A Theory of Structural Functionality in Neural Nets

E. Engeler

Abstract
We propose Neural Algebra as a model of the brain. The goal is to develop mathematical tools for discussion of neural correlates of mental notions and their analysis and synthesis.
– Neural Algebras consist of formal objects that represent sets of cascades of firing neurons; the binary operation reflects their causal relation. Neural algebras are derived from graph algebras (of which the Plotkin-Scott-Engeler models of Combinatory Logic are an example), enriched by internal structure of the base elements. Functions, abilities and mental concepts are realized as specific types of objects in the algebra; their interoperation is described by equations. Consciousness presents as a recursion equation, reflecting its self-referential character. The lattice of solutions describes its different phases and moving context. The model is related to evidence from the neurosciences.

1 Introduction

How does the brain think?
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 here is patterns of firings. Indeed, this is the key word in our aim to find the right mathematics for representing brain functions.

What is the right mathematics for the brain?
Obviously, the net of interconnected neurons in the brain constitutes a sys-
tem with a great number of parallel processes, linked as cascades of firing neurons. A firing pattern in the brain consists of a set of cascades.

The idea of considering firing patterns as objects of an algebra forms the basis for our construction of neural algebras as a structural theory of neural nets. Neural algebras, accordingly have only one kind of objects, formal representation of firing pattern in the brain, formally representing sets of cascades of firing neurons. The web of mutual interactions of these objects is represented in the algebra as its basic composition operation. This gives rise to our proposed Neural Algebras.

Neural algebras in this way link states and activities of the brain to objects in a mathematical model thereof. This is reasonable under the widely held conviction, that all mental concepts and activities are accompanied, represented by, embodied as, or simply: are identical to patterns of firing neurons. Such patterns typically involve a great number of neurons, linked over considerable distances and are active for considerable time relative to the time scale of the individual neuron. Many of the mental concepts and activities are episodic in character, in particular in the way in which they are activated and used.

2 Basic Definitions

Artificial neural nets

An artificial neural net $A$ is a directed graph whose edges are weighted by rational numbers $w \in \mathbb{Q}$, (1)–(4). The nodes of the graph correspond to "neurons", the edges to "synapses" whose weights represent the strength of their contribution to the firing activity. If $a_1, \ldots, a_n$ are nodes connected
to node $b$ with weights $w_1, \ldots, w_n$, and $a_1 \ldots, a_n$ fire at time $t - 1$, then $b$ fires at time $t$ if the sum of weights exceeds a given threshold, generally put at 1.

A firing function $f$ is defined for some $a \in A, t \in \mathbb{Z}$ by: $f(a, t) = 1$ if $a$ fires at time $t$, and 0 otherwise; its domain of definition determines a set of firings. This set is consistent if all the firings satisfy the firing condition stated above.

Cascades of firings

By connectivity, the firing of neurons progresses through a neural net and produces cascades of firings. Such cascades are formally represented by track expressions.

Given a neural net $A$, the basic cascade consists of the activation of a single neuron $a$ at time $t$. The corresponding track expression is simply $a(t)$.

From this, composite track expressions $x(t)$ are built up recursively. They all have the form

$$x_b(t) = \{\{x_{a_1}(t-1)\}, \ldots, \{x_{a_n}(t-1)\}\} \xrightarrow{t} x_{a_{n+1}}(t+1).$$

This expression denotes the fact that the activation of the “input” cascades $x_{a_1}, \ldots, x_{a_n}$ at time $t - 1$, by firing $b$ at time, $t$ activates the “output” cascade $x_{a_{n+1}}$ at time $t + 1$. 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 $x_{a_1}, \ldots, x_{a_{n+1}}$.

If the context allows, we shall drop subscripts and time.

Composite track expressions, denoting prolonged cascades, result by substituting track expressions for some or all of the key neurons in $x(t)$ as follows:
If \( c(t') \) is a basic track expression occurring in an input cascade anywhere in \( x(t) \), then the result of substituting a track expression \( y_c(t') \) for \( c(t') \) is again a track expression.

If \( c(t' + 1) \) is the output neuron of a track expression whose key neuron is \( d \), then a track expression \( y_c(t' + 1) \) may be substituted for \( c(t' + 1) \), provided that the track expression \( d(t') \) is one of its inputs.

A track expression is consistent, if the firing set given by the subscript notation is consistent with the given neural net \( A \).

**Firing Patterns**

Our formal model presupposes a given neural net \( A \).

A firing pattern is an set of track expressions whose combined firings are consistent with \( A \), representing a set of possibly interrelated cascades of firings. Firing patterns are the basic objects of our theory.

There is a formal similarity between the graphs of functions in analysis (considered as pairs of arguments and values), and firing patterns; these also embody functions, mental functions.

**Composition**

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 a firing pattern to another. Recall 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 denotes what new firings this firing produces. The same is true for sets of 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 $M$, on $N$ as follows:

$$M \cdot N = \{ x : \text{there is an element } \{x_1(t-1), \ldots, x_n(t-1)\} \xrightarrow{t} x(t) \text{ in } M \text{ such that } \{x_1(t-1), \ldots, x_n(t+1)\} \subseteq N \}.$$ 

By this definition $M \cdot N$ determines a consistent firing pattern.

Applying a firing pattern $M$ to more than one firing patterns, say $N_1$ and $N_2$ may be accomplished in various ways:

$(M \cdot N_1) \cdot N_2$ turns $M$ into a binary operation;

$M \cdot (N_1 \cup N_2)$ is what is most often understood as applying $M$ to both $N_1$ and $N_2$.

Observe however, that $M \cdot (N_1 \cup N_2)$ may be larger than $M \cdot N_1 \cup M \cdot N_2$. Note also, that $N_1 \cup N_2$ is not necessarily consistent if $N_1$ and $N_2$ are.

Conforming to the motivation for the model, namely that all objects represent firing patterns of one active brain, we demand closure under union. This is the case in our model for example under the stronger assumption that the firing function $f$ is simultaneously defined over all of $A$ and all relevant time instances.

**Definition 1 (Neural Algebra)** A set of consistent firing patterns over a weighted directed graph $A$ gives rise to an algebraic structure. If this structure is closed under union it is called a Neural Algebra and denoted by $N_A$. 


3 Conceptual Structures in Neural Nets

What is the relation between the structures in a neural net and their function?

Even before that: How do we decide on the conceptual definition and selection of such structures and functions?

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 basis of the **neural algebra approach** is the idea that mental processes are distinguished by giving them well accepted names, names of *concepts* if they have proved to be stable and express the gist of the matter. In our approach, such concepts correspond to *firing patterns* and thus become available to investigation as objects in a neural algebra.

This 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 those firing patterns that have a chance to be truly relevant. We approach this by using a mathematical criterion, namely of being stable concepts in the sense that they act as retractions, as specified below.

**Predication.**

Elements $R$ of the neural algebra $\mathcal{N}_A(F)$ 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 on the objects to which it is to be applied. 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 accidental or irrelevant aspects.

We identify abstract concepts in $\mathcal{N}_A$ with the corresponding predication, considered as an abstraction operation: 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 conceptual content of $X$.

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 vain, a predication $R$ makes only sense if it is *sustained* for a time
interval \([t_1, t_2]\), called appreciable if \(t_2 - t_1 > \nu\) for some arbitrarily fixed number \(\nu\), say \(10^5\). This means that the set of firing times of the key neurons of \(R\) cover an appreciable time interval, the sustension interval of \(R\).

The composition of sustained firing patterns may not be sustained.

Two sustained firing patterns \(X\) and \(Y\) are proximate, if their sustension intervals overlap for an appreciable subinterval. We write \(X \approx Y\) in this case.

Accordingly, we define:

**Definition 2 (Neural Concepts)** A firing pattern \(R\) is a abstract concept if it satisfies the retraction equation \(R \cdot (R \cdot X) = R \cdot X\) for all \(X\).

A firing pattern \(R\) represents an embodied concept if it is a sustained firing pattern \(R\) and has the following property:

All sustained inputs \(X\) for which both \(R \cdot X\) and \(R \cdot (R \cdot X)\) are sustained satisfy \(R \cdot (R \cdot X) \approx R \cdot X\).

In most examples below, the sustension intervals of firing patterns are given in the following context:

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.

Given a time interval \(t_0 \leq t \leq t_1\) and a track expression \(x(t)\) then

\[
\{x(t)\}_{t_0}^{t_1} = \{x(t) : t_0 \leq t \leq t_1\}.
\]

Familiar concepts are typically based on sustained firing patterns, having an episodic character, and can be described as scripts or as memories:

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.

Fleeing upon being threatened may serve here as a simple example of a script; it is based on the embodiment of the following instinctive reaction, as expressed by the track expression

\[ s_c(t) = \begin{cases} 
(u(t-2), v(t-2)) \overset{t-1}{a} w(t), & \{p(t-2)\} \overset{t-1}{b} (t)q \overset{t}{c} m(t+1) \\
\end{cases} \]

\( u \): it’s big, \( v \): it moves fast towards me, \( a \): it’s dangerous, \( w \): watch carefully.

\( p \): its an enemy, \( b \): I’m in danger, \( q \): cry alert,

\( c \): decide to flee, \( m \): flee!

The corresponding firing pattern, the script of this instinct, is simply \( \{s_c(t)\}_{t=0}^t \). Figure 1 highlights the cascade which correspond to the track expression \( s_c \) of the example above, the reaction upon a threat.

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

**Theorem 1 (Structure Theorem for Concepts in Neural Nets)** Given a firing pattern \( S \) with distinguished key neurons \( s_1, \ldots, s_k \), there exists a solution \( R \) of the retraction equation \( R \cdot (R \cdot X) \approx R \cdot X \), where \( R \) is the firing pattern with key neuron \( r \), resulting from augmenting the net for \( S \) by attaching a cycle linking \( r \) to each \( s_i, i = 1, \ldots, k \).

Proof part 1. Consider the firing pattern of a concept based on a single track expression \( s \) (for example \( s_c \) from above), and choose some neuron \( r \). Define
Figure 1: Fleeing on a Threat.

\[ S = \{(s(t))^{t+1} \rightarrow s(t + 2)\}_{t_0}^{t_1}, \]

and let \( X \) be sustained for a subinterval \([t'_0, t'_1]\) of \([t_0, t_1]\). Then clearly \( S \cdot (S \cdot X) \approx S \cdot X \), since

\( S \cdot X = \emptyset, \) if \( s(t'_0) \notin X; \)

\( S \cdot X = \{s(t')\}_{t'_0}^{t'_1}, \) if \( s(t'_0) \in X. \)

Illustration,(fig.1) : Consider the track expression \( s \) as being activated for a time interval \([t_0, t_1]\), e.g. intended to teach a movement or presenting a picture. In the example above: Attaching a cycle to \( c \) through a new neuron \( r \) produces the concept of fleeing upon threat, (red arrows.)

Proof, part 2: Higher-Order Concepts

Conceptual circuits may themselves become involved in other complex
neural circuits and may in this way give rise to second- and higher-order concepts.

The proof of the retraction equation in these cases follows a simple pattern: Let $M$ be the object to be conceptualized (it may itself include conceptual elements), and let $r$ be a new neuron, the *reference neuron* to $M$. This neuron is connected to the key neuron of each element of $M$, (including the reference neurons of its conceptual elements) by a simple cycle as above.

Let $R = \{\{s(t)\} \xrightarrow{t+1}{r} s(t+2) : s(t), s(t+2) \in M\}$, Then $R \cdot (R \cdot X) \approx R \cdot (X \cap M) \approx (X \cap M) \cap M = X \cap M = R \cdot X$.

A simple example of a second order concept is the concept $R$ of the association of two concepts $S_1$ and $S_2$ : ”upon $S_1$ follows $S_2 ”$, e.g. one script follows another.

Consider $S_1$ as the conceptualization of the script $s_1$, $S_2$ of $s_2$, and let the concatenation of these concepts be established by a neuron $c$ which links the reference neurons $a$ and $b$ of these two concepts:

Denote

\[ x_c(t+2) = \{\{s_1(t)\} \xrightarrow{t+1}{a} s_1(t+2)\} \xrightarrow{t+2}{c} \{\{s_2(t+2)\} \xrightarrow{t+3}{b} s_2(t+4)\}. \]

The association of $S_1$ with $S_2$ is realized by a new key neuron $r$. The *association concept* is therefore defined by

\[ R = \{\{x_c(t+2)\} \xrightarrow{t+3}{r} x_c(t+4)\}_{t_0}^{t_1} \]

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

As an example, we return to the flight-upon-being-threatened example. Choosing a firing track with key neuron $b$, and one with the key neuron $m$,
we recruit corresponding concept neurons \textit{danger} and \textit{flight}. These are then linked by an intermediate neuron which is then the key neuron of the concept of this association with concept neuron \textit{flight} \textit{on} \textit{danger}, (Fig.2, red arrows).

To the corresponding figure we have added a component that illustrates a hypothesis on so called mirror neurons: Reading an illustrated scary story which introduces a dangerous situation, a certain neuron is activated which we link to the neuron \textit{a} in the "flight upon threat" net. As such, \textit{a} mirrors the reading experience of being scared. The said net then activates \textit{w},
which mirrors the experience of feeling scared. Its link to a reader’s neuron induces the sought-for pleasurable experience, (Fig.2, green arrows).

This model of mirroring, to reflect the conceptually higher order of the reading activity, should perhaps be lifted to connecting the reader’s neurons to conceptual neurons based on key neurons $a$ and $w$ rather than to the ”instinctive” neurons themselves.

Concepts and scripts can be enchained, conceptual memories can be associative, and both can be combined to create more complex entities, which we will call scenarios. Scenarios are the main building blocks for the activities of the brain, conscious and unconscious. An example is described in the appendix, the scenario of an arithmetic lesson in first grade.

4 Consciousness in Neural Nets

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” (5)–(12): it has been called ”the ultimate intellectual challenge in the new millenium”. 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 (13) to quantum theory (14), geometry (15, information theory (16)) and statistics (17), and relating them to neurological facts and psychological experiments.
Consider the brain as an – enormously complex – mechanism of interlocking processes. Consciousness may then 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 (18) as one of the major applications of his cybernetics.

In our context: If the said mechanisms in the brain are firing patterns, we are led to the following.

**Definition 3 (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:

Let $B$ comprise the firing patterns corresponding to the context of observing, acting, planning, moving, etc., and let $C$ be the prospective firing pattern of ”consciousness”. Then $B \cdot C$ is the results of observing, acting, etc. as dependent on consciousness, and $C \cdot B$ represents the action of consciousness on these activities. To the sum of these results, together with $C$ itself, $C$ is again applied.

*Abstract Consciousness*. Translated into neural algebra, the above definition of consciousness transforms into an equation of the form

$$C \cdot (C \cup B \cdot C \cup C \cdot B) = C.$$
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” (19). 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.

**Theorem 2 (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

\[
\varphi\left(\bigcup D\right) = \bigcup_{X \in D} \varphi(X).
\]

These set are consistent firing pattern since the union of directed consistent sets is again consistent with \( A \). 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 for a given firing pattern $B$ in $\mathcal{N}_A$ constitute the lattice of abstract states of consciousness of the active ”brain” $B$.

**Sustained Consciousness.**

The lattice of solutions of the consciousness equation can best be interpreted if we consider them as temporarily sustained activities. In this sense, consciousness may move from one temporary state to another in the lattice, analogous to the familiar experience of the shifting focus of consciousness. Thus, *sustained consciousness* is based on a sustained brain $B$ and is constituted by sustained solutions of the approximate form of the consciousness equation above.

Again, the question arises how to characterize firing patterns and their neural correlates for solutions to the consciousness equation.

**Theorem 3 (Structure Theorem for Consciousness in Neural Nets)**  

*The set of solutions $C$ of the consciousness equation*

\[
C \approx C \cdot (C \cup B \cdot C \cup C \cdot B).
\]

*has the following properties:*

1. $C$ has a base in one or more cycles of the directed graph.
2. $C$ can be expanded along any outgoing edge.
3. $C$ 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 sequentially with weights 1, and firing at all times \(t, t \in [t_0, t_1]\).

The firing patterns \(C_i\) of the neurons \(a_i\) are therefore defined recursively by:

\[
C_i = \{a_i(t)\}_{t_0}^{t_1} \cup \{(x_{a(i-1)}(t-1)) \xrightarrow{t_{a_i}} x_{a(i+1)}(t+1)\}_{t_0}^{t_1},
\]

with \(x_{a(i-1)}(t-1) \in C_{i-1}, x_{a(i+1)}(t+1) \in C_{i+1}\).

Observe that \(C_2 \cdot C_1 = C_3\), etc.

Restrict the brain \(B\) to this cycle.

Taking \(C\), and here also \(B\) as the union of the \(C_i\), we obtain \(C \cdot C \approx C\) and therefore

\[
C \cdot (C \cup B \cdot C \cup C \cdot B) \approx C \cdot (C \cup C \cdot C) = C.
\]

Proof of part 2: It suffices to consider a cycle consisting of just one node \(a\) with a second node \(c\) connected to \(a\) via a node \(b\), all by edges of weight 1. The corresponding firing pattern is defined recursively by:

\[
A = \{a(t), c(t) \xrightarrow{t+1_b} a(t+2), x_a(t) \xrightarrow{t+1_a} x_a(t+2) : x_a \in A\}_{t_0}^{t_1}.
\]

Note that \(c(t) \xrightarrow{t+1_b} a(t+2)\) is not in \(A \cdot A\) for any \(t\). Therefore the consciousness equation does not hold for the extended cycle \(A\).

The proof of part 3 for the cycle at \(a\) with an edge leading away from the
cycle can be illustrated by having an edge from \( a \) leading to a cycle at \( b \).

The cycle at \( b \) has the firing pattern

\[
B = \{ (t), x_b(t-1) \xrightarrow{t}{b} x_b(t+1) : x_b \in B \}_{t_0}^{t_1}.
\]

Attaching the cycle at \( b \) to the cycle at \( a \) we obtain the firing pattern

\[
A = \{ (a(t), x_a(t) \xrightarrow{t+1}{a} x_a(t+2), x_a(t) \xrightarrow{t+1}{b} x_b(t+2), x_b(t) : x_a \in A, x_b \in B \}_{t_0}^{t_1}.
\]

For this \( A \) we do have \( A \cdot A \approx A \), and \( A \) solves the consciousness equation.

*Consciousness, concepts and the neural mind.*

Concepts are, by their connectional structure, candidates for inclusion in solutions to the consciousness equation, attaching them at various points of a basic cycle of consciousness, and activating them at various times by convenient triggers. The lattice structure of the set of solutions thus 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 connections according to the firing history, constituting what one might reasonably call the *neural mind*.

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

5 Concepts and Consciousness in the Large

The notions of concept and consciousness were modeled above on neural nets whose level of abstraction from the psychophysical brain was rel-
atively modest; they related more closely to the individual neurons and their connections rather than to any kind of overall organization. However, much of present neuroscience is concerned with a higher level of organization (20), (21): One considers brain areas that have been identified as being involved in specific functions (22), and investigates their connectivities and functional dependencies (23). Such a view of the organization of the brain can be obtained by suitable abstraction from a neural algebra $\mathcal{N}_A$:

First, select a subset $A'$ of $A$, for example conceptual neurons of an appropriate order. Let $\pi$ be a partition of $A'$ into neural ensembles. The track expressions $x(t)$ of $\mathcal{N}_A$ are lifted to track expressions on these ensembles as follows:

If $a$ is any neuron in $A$, let $\bar{a}$ denote that assembly in the partition of $A'$ to which $a$ belongs.

Time variables $t$ are now understood to denote intervals of sustension, so $\bar{a}(t)$ expresses that the brain area $\bar{a}$ is activated during the time interval $t$.

The partition $\pi$ is assumed to respect the firing laws of $A$. For example, the track expression

$$\{\bar{a}(t'), \ldots, \bar{b}(t')\} \xrightarrow{e} \bar{d}(t''')$$

assumes that the track expressions for all elements of the corresponding ensembles hold uniformly and for all appropriate time instances in the corresponding time intervals $t', t'', t'''$. The definitions of composite track expressions and firing patters are lifted the same way and determine the quotient algebra $\mathcal{N}_{A'}/\pi$, the functional connectome, which now models the structural/functional dependencies between the chosen neural ensembles.
The notions of concept and consciousness transfer without change to the functional connectome.

Consider the story of Chicken Little, (fig.3). Attached to a cycle, which we may call "core consciousness", three extensions of consciousness are depicted, namely "feeding", "brooding" and the story: "Chicken Little, when something fell on its head, was sorely frightened, called out that the sky was falling; by which it was convinced that the sky was indeed falling and continued to be frightened." –The lattice of the solutions of the
consciousness equation for this chicken brain is what would be called its mind.

*The Lattice of Functionality.* Note that the choice of sets $A'$ and their partition $\pi$ constitutes a hierarchy of possible neural algebras $\mathcal{N}_A/\pi$, indeed again a lattice, typically determined by increasingly detailed knowledge of the structure of the underlying brain. This allows some freedom in the interpretation of the model. It serves to illustrate these notions by graphic examples, and to relate them to proposals for brain functionalities in the literature.

### 6 Challenges

#### Time Development and Learning

The concept of learning (24) would merit more than the following few remarks.

Technically, we have so far tacitly subsumed learning in the firing history of $\mathcal{N}_A$: First, the neural net $A$ may comprise the totality of all neurons that are ever considered in the model. Second, the firing function $f$ is defined pointwise; this 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.

The proposed solution is to modify the definition of $\mathcal{N}_A$ and make the weight of synapses dependent on the previous firing history: Let $w_{a,b}(t)$ denote the weight given to the edge from $a$ to $b$ at time $t$. Recall the critical length $\nu$ of ”appreciable” time. The value of $w_{a,b}(t)$ is determined by some learning function $F(\mu_t)$ dependent on the course of values $\mu_t$ of $f(a, t')$ in
the time interval $[t - \nu, t]$:

$$w_{a,b}(t) = F(\mu_t).$$

This firing function is consistent if it is consistent with the firing laws of $A$, (now of course dependent on the timely values of the weights computed by $F$.)

The neural algebra is thus given by a directed graph $A$, a learning function $F$ and a firing function $f$ assumed consistent with $F$ and $A$.

By restricting the neural algebra $\mathcal{N}_A$ to subsets $A_i$ (with adapted weights) and to periods $T_i$, we may describe the progression of learning and forgetting in the brain as a continuation:

$$\mathcal{N}_{A_1}(T_1) \Rightarrow \mathcal{N}_{A_2}(T_2) \Rightarrow \mathcal{N}_{A_3}(T_3) \Rightarrow \ldots$$

and using corresponding partitions $\pi_i$.

Again, the choice of the learning function $F$, the sets $A_i$ and their partition $\pi_i$ can only be determined by increasingly detailed knowledge of the structure of the underlying brain.

Animal, Social and Artificial Consciousness

The consciousness of animals, of which Chicken Little is a curious example 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 (25); much additional information has been obtained, approximating total neural modeling (26). 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 (27), 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 (28).

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 interacting highly parallel processes. One may speculate about corresponding future architectures for interlaced memories and distributed programs.

*Algebras of the mind: Logical Challenges*

Algebras of the mind, such as Booles Laws of Thought, have as their objects mental concepts such as propositions and link them by operations using copula inherited from linguistics such as *and*, *or*, *not*. Mathematical logic has since developed this conception in various ways, including non monotonic logics and logics of knowledge.

Curry (29) in a quite different direction, attempting to deal with the so-called crisis in the foundations of mathematics, invented Combinatory Logic, whose objects may be interpreted for the purposes of this paper as thoughts. There is only one operation, that of applying thoughts to thoughts; thus, mental objects and mental activities are conjoined into one category of
mathematical objects, a concept taken up in this paper.

As algebraic structures, neural algebras $\mathcal{N}_A$ are closely related to models of combinatory logic. Indeed, for rich graphs $A$ 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 the logic of *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 selection of concepts to be considered raises a basic epistemological problem: There is the danger to be entrapped 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 (consciously) acting entities.

6 **Looking for Confirmation**

The present author, fascinated by the challenges of neuroscience, but greatly intimidated by the enormous literature – remaining within his field of com-
petence and scientific background, applications of combinatorial algebra –
developed the present mathematical model of mental functions and their
neural embodiment.
But all that mathematical models such as ours can provide for neuroscience
is explanation and prediction of selected aspects.

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:

1. The neural algebra construction of ”concept” consists of centering it on
a single neuron, (Theorem 1):
Single neurons have been identified as the key to recognize a face, (30).
Single neurons serve as mirror neurons, called upon when a concept, e.g.
a feeling, needs to be associated to a concept pertinent to it (31).
Recruiting new neurons to create new concepts and abilities has been iden-
tified (32),(33), and shown to be involved in the learning of bird songs (34)
– (36), and in reading.

2. The characterization of consciousness as based on linked cycles of par-
tial consciousness and concepts, (Theorem 3, parts (1) and (2)):
Recurrent or reentrant connectivities in the brain have been recognized to
be involved in conscious activities, e.g. in the visual cortex (37), (38),
and more generally in linked convergence–divergence–zones and regions,
(39),(11, Chapt.6).

3. Ignoring inputs, (Theorem 3, part (3)).:
Multiple experiments have shown that incoming signals, e.g. from the
primary visual cortex, are not integrated into consciousness.
The explanatory power of our model is surely helpful, in particular if it is combined with mathematical techniques, developed in combinatorial algebra for the formulation and solution of equations. This would allow to formally express hypotheses on the interrelations between brain functions as represented by objects in the neural algebra of a brain and their embodiment.

Acknowledgment. The author thanks his friends, collaborators (40), anonymous referees (41) and various audiences of presentations for their interest, patience and help.

Appendix: First Grade Arithmetic, a Scenario.

Narrative

Imagine a classroom in first grade. The teacher intones "two plus three" and then, turning to little Alice, prompts her "equals what?"

This is how this scenario may play in little Alice’s little brain, (fig.4, fig.5): The relevant part of the brain is shown in layers reminiscent of cortical layers. The top layer consists of neurons corresponding to phonemes, activated by and when these phonemes are presented by auditory circuitry. By a network of cascades, roughly sketched here, these phonemes are integrated and activate the neurons \( c_1, c_2, c_3 \) on the next layer. Their activation means that the word or phrase constituted by the corresponding sequence of phonemes is recognized. They are the reference neurons of the corresponding word concepts two, plus three, equals?. The operations neurons \( o_1, o_1, o_4 \) use the reference neurons \( c_1 \) and \( c_2 \) as input to activate the reference neuron (5) of the desired arithmetical result, the concept 5.
The operation neuron $o_3$ activates the reference neuron (!) of the prompting concept !. Finally, $o_3$ activates the cascades attached to the concepts 5, 2, 3, !. Because only 5 and ! are active, the output generated is the desired "five, Miss". This is what Alice responds. "Well done", says the teacher and Alice blushes prettily (not shown.)

Figure 4 shows (part of) the brain structure, and Figure 5 the activation
history detailed in the above narration. The weights are omitted in fig.4 but are understood to legalize the activation history in fig.5.

The formal counterpart to our narration is a set of neural algebra objects of which we give the following examples.

The objects $A_k$ represent the activation of the recognition neurons $c_k$, $k = 1, 2, 3$. They integrate the input phonemes by the cascades represented by track expressions $x_{b_{i,k},j}$, $k = 1, 2, 3$, $i = 1, \ldots m_k$, $j = 1, \ldots n_{i,k}$.

$$A_k = \{\{x_{b_{i,k,1}}, \ldots, x_{b_{i,k,n_{i,k}}}\} \rightarrow c_{i,k}\}, i = 1, \ldots, m_k.$$ 

Not to overload these expressions, we omitted all time arguments of track expressions and time indices of arrows and sets. Thus, $A_i$ designates the set of all expressions with time notations supplied.

$$C_1 = \{A_1 \rightarrow y : y \in \text{two}\},$$

with

$$\text{two} = \{\{w\} \rightarrow w : w \in T\},$$

where $T$ is the set of track expressions constituting the concept of the word ”two”. Similarly for the other word and number concepts. The arithmetic operation of addition is represented by the object

$$Add = \{\{c_1\} \rightarrow [c_1][o_3], \{c_1, c_2\} \rightarrow (5)\},$$

and so on.

**Pedagogy**

The pedagogue appreciates the difficulty of designing the neural net for first grade arithmetic. Our teacher has trained her pupils to do addition by
the method of fixed addends: \texttt{plusthree}, \texttt{plusfour}, \ldots; perhaps she has in mind to later teach minus three, minus four, etc.

Dependent on current methodology taught at her college, she might have opted for rote learning, creating phrase concepts \texttt{twoplusthree}, \texttt{threeplusfour}, \ldots, linking these to the corresponding number concepts 5, 7, \ldots. Or, she may be an adherent to the ”set theory” trend, and teach the concept \texttt{plus} with links from two, three, \ldots to the corresponding number concepts, thinking of teaching multiplication the same way. – These choices have an influence on student performance, to which have to be added the effects of missing links and incorrect weights resulting from misunderstandings and inattention. Thus we could add another neural–whatever to academic disciplines, (if it does not already exist.)

\textit{On speech learning}

Observe that the number five appears twice as a concept: as the concept of the spoken word and as an arithmetical concept. To learn to speak ”five”, one could imagine a learning feedback in the brain which compares vocalizations and auditory inputs in order to adapt the weights in the speech tracts of the number concept 5. Appreciating the technological advances in speech learning and speech generation on the one hand, and research in, say, bird song learning, it is not improbable that the proposed dualism between five and 5 etc. may be elucidated one day.

\textbf{References}


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


