Manuscript draft, Dec.31, 2016

Neural Algebra

A MODEL OF INTERACTING BRAIN FUNCTIONS

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 derives from 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 as 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. In particular we analyze the representation of perception and of control in its various forms, distributed, hierarchical, recursive and especially reflexive control, the latter modeling the concept of self-reflecting control.

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.

0 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.

1 Brain Functions

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

We approach this problem by proposing an algebraic system, *Neural Algebra*. The algebraic framework then allows to distinguish types of brain functions by their algebraic properties, equations as it were. Since the algebraic elements are interdefinable with corresponding neural structures, the model supports investigations of specific hypotheses about the interrelation between brain-function and brain-structure.

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 a specific *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. This is of course disregarded in our model as well as the more advanced insights of neurology.– However, the firing law of the model may include distinctions on the type of messages from neuron to neuron such as the strength of the signal, spiking rates etc.; it may also incorporate aspects of learning, e.g. by coding the firing histories of some or all of the connections.

Definition 1 (Brain Model) *A directed graph A together with a firing law and a firing function f conforming to it constitute a brain model $\mathcal{A}$.*

Observing the activation history, we may be able to distinguish episodes of firings of subpopulations of the neurons in the brain: *firing patterns* that specify the firing of some set of neurons a in A during some time interval in $\mathbb{Z}$. 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 finite branched sequence of firings of neurons which causally follow from these original activations as determined by the firing law. Any firing pattern can be viewed as a set of such cascades.

FIRING TRACKS AND BRAIN FUNCTIONS

The overall goal of modeling is to ascribe interpretations to firing patters such as recognizing a shape or initiating an activity, in other words, to view them as *brain functions*. In analogy to the usual set-theoretic definition of functions as sets of pairs, each individual cascade is to be understood as a pair of input-cascade/output-cascade. This is accomplished by *freely choosing* the firing of a specific neuron in the cascade as the key point of causality: in essence the parts of the cascade that are its causal antecedents are understood as inputs; a cascade that follows it is understood as output. Note: By choosing key neurons in a set of cascades a firing pattern obtains a causal interpretation.

This analysis of cascades is formalized below by representing cascades as *track expressions*. Note that the input- and output-cascades should reasonably also be represented by track expressions, thus structuring the whole cascade. This leads to the following recursive definition of track expressions, given a brain model $\mathcal{A}$.

The *basic track expression* $\langle a, t \rangle$ denotes the activation of a single neuron a at an integer time instance t.

Composite track expressions denote the activation of the neurons of a cascade in A. Each track expression has 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 $\langle a, t \rangle$ is a.

Consider antecedent neurons $a_1, \ldots, a_n$ connected to neuron b along paths of one or more directed 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 $x_{a_1}, \ldots x_{a_{n+1}}$ to form $x_b(t)$. 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 the timing of the key neurons, describes a firing of the neurons occurring in it, thus defining a cascade of firings. This leads to:

Definition 2 (Causal Track Expressions) *A track expression $\langle a, t \rangle$ is causal, if $f(a, t) = 1$.*

A 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 of all neurons on the paths and at the times denoted by the track expression.*

Remember that any cascade may be interpreted by different causal track expressions, depending on what successive choices are made of key neurons in the coding. Also note that the same neuron may occur repeatedly in a track expression, reflecting the fact that activations may be cyclic.
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.

Definition 3 (Brain Functions) *Any set of track expressions which are causal with respect to the firing law is called a brain function.*

Brain functions are the basic objects of our theory. Brain functions, firing patterns, are quite complex sets, a fact to which we have become quite oblivious in the case of functions in analysis when functions are defined set-theoretically. 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 brain functions.

2 The Neural Algebra $\mathcal{N}_{\mathcal{A}}$

Brain functions 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 brain function 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, such sets thus constitute functions, brain functions.

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

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

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

Definition 4 (Neural Algebras) *Given $\mathcal{A} = (A, f)$, a firing law and the operation of composition, and a set $\mathbf{B}$ of subsets of the set of causal firing tracks, closed under this operation and union, defines an algebraic structure, the* neural algebra $\mathcal{N}_{\mathcal{A}} = \langle \mathbf{B}, \cdot, \cup \rangle$.

Brain functions are 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 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. A set R of track expressions $x_a(t)$ makes only sense as a brain activity if the firing of its neurons covers an *appreciable* time interval $[t_0, t_1]$, given by the smallest and largest time indices t occurring in expressions $x_a(t)$ in R, for example 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 $x_{a_1}(t), \ldots x_{a_n}(t)$ we denote by $R = \{x_{a_1}(t), \ldots x_{a_n}(t)\}_{t_0}^{t_1}$ the firing pattern of which the set of firing times t 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 $x_a(t)$ is a track expression, then $x_a(t')$ is the result of substituting t' for t everywhere in $x_a(t)$, 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 and they are equal there. 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 an external source of signals at one or more neurons which continue to activate the input neurons of R during a given time interval, thinking of 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.

The concept of *learning* 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. Our model may include Hebbian learning familiar from artificial neural nets, 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.

2 Combinatory Algebra, Neural Algebra, Logic and Language: Historical Note

Neural Algebra has two historical roots, combinatory logic and computability theory.

Combinatory Logic was invented by Haskell B.Curry in his 1929 Göttingen thesis (1), directed by Paul Bernays (who is also the Doktorvater of the present author as well as Sauders MacLane and Gerhard Gentzen) in the context of foundations of mathematics. In the following we present com-

binatory logic by relating it to models of computation.
Computability theory was established by Alan Turing in his famous 1936 paper (2).Since then, i.e. since Universal Turing Machines and the von Neumann Machines, inputs and programs of computers are of the same nature and we may simply call them ”data”. They are perhaps marks on the Turing tape or bits in the computer hardware. The basic operation on data is *application*: Programs may be applied to data and of course result in data, which may again be programs. Programs may also be applied to programs, resulting in data, etc. Indeed, we may admit all combinations of applications on data, resulting always in data (e.g. error messages).

Example: Combining data and programs

x, y	are programs
z	is data
y	yields data
x	yields a program
$(x \cdot z) \cdot (y \cdot z)$	applies the new program to the new data

This combination of x, y, z is regarded as a new program $\mathbf{S}$ with
$\mathbf{S} \cdot x \cdot y \cdot z = (x \cdot z) \cdot (y \cdot z)$. Observe that we used the convention of parenthesizing expressions to the left.
Objects defined in this manner are called *combinators* and the general principle for their introduction is formulated as an
Definition- Scheme: For every combination $\phi(x_1, \ldots x_n)$ of data there is a data t_ϕ such that $t_\phi x_1 \ldots x_n = \phi(x_1, \ldots x_n))$

All such t_ϕ are called "combinators". They arose in the context of simplifying the technical apparatus of symbolic logic in the 1920s. There is one basic result:

Theorem (Schönfinkel (3)):

Two combinators suffice for expressing all combinators, namely the above $\mathbf{S}$, characterized by its equation $\mathbf{S} \cdot x \cdot y \cdot z = (x \cdot z) \cdot (y \cdot z)$,
together with $\mathbf{K}$, characterized by $\mathbf{K}xy = x$.

Combinatory Logic is the formal theory of equations with one binary operation and the axioms $\mathbf{S} \cdot x \cdot y \cdot z = (x \cdot z) \cdot (y \cdot z)$, $\mathbf{K}xy = x$ and $\mathbf{K} \neq \mathbf{S}$. The status of this theory was questionable at the beginning; a version of it had been shown formally inconsistent. A formal consistency proof in the spirit of the Hilbert Program was established by Alonzo Church and Barkeley Rosser (4) using the same finitistic proof-theoretic tools which had failed for number theory. We are therefore enabled to talk consistently about *Combinatory Algebras* as models for the axioms of combinatory logic.

A combinatory algebra $\mathcal{D} = \langle \mathcal{D}, \cdot, \mathbf{K}, \mathbf{S} \rangle$ consist of a set D with a binary application operation $\cdot$ and two distinguished elements of D satisfying the axioms of combinatory logic.

But for forty years the only model was the model consisting of equivalence classes of combinatory expressions (in $\mathbf{K}$ and $\mathbf{S}$) which exists on the basis of formal consistency. Then, in the 1970's, Gordon Plotkin (5) and Dana Scott (6), looking for models of the closely related Lambda Calculus, invented the set theoretic model P_ω and D_ω (vide 7), the latter constructed to be isomorphic to its function space. Thereafter this author introduced the *Graph Model* for combinatory logic (8) whose transparent and explicit structure lends itself to appications in mathematics and modelling (9), in particular in biology (10). Graph models have this pleasing representation theorem:

Theorem: Any binary algebraic structure can be isomorphically embedded in a graph model of combinatory logic.

Using this suggestion, Neural Algebras, as presented above, are constructed as a much enriched type of graph models.

Combinatory logic rarely forms a core subject of courses in mathematical logic; an early example is the 1950 book of my former colleague Paul C.Rosenbloom (11) and my own (12). Better represented in the literature is the formal system of the Lambda Calculus; most work done on these connected subjects is done in the context of lambda calculus and its descendants.

Looking for some orientation, let us now turn to language and thinking because it is obviously that which seems to happen in the brain and about which we believe to have common insights (and a sufficiency of linguistic and philosophical literature). Thoughts, then, may be understood as brain functions and thinking means applying thoughts to thought: If the thought M is applied to the thought N, then the result $M \cdot N$ 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” applied to propositions. 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 as modelled algebraically.

Such a project has venerable historical roots: there was the *ars magna* of Ramon Llull (1232 - 1316) in which individual concepts are represented formally as entries in a graphical representation (13). Such objects could be considered as subjects or predicates (in the grammatical sense) or as logical intermediaries. Their combinations (by manipulating the graphical representation) result in judgements, questions, etc.. Umberto Eco (1932 -2016)) relates Lullus' art to other universalist proposals (14), for example that of Gottfried W. Leibniz (1646 - 1716) who, four centuries later, attempted another project, his *Characteristica Universalis*. This was to be an artificial symbolic language, universal in the sense that all human languages can be represented, even derived from it, and that it offers a technique of formal manipulations, "calculemus". *Remarks:* One such attempt was to map predicates onto natural numbers, the inclusion operation into divisibility and the combination of predicates into multiplication. But predicates are neither commutative nor associative under application; as our model demonstrates. Another attempt goes in the direction of finding universality in existing languages. Here, this author, visiting the Hanover and the Wolfenbüttel libraries, was impressed by Leibniz's appreciation of Chinese culture and in particular by his suggestion that Chinese characters, considered as ideographs, and their textual combinations are close to the idea of a characteristica universalis. Louis Couturat (1868 -1914) is one who attempted to relate Leibniz to modern logic (16).

3 Predicates, Concepts and their Structure in Neural Algebras

Neural Algebra is this author's attempt at relating brain functions to thought and language in a simple model of the brain. This algebraic framework allows to distinguish types of thoughts by their algebraic properties, equations as it were. For example, conceptual thoughts, concepts, will be 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 reduced to solving equations for unknown brain mechanisms.

Our first aim is to capture the notion of concept. To arrive at its formal definition we look at brain functions that correspond 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] $\cdot$ [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. It is tempting to investigate ancient texts on sophisms, for example the famous Johannes Buridan (17) and modern treatises on Rhetorics such as (18) in the light of our approach; but this is another chapter.

CONCEPTS

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 the predication R acts as 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 5 (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 and 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 set $\hat{R}$ of characteristic neurons r for a given predication R which we wish to identify as a concept. This $\hat{R}$ *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 $\hat{R}$, can be presented in the form*
$\{\alpha_i \xrightarrow[r]{t} x_{a_i}(t_i) : x_{a_i}(t_i) \in \alpha_i \subseteq \{x_{a_j}(t_j) : j \in I\}, i \in I, r \in \hat{R}\}_{t_0}^{t_1}$,
and realized by a connectome centered at the common neurons $r \in \hat{R}$ with all paths returning to it.

Proof: Let $R = \{\alpha_i \xrightarrow[r]{t} x_{a_i}(t_i) : i \in I, r \in \hat{R}\}_{t_0}^{t_1}$ be a concept. 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 $\alpha_i = \{x_a(t), y_b(t)\}_{t_0}^{t_1}$,

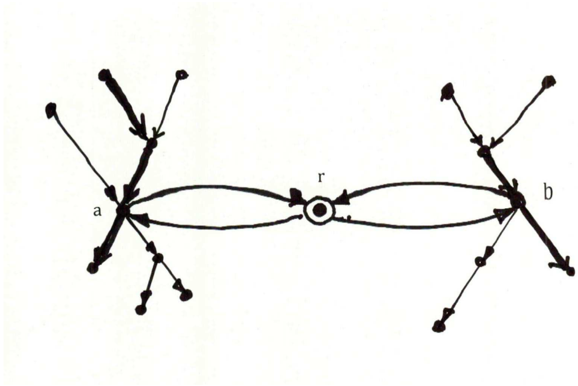

Figure 1: Connectome of a Concept.

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 satisfies the equation by construction and is therefore a concept.
Fig.1 shows the schematic view of the very rudimentary concept R; the constituting tracks are highlighted:

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

Higher-Order 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 of these concepts is established by a set of neurons which links the reference neurons of these two concepts. Generally, the set of concepts is not closed under application.

PERCEPTION AND APPERCEPTION

The question arises how $\mathcal{N}_{\mathcal{A}}$ could serve to model familiar brain functions and to represent their properties by equations. A central such notion is that of *apperception*. First, think of *perception* P as the operation of activating a set of neurons following some input V, e.g. visual input, through some connected layers of neurons. The brain function of apperception means additionally that P applied to V results in a specific activation – think of recognizing a shape. Let $R = P \cdot V$ denote a brain function to be identified with an apperception. As such, it needs to be a conceptual entity. Hence:

Definition 6 (Apperception) *The brain function P is apperceptive if $P \cdot V$ is a concept for all V, that is it satisfies the equation*

$$(P \cdot V) \cdot ((P \cdot V) \cdot X)) \approx (P \cdot V) \cdot X \quad \textit{for all} \quad V, X.$$

Fleeing upon being threatened may serve here as a simple example of an apperception; it models the embodiment of a familiar instinctive reaction pattern. The modeling is based on hypothesizing conceptual objects S_0 of threat, D_0 of danger, L_0 of the lack of cover, and F_0 of realizing the necessity of flight, all of them eventually depending on some visual input V: S_0 is the result of the perception S of the visual input V as a threat, $S_0 = S \cdot V$. Correspondingly $D_0 = D \cdot V$, $L_0 = L \cdot V$. For example, (cf. fig.2),

$$S_0 = \{\{a_2(t'), a_3(t'')\} \xrightarrow[s]{t} a_2(t''')\}_{t_1}^{t_2},$$

and S consists of all $\{v\} \xrightarrow[a_1]{t} u$ and $\{v\} \xrightarrow[a_2]{t} u$ with $v \in V, u \in S_0$.
With $F_0 = \{\{f\} \xrightarrow[c]{t} f\}_{t_1}^{t_2}$, the concept of being forced to flee, the equation

$$F_0 \approx F \cdot ((S \cup D \cup L) \cdot V)$$

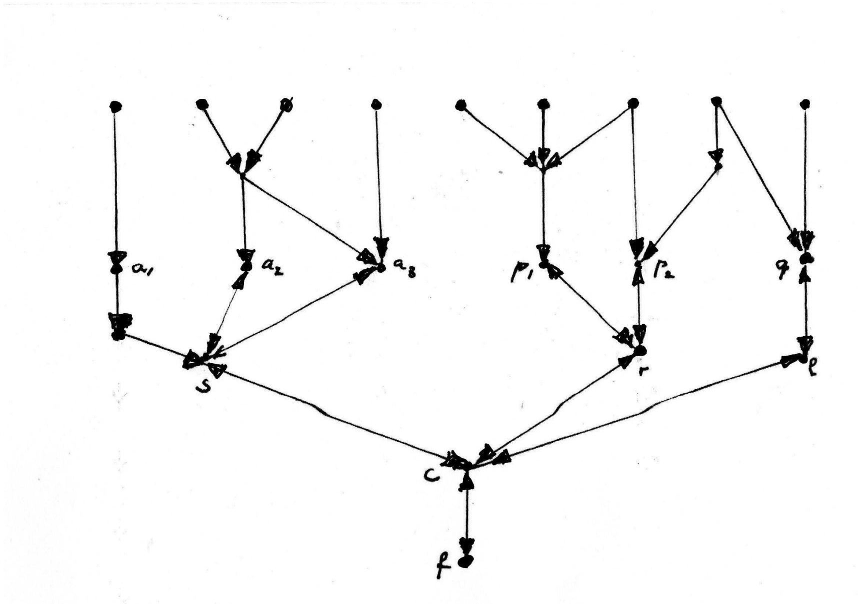

Figure 2: Fleeing upon a Threat.

represent a possible hypothesis for the brain function in question. Figure 1 shows a corresponding connectome.

The above example illustrates the discipline of modeling by neural algebra, progressing from equations to solutions to connectome. It is rudimentary on several aspects:
It is not plausible that a real brain has individual neurons such as s, r, l, c above that completely and exclusively correspond to familiar classifications of conceptual objects, 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. Second, the activation of concepts may depend on more complex networks than the feed-forward network depicted. Indeed, one should envision extensive, recurrent and deeply stacked collaborating networks, some perhaps corresponding to logical connectives.

4 Control

Controling is another important aspect of the functionality of the brain. This was recognized early by Norbert Wiener in his ”Cybernetics”, a mathematical theory of control, which he applied to the brain (19).

In Neural Algebras control is a form of interaction between brain functions and is therefore expressible by equations. We illustrate this by a small number of examples.

Generally, a *law of interaction* between brain functions $A, C, B_1, \ldots, B_n$ is an equation $A \approx \phi(C, B_1, \ldots, B_n)$ where ϕ is some expression built up using the operations $\cdot$ and $\cup$ admitted in the model. If ϕ is of the form $C \cdot \psi(B_1, \ldots, B_n)$ we say that C *controls* this interaction:

$$A \approx C \cdot \psi(B_1, \ldots, B_n).$$

Control is effected by applying C to the inputs B_i, typically concepts, resulting in the controlled output A.

Simple Control. The controlling object C controls an output B by the operation $C \cdot B$ with the goal to stabilize the output by the recursion

$$B \approx C \cdot B,$$

starting from an initial state B_0. If this state satisfies the initial condition $C \cdot B_0 \supseteq B_0$ there is indeed a solution, obtained as follows: Let

$$B_{i+1} = B_i \cup C \cdot B_i, i = 0, 1, \ldots.$$

Note $B_{i+1} \supseteq B_i$ for all i since application is monotonic in both arguments. If $\mathcal{N}_\mathcal{A}$ is complete, i.e. closed under unions of directed sets, then $B = \bigcup_i B_i$ is a solution. The completeness of $\mathcal{N}_\mathcal{A}$ follows from the fact that the time indices in A are bounded by the finiteness of the model, cf. section 2 above.

Coupled control is a simple control which arises if two controlling objects C_1 and C_2 act in a coupled manner, e.g.

$$A \approx C_1 \cdot B, \qquad B \approx C_2 \cdot A;$$

a situation encountered in neurology (e.g. hand-eye movement) just as in mechanics (e.g. coupled pendulum). The solutions A and B are again obtained by iteration.

Joint control is a special case of coupled control, described by the equations

$$C_1 \cdot A \approx B \approx C_2 \cdot A.$$

The above forms of control are represented by first-order equations.

Higher-order equations represent *hierarchical control*: the controlling object C is itself controlled by a separate controlling object D. For example

$$A \approx C \cdot A, \qquad C \approx D \cdot C.$$

Adaptive control is another second-order control:

$$A \approx C \cdot A, \qquad C \approx D \cdot (A \cup C).$$

By augmenting these control equations by another object, external control

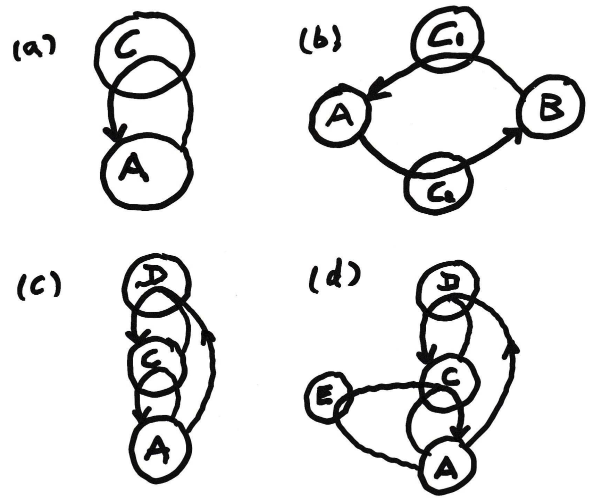

Figure 3: Control diagrams: (a) simple control, (b) coupled control, (c) ad adapive control, (d) externally adaptive control

E, we obtain controls that seems to be omnipresent in biology:

$$A \approx C \cdot (A \cup E \cdot A), \qquad C \approx D \cdot (A \cup C).$$

Functional Diagrams of control are a useful visual aid to understand the various regimes of control. They depict the structure of an applicatory expression involving objects $A, B, C, \ldots$ by blobs. The convention is that an arrow originating at A passing through C and terminating at B stands for the equation $B \approx C \cdot A$. Fig. 3 illustrates some of the above notions of control by corresponding functional diagrams.

To illustrate the relation between functional diagrams and corresponding connectomes consider the simplest case, the application $B \approx C \cdot A$ for

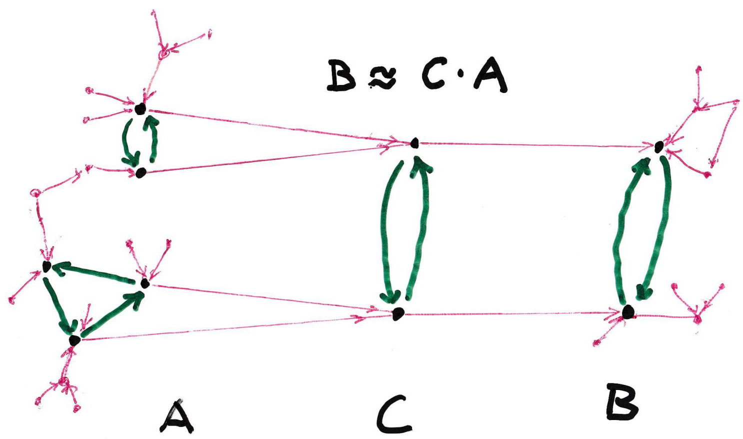

Figure 4: Connectome of application of concepts

concepts A, B, C. By Theorem 1 the key neurons $\hat{A}, \hat{B}, \hat{C}$ are connected by recurrent circuits; in Fig.4 these are indicated by green cycles. The individual firing tracts constituting the three concepts are in red.

The *mathematical challenge* is to pass from specific control equations to connectomes representing their solutions. We illustrate this by the interesting equation of reflexive control.

Reflexive Control is the kind of control which reflects on the controlling process itself. The equation of reflexive control is obtained by analyzing this concept: Let us understand reflexive control as the ability of the brain to observe itself as it is planning, acting and reacting. This definition, at first sight, appears circular. Interpreted in $\mathcal{N}_\mathcal{A}$ it is simply self-referential: The model $\mathcal{N}_\mathcal{A}$ comprises firing patters corresponding to perceiving, acting, planning, moving, etc. Let C be the prospective firing pattern of reflexive control. Let B be some set of causal firing tracks, what might be called ”the active brain” in the model. Then $B \cdot C$ is the result of observing, acting, etc. as dependent on the content of C, and $C \cdot B$ represents the reaction to the control to such activities. To these objects, including C itself, C is again applied; as in ”observing itself” above, i.e. reflexively. This characterization of reflexive control transforms into an equational definition as follows:

Definition 7 (Reflexive Control) *A sustained set C of track expressions in the brain B represents reflexive control 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 causal sequence $\{y_{c_0}(t_0), y_{c_1}(t_1), y_{c_2}(t_2), \dots\}$ 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 $i = 0, 1, \dots$, which is cyclic in the indices i modulo some period n.

Theorem 2 (Connectomes of Reflexive Control) *A neural algebra admits nontrivial reflexive control if and only if it contains at least one sustained causal cycle.*

Proof:

Assume that the reflection equation has a nonempty sustained solution C of sufficient length (see below) and consider a track expression $x_c(t) = \alpha \xrightarrow[c]{t} y(t')$ 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 $x = \alpha \xrightarrow[c]{t} y$ define the *input structure* $\sigma(x)$ as the tuple consisting of the key neuron of x and the key neurons of the elements of α. The solution C contains a sequence $x_{c_0}(t_0), x_{c_1}(t_1), \ldots$ 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 $i = 0, 1, \ldots$. If C is sustained and of sufficient length, 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 $\sigma(x_{c_{n-1}}(t_{n-1}))$ and therefore C contains a sustained 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 sustained causal cycle of length n. By recursion construct

$$C_{j+1} = C_j \cup \{\alpha_i \xrightarrow[c_i]{t_i} x_{c_{i+1}}(t_{i+1}), i = 0, 1, \ldots i-1 \mod n,$$

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

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 reflection equation

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

based on C_0. Generally, any set of causal cycles generates such a solution; they form a lattice by set-inclusion.

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, e.g. concepts, to another. Again, one should envision connectomes larger by several orders of magnitude.

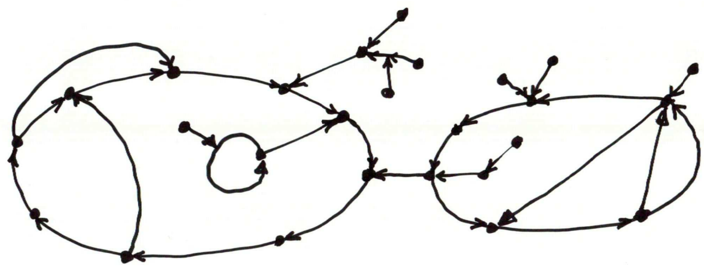

Figure 5: Connectome for Reflexive Control.

Remark: The theorem indicates that there is a condition on available brain functions to support reflexive control; it is required by the ”sufficiently large sustention” condition used in the proof. The lattice structure of the set of solutions corresponds to phases or states of control, and their contextual movement depends on the inclusion/exclusion of the various inputs available from present states. In other words: reflexive control expands/contracts by attaching/releasing connections to perceptions, memories etc. according to the firing history.

5 Discussion: Neurology, Consciousness and Thinking

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:
For $\mathcal{N}_{\mathcal{A}}$ the notions of thought, concept and consciousness are modelled 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 (20, 21): 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 (22, 23). This is exactly what Neural Algebra attempts to provide a mathematical model for.

Sets of Key Neurons may remind us of the following:
Single neurons have been identified as the key to *recognize a face*, (24). 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, (25). Sets of key neurons constitute the basis for neural representations of concepts. Some concepts have been found to correspond to specific locations in the human cortex. Connections between them and to what we called higher-order concepts such as planning at other locations have also be distinguished. This is a main subject of neuroscience.
The above analysis of control results in equations and circuits that may have their image in the wiring of the cortex. Indeed our adaptive second-order control seems to correspond to findings in the cat brain (26), and more generally to a structure that is omnipresent in the neocortex. This has been called the *canonical circuit* of dominant interactions; they go down cortical layers and across cortical sheets (23) as predicted in the model by their second-order nature. It has been conjectured (27) that cortical lamination is providing a general scaffold and that the canonical circuits may allow neurons to connect with each other with a minimum of wires.

Recruiting new neurons and synapses to create new abilities has been identified, and shown to be involved in the learning of bird songs (28) and speech (29), and in reading. Again, like in the formation of concepts, in our model this corresponds to the introduction and recruiting of key neurons in the firing history.

Understanding Consciousness has been termed ”the most challenging task confronting science”, and what has been a philosophical mainstay has become a legitimate question of ”hard science” (30, 31), now turning into ”big science”. Motivated by its depth and range, some rich, beautiful and touchingly personal books came to be written by some of the pioneers (32) – (37). And, not surprisingly, we observe an enormous production of papers on brain and consciousness in neuroscience alone: about six papers per day according to a citation search. There have also been some interesting attempts at theoretical synthesis under different viewpoints: proposing mathematical approaches, ranging from dynamical systems (38) to quantum mechanics (39), information theory (40) and statistics (41), and relating them to neurological facts and psychological experiments.
Some scientists have denied the have denied the possibility of any overarching mathematical ”theory of the mind”. For some, this may be on theological grounds, reminiscent of Georg Cantor’s spiritual conflict with the actual infinite, a central aspect of the God of Christianity. Others, notably Misha Gromov (42), have doubted it because the envisioned goal seemed to transcend the power of mathematical imagination. – What we have attempted here is of course considerably more modest, for which we beg the reader’s indulgence.

Human consciousness. To embody a solution of the ”consciousness equation” (Definition 7) in $\mathcal{N}_{\mathcal{A}}$ we arrived at set of causal cycles (Theorem 2). This characterization of ”consciousness” as based on linked cycles of partial consciousness and concepts, 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, and more generally in linked circuits (43) and so-called convergence–divergence–zones and regions, (44) and (35, chapt.6). The 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 Hofstadter's ”I am a Strange Loop” (45). If we wish to visualize the activity of conscious thinking we may distinguish parts of this network to correspond to specific concepts, others to perceptual operation, partly linked to the outside, and other operations that lead from concepts to concepts such as abstraction- and generalization-operations whose connectional correlates should be reasonably clear.
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; much additional information has been obtained, approximating total neural modeling (46). 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.
Artificial consciousness may be an utopian goal (47, 48), although it has been studied in the context of artificial intelligence, not least in the hope of modeling the perceived advantage of ”conscious” over ”mechanistic” robots (49), including swarms of robots (49). Even plants may have a sort of consciousness (50).

Much remains to be done, and this author, fascinated by the challenge and the enormous literature is greatly intimidated.

References

(1) Curry, H.B.,Grundlagen der kombinatorischen Logik, Amer. J. Math. 52 (1930) 509-536, 589-834
(2) Turing, A.,On computable numbers with an application to the Entscheidungsproblem, Proc. London Math.Soc. 42 (1936), 230-265
(3) Schönfinkel, M., Ueber die Bausteine der mathematischen Logik, Math. Annalen 92 (1924), 305-316
(4) Church, A. and Rosser, J.B., Some properties of conversion, Trasns. Amer. Math.Soc. 39 (1936), 472-482
(5) Plotkin, G.D., A set-theoretical definition of application, manuscript, Edinburgh 1972
(6) Scott, D.S.,Models foe the λ calculus, manuscript 1969

(7) Baendreght,H.P., The Lambda Calculus, North Holland 1984
(8) Engeler, E., Algebras and Combinators. Algebra Universalis, 13 (1981), 389 - 392
(9) Engeler, E.et al., The Combinatory Programme Birkhäuser Boston, 1995
(10) Engeler, E., Neural Algebra and Consciousness. Algebraic Biology, Lecture Notes in Computer Science 5147, 96-109 (2008)
(11) Rosenbloom, P.C., The Elements of Mathematical Logic, Dover 1950
(12) Engeler, E., Foundations of Mathematics, Springer Verlag, Berlin Heidelberg 1993

(13) Lullus, R.., Ars Brevis (latin/german). Meiner, Hamburg 1999
(14) Eco, U., Die Suche nach der vollkommenen Sprache (vide p.88-29), C.H.Beck, München 1994
(15) Hirsch, E.C., Der berühmte Herr Leibniz, C.H.Beck, München 2016
(16) Couturat, L., La Logique de Leibniz, Paris 1903, (see also www.leibnizedition.de)
(17) Buridan, J., Sophismes, J.Vrin, Paris 1993
(18) Marciszewski, W., Logic from a Rhetorical Point of View, de Gruyter, Berlin 1993
(19) Wiener, N., Cybernetics; or Control and Communication in the Animal and the Machine. MIT Press, Cambridge, Mass. 1948
(20) Sporns, O., Networks of the Brain. MIT Press, Cambridge,Mass. 2011
(21) Seung, S., Connectomes; how the brain's wiring makes us who we are. Houghton Mifflin Harcourt, New York (2012)
(22) Douglas, R.J. and Martin, K.A.C., Neuronal circuits of the neocortex,

Ann.Rev.Neurosc. 27, (2004), 419-451
(23) Douglas, R.J. and Martin, K.A.C., Behavioral achitecture of the cortical sheet, Current Biology 22, (2012), 1033-1038
(24) 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)
(25) Quiroga, R.Q., Fried, I., Koch, C., Brain cells for grandmother. Sci.American, Febr.2013, 31-35.
(26) Heinzle, J., Hepp, K., Martin, A.C., A microcircuit model of the frontal eye field. J.Neuroscience 27, 9341 – 9353 (2007)
(27) Chklovskii, D.B. e al. Wiring optimization in cortical circuis, Neuron 34, (2002), 341-347
(28) Deahaene,St. et al., How learning to read changes the cortical networks for vision and language. Science 330, 1359 – 1364 (2010)
(29) Bolhuis, J.J., Okanoya, K., Scharff, C., Twitter evolution; converging mechanisms in birdsong and human speech. Nature Review Neuroscience 11, 747 – 759 (2010)
(30) Hahnloser, R.H.R., Kotowicz, A., Auditory representations and memory in birdsong learning. Current Opinion in Neurobiology 20, 332 – 339 (2010)
(31) Crick, F.,The Astonishing Hypothesis. Simon & Schuster, London (1990, 1994)
(32) Crick, F., Koch,C., Towards a neurobiological theory of consciousness. Seminars in Neuroscience 2, 263 – 275 (1990)
(33) Penrose, R., The Emperor's New Mind. Oxford 1989.
(34) Churchland, P.S., Sejnowski, T.J., The Computational Brain. MIT

Press, Cambridge, Mass. (1992)
(35) Valiant, L.G., Circuits of the Mind. Oxford University Press, New York (1994)
(35) Damasio, A., Self comes to Mind; Constructing the Conscious Brain. Vintage Books London (2012)
(36) Koch, C., Consciousness; Confessions of a Romantic Reductionist. MIT Press, Cambridge, Mass. (2012)
(37) Tononi, G., Phi, A Voyage from the Brain to the Soul. New York 2012.
(38) Izhikevich, E.M.:,Dynamical Systems in Neuroscience. MIT Press, Cambridge, Mass. (2007)
(39) Koch, C., Hepp, K.:Quantum mechanics in the brain. Nature 440,611 – 612(2006)
(40) Tononi, G.: Consciousness as Integrated Information; a Provisional Manifesto. Biol.Bull. 215, 216 – 242 (2008)
(41) Grenander, U., A Calculus of Ideas: A Mathematical Study of Human Thought. World Scientific, Singapore (2012).
(42) Gromov, M., Great curcle od mysteries: mathematics, the world, the mind, manuscript 2015
(43) Singer, W., Consciousness and the binding problem. Annals NY Acad.Sci., 929 (2001), 123-146.
(44) Meyer, K., Damasio, A., Convergence and divergence in a neural architecture for recognition and memory. Trends in Neuroscience 32, 376 – 382 (2009)
(45) 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.
(46) Ardiel, E.L., Rankin, C.H.: An elegant mind; learning and memory in *Caenorhabditis elegans*. Learning & Memory 17, 191 – 2001 (2010)
(47) Sejnowski, T.: Learning optimal strategies in complex environments. Proc.Natl.Acad.Sci.USA 107, 20151 – 20152 (2010)
(48) 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)
(49) Sejnowski,T.: When will we be able to build a brain like ours? Scientific American, (2010)
(50)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.
(51)McCarthy,J., Making robots conscious of their mental states. In: Muggleton,S. (ed.), Machine Intelligence15, Oxford Univ.Press, Oxford (1996)
(52) Chamowitz, D., What a Plant Knows. Oneworld Publ.C.. 2013.