdot X$ denotes the extent to which the predication R applies to the thought X; here: to what extent does [whiteness] apply to [snow].

Language also aims to describe composition of thoughts, such as predicating about a predication or about the result of a predication, e.g. by qualifying it. There may be confusion: To say that D is a philosopher king can be understood as expressing quite different thoughts. Is D a king who is also a philosopher, a philosopher who is also a king, or, remembering Plato, is he an example of the best way to govern the state ? By specifying the mode of applications of thoughts to thoughts this translates into: Is it [philosopher] · ([king] · D), or [king] · ([philosopher] · D), or ([philosopher] · [king]) · D, or ([king] · [philosopher]) · D ?

Indeed, many of the conundrums of communication, in fact many of the traditional sophisms, are based on language lacking (or not using, or abusing) precision in expressing the exact structure of the application of thoughts to thoughts. In particular, thoughts are not associative.

Another type of qualification is context. In the context of politics [blueness] $\cdot R$ for a person R calls R a democrat, while in a medical context it expresses an emergency. This underlines the insight that [blueness] as any other predication, is extensive, persistent and broadly applicable.

If a predication is to be conceptually relevant (and a stable component of the activities of the brain), the main requirement is that it should be general, or abstract, enough not to depend on accidental, extraneous, conditions on the objects to which it is to be applied. This corresponds to the traditional notion of a concept. Since Aristotle, concepts are arrived at by abstraction: by taking a thought and eliminating all extraneous elements, the *accidentia*, its accidental or irrelevant aspects.

We base abstract concepts in $\mathcal{N}_{\mathcal{A}}$ on corresponding predications, considered as abstraction operations: If R is a concept applied to a thought X which belongs to the conceptual field of R, then $R \cdot X$ removes from X all aspects that are irrelevant with respect to the predication R. Thus, if applying R again returns the same result, this is the pure abstract, the R -conceptual content of X.

Accordingly, we define:

Definition 4 (Concepts) *A predication R is a concept if all sustained inputs X for which both $R \cdot X$ and $R \cdot (R \cdot X)$ are sustained satisfy $R \cdot (R \cdot X) \approx R \cdot X$.*

A given brain model $\mathcal{N}_{\mathcal{A}}$ may or may not admit firing patterns that are conceptual. The immediate question is therefore: What are the connectome structures corresponding to concepts ?

An important aspect of the usual notion of ”concept” is the fact that it can be called by a name. This aspect is realized in our model by choosing a characteristic neuron r for a given predication P which we wish to identify as a concept. This r *names* the concept.

To simplify notation, let lower case greek letters denote finite sets of causal track expressions, the involved time instances for its members are tacitly understood.

Theorem 1 (Connectomes of Concepts) *Every concept* R*, named by* r*, can be presented in the form*
$\{\alpha_i \underset{r}{\overset{t}{\rightarrow}} x_{a_i}(t_i) : x_{a_i}(t_i) \in \alpha_i \subseteq \{x_{a_j}(t_j) : j \in I\}, i \in I\}_{t_0}^{t_1}$,
and realized by a connectome centered at one common neuron r *with all paths returning to it.*

Proof: Let $R = \{\alpha_i \underset{r}{\overset{t}{\rightarrow}} x_{a_i}(t_i) : i \in I\}_{t_0}^{t_1}$ be a concept, $R \in \mathbf{B}$. From the defining equation $R\cdot(R\cdot X) \approx R\cdot X$ it follows at once that R maps $R\cdot X$ onto itself for any sustained X. Therefore $M = R\cdot X \subseteq \{x_{a_i}(t) : i \in I\}_{t_0}^{t_1}$. Note that $\alpha_i \subseteq R \cdot X$ for all sustained X: Assume $\alpha_i = \{x_a(t), y_b(t)\}_{t_0}^{t_1}$, with $x_a(t) \in M$ and $y_b(t) \notin M$. Then $R \cdot \{x_a(t), y_b(t)\}_{t_0}^{t_1} = \{x_a(t)\}_{t_0'}^{t_1'}$ and $R \cdot (R \cdot \{x_a(t), y_b(t)\}_{t_0}^{t_1} = \emptyset$. Hence α_i must be a subset of M.

Conversely, if S is of the above form, then it is obviously a retraction and therefore a concept.

Fig.4 shows the schematic view of the very rudimentary concept R; the constituting tracks are highlighted:

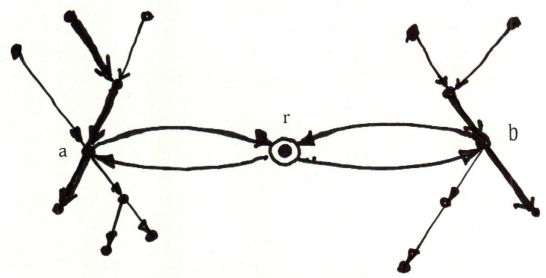


Figure 4: Connectome of a Concept.

$$R = \{\{x_a(t'), x_b(t'')\} \xrightarrow[r]{t} x_a(t''), \{x_b\} \xrightarrow[r]{t} x_b\}_{t_0}^{t_1}.$$

CONCEPTUALIZATION

To turn a predicate P into a concept R means to specify those aspects of P which constitute its abstract, non-accidental, components. Let $P = \{\alpha_i \xrightarrow[s_i]{t} x_{a_i}(t') : i \in I\}_{t_0}^{t_1}$ and choose a subset $S \subseteq \{x_{a_i}(t) : i \in I\}_{t_0}^{t_1}$ of the right-hand side of P and a key neuron r. Then we may set $R = \{\{x_{a_i}(t')\} \xrightarrow[r]{t} x_{a_i}(t'') : x_{a_i} \in S\}_{t_0}^{t_1}$, which is obviously a concept. If S contains all of the second elements of P the we would call R the full *abstract concept* of P, otherwise R constitutes some indications or marks about the perception P chosen by specifying S.

If, instead of choosing from the second elements of P to construct a concept, we may use some of its antecedents α_i to form $A \subseteq \{\alpha_i \xrightarrow[s_i]{t} \perp : i \in I\}_{t_0}^{t_1}$. By setting $R = \{\{\alpha_i \xrightarrow[s_i]{} \perp\} \xrightarrow[r]{t} (\alpha_i \xrightarrow[s_i]{} \perp) : (\alpha_i \xrightarrow[s_i]{} \perp) \in A\}$, we obtain a conceptual object which we could call a *pre-concept*. It embodies, depending on the choice of A, the ”causes” ,”traits” or ”clues” that contribute to the predication P.

Concepts themselves do not constitute a subalgebra of the neural algebra: Consider two concepts R and S and compose $R \cdot S = \{x : \exists(\alpha \underset{r}{\rightarrow} x) \in R, \alpha \subseteq S\}$. Observe $x \in \alpha$ by theorem 1, hence $x \in S$. Therefore $R \cdot S \subseteq S$ and $x = \beta \underset{s}{\rightarrow} y$ in $R \cdot S$ for some β and y. According to theorem 1 we would need $\beta \subseteq R \cdot S$ which is not guaranteed (as it would be if β were a singleton, as is the case for the composition of abstract concepts and pre-concepts.)

It is straightforward to construct concepts that act on given concepts (that is second order and higher order concepts), such as other peoples concepts; planning reaction patterns and the like.
A simple example of a second order concept is the concept R of the causal concatenation of two concepts S_1 and S_2 : "upon S_1 follows S_2 ", e.g. one script follows another.
The concatenation S of these concepts is established by a neuron s which links the reference neurons a and b of these two concepts:

$$S = \{\{x_a(t')\} \overset{t}{\underset{s}{\rightarrow}} x_b(t''), x_a \in S_1, x_b \in S_2\}_{t_0}^{t_1}.$$

for some time interval $[t_0, t_1]$ of sustension. We might call a and b the reference or conceptual neurons of order one, s of order two.

7 Self-Reference and Reflexive Control in $\mathcal{N}_\mathcal{A}$

Fixpoints in $\mathcal{N}_\mathcal{A}$

Among the equations that are important for the understanding of the interactions of brain activities, fixpoint equations, as so often in key places in mathematics, play an important role. Consider arbitrary elements of the neural algebra $\mathcal{N}_\mathcal{A}$ and a variable X, and combine them by the operations of composition and of union into an expression $\varphi(X)$. Then $\varphi(X) = X$ is a fixpoint equation.

Theorem 2 (Fixpoint Theorem) *In $\mathcal{N}_\mathcal{A}$ all fixpoint equations have a approximate solution; the solutions form a lattice by inclusion. If $\varphi(X_0) \supseteq X_0$ then there is a solution which includes X_0 .*

Proof : Because composition and union are monotonic operation with respect to set inclusion, $X' \supseteq X$ implies $\varphi(X') \supseteq \varphi(X)$. Also, if D is a set of sustained elements of $\mathcal{N}_\mathcal{A}$, directed by inclusion, then

$$\varphi(\bigcup D) = \bigcup_{X \in D} \varphi(X).$$

From this follows, that the fixpoint equation $\varphi(X) = X$ has a least solution

$$\bigcup_n \varphi^n(\emptyset),$$

where $\varphi^0(X) = X$ and $\varphi^{n+1}(X) = \varphi(\varphi^n(X))$:

$$\varphi(\bigcup_n \varphi^n(X_0)) \approx \bigcup_{n+1} \varphi^n(X_0) = \bigcup_n \varphi^n(X_0).$$

In the same way, if $\varphi(X_0) \supseteq X_0$, then

$$\bigcup_n \varphi^n(X_0)$$

is the least fixpoint including X_0.

SELF-REFERENTIAL CONCEPTS

Many of the famous sophisms may be considered as being based on self-reference. This is not the place to dwell on the history of paradoxes and sophisms linked to self-reference.

Let me mention just one, theorem 66 in Dedekind's important and famous "Was sind und was sollen die Zahlen", where he proved the existence of the infinity of natural numbers by considering "the totality of objects of thought" as follows: The thought that I am thinking is itself an object of thought. Taking any thought object N_0, e.g. the thought "I am thinking of a number". Reflecting i the thought of this object is again an object of thought, etc, yielding an infinity of objects of thought. If R is the mental operation of reflecting on a thought, the Dedekind's construction step is $N_{i+1} = N_i \cup R \cdot N_i, i = 0, 1, \ldots$. Collecting up yields $N = \bigcup_i N_i$ which solves the recursion $N = N \cdot N$ with initial condition N_0. [1] Realize, however, that the object Z, because of the restricted sustensions, only approximates the desired object. The infinity of the natural number, while non-contradictory, is here something like an illusion. But convincing to the finite brain.

Compare this with the familiar conundrum of "the painting that shows the painter executing this painting".This work of art, even if it existed somewhere, can only be approximately thought of.

Reflexive Control in $\mathcal{N}_{\mathcal{A}}$

The notion of reflexive control is derived here by an analysis of what is experienced as consciousness. Whether the result is indeed a sort of artificial consciousness operating in the model brain is of course debatable.

[1]Dedekind's conception was criticized by the co-editor Emmy Noether of his collected works (Braunschweig, 1932) as being based on the contradictory notion of the totality of thought objects, see p.391.

For the purposes of this analysis let us understand *consciousness as the ability of a " brain" to consciously observe itself as being conscious and as consciously planning and acting*. This definition, at first sight, appears circular. Interpreted in $\mathcal{N}_{\mathcal{A}}$ It is simply self-referential:

The model $\mathcal{N}_{\mathcal{A}} = \langle \mathbf{B}, \cdot \rangle$ comprises firing patters corresponding to observing, acting, planning, moving, etc. Let C be the prospective firing pattern of "consciousness". Let the *brain* B of the model be the set of all its causal firing tracks. Then $B \cdot C$ is the result of observing, acting, etc. as dependent on consciousness, and $C \cdot B$ represents the action of consciousness on such activities. To these objects, including C itself, consciousness C is again applied; as in: "observing itself ..." above, i.e. reflexively.

This characterization of consciousness transforms into an equational definition as follows:

Definition 5 (Reflexive Control) *A sustained set C of track expressions in the brain B represents a reflexive control mechanism if it satisfies the equation*

$$C \cdot C \cup C \cdot (B \cdot C) \cup C \cdot (C \cdot B) \approx C.$$

The question arises how to characterize firing patterns and their connectional correlates corresponding to solutions of the above equation. For this we need the notion of a *causal cycle*. This is a *sustained* sequence $\{x_{c_0}(t_0), x_{c_1}(t_1), \ldots x_{c_{n-1}}(t_{n-1})\}_{t'}^{t''}$ of causal track expressions of the form $\alpha_i \xrightarrow[c_i]{t_i} x_{c_{i+1}}$ with $x_{c_{i-1}} \in \alpha_i$ for $i = 0, 1, \ldots n-1$, where the indices are of course understood modulo n.

Theorem 3 (Structure Theorem) *A neural algebra admits nontrivial reflexive control if and only if it contains at least one causal cycle.*

Proof:

By the fixpoint theorem, the reflexive control equation has a set of solutions in $\mathcal{N}_{\mathcal{A}}$ forming a lattice; certainly the empty set $\emptyset$ is a solution. Assume now that there is a nonempty sustained solution C and consider track expressions $x_c(t) = \alpha \underset{c}{\overset{t}{\rightarrow}} y(t')$ in C. By the equation, C being a left factor, $y(t')$ is an element of C and is therefore also of this form. For $x = \alpha \underset{c}{\overset{t}{\rightarrow}} y$ define the *input structure* $\sigma(x)$ as the set consisting of the key neuron of x and the key neurons of the elements of α. Then C contains a nonfinite sequence $x_{c_0}(t_0), x_{c_1}(t_1), \ldots$ of causal track expressions of the form $\alpha_i \underset{c_i}{\overset{t_i}{\rightarrow}} x_{c_{i+1}}$ with $x_{c_{i-1}} \in \alpha_i$ for $i = 0, 1, \ldots$. The corresponding sequence of input structures is eventually cyclic, (the model being finite). By disregarding a non-cyclic initial segment of this sequence, we assume that it repeats after $x_{c_{n-1}}(t_{n-1})$. Thus, C contains a causal cycle.

Conversely, assume that $C_0 = \{\{x_{c_0}(t_0), x_{c_1}(t_1), \ldots x_{c_{n-1}}(t_{n-1})\}_{t'}^{t''}$ is a causal cycle. By recursion construct

$$C_{j+1} = C_j \cup \{\alpha_i \underset{c_i}{\overset{t_i}{\rightarrow}} x_{c_{i+1}}(t_{i+1}, i = 0, 1, \ldots i-1 \mod n,$$

$$x_{c_{i-1}} \in \alpha, x_{c_{i-1}}, x_{c_i}, x_{c_{i+1}} \in C_j\},$$

resulting in

$$C = \bigcup_j C_j.$$

From the structure of C we conclude

$$C \cdot C \approx C, B \cdot C \subseteq B, C \cdot B \subseteq C, C \cdot (B \cdot C) \subseteq C,$$

and therefore C is a nontrivial solution of the consciousness equation

$$C \cdot C \cup C \cdot (B \cdot C) \cup C \cdot (C \cdot B) \approx C,$$

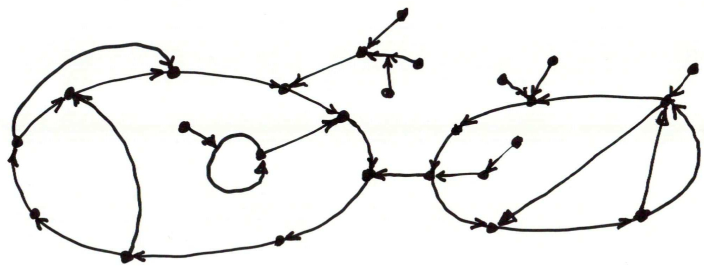


Figure 5: Connectome for States of Reflexive Control.

based on C_0. Correspondingly, any set of causal cycles generates such a solution.– Note that each causal cycle represents a concept since $C \cdot (C \cdot X) \approx C$ for all X by the above.

Fig. 5 shows a very rudimentary scheme of reflexive control; activation, triggered by some of the links shown, may migrate from one of the possible cycles to another. Again, one should envision connectomes larger by several orders of magnitude.

If we allow ourselves to speak here loosely of ”consciousness”, the above equations could be called the ”consciousness equation”. The lattice structure of the set of solutions then reflects the phases or states of consciousness, and their contextual movement depends on the inclusion/exclusion of the various cycles, concepts, available from present states. In other words: consciousness expands/contracts by attaching/releasing connections to perceptions, memories etc. according to the firing history, constituting what one might reasonably call the *neural mind*. The sustained thoughts $\{x_i\}_{t'}^{t''}$ constituting C are the ”content” of consciousness. Only if their sustention $[t', t'']$ exceeds the length of the cycle considerably, consciousness is ”aware” of them.

Example: In Fig.2 (Fleeing on a Threat), the cascade in F leading to the flight decision (neuron c) is a possible content of the consciousness C of the individual, assuming that the neuron c is on the cycle. The object $F \cdot C$ is the conscious reaction of flight.

7 Thinking about Episodes in a Scenarios

EXPLANATIONS

A scenario is what happens in the brain. It consists of a collection of thoughts, memories, ongoing experiences and activities, states of consciousness, etc., in short of objects in $\mathcal{N}_{\mathcal{A}}$. To explain how such an object is present in the history of this brain, we rely on declarations $\models P$ of acceptance of P and on insights about causality: P directly explains Q if $Q \subseteq P$, a fact denoted by $P \models Q$.

An *explanation* is a tree-like structure, reminiscent of formal proofs in logic. Starting with some accepted declarations such as $\models P$, we use rules with which to proceed, such as: from $P \models Q$ and $\models P$ conclude $\models Q$, and $P \cdot R \models Q \cdot R$, etc.

LOGICAL THOUGHTS

Explanations are of essence when dealing with composite objects, composite thoughts. By dealing with operations on ”thoughts” we are invading the realm of logic. In the following we only deal with a small aspect, namely propositional logic.

Since the predicate $P \cdot X$ describe the extent to which the predication P applies to the ”thought” X, propositions — which are either true or false — would be predicates with values $P \cdot X = X$ (for ”true”), respectively

$P \cdot X = \emptyset$ (for ”completely false”) . Logical combinations of thoughts P and Q such as conjunction $P\&Q$ and disjunction $P \vee Q$ may be realized by convenient combinations of the connectomes that realize P and Q.
Negation of $\bar{P}$ of P means to detail the grounds for rejection of P by listing choices of rejecting specific track expressions in P and introducing ”opposing” track expressions. Thus, negation relates predications and their possible opposites. We may also construct contradictory thoughts such as $P\&\bar{P}$. Far from exploding thoughts X (in the sense of ”ex contradictio quodlibet”), $(P\&\bar{P}) \cdot X$ resolves the contradiction by rescuing the common aspects of the predications P and $\bar{P}$, reminiscent of dialectics. This embedding of logic into the study of predications would be an interesting (unwritten) chapter.

To experiment with the above notion of explanation, the reader may be amused by following the logic of the superior brain of Hercule Poirot in analyzing the closing scenario of a detective story, e.g. in Agatha Christie’s ”Five Little Pigs”. Considering objects that represent actions, motivations, observations and conclusions by diverse people, and thoughts of people about other people’s thoughts, Poirot explains why Elsa must be the murderess.
Neuroscientists, rather than in fiction, are interested in scenarios happening in the brain, in specified functions and structures. The discipline of expressing these notions as objects in a neural algebra, and investigating them using equations among the, would be illuminating, I expect.

PROBABLE THOUGHTS

Starting from his work on pattern recognition, U.Grenander [25] investigates the association of probabilities to patterns of neural activations, i.e. the probability of one pattern to relate to another. It would be interesting to identify Grenander's patterns with elements of a neural algebra. As observed above, $\mathcal{N}_\mathcal{A}$ is essentially a submodel of a Plotkin-Scott-Engeler graph model of the untyped λ-calculus. Hence the recent proposal by Dana Scott (personal communication) of stochastic λ-calculi would lend itself to a promising refinement of this approach.

9 Apologia

Retired people, emeriti, love to travel to new places. I do. The friendly natives take pleasure to show the impressive sights and amusing curiosities of the country and listen with tolerance to the strange, mathematical, accents of their visitor when he describes his perception of what they show him. In telling this story as a piece of mathematics, I could have more soberly, and perhaps more wisely, have chosen a more neutral terminology: ”reactive nets” for ”brain”, ”net-functions” for ”thoughts”, ”retractive functions” for ”concepts” and ”auto-reflecting” functions for ”consciousness”. But an author does have the liberty of naming dramatis personas. Anyway, I really did think about brains.

10 Discussion

1

”How does the brain work?” Asked by Colbert in his show to answer this in five words or less, Steve Pinker rose to the challenge and quote ”Brain cells fire in patterns”, (1). This encapsulates the direction of research that is on the point of turning into Big Science (2) - (6) by the Allen Foundation, the European Union and NIH and others with impressive funding and many laboratories and collaborators involved. The challenge also goes to mathematics and computer science: How, and to what purpose should we model the brain and its activities?

There is abundant reason to be sceptic, especially when it comes to higher cognitive functions (7), or confronted with perceived irreducibility of the mental to the physical (8), or with the innate complexity of models faced with enormous growth of observational and experimental data (9). The discussion is far from closed and interesting new mathematical models are constantly being proposed (10).

This author's contribution to modeling, neural algebra, was first introduced at lecture in 2005 in answer to a challenge by my friend K.Hepp, published in provisional form in (11, 12), based on the author's version of the Plotkin-Scott-Engeler graph model of the λ-calculus (13, 14, 15).

2

Understanding consciousness has been termed ”the most challenging task confronting science”, and what has been a philosophical mainstay has turned into a legitimate question of ”hard science” (16), (17). Motivated by its depth and range, some rich, beautiful and touchingly personal books came to be written by some of the pioneers (18) - (23).

And, not surprisingly, we observe an enormous production of papers on brain and consciousness in neuroscience alone: about six papers per day,

(2101 titles in 2010 according to a citation search.) There have also been some notable attempts at theoretical synthesis, under different viewpoints, proposing mathematical approaches, ranging from dynamical systems (24) to quantum mechanics (25), geometry (26), information theory (27) and statistics (28), and relating them to neurological facts and psychological experiments.

Consciousness may be perceived as an internal mechanism of the brain which seeks a balance between processes that are caused by outside sources and by diverse internal processes, conscious and unconscious. This homeostatic behavior was first described by Wiener (29) as one of the major applications of his *cybernetics*. Our consciousness equation formulates the self-referential character of consciousness, an aspect that has been formulated and investigated throughout the history of the concept, from Descartes' ”cogito ergo sum” to ”I am a Strange Loop” (30).

3

For $\mathcal{N}_\mathcal{A}$ the notions of thought, concept and consciousness are modeled on neural nets whose level of abstraction from the psychophysical brain is relatively modest; it starts with the individual neurons and bases overall organization on these. Much of present neuroscience is in fact concerned with such higher levels of organization (31), (32): One considers neurons, and more generally brain areas, that have been identified as being involved in specific functions, and investigates their connectivities and functional dependencies (33, 34). This is exactly what Neural Algebra attempts to provide a mathematical model for.

The concept of *learning* (35) would merit more than the following few remarks. Our model tacitly subsumes learning in the *firing history* of $\mathcal{A}$: First, the neural net A of the brain model comprises the totality of all neurons that are ever considered. Second, the firing function f permits periods where different subsets of A are involved or dismissed from activity. This leaves out Hebbian learning, effected by changes in the weights of synapses over time, thereby affecting the definition of legality for firing functions. One possible solution is to make the weights of synapses dependent on some of the previous firing history.

4

The consciousness of animals is a much debated concept. A technical approach may conceivably start with the knowledge, obtained laboriously, of the actual neural net of some species. The famous nematode *caenorhabditis elegans* had its complete neural network mapped with all its synapses (36); much additional information has been obtained, approximating total neural modeling (37). In principle, we could eventually ask for the consciousness of that animal. In other words: ”How does it feel to be a worm ?” This remains to be done, and not only for worms. But, judging from a possible lower bound on the number of neurons required for consciousness to be initiated and sustained, c. elegans may not qualify.

Social consciousness, in a technical sense, would consist of understanding individuals (people or ants etc.) as nodes in a (social) net, their interactions as edges in the net and the strength of these interactions as the weights of these edges (38).

Artificial consciousness may be an utopian goal (39), (40), although it has been studied in the context of artificial intelligence, not least in the hope of modeling the perceived advantage of ”conscious” beings over ”mechanistic” robots (41), including swarms of robots (42). Even plants may have a sort of consciousness (43).

More generally, it would appear that the neural algebra approach could contribute to computer science in providing templates for the realization of memory structures and of interacting highly parallel processes. One may speculate about corresponding *future architectures* for interlaced memories and distributed programs.

5

Taking the risk to throw glances over the fence, I find some reassurance for the present model, hoping that others would perhaps share it. They may wish to consider the following instances:

Single neurons have been identified as the key to *recognize a face*, (44). Such ”grandmother cells” appear to act as codes for concepts. This is reflected in our characterization of ”concept” in Theorem 1, by introducing corresponding key neurons. Similarly for *mirror neurons*, called upon when a concept, e.g. a feeling, needs to be associated to a concept pertinent to it, (45).

Recruiting new neurons and synapses to create new abilities has been identified, and shown to be involved in the learning of bird songs (46) – (51), and in reading (52). Again, like in the formation of concepts, in our model this corresponds to the introduction of key neurons.

Our characterization of ”consciousness” as based on linked cycles of partial consciousness and concepts, (Theorem 3), also has some parallels: Recurrent or reentrant connectivities in the brain have been recognized to be involved in conscious activities, e.g. in the visual cortex (53), and more generally in linked circuits (54) and so-called convergence–divergence–zones and regions, (55) and (21, chapt.6).

6

The present author, fascinated by the challenges of neuroscience, is greatly intimidated by the enormous literature of which he encountered the following.

References

[1] Guehenno, C.M., Pinker's brain picked on ”Colbert Report. The Harvard Crimson, February 8, 2007.

[2] Koch, C. and Reid, R.C., Observatories of the Mind, Nature 483 (2012), 397- 398.

[3] Makram, H., The blue brain project. Nature Reviews Neuroscience, 7 (2006), 153 - 160.

[4] Miller, G., Blue brain founder responds to critics, clarifies goals. Science 334 (2011), 748-749.

[5] Alvisato, A.P. et al., The brain activity map. Science 339 (2013), 1284 - 1285.

[6] Underwood, E., Brain projects draws presidential interest, but mixed reactions. Science 339 (2013), 1022 - 1023.

[7] Hepp, K., Reading in the brain. preprint 2012, p.13.

[8] Nagel, Th., Mind and Cosmos, Oxford 2012. Review by Musholt, K., Science 339 (2013), 1277.

[9] Koch, C., Modular biological complexity. Science 337 (2012), 531-532.

[10] Smolensky, P., Symbolic functions from neural computation. Phil.Trans.R.Soc.A ,370 (2012), 3547-3569.

[11] Engeler, E., Neural Algebra and Consciousness. Algebraic Biology, Lecture Notes in Computer Science 5147, 96–109 (2008)

[12] Engeler, E., Algebras of the brain and algebras of the mind. Abstracts, 14th Congress of Logic, Methodology and Philosophy of Science. Nancy 2011, p.204.

[13] Scott ,D.S., Data Types as Lattices. SIAM Journal of Computing, 5 (1976), 522 - 587.

[14] Engeler, E., Algebras and Combinators. Algebra Universalis, 13 (1981), 389 - 392.

[15] Plotkin, G.D., Set-theoretical and other elementary models of the λ-calculus. Theoretical Computer Science, 121, (1993), 351 -409-

[16] Crick, F.,The Astonishing Hypothesis. Simon & Schuster, London (1990, 1994)

[17] Crick, F., Koch,C., Towards a neurobiological theory of consciousness. Seminars in Neuroscience 2, 263 – 275 (1990)

[18] Penrose, R., The Emperor’s New Mind. Oxford 1989.

[19] Churchland, P.S., Sejnowski, T.J., The Computational Brain. MIT Press, Cambridge, Mass. (1992)

[20] Valiant, L.G., Circuits of the Mind. Oxford University Press, New York (1994)

[21] Damasio, A., Self comes to Mind; Constructing the Conscious Brain. Vintage Books London (2012)

[22] Koch, C., Consciousness; Confessions of a Romantic Reductionist. MIT Press, Cambridge, Mass. (2012)

[23] Tononi, G., Phi, A Voyage from the Brain to the Soul. New York 2012.

[24] Izhikevich, E.M.:,Dynamical Systems in Neuroscience. MIT Press, Cambridge, Mass. (2007)

[25] Koch, C.,Hepp, K.:Quantum mechanics in the brain. Nature 440,611 – 612(2006)

[26] van Weeden et al.: The geometric structure of the brain fibre pathways. Science 335, 1628–1634 (2012)

[27] Tononi, G.: Consciousness as Integrated Information; a Provisional Manifesto. Biol.Bull. 215, 216 – 242 (2008)

[28] Grenander, U., A Calculus of Ideas: A Mathematical Study of Human Thought. World Scientific, Singapore (2012).

[29] Wiener, N., Cybernetics; or Control and Communication in the Animal and the Machine. MIT Press, Cambridge, Mass. (1948)

[30] Hofstadter, D., I am a Strange Loop. Basic books, New York (2007). Reviewed by Martin Gardner, Do loops explain consciousness ? Notices of the Amer.Math.Soc. 54, (2007), 852-854.

[31] Sporns, O., Networks of the Brain. MIT Press, Cambridge,Mass. (2011

[32] Seung, S., Connectomes; how the brain's wiring makes us who we are. Houghton Mifflin Harcourt, New York (2012)

[33] Ruths, J. and Ruths, D., Control Profiles of Complex Networks. Science 343 (2014), 1373 - 1373; plus suppl. material.

[34] Vogelstein, J.T. et al., Discovery of Brainwide Neurol-Bahaviour Maps via Multiscale Unsupervised Structure Learning. Science 344 (2014) 386 - 392; plus supple. material. (Introduced in the same issue by O'Leary and Marder, Mapping Neural Activation onto Behaviour in an Entire Animal.)

[35] Sejnowski, T.: Learning optimal strategies in complex environments. Proc.Natl.Acad.Sci.USA 107, 20151 – 20152 (2010)

[36] White, JB., Southgate, E., Thomson, J.N., Brenner,S., The structure of the nervous system of *Caenorhabditis elegans*. Philos.Trans. R.Soc., London, Biol.Sci. 314, 1 – 340 (1986)

[37] Ardiel, E.L., Rankin, C.H.: An elegant mind; learning and memory in *Caenorhabditis elegans*. Learning & Memory 17, 191 – 2001 (2010)

[38] Malone, Th.W., Collective Intelligence. Edge website, downloaded Dec.3, 2012.

[39] Sejnowski,T.: When will we be able to build a brain like ours? Scientific American, (2010)

[40] Kurzweil, R., How to Create a Mind. New York 2012. Reviewed by C.Koch, The end of the beginning of the brain, Science 339 (2013), 759-760.

[41] McCarthy,J., Making robots conscious of their mental states. In: Muggleton,S. (ed.), Machine Intelligence15, Oxford Univ.Press, Oxford (1996)

[42] More, G., The rise of the swarm. Comm.ACM, 56 (2013), 16-17.

[43] Chamowitz, D., What a Plant Knows. in print 2013.

[44] Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C., Fried, I.: Invariant visual representation by single neurons in the human brain. Nature 435, 1102– 1107 (2005)

[45] Quiroga, R.Q., Fried, I., Koch, C., Brain cells for grandmother. Sci.American, Febr.2013, 31-35.

[46] Rizzolatti, G., Craighero, L., The mirror neuron system. Annual Rev. Neurosci. 27, 168 – 192 (2004)

[47] Meyer, K., Another remembered past. Science 335 (2012), 215-216.

[48] Nottebohm, F., Liu, W., The origins of vocal learning: new sounds, new cuircuits, new cells. Brain & Language 115, 3 – 17 (2010)

[49] Bolhuis, J.J., Gahr, M., Neural mechanisms of birdsong memory. Nature Review Neuroscience 7, 347 – 357 (2006)

[50] Bolhuis, J.J., Okanoya, K., Scharff, C., Twitter evolution; converging mechanisms in birdsong and human speech. Nature Review Neuroscience 11, 747 – 759 (2010)

[51] Hahnloser, R.H.R., Kotowicz, A., Auditory representations and memory in birdsong learning. Current Opinion in Neurobiology 20, 332 – 339 (2010)

[52] Deahaene,St. et al., How learning to read changes the cortical networks for vision and language. Science 330, 1359 – 1364 (2010)

[53] Heinzle, J., Hepp, K., Martin, A.C., A microcircuit model of the frontal eye field. J.Neuroscience 27, 9341 – 9353 (2007)

[54] Singer, W., Consciousness and the binding problem. Annals NY Acad.Sci., 929 (2001), 123-146.

[55] Meyer, K., Damasio, A., Convergence and divergence in a neural architecture for recognition and memory. Trends in Neuroscience 32, 376 – 382 (2009)