Talks at:
ICSI, Berkeley, January 18, 2011
ICALP, Zurich, July 3, 2011

Algebras of the Brain

E. Engeler

How does the brain compute ?

Here is the answer that Steve Pinker gave in a TV interview when challenged to answer in a short sentence: *By neurons firing in patterns.*
The operative word is *patterns of firings*. Indeed, it is the key word in our aim to find *the right mathematics for representing brain activities*.

Obviously, the net of interconnected neurons in the brain constitutes a system with a great number of parallel processes. Mathematics has dealt with such systems before:
The traditional and enormously successful approach is by systems of differential equations. It all starts with the insight, that the basic laws operate on the infinitesimal scale. Local infinitesimal changes propagate from the given boundary values. Calculus turns infinitesimal laws typically into ordinary or partial differential equations, thus mathematizing the whole of the system: vibrating strings, heat equation, reaction-diffusion equation, Maxwell equations, . . .
The fact is, that this approach succeeds precisely because at all points in the domain of the system it is always the same local dependence that obtains. But the dynamical systems approach, while retaining its paradigmatic power (vz. attractors, etc.), meets its limits as soon as this uniformity requirement is violated. And this is what happens in the case of the neural system of the brain: While the basic building principles of neurons remain the same, there are enormeous differences in the size and extent otf their connections and therefore of their mutual dependencies.

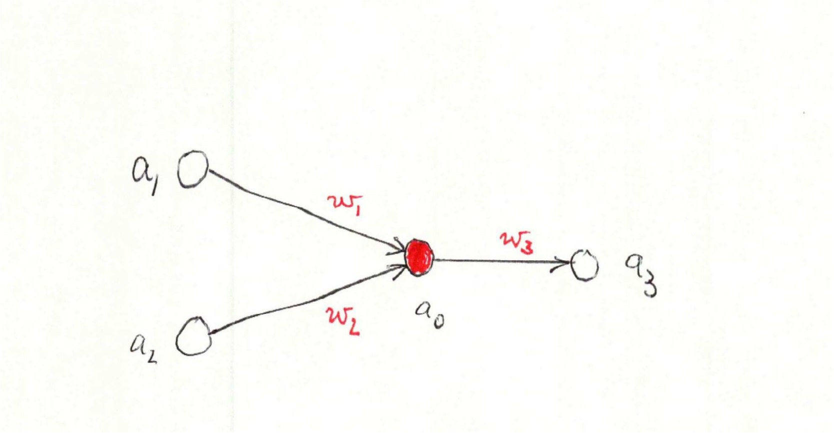

Figure 1: Simple Track.

Artificial neural nets

The mathematical models of the ”brain” on which we base the development of Neural Algebra are so-called artificial neural nets. Of these there are many variants, abstracted from increasingly detailed knowledge of biological neural nets and their function. We choose a very simple kind of model; but it turns out, that adaptations could be made to encompass much richer neurological details on the one hand, or to consider nets based on functional connectivity based on correlated activities of segmentations of the brain. Indeed, it seems possible to apply the model to nets of interactive processes far removed from neurology.
An *artificial neural net* A is a directed graph whose edges are weighted by rational numbers $w \in \mathbb{Q}$. The nodes correspond to ”neurons”, the edges to ”synapses” whose weights represent the strength of their contribution to the activity: if neurons $a_1, \ldots, a_n$ to node a_0 have edges of weights $w_{i,0}$ then the firing of a_0 is induced if the sum of these weights exceeds a threshold which we generally put at 1. All firings take place at discrete time instances $t, t \in \mathbb{Z}$, the set of integers.
Our formal model is based on representing the local laws that govern neural nets (fig.1) by *track expressions* as follows:

Simple track expression

$$x_0 = \{a_1, a_2\} \xrightarrow[a_0]{t} a_3$$

The track expression x_0 is read as: neurons a_1, a_2 fire at time $t - 1$. The sum of weights

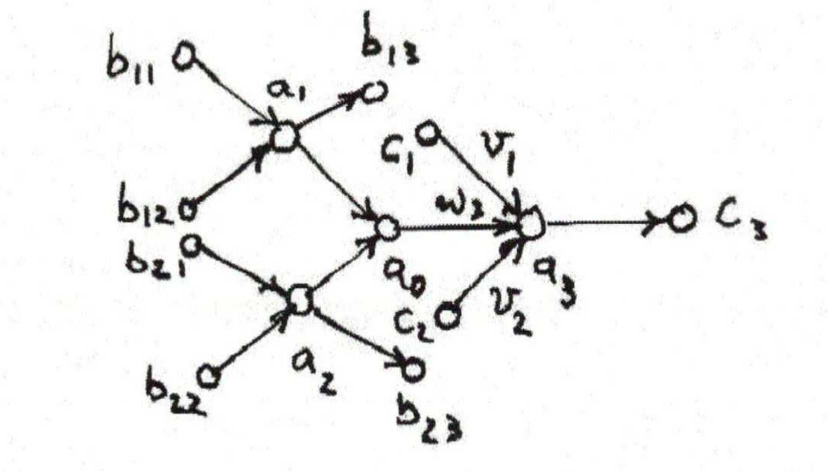

Figure 2: A Cascade.

w_1, w_2 exceed the threshold, a_0 fires at time t ; w_3 also exceeds the threshold, and a_3 fires at time $t+1$.
a_0 is called the *key neuron* of this expression.

Track expressions: Cascades if firings

By connectivity, the firing of neurons progresses through a neural net and produces cascades of firings. Such cascades (fig.2) are formally represented by iterating the formation of track expressions, starting with simple track expressions. For example, the track expression x_1 below arises by substitution of track expressions for individual neurons in the track expression x_0 such that the substituted expressions have these neurons as their key neurons. These substitutions are legal, if the relative sums of weights exceed the threshold.

Iterated track expressions

$$x_0 = \{a_1, a_2\} \xrightarrow[a_0]{t} a_3$$

$$x_1 = \Big\{\{b_{1,1}, b_{1,2}\} \xrightarrow[a_1]{t-1} b_{1,3}, \{b_{2,1}, b_{2,2}\} \xrightarrow[a_2]{t-1} b_{2,3}\Big\} \xrightarrow[a_0]{t} \Big\{\{c_1, c_2\} \xrightarrow[a_3]{t+1} c_3\Big\}.$$

Neuron a_0 remains the key neuron of x_1.

Firing patterns

The firing of a neuron a at time t is denoted by the value 1 of the firing function f with $f(a,t) = 1, a \in A, t \in \mathbb{Z}$. We say that a set of firings, determined by f is *consistent* with the neural net A , if, whenever $a_1 \dots a_n$ are edges in A leading to b with weights $w_1 \dots w_n$, $\Sigma_i w_i$ exceeds a given threshold and all $f(a_i, t-1) = 1$ then $f(b, t+1) = 1$-

A *firing history* $F(A)$ for A is any consistent set of firings.

The set of all legal track expression corresponding to the firing history $F(A)$ is denoted by $S(A)$.

A *Firing pattern) is any subset of* $S(A)$.

Convention: In what follows, we tacitly assume that all firings occurring in definitions, proofs and examples belong to a fixed firing set, a possible history of firings in a brain.

Firing patters are the basic objects of our theory. They embody mental functions. Functions in analysis and firing patterns are both quite complex infinite sets, a fact to which we have become quite oblivious in the case of analysis. We can add, multiply, differentiate and integrate functions. Our goal is to develop a corresponding operability with firing patters.

Interaction of firing patterns

Firing patterns are related by acting on each other as determined by the structure of the net. We untangle these interactions by basing them on the concept of *applying* firing a pattern to another. Observe that in each individual track expression the expression to the left of the main arrow represents the cascade that prompts the key neuron to fire. The cascade denoted by the expression on the right is what new firings this firing causes. The same is true for sets of track expressions. In this view, a track expression corresponds to an element of the graph of a function (considered as a set of ordered pairs, and a firing pattern is, as we suggested, a mental function.

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

Composition

A set M composed with a set N applies the causation, represented by M, on N as follows:

$$M \cdot N = \{\ x :\ \ \textit{there is an element}\ \{x_1, \ldots, x_k\} \underset{a}{\overset{t}{\rightarrow}} x\ \textit{in}\ \ M \\ \textit{such that}\ \{x_1, \ldots, x_k\} \subseteq N\}\ .$$

Composition of firing patterns give rise to an algebraic structure, the neural algebra $\mathcal{N}_A$. To summarize:

A neural Algebra $\mathcal{N}_A$ consists of a weighted directed graph A

”the neural net of the brain”

a firing history $F(A)$,

”the activity of the brain”

the set $S(A)$ of track expressions determined by $F(A)$,

”cascades of firing neurons”

a set of subsets of $S(A)$,

”firing patterns”

closed under the binary operation of composition.

”neural algebra”

Challenges

What is the relation between the structures in a neural net and their function ? But: How do we decide on the definition and selection of such functions or concepts ?

The *biological approach* is exemplified by brain imaging. There is always the statistical approach: various techniques of brain imaging can be used to show that experiments on (sometimes large samples of) animal or human subjects exhibit a clear correlation between parts of the brain structure and a particular concept or function. In this way one is able to isolate what is worthwhile. This is enormously successful scientifically, it also produces beautiful pictures and serves many derived disciplines of neuroscience.

The *neural algebra approach* to the structure/function problem profits from the fact that the objects, firing patterns, serve at the same time functionally - by composition of mental functions - and structurally - by reading parts of the neural net off the track expressions representing such functions.
But then there are uncountably many possible elements of the neural algebra $\mathcal{N}_A$, and we are faced with the problem of identifying the truly relevant ones among them.

Predication
Elements R of the neural algebra $\mathcal{N}_A$ are always operations, as left factors. Some of them may be considered as predicates in the sense of *predication operations*:
$R \cdot X$ computes the extent to which the ”predicate” R applies to X.
If a predication is to be conceptually relevant, the main requirement is that it should be general, abstract, enough not to depend on accidental, extraneous, conditions around it. This corresponds to the traditional notion of a concept. Since Aristoteles, concepts or universals are arrived at by abstraction: by taking a thought and eliminating all extraneous elements, the *accidentia*, the accidential or irrelevant aspects.

Concepts
We identify concepts in $\mathcal{N}_A$ with the corresponding abstraction operation. If R is a conceptual abstracting operation, applied to a thought X which belongs to the conceptual field of the concept, then $R \cdot X$ removes from X all aspects that are irrelevant with respect to the predication R.
If applying R again returns the same result, this is the pure abstract, the conceptual content of X. Accordingly, we define:
R is a *concept* if it satisfies the equation $R \cdot (R \cdot X) = R \cdot X$ for all X.
Familiar concepts are typically based on a firing pattern that has an *episodic character*. Familiar examples are notions such as skripts and memories.
Scripts act situationally and are templates for procedures, projects, processes ...

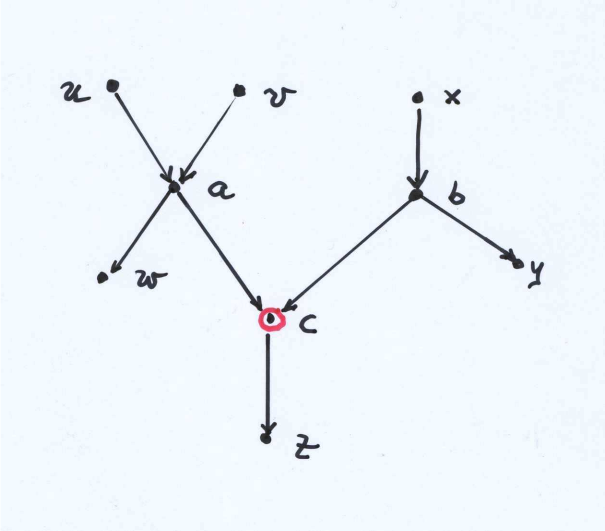

Figure 3: Fleeing on a Threat.

Memories are invoked by triggers and store auditory and visual perceptions, thoughts, emotions ...

Example

Fleeing upon being threatened may serve here as a simple example of a script:

$$s = \{\{u, v\} \xrightarrow[a]{t-1} w, \{x\} \xrightarrow[b]{t-1} y\} \xrightarrow[c]{t} z$$

u : it's big, v : it moves fast towards me, a : it's dangerous, w : watch carefully.

x : no cover, b : I'm exposed, y: I'm in danger.

c : decide to flee, z : flee !

The imeditate question is how to characterize firing patterns that correspond to concepts; what is the structure of the neural correlates of concepts ?

Firing pattern of a basic concept

Let s be any track expressions (e.g. the one above), and choose some neuron r. Define

$$S = \{\{s\} \xrightarrow[r]{t} s : t \in \mathbb{Z}\},$$

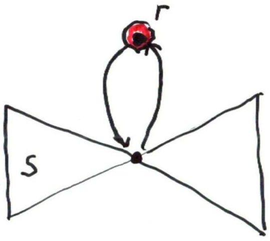

Figure 4: Conceptual Net.

then clearly $S \cdot (S \cdot X) = S \cdot X$ for all X, since

$$S \cdot X = \emptyset, \text{ if } s \notin X;$$

$$S \cdot X = \{s\}, \text{ if } s \in X.$$

Note that the cycle at r serves as a sort if pacemaker, and observe that S is the identity operator restricted to s, in a sense it ”identifies” it.

Neural net of a basic concept

The figure (fig.4) illustrates a neural net which realizes the firing pattern S.
It may be argued that in reality the brain does not work on the time scale from minus to plus infinity, that is in $\mathbb{Z}$, but during a finite time interval I. If in the definition of a script S, we replace $\mathbb{Z}$ by I, the defining equation for concepts is only approximately satisfied: mental concepts tend to be a bit fuzzy around the edges.

Scripts can be enchained, memories can be associative, and both can be combined to create more complex concepts; in a bigger context, the role of a pacemaker is taken over by scripts being embedded in larger cycles, as illustrated in fig.5.

Creation of concepts

Consider the track expression s as an input at time t_0, (e.g. the teaching of a movement

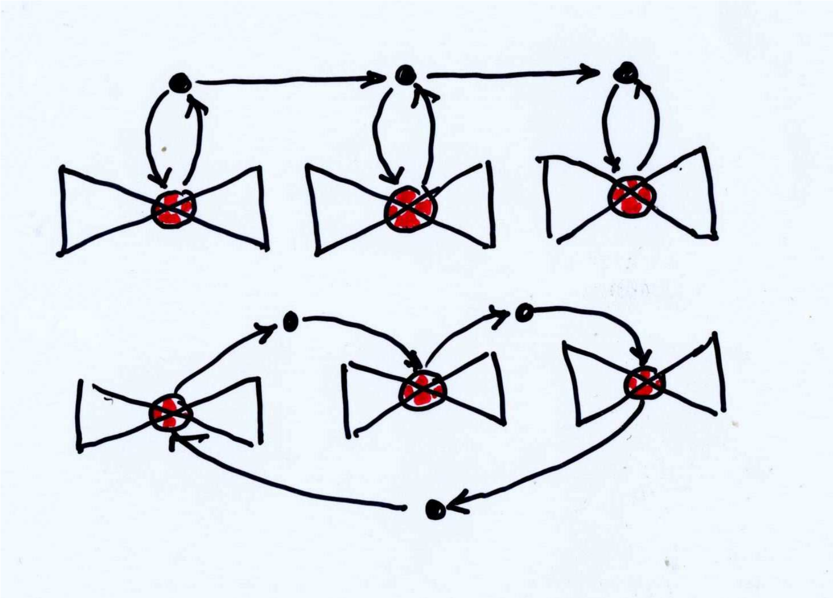

Figure 5: Key Chain and Key Ring.

or the presentation of a picture.) *Perceiving* the input s should produce a concept, the *perception* P of s, with $P \cdot (P \cdot X) = P \cdot X$ for all X.
Perception is *realized* in a neural net by recruiting a neuron p and setting
$P = \{\{s\} \xrightarrow[p]{t} s : t > t_0\}$. Then for all X

$$P \cdot X = \{s(t) : t > t_1\}, \text{if } s(t_1) \in X \text{for some } t_1 \geq t_0, \emptyset \text{ otherwise.}$$

$$P \cdot (P \cdot X) = \{s(t) : t > t_1 + 1\}, \text{if } s(t_1) \in X \text{for some } t_1 \geq t_0, \emptyset \text{ otherwise.}$$

Clearly, the conceptualization of a perception is again only approximately obtained.

Instead of mobilizing single neurons to conceptualize scripts or memories, perception may consist more generally in attaching them to an existing key ring.

Consciousness

Let us understand neural consciousness as
the ability of a neural net B ("the brain") to consciously observe itself as being conscious and as consciously planning and acting.

These abilities are embodied as activities in sub-populations of the “brain”, represented by firing patterns; their interrelation is expressed by their composition: If C is the firing pattern corresponding to “consciousness”, and M_1, M_2, etc. are the firing patterns corresponding to the context of observing, acting, planning, moving, etc. then $M_1 \cdot C$, $M_2 \cdot C$, etc. are the results of observing, acting, etc. as dependent on consciousness. To the sum of these results, together with C itself, C is again applied.

Translated into neural algebra, our definition of consciousness transforms into an equation of the form

$$C \cdot (C \cup \bigcup_i M_i \cdot C) = C\,.$$

This equation formulates the self-referential character of consciousness, an aspect that has been formulated and investigated throughout the history of the concept, witness ”cogito ergo sum” to ”I am a strange loop”. Algebraically, we have here a fixpoint equation, such as encountered quite frequently in key places in various parts of mathematics:

Let $\varphi(X)$ be any algebraic composition of X with elements of the neural algebra $\mathcal{N}_A$, then

$$\varphi(X) = X$$

is a fixpoint equation.

Fixpoint Theorem

In $\mathcal{N}_A$ all *fixpoint equations* have a 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

If $N_1 \supseteq N_2$ then $M \cdot N_1 \supseteq M \cdot N_2$ by the definition of composition; equally $M_1 \cdot N \supseteq M_2 \cdot N$ for $M_1 \supseteq M_2$. Hence, if $\varphi(X)$ is any algebraic composition of X with elements of $F(A)$ then $X' \supseteq X$ implies $\varphi(X') \supseteq \varphi(X)$. More generally, if D is a directed set of elements of $\mathcal{N}_A$ 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 $\varphi^{n+1}(X) = \varphi(\varphi^n(X))$. 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.

The solutions of the fixpoint equation for consciousness constitute the set of persistent activity patterns in a net of neurons that may be understood as ”states of consciousness”. (The apparent circularity of our non-formal definition thus resolves itself as multiple entry of the unknown in a single equation.)
Again, the question arises how to characterize firing patterns and their neural correlates of solutions to the consciousness equation.

Structure Theorem of Consciousness
(1) Consciousness has a base in one or more cycles of the directed graph.
(2) Consciousness can be expanded along any outgoing edge.
(3) Consciousness never expands backwards into cycle free “stimulus and response” subgraphs.

To illustrate the proof of part 1 of this theorem, consider a cycle of neurons $a_0, \ldots, a_{n-1}$, connected with weights 1 , and firing at times t, $f(a_i, t) = 1$ iff $t \equiv 1 \ mod(n)$.
Let each set C_i consist of a_i and all terms

$$x_i = \{x_{i-1}\} \xrightarrow[a_i]{t} x_{i+1},$$

with $x_{i-1} \in C_{i-1}$, $x_{i+1} \in C_{i+1}$ and $t \in \mathbb{Z}, t \equiv i \ mod(n)$
and let us restrict the ”mind” M provisionally to this cycle.
Observe that $C_2 \cdot C_1 \ = \ C_3$, etc. Taking C, and here also M as the union of the C_i, we obtain $C \cdot C = C$ and therefore

$$C \cdot (C \cup M \cdot C) = C \cdot (C \cup C \cdot C) = C.$$

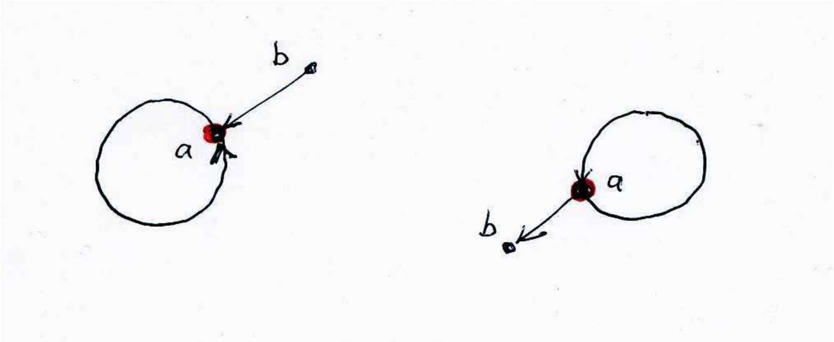

Figure 6: Input-Output to Consciousness.

Remark: Just as in the case of concepts, if firing times are restricted to an interval, say $-10^5 < t < 10^5$, the equation still holds with the exception of a few terms ”around the edges”. Consciousness is always temporary, and somewhat fuzzy ant the start and the end...

Proof of part 2: It suffices to consider cycles consisting of just one node as in the figure 6. For the example on the left the firing pattern restricted to an interval I of $\mathbb{Z}$ is defined recursively as

$$A_I = \{a \xrightarrow[a]{t} a, b \xrightarrow[a]{1} a : t \in I\},$$

$$\cup\{\{x_1, \ldots, x_k\} \xrightarrow[a]{t} y\} : x_1, \ldots, x_k, y \in A_I\}$$

$$\cup\{b \xrightarrow[a]{1} y : y \in A_I\}$$

Note that for $I = \mathbb{Z}$ we have $A_\mathbb{Z} \cdot A_\mathbb{Z} \neq A_\mathbb{Z}$.
The proof of part 3 for the figure on the right is similar:

$$B_I = \{a \xrightarrow[a]{t} a, a \xrightarrow[a]{t} b : t \in I\},$$

$$\cup\{\{x_1, \ldots, x_k\} \xrightarrow[a]{t} b\} : x_1, \ldots, x_k \in B_I\}$$

But now $B_\mathbb{Z} \cdot B_\mathbb{Z} = B_\mathbb{Z}$.

Consciousness and concepts

Concepts are by their connectional structure candidates for inclusion in solutions to the consciousness equation, as illustrated in fig.7. The concept ”fleeing upon danger” could

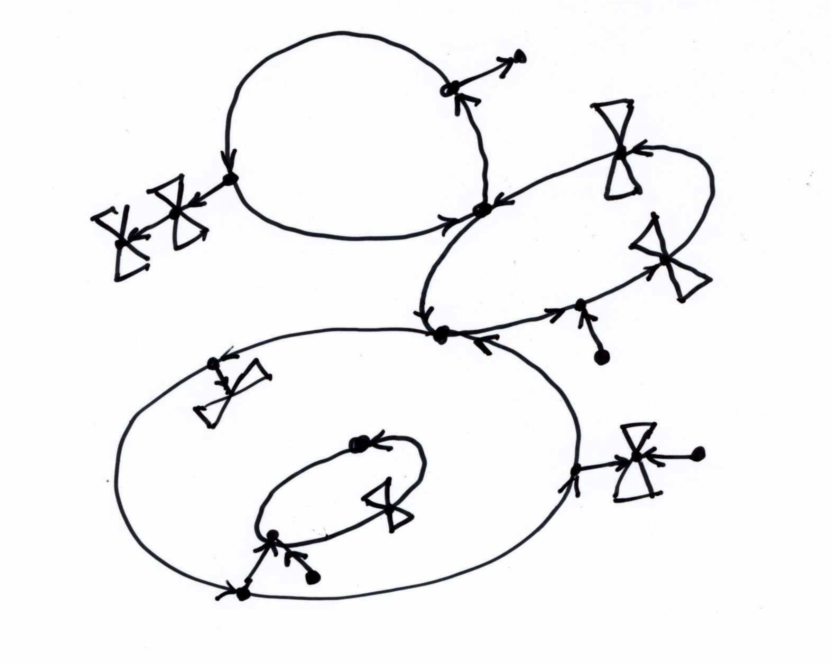

Figure 7: A Ring of Consciousness.

be one of the entries.

The lattice structure of the set of solutions reflects the phases of consciousness and their contextual movement depends on the inclusion/exclusion of the various concepts available from present states. In other words: consciousness expands/contracts by attaching/releasing key rings according to the firing history.

Of course, most of the the conceptual key rings in a brain would represent subconscious scripts and memories, indeed what are called instincts, some of them inherited, some acquired.

Animal, Social and Artificial Consciousness

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 their synapses, and much additional information has been obtained, approximating total neural modelling. 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 ...

Social consciousness, in a technical sense, would consist of understanding individuals 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, 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” ones.

John McCarthy, one of the pioneers in designing reasoning and consciousness for robots, distinguishes between an AI approach and the neuroscience approach. According to him, the AI approach considers consciousness composed of stratified levels of self-awareness. Awareness is represented by set of sentences (in a formal or better in a natural language) available to the robot's reasoning system. He does not claim a theory of human consciousness, in particular he does ”not claim that the human brain uses sentences as its primary way of representing information”.

I beg to differ. It seems both possible, and indeed very promising, to translate track expressions into McCarthy's LISP and vice versa, and thus to represent AI models, formulated in LISP, directly in neural algebra, and to take advantage of its non-stratified character and the obvious formal similarities between LISP grammar and track exprtessions, (and also with PROLOG statements.)

Outlook

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

As algebraic structures, neural algebras $\mathcal{N}_A$ are closely related to models of combinatory logic. Indeed, for rich graphs A and firing histories $F(A)$ they are such models. Logic deals with laws of thought, combinatory logic with the laws of applying thoughts to thoughts. If we now identify thoughts with concepts in our technical sense, then neural algebra may be regarded as a model of *neural logic*, relating algebras of the mind with algebras of the brain.

To develop this theme, consider:

1. What do we learn about *composite concepts* ?

2. Equations in neural algebras containing one or more unknowns correspond to *conceptual mental problems*. What is the relation between algorithms for solving equations and processes in corresponding neural structures ?

3. The question of concepts raises a basic epistemological problem: There is the danger to be trapped by cultural preconceptions in the widest sense, by notions that are supported by diverse scientific, linguistic and other (partially unreflected) traditions.

4. Conversely, concepts that have established themselves by convention may well be structurally representable. This is particularly attractive in the more general context of applying neural algebra models, e.g. in sociology, or when one speaks of the market or of nature as of (consciously) acting entities.

Coda (September 2011)

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”: it has been called ”the ultimate intellectual challenge in the new millenium”. Not surprisingly, there has been 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 been some notable attempts at theoretical synthesis, under different viewpoints, proposing mathematical approaches (ranging from dynamical systems to quantum theory), and relating them to neurological facts and psychological experiments.

The present author, also fascinated by the challenge, remains within his field of competence and scientific background, (E.Engeler et al., The Combinatory Programme, Birkhäuser, 1995), and, using this experience, developed the present mathematical model of mental functions and their neural embodiment, (E.Engeler, Neural Algebras and Consciousness, a Theory of Structural Functionality in Neural Nets, in: Algebraic Biology, Springer LNCS 5147, 2008, p. 96-109.)

Consciousness seems consistently evasive to strict characterization and exact localization. All that that mathematical models such as ours can provide is explanation and prediction of selected aspects, increasing their plausibility but remaining short of definitive validation.

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:

The single neuron identified as the key to recognize a face (cf. the key neuron of that concept).

Mirror neurons (cf. the key neuron to identify with a related concept).

Perception and learning as investigated in bird songs, reading and early development of the brain (cf. recruiting neurons to create concepts and abilities).

Recurrent or reentrant connectivity in the brain, e.g. in the visual cortex (cf. key rings of consciousness) and corresponding EEG measurements.

The explanatory power of our model is surely helpful, in particular if it is combined with the mathematical techniques, developed for the formulation and solution of equations, to formally express hypotheses on the interrelations between brain functions as represented by objects in the neural algebra of a brain.

We hope to be able one day to finalize this paper, which now retains the aspects of the original oral presentation (being short on references and acknowledgments), and to expand the necessary mathematical background from the original lecture notes publication cited above. We also should develop the algebra of concepts into a neural logic, (as proposed in: E.Engeler, Algebras of the Mind and Algebras of the Brain, 14th Congress of Logic, Methodology and Philosophy of Science 2011. Abstracts, p.204, extended abstract http://www.lmps2011.org/en/editorial.html last viewed Sept.2011.)