Neuronal Synchronization, Covariation, and Compositional Representation

Markus Werning

This paper tries to unify a widely held semanticist view on the nature of meaning and content with two central neuro-scientific hypotheses on the role of cortical neurons. The semanticist view presupposes the covariation of concepts with their contents as well as the compositionality of meaning and content. On the side of neuroscience, the existence of cortical feature maps and the mechanism of neural synchronization is assumed. A neural correlate of a semantics is proposed that covers not only the propositional case, the case of first order predicate languages, but also modal-logical structures.[1]

1 Covariation and Compositionality

The semanticist view I appeal to characterizes the triangle between language, mind and world roughly as follows: Linguistic utterances are expressions of meaning. Those meanings are mental representations, which are often called concepts. Concepts again have an external content and this content is responsible for an utterance having reference or denotation. Here is an example: The utterance ‘dog’ expresses a mental concept – let’s call it [dog]. This concept has a certain content and thereby relates the utterance to its denotation in the world: dogs, doghood, sets of dogs or sets of possible dogs, depending on your favorite semantic theory. This story tells us how utterances can be *about* things in the world or, in other words, how one can speak of dogs by means of the word ‘dog’.

Leaving aside what mechanism underlies the relation between mental representation and the production of phonological sequences, semanticists of the kind described endorse the view that the relation between concepts and their

Address for correspondence: Department of Philosophy, Heinrich-Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany.
E-mail: werning@phil.uni-duesseldorf.de.

[1] This paper is designed to supplement and extend earlier publications on this issue (Werning, 2001, 2003, 2005b) and, in parts, builds on them.

The Compositionality of Meaning and Content.
Volume II: Applications to Linguistics, Psychology, and Neuroscience.
Edited by Edouard Machery, Markus Werning & Gerhard Schurz.

content is some relation of covariation. A concept has the content it has because it co-varies with certain thing and not with others. Co-variation between concepts and their contents is a causal-informational relation of sorts and has been explored in the literature (Fodor, 1992). The sense in which conceptual contents are *responsible* for linguistic expressions having denotation – I assume for reasons of simplicity – is identity (More complex relations between content and denotation may be viable, too). The denotation of an utterance is identical to the content of the concept the utterance is an expression of. This view is captured by our first hypothesis:[2]

Hypothesis 1 (Covariation with Content). *An expression has the denotation it has because the concept it expresses reliably co-varies with a content that is identical to the expression's denotation.*

Since natural languages have a rich constituent structure, it is rather plausible to a assume that the structure of their meanings is complex, too, and that the structure of meanings in some way or another resembles the structure of their expressions. Now, the most simple way to spell out this relation of resemblance is by means of a structural match, in technical terms: a homomorphism. This homomorphism is spelled out by the principle of the compositionality of meaning:

Hypothesis 2 (Compositionality of Meaning). *The meaning of a complex expression is a syntax-dependent function of the meanings of its syntactic constituents.*

It would be surprising, furthermore, if the covariation relations between primitive concepts and their contents should not in some way or another contribute to the covariation relations between complex concepts and their contents. The quest for simplicity again leads us to the hypotheses that the contents of the primitive concepts are the *sole* factors to determine the content of a therefrom combined complex concept. Again, this is just what the principle of compositionality says for contents, our third and last semanticist hypothesis:

Hypothesis 3 (Compositionality of Content). *The content of a complex concept is a structure-dependent function of the contents of its constituent concepts.*

[2] I conceive of denotation in a broad sense as *modal denotation* and distinguish between reference and denotation. The denotation of an expression is a function from possible worlds to the referents of the expression in those possible worlds. The denotation of a sentence p, e.g., is the set of pairs $\{\langle w, t\rangle | p$ has the truth-value t in the world $w\}$, while its referent is the truth-value it has in the actual world. Little depends on the particular view one assumes with regard to the nature of denotation in our context.

The three semanticist hypotheses, though not uncontroversial, are at the core of many contemporary semanticist theories. They may thus serve us as a starting point for our reductive project.[3]

2 Neuronal Reduction

The aim of this paper is to make out a neuronal structure $\mathscr{N} = \langle N, \Sigma_N \rangle$ that fulfills the three semanticist hypotheses: co-variation with content as well as compositionality of meaning and content. The neuronal structure shall consist of a set of neuronal states N and a set of thereon defined operations Σ_N. Since the three semanticist hypotheses may serve as (minimal) identity criteria for concepts, their fulfillment by a neuronal structure will justify us in identifying the neuronal structure with a structure of concepts. The three hypotheses hence form the adequacy conditions for a neuronal reduction of concepts.

Since a large part of natural language can be paraphrased by means of a formal monadic first order predicate language with identity, the adequacy conditions may be formalized for that case in the following way (In the formalization of compositionality I follow Hodges, 2001; see also Werning, 2004):

Principle 1 (Adequacy Conditions). *Let*

$$PL^{=} = \langle L, \Sigma_L \rangle$$

be a monadic first order predicate language with identity. Let it comprise the set of grammatical terms L and the syntactic operations of identity, predication, negation, conjunction, disjunction, implication and existential as well as universal quantification:[4]

$$\Sigma_L = \{\sigma_{=}, \sigma_{pred}, \sigma_{\neg}, \sigma_{\wedge}, \sigma_{\vee}, \sigma_{\rightarrow}, \sigma_{\exists}, \sigma_{\forall}\}.$$

Let there furthermore be a denotation function

$$\nu : L \rightarrow W$$

that maps the grammatical terms of $PL^{=}$ onto their denotations and let this function of denotation be compositionally evaluable by a worldly structure of denotations

$$\mathscr{W} = \langle W, \Sigma_W \rangle.$$

[3] My appeal to simplicity arguments in favor of compositionality has to do with my reluctance to accept the three most often cited reasons for compositionality, namely productivity, systematicity, and inferentiality (Werning, 2005a). I am aware that the simplicity arguments lack the force of a strict argument or even a proof. They are therefore marked as hypotheses.

[4] For an exemplification of the syntactic operations see the mappings (28) below.

That is: For every syntactic operation $\sigma \in \Sigma_L$, there is a function $\nu_\sigma \in \Sigma_W$ such that for every non-atomic grammatical term $\sigma(t_1, ..., t_n) \in L$ the following equation holds:

$$\nu(\sigma(t_1, ..., t_n)) = \nu_\sigma(\nu(t_1), ..., \nu(t_n)).$$

Then any neuronal structure

$$\mathscr{N} = \langle N, \Sigma_N \rangle$$

is a structure of internal representations expressible by $PL^=$ if and only if the following three conditions hold:

(a) $\mathscr{N}$ is a compositional semantics of meanings for $PL^=$, that is: There is a function

$$\mu : L \rightarrow N$$

and for every syntactic operation $\sigma \in \Sigma_L$, there is a function $\mu_\sigma \in \Sigma_N$ such that for every non-atomic grammatical term $\sigma(t_1, ..., t_n) \in L$ the following equation holds:

$$\mu(\sigma(t_1, ..., t_n)) = \mu_\sigma(\mu(t_1), ..., \mu(t_n)).$$

(b) $\mathscr{N}$ is compositionally evaluable with respect to content, that is: There is a function

$$\kappa : N \rightarrow W$$

and for every operation $h \in \Sigma_N$, there is a function $\kappa_h \in \Sigma_W$ such that for every neuronal element $h(m_1, ..., m_n) \in N$ the following equation holds:

$$\kappa(h(m_1, ..., m_n)) = \kappa_h(\kappa(m_1), ..., \kappa(m_n)).$$

(c) The elements of $\mathscr{N}$ reliably co-vary with their contents, the elements of $\mathscr{W}$, such that for every grammatical term $t \in L$ the following holds:

$$\nu(t) = \kappa(\mu(t)).$$

3 Neurobiological Evidence

For many feature dimensions (color, orientation, direction, size, etc.) involved in the course of visual processing one can anatomically identify so-called neuronal feature maps (Hubel & Wiesel, 1968). These are clusters of clusters of neurons that exhibit a certain topological organization (see Fig. 1). With regard to one feature dimension, one finds a pinwheel-like structure for each receptive

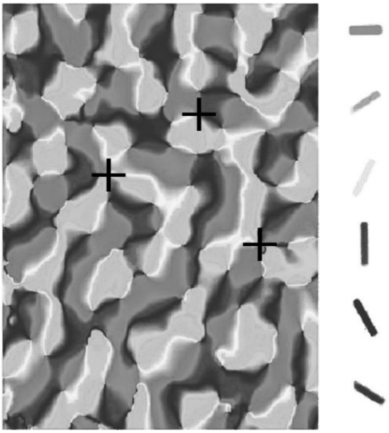

Figure 1: Feature map. Optical image of an orientation map in the primary visual cortex of a macaque. Shadings code the preferred orientations of neurons as indicated by the bars on the right. Three exemplary pin-wheel centers are marked by black crosses. The horizontal extent is 3.3 mm. Adapted from Obermayer and Blasdel (1993).

field (i.e., a specific region of the stimulus). This structure is called a hypercolumn. It typically has an extent of about $1\,\mathrm{mm}^2$. For each receptive field or, correspondingly, each hypercolumn, neurons for the entire spectrum of features in the respective feature dimension (e.g., orientation) fan out around a pin-wheel center. Neurons of a hypercolumn with a tuning for one and the same feature (e.g., verticality) form a so-called column.

A features map thus is an assembly of hypercolumns, one per receptive field. Neurons of neighboring hypercolumns are selective for properties that are instantiated in neighboring receptive fields on the stimulus. This means, there is some topological correspondence between the neighbor relations of hypercolumns in a feature map and the neighbor relations among receptive fields in the stimulus. Within one hypercolumn we, furthermore, have a topological correspondence between the neighbor relations of columns and the similarity relations of the features for which the neurons of each column select. Neurons of neighboring columns select for similar features.

More than 30 so organized cortical areas, which occupy approximately one-half of the total cortex, are experimentally known to be involved in the visual processing of the monkey (Felleman & van Essen, 1991), less are known for humans. These findings justify the following hypothesis:

Hypothesis 4 (Feature maps). *There are many cortical areas that function as topologically structured feature maps. They comprise clusters of neurons whose function it is to show activity only when an object in their receptive field instantiates a certain property of the respective feature dimension.*

The fact that features which belong to different feature dimensions, but may be properties of the same stimulus object are processed in distinct regions of cortex, poses the problem of how this information is integrated in an object-specific way. How can it be that the horizontality and the redness of a red horizontal bar are represented in distinct regions of cortex, but still are part of the representation of one and the same object? This is the binding problem in neuroscience (Treisman, 1996).

A prominent and experimentally well supported solution postulates neuronal synchronization as a mechanism for binding (von der Malsburg, 1981; Gray, König, Engel, & Singer, 1989): Neurons that are indicative for different properties sometimes show synchronous activation, but only when the properties indicated are instantiated by the same object in the perceptual field; otherwise they are firing asynchronously. Synchrony, thus, might be regarded to fulfill the task of binding together various property representations in order to form the representation of an object as having these properties.

The fact that object-specific synchrony has been measured within columns, within and across hypercolumns, across different feature maps, even across the two hemispheres and on a global scale (for a review see Singer, 1999) supports the following hypothesis:

Hypothesis 5 (Synchrony). *Neurons of different feature clusters have the function to show synchronous activation only if the properties indicated by each feature cluster are instantiated by the same object in their receptive field.*

4 Oscillatory Networks

The two neurobiological hypotheses on neuronal feature maps and synchrony allow us to regard oscillatory networks (see Fig. 2) as a plausible model of informational processes in the visual cortex. The design of oscillatory networks is also supported by principles of *Gestalt* psychology that govern the representation of objects.

According to some of the *Gestalt* principles, spatially proximal elements with similar features (similar color/orientation/direction/size, etc.) are likely to be perceived as one object or, in other words, represented by one and the same object concept. If, for example, in a field of densely arranged, randomly moving dots a bunch of neighboring dots are moving in the same direction, you are likely to perceive them as one object. If in a field of randomly arranged, varicolored bars, a group is in parallel and of the same color, we see them as belonging together and forming an object of its own.

The *Gestalt* principles are implemented in oscillatory networks by the following mechanism: Oscillators that select input from proximal stimulus elements with like properties tend to synchronize, whereas oscillators that select input from proximal stimulus elements with unlike properties tend to de-synchronize. As a consequence, oscillators selective for proximal stimulus elements with like properties tend to form out a synchronous oscillation when stimulated simultaneously. This oscillation can be regarded as one object concept. In contrast, inputs that contain proximal elements with unlike properties tend to cause anti-synchronous oscillations, i.e., different object concepts.

A single oscillator consists of two mutually excitatory and inhibitory neurons, each of which represents a population of biological cells. If the number of excitatory and inhibitory biological cells is large enough, the dynamics of each oscillator can be described by two variables x and y. They evolve over time according to the following differential equations:

$$\begin{aligned}\frac{dx}{dt} &= -\tau_x x - g_y(y) + L_0^{xx} g_x(x) + I_x + N_x \\ \frac{dy}{dt} &= -\tau_y y + g_x(x) - I_y + N_y.\end{aligned} \tag{1}$$

Here, τ_ξ ($\xi \in \{x, y\}$) are constants that can be chosen to match refractory times of biological cells. The g_ξ are transfer functions that tell how much of the activity of a neuron is transferred to other neurons. The constant L_0^{xx} describes self-excitation of the excitatory cell population. I_ξ are static external inputs and N_ξ variable white noise, which models fluctuation within the cell populations. With I_ξ above threshold, the solutions of the system of equations (1) are limit-cycle oscillations. For a more detailed description of the network see Maye (2003).

Stimulated oscillatory networks, characteristically, show object-specific patterns of synchronized and de-synchronized oscillators within and across feature dimensions. Oscillators that represent properties of the same object synchronize, while oscillators that represent properties of different objects de-synchronize. In the simulation we observe that for each represented object a certain oscillation spreads through the network. The oscillation pertains only to oscillators that represent the properties of the object in question.

A considerable number of neurobiological studies have by now corroborated the view that cortical neurons are rather plausibly modelled by oscillatory networks (cf. Singer & Gray, 1995; Schillen & König, 1994; Werning, 2001; and Maye, 2003). Together with the simulations described, these studies suggest that the synchrony of oscillations indicates the sameness of objects and that an oscillation pertaining to a neuronal feature cluster indicates that the object indicated by the oscillation has the featured property.

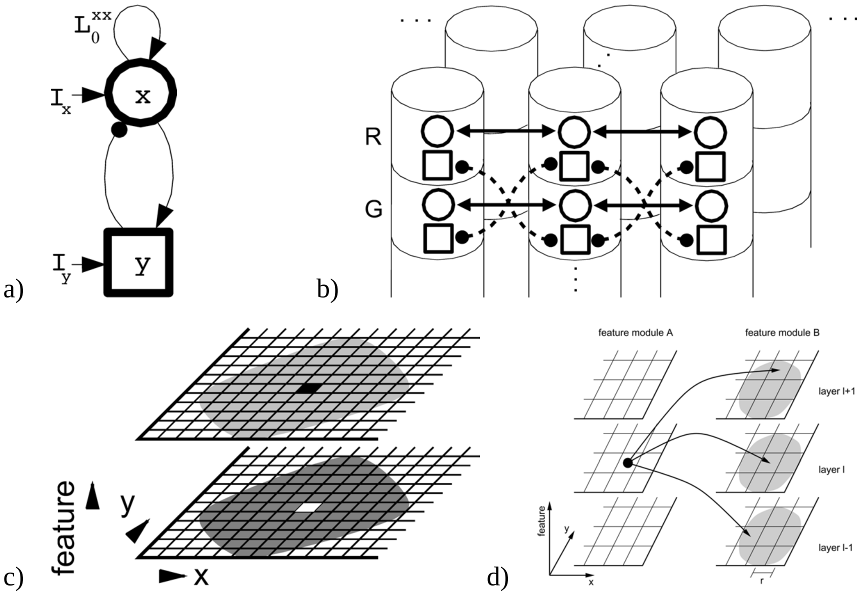

Figure 2: Oscillatory network. a) A single oscillator consists of an excitatory (x) and an inhibitory (y) neuron. Each model neuron represents the average activity of a cluster of 100 to 200 biological cells. L_0^{xx} describes the self-excitation of the excitatory neuron. I_x and I_y amount to external input. b) Synchronizing connections (solid) are realized by mutually excitatory connections between the excitatory neurons and hold between oscillators within one layer. Desynchronizing connections (dotted) are realized by mutually inhibitory connections between the inhibitory neurons and hold between different layers. ‘R’ and ‘G’ denote the red and green channel. The cylinder segments correspond to Hubel and Wiesel’s (1968) columns, whole cylinders to hypercolumns. c) A module for a single feature dimension (e.g., color) consists of a three-dimensional topology of oscillators. There is one layer per feature and each layer is arranged to reflect two-dimensional retinotopic structure. The shaded circles visualize the range of synchronizing (light gray) and desynchronizing (dark gray) connections of a neuron in the top layer (black pixel). d) Two coupled feature modules are shown schematically. The single oscillator in module A has connections to all oscillators in the shaded region of module B. This schema is applied to all other oscillators and feature modules. Reprinted from Werning (2005b) and Maye and Werning (2004).

5 Hilbert Space Analysis

Fig. 3a shows a stimulus we presented to an oscillatory network. The network consists of a color module with layers for redness and greenness and an orientation module with layers for verticality and horizontality. In the stimulus the human observer perceives two objects: A red vertical object and a green horizontal objects. Now, how does the network answer to the stimulus? What can we say about the network dynamics?

The oscillations spreading through the network can be characterized mathematically: An oscillation function, or more generally the activity function $x(t)$ of an oscillator is the activity of its excitatory neuron as a function of time during a time window $[-\frac{T}{2}, +\frac{T}{2}]$. Mathematically speaking, activity functions are vectors in the Hilbert space $L_2[-\frac{T}{2}, +\frac{T}{2}]$ of in the interval $[-\frac{T}{2}, +\frac{T}{2}]$ square-integrable functions. This space has the inner product

$$\langle x(t)|x'(t)\rangle = \int_{-T/2}^{+T/2} x(t)\, x'(t) dt. \tag{2}$$

The degree of synchrony between two oscillations lies between -1 and $+1$ and is defined as their normalized inner product

$$\Delta(x,x') = \frac{\langle x|x'\rangle}{\sqrt{\langle x|x\rangle\langle x'|x'\rangle}}. \tag{3}$$

The degree of synchrony, so defined, corresponds to the cosine of the angle between the Hilbert vectors x and x'. The most important cases are:

$$\begin{aligned}
\Delta(x,x') = 1 &\Leftrightarrow x \text{ and } x' \text{ are parallel (totally synchronous)},\\
\Delta(x,x') = 0 &\Leftrightarrow x \text{ and } x' \text{ are orthogonal (totally uncorrelated)},\\
\Delta(x,x') = -1 &\Leftrightarrow x \text{ and } x' \text{ are anti-parallel (totally anti-synchronous)}.
\end{aligned}$$

6 Eigenmodes

From synergetics it is well known that the dynamics of complex systems is often governed by a few dominating states. These states are the eigen- or principal modes of the system. The corresponding eigenvalues designate how much of the variance is accounted for by a mode.

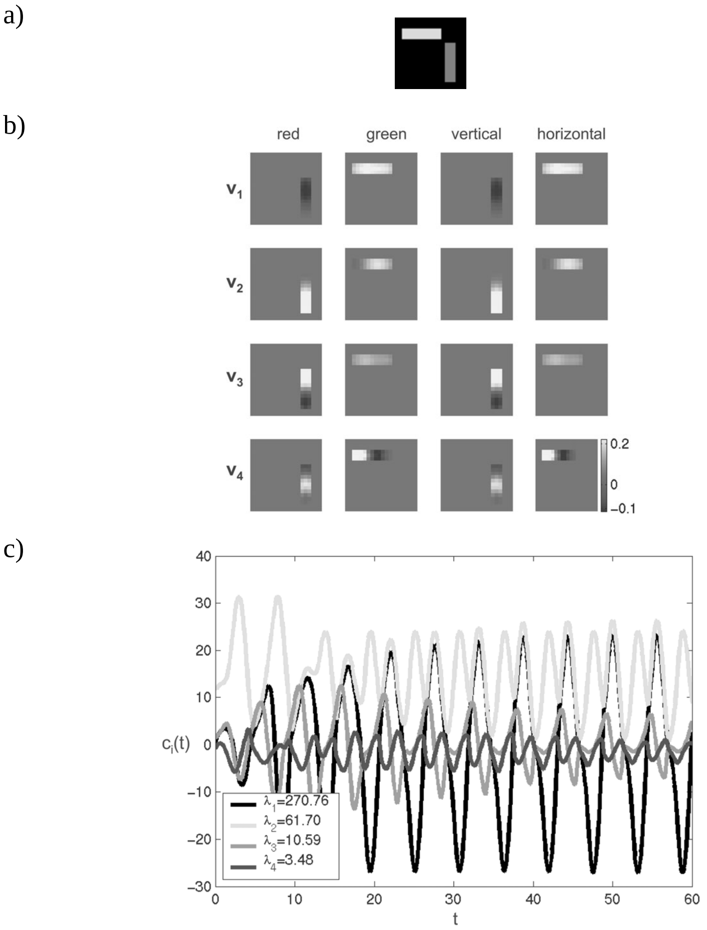

Figure 3: a) Stimulus: one vertical red bar and one horizontal green bar. b) The eigenvectors $\mathbf{v}^1, \ldots, \mathbf{v}^4$ of the four eigenmodes $1, \ldots, 4$ with the largest eigenvalues are shown in one line. The values of the vector components are coded by colors. The four columns correspond to the four feature layers. Dark shading signifies negative, mid-gray zero and light shading positive components. c) The characteristic functions and eigenvalues for the first four eigenmodes. Reprinted from Werning (2005b).

The overall dynamics of the network is given by the Cartesian vector

$$\mathbf{x}(t) = \begin{pmatrix} x_{A1}(t) \\ \vdots \\ x_{A2}(t) \\ \vdots \\ x_{B1}(t) \\ \vdots \\ x_k(t) \end{pmatrix}. \tag{4}$$

$A, B, \dots$ signify the feature dimensions and the subsequent numbers enumerate particular features of the feature dimension in question. The index $A1 = 1$ marks the beginning of the first feature cluster (e.g., red) in the first feature module (e.g., color), $A2$ the beginning of the second feature cluster (e.g., green) in the first module, $B1$ the first cluster (e.g., vertical) in the second module (orientation) and so on. The vector $\mathbf{x}(t)$ comprises the activities of the excitatory neurons of all k oscillators of the network after a transient phase and is determined by a solution of the system of differential equations (1).

For each eigenmode, the eigenvalue λ and its corresponding eigenvector $\mathbf{v}$ are solutions of the eigen-equation for the auto-co-variance matrix $\mathbf{C} \in \mathbb{R}^{k \times k}$:

$$\mathbf{C}\mathbf{v} = \mathbf{C}\lambda, \tag{5}$$

where the components C_{ij} of $\mathbf{C}$ are determined by the network dynamics $\mathbf{x}(t)$ as:

$$C_{ij} = \langle x_i | x_j \rangle. \tag{6}$$

The eigenvector $\mathbf{v}^1$ of the strongest eigenmode is shown in Fig. 3b and exhibits a significant difference between the two objects in the stimulus. To assess the temporal evolution of the eigenmodes, the notion of a characteristic function $c_i(t)$ is introduced. The network state at any instant can be considered as a superposition of the eigenvectors $\mathbf{v}^i$ weighted by the corresponding characteristic functions $c_i(t)$ of Fig. 3c:

$$\mathbf{x}(t) = \sum_i c_i(t)\mathbf{v}^i. \tag{7}$$

The eigenmode analysis separates spatial from temporal variation in the network dynamics. The eigenvectors are constant over time, but account for the varying behavior of the spatially distributed oscillators of the network. In contrast, the characteristic functions are the same for all oscillators, but account for the temporal dynamics of the network as a whole.

In the long run the network exhibits stable oscillatory behavior. As one can guess from Fig. 3c only the first two eigenmodes are stable because the amplitudes of their characteristic functions do not decrease. In contrast, the amplitudes of the characteristic functions of the other eigenmodes in the long run apparently converge to zero.

The eigenmodes, for any stimulus, can be ordered along their eigenvalues:[5]

$$\lambda_i > \lambda_{i+1}. \tag{8}$$

For this reason, I will introduce the useful convention of signifying each eigenmode by the index $i \in \mathbb{N}$. For any stimulus we have the mapping:

$$i \mapsto \langle \mathbf{v}^i, c_i(t), \lambda_i \rangle,$$

which, for each eigenmode i, renders the i-th eigenvector $\mathbf{v}^i$, the corresponding characteristic function $c_i(t)$ and the eigenvalue λ_i.

7 First Steps Into Semantic Interpretation

In this section, I will develop a heuristics that allows us to interpret the dynamics of oscillatory networks in semantic terms. Oscillatory networks that implement the hypotheses of feature maps (Hyp. 4) and synchrony (Hyp. 5), I argue, realize a structure of internal representations expressible by a monadic first order predicate language with identity $PL^=$.

Because of Hyp. 5 we are allowed to regard oscillation functions as internal representations of individual objects. They may thus be assigned some of the individual terms of the language $PL^=$. Let

$$Ind = \{a_1, \ldots, a_m, z_1, \ldots, z_n\} \tag{9}$$

be the set of individual terms of $PL^=$, then the partial function

$$\alpha : Ind \to L_2[-\frac{T}{2}, +\frac{T}{2}] \tag{10}$$

be a constant individual assignment of the language. By convention, I will assume for the domain of α, unless indicated otherwise, that

$$\mathrm{dom}(\alpha) = \{a_1, \ldots, a_m\} \tag{11}$$

so that the $a_1, \ldots, a_m$ are individual constants and the $z_1, \ldots, z_n$ are individual variables. Sometimes I will use a, b as placeholders for $a_1, \ldots, a_m$.

[5] I assume that the ordering is strict, i.e., none of the eigenvalues is degenerate.

Following equation (3), the synchrony of oscillation functions is a matter of degree. The sentence $a = b$ expresses a representational state of the system to the degree the oscillation functions $\alpha(a)$ and $\alpha(b)$ of the system are synchronous. Provided that Cls is the set of sentences of $PL^{=}$, the degree to which a sentence expresses a representational state of the system, for any eigenmode $i \in \mathbb{N}$, can be measured by the (in $\mathbb{N}$ possibly partial) function

$$d : Cls \times \mathbb{N} \rightarrow [-1, +1].$$

In case of identity sentences, for every eigenmode i and any individual constants a, b, we have:

$$d(a = b, i) = \Delta(\alpha(a), \alpha(b)). \tag{12}$$

When we take a closer look at the first eigenmode of Fig. 3b, we see that most of the vector components are exactly zero (marked by mid-gray). However, few components $v^1_j, v^1_{j'}, \ldots$ in the greenness and the horizontality layers are positive (marked by light shading) and few components $v^1_l, v^1_{l'}, \ldots$ in the redness and the verticality layers are negative (marked by dard shading). Since the contribution of the eigenmode to the entire network state is weighted by its characteristic function, the positive component v^1_j contributes to the activity of $x_j(t)$ with $+|v^1_j|c_1(t)$, while the negative component v^1_l contributes with $-|v^1_l|c_1(t)$ to $x_l(t)$. Since the Δ-function is normalized, only the signs of the constants matter. The weighted positive components of the eigenmode are all exactly parallel with one another, the weighted negative components are all exactly parallel with one another, but any weighted positive component is exactly anti-parallel to any weighted negative component:

$$\Delta(v^1_j c_1(t), v^1_{j'} c_1(t)) = 1, \tag{13}$$

$$\Delta(v^1_l c_1(t), v^1_{l'} c_1(t)) = 1, \tag{14}$$

$$\Delta(v^1_j c_1(t), v^1_l c_1(t)) = -1. \tag{15}$$

We may interpret this by saying that the first eigenmode represents two objects as distinct from one another. The representation of the first object is the positive characteristic function $+c_1(t)$ and the representation of the second object is the negative characteristic function $-c_1(t)$. Both, the positive and the negative function can be assigned to individual constants, say a and b, respectively. In the eigenmode analysis we can thus identify sharp representations of objects in the network: the characteristic functions and their negative mirror images. These considerations, for every eigenmode i, justify the general definition:

$$d(\neg a = b, i) = \begin{cases} +1 \text{ if } d(a = b, i) = -1, \\ -1 \text{ if } d(a = b, i) > -1. \end{cases} \tag{16}$$

Notice that unlike identity, its negation is represented by the network as sharp, i.e., non-gradual. Within each eigenmode, at most two objects can be represented as non-identical. As we will see later on, sharpness is a general feature of negation in our semantics as such.[6]

Following Hyp. 4, clusters of feature selective neurons function as representations of properties. They can be expressed by monadic predicates. I will assume that our language $PL^=$ has a set of monadic predicates

$$Pred = \{F_1, ..., F_p\} \tag{17}$$

such that each predicate denotes a property represented by some feature cluster. To every predicate $F \in Pred$ I now assign a diagonal matrix $\beta(F) \in \{0,1\}^{k\times k}$ that, by multiplication with any eigenmode vector $\mathbf{v}^i$, renders the sub-vector of those components that belong to the feature cluster expressed by F:

$$\beta : Pred \rightarrow \{0,1\}^{k\times k}. \tag{18}$$

With respect to our particular network, the matrix $\beta(red)$, e.g., is zero everywhere except for the first $\frac{k}{4}$ diagonal elements:

$$\beta(red) \;=\; \begin{pmatrix} 1 & 0 & & \cdots & & 0 \\ 0 & \ddots & & & & \\ & & 1 & & & \\ \vdots & & & 0 & & \vdots \\ & & & & \ddots & \\ 0 & & & \cdots & & 0 \end{pmatrix}. \tag{19}$$

The multiplication of $\beta(red)$ with the first eigenmode vector $\mathbf{v}^1$ gives us the components of $\mathbf{v}^1$ for the redness-layer in the color module of the network:

$$\beta(red)\mathbf{v}^1 \;=\; \begin{pmatrix} v_1^1 & \cdots & v^1_{i_(green)-1} & 0 & \cdots & 0 \end{pmatrix}. \tag{20}$$

Since $\beta(F)$ is a hardware feature of the network and does neither vary from stimulus to stimulus, nor from eigenmode to eigenmode (and is, model-theoretically speaking, hence constant in all models), it is sensible to call it the *neuronal intension* of F.

The neuronal intension of a predicate, for every eigenmode, determines what I call its neuronal extension, i.e., the set of those oscillations that the neurons on

[6] This evaluation of non-identity is chosen for reasons of consistency with the Gödel system introduced below.

the feature layer contribute to the activity the eigenmode adds to the overall networks dynamics. Unlike the neuronal intension, the neuronal extension varies from stimulus to stimulus and from eigenmode to eigenmode (just as extensions vary from possible world to possible world). Hence, for every predicate F its *neuronal extension* in the eigenmode i comes to:

$$\{f_j|\mathbf{f} = c_i(t)\beta(F)\mathbf{v}^i\}. \tag{21}$$

Here, the f_j are the components of the vector $\mathbf{f}$. The neuronal extension of the predicate *red* in the first eigenmode in our experimental setting thus comes to the following set of functions – it comprises all those temporally evolving activities the redness-components contribute to the overall network dynamics in the first eigenmode:

$$\{f_j|\mathbf{f} = \beta(red)\mathbf{v}^1 c_1(t)\} \;\; = \;\; \{v_1^1 c_1(t), ..., v_{i(green)-1}^1 c_1(t), 0\}. \tag{22}$$

To determine to which degree an oscillation function assigned to an individual constant a is in the neuronal extension of a predicate F, we have to compute how synchronous it maximally is with one of the oscillation functions in the neuronal extension. We are, in other words, justified to evaluate the degree to which a predicative sentence Fa expresses a representational state of our system, with respect to the eigenmode i, in the following way:

$$d(Fa,i) = \max\{\Delta(\alpha(a), f_j)|\mathbf{f} = c_i(t)\beta(F)\mathbf{v}^i\}. \tag{23}$$

Having by now provided a semantic evaluation for every atomic sentence of $PL^=$, how can we evaluate the truth-functional connectives? Since we are here dealing with an infinitely many-valued semantics, we have to look at the broader spectrum of fuzzy logics. In those logics the conjunction is semantically evaluated by a t-norm.[7] Having once made a choice for a certain t-norm as the semantic correlate of conjunction, the functions of semantic evaluation for most of the other connectives can be derived by systematic considerations (cf. Gottwald, 2001).

As will become obvious in the course of the remaining sections, the system that fits my purposes best is Gödel's (1932) min-max-logic. Here the conjunction is evaluated by the minimum of the values of the conjuncts, which is a t-norm. Let ϕ, ψ be sentences of $PL^=$, then, for any eigenmode i, we have:

$$d(\phi \wedge \psi, i) = \min\{d(\phi,i), d(\psi,i)\}. \tag{24}$$

[7] A binary operation $\mathbf{t}$ in the real interval $[-1,+1]$ is called a t-norm if and only if it is ($d,d',d'' \in [-1,+1]$): (i) associative, i.e., $\mathbf{t}(d,\mathbf{t}(d',d'')) = \mathbf{t}(\mathbf{t}(d,d'),d'')$; (ii) commutative, i.e., $\mathbf{t}(d,d') = \mathbf{t}(d',d)$; (iii) non-decreasing in the first element, i.e., $d \leq d' \Rightarrow \mathbf{t}(d,d'') \leq \mathbf{t}(d',d'')$; and (iv) has 1 as neutral element, i.e., $\mathbf{t}(d,1) = d$.

The evaluations we have so far introduced allow us to regard the first eigenmode of the network dynamics, which results from stimulation with one red vertical object and one green horizontal object (see Fig. 3), as a representation expressed by the sentence

This is a red vertical object and that is a green horizontal object.

We only have to assign the individual terms *this* $(= a)$ and *that* $(= b)$ to the oscillatory functions $-c_1(t)$ and $+c_1(t)$, respectively, and the predicates *red* $(= R)$, *green* $(= G)$, *vertical* $(= V)$ and *horizontal* $(= H)$ to the redness, greenness, verticality and horizontality layers as their neuronal intensions. Simple computation then reveals:

$$d(Ra \wedge Va \wedge Gb \wedge Hb \wedge \neg a = b, 1) = 1. \tag{25}$$

8 Eigenmodes as Alternative Epistemic Possibilities

So far I have concentrated on a single eigenmode, only. The network, however, generates a multitude of eigenmodes. We tested the representational function of the different eigenmodes by presenting an obviously ambiguous stimulus to the network. The stimulus shown in Fig. 4a can be perceived as two red vertical bars or as one red vertical grating. It turned out that the network was able to disambiguate the stimulus by representing each of the two epistemic possibilities in a stable eigenmode of its own (see Fig. 4b).

Eigenmodes, thus, play a similar role for neuronal representation as possible worlds known from Lewis (1986) or Kripke (1980) play for semantics. Like possible worlds, eigenmodes do not interfere with each other because they are mutually orthogonal. Moreover, the identity of oscillation functions (as for rigid designators in Kripke semantics) and of the neuronal intensions of predicates pertains across eigenmodes.

We now see that both of the two stable eigenmodes shown in Fig. 4b can be expressed by a disjunctive sentence, if we semantically evaluate disjunction as follows:

$$d(\phi \vee \psi, i) = \max\{d(\phi, i), d(\psi, i)\}, \tag{26}$$

for any sentences ϕ and ψ of $PL^=$ and any eigenmode i. Either of the two eigenmodes $i = 1, 2$ makes $d(\phi, i)$ assume the value +1 if ϕ is set to the following disjunctive sentence, which says that there is one red vertical object – denoted by a – or two red vertical objects – denoted by b and c:

$$(Ra \wedge Va) \vee (Rb \wedge Rc \wedge Vb \wedge Vc \wedge \neg b = c).$$

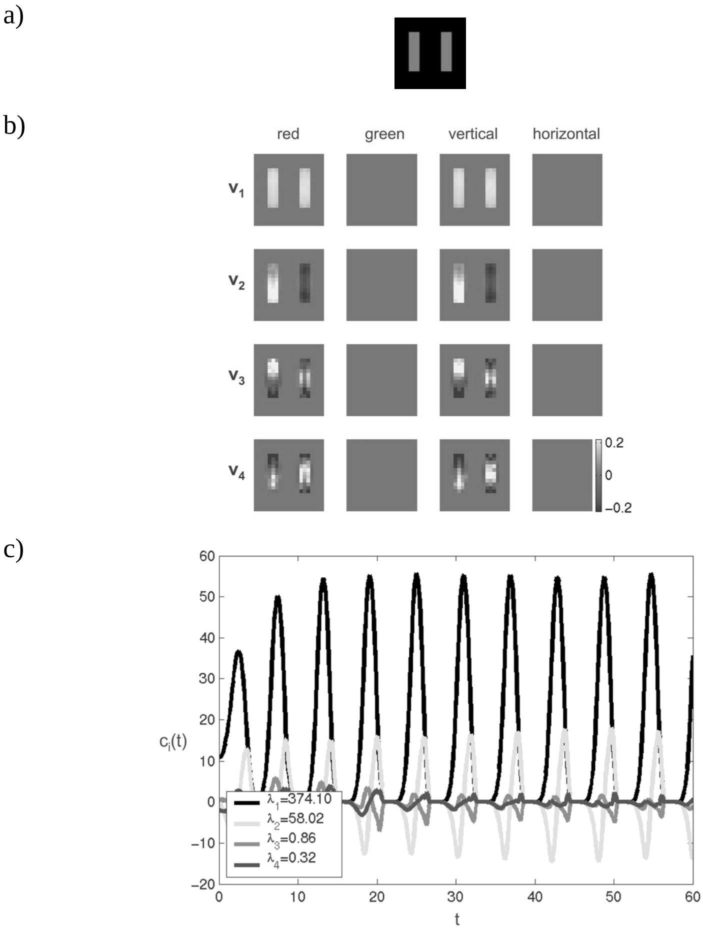

Figure 4: a) Stimulus: two vertical red bars or one red vertical grating. b) The eigenvectors $\mathbf{v}^1, \ldots, \mathbf{v}^4$ of the four eigenmodes $1, \ldots, 4$ with the largest eigenvalues are shown in one line. The first mode represents the stimulus as one red vertical object, while the second mode represents it as two red vertical objects. c) The characteristic functions show the temporal evolution of the first four modes. Only the first two are non-decreasing and thus belong to stable eigenmodes. Reprinted from Werning (2005b).

One only needs to make the following assignments of individual constants to oscillation functions:

$$\alpha(a) = +c_1(t), \quad \alpha(b) = +c_2(t), \quad \alpha(c) = -c_2(t).$$

The choice of the maximum as the semantic evaluation of disjunction is the primary reason for me to prefer the Gödel system over alternative systems of many-valued logic. In general, many-valued logics semantically evaluate disjunction by a t-conorm of which the maximum-function is an instance.[8] For our purposes the maximum is the best choice of a t-conorm because it is the only continuous t-conorm that always takes the value of one of the disjuncts as the value of the disjunction (The proof has to do with the particular non-Archimedian character of the Gödel t-norm, see Gottwald, 2001, p. 75). Other continuous t-conorms would hence not allow us to treat eigenmodes as independent alternative possibilities. We would not be able to say that a certain disjunction is true because a possibility (i.e., an eigenmode) expressed by one of its disjuncts exists.

9 Making Syntax and Semantics Explicit

We are leaving the heuristic approach now and turn to a formally explicit description of the neuronal semantics realized by oscillatory networks. Let the oscillatory network under consideration have k oscillators. The network dynamics is studied in the time window $[-\frac{T}{2}, +\frac{T}{2}]$. For any stable eigenmode $i \in \mathbb{N}$, it renders a determinate eigenvector $\mathbf{v}^i$, a characteristic function $c_i(t)$ and an eigenvalue λ_i after stimulation. The language to be considered is a monadic first order predicate language with identity ($PL^=$). Besides the individual terms of Ind and the monadic predicates of $Pred$, the alphabet of $PL^=$ contains the logical constants $\wedge, \vee, \rightarrow, \neg, \exists, \forall$ and the binary predicate $=$. Provided we have the constant individual and predicate assignments α and β of (10) and (18), the union

$$\gamma = \alpha \cup \beta \tag{27}$$

is a comprehensive constant assignment of $PL^=$. The individual terms in the domain of α are individual constants, those not in the domain of α are individual variables. The syntactic operations of the language $PL^=$ and the set SF of sentential formulae as their recursive closure can be defined as follows, for

[8] A binary operation $\mathbf{s}$ in the real interval $[-1, +1]$ is a t-conorm if and only if it is (i) associative, (ii) commutative, (iii) non-decreasing in the first element, and (iv) has -1 as neutral element.

arbitrary $a,b,z \in Ind, F \in Pred$, and $\phi, \psi \in SF$:

$$\begin{array}{lll} \sigma_{=} : (a,b) \mapsto a = b; & \sigma_{pred} : (a,F) \mapsto Fa; & \sigma_{\neg} : \phi \mapsto \neg\phi; \\ \sigma_{\wedge} : (\phi,\psi) \mapsto \phi \wedge \psi; & \sigma_{\vee} : (\phi,\psi) \mapsto \phi \vee \psi; & \sigma_{\rightarrow} : (\phi,\psi) \mapsto \phi \rightarrow \psi; \\ \sigma_{\exists} : (z,\phi) \mapsto \exists z\phi; & \sigma_{\forall} : (z,\phi) \mapsto \forall z\phi. & \end{array} \qquad (28)$$

The set of terms of $PL^{=}$ is the union of the sets of individual terms, predicates and sentential formulae of the language. A sentential formula in SF is called a *sentence* with respect to some constant assignment γ if and only if, under assignment γ, all and only individual terms bound by a quantifier are variables. Any term of $PL^{=}$ is called γ-*grammatical* if and only if, under assignment γ, it is a predicate, an individual constant, or a sentence. Taking the idea at face value that eigenmodes can be treated like possible worlds (or more neutrally speaking: like models), the relation 'i neurally models ϕ to degree d by constant assignment γ', in symbols

$$i \models^{d}_{\gamma} \phi,$$

for any sentence ϕ and any real number $d \in [-1, +1]$, is then recursively given as follows:

Identity: Given any individual constants $a,b \in Ind \cap \mathrm{dom}(\gamma)$, then $i \models^{d}_{\gamma} a = b$ iff $d = \Delta(\gamma(a), \gamma(b))$.

Predication: Given any individual constant $a \in Ind \cap \mathrm{dom}(\gamma)$ and any predicate $F \in Pred$, then $i \models^{d}_{\gamma} Fa$ iff $d = \max\{\Delta(\gamma(a), f_j) | \mathbf{f} = \gamma(F)\mathbf{v}^i c_i(t)\}$.

Conjunction: Provided that ϕ, ψ are sentences, then $i \models^{d}_{\gamma} \phi \wedge \psi$ iff $d = \min\{d', d'' \,|\, i \models^{d'}_{\gamma} \phi$ and $i \models^{d''}_{\gamma} \psi\}$.

Disjunction: Provided that ϕ, ψ are sentences, then $i \models^{d}_{\gamma} \phi \vee \psi$ iff $d = \max\{d', d'' \,|\, i \models^{d'}_{\gamma} \phi$ and $i \models^{d''}_{\gamma} \psi\}$.

Implication: Provided that ϕ, ψ are sentences, then $i \models^{d}_{\gamma} \phi \rightarrow \psi$ iff $d = \sup\{d' \in [-1, +1] \,|\, \min\{d', d''\} \leq d'''$ where $i \models^{d''}_{\gamma} \phi$ and $i \models^{d'''}_{\gamma} \psi\}$.

Negation: Provided that ϕ is a sentences, then $i \models^{d}_{\gamma} \neg\phi$ iff (i) $d = 1$ and $i \models^{-1}_{\gamma} \phi$ or (ii) $d = -1$ and $i \models^{d'}_{\gamma} \phi$ where $d' < 1$.

Existential Quantifier: Given any individual variable $z \in Ind \setminus \mathrm{dom}(\gamma)$ and any sentential formula $\phi \in SF$, then $i \models^{d}_{\gamma} \exists z\phi$ iff $d = \sup\{d' \,|\, i \models^{d'}_{\gamma'} \phi$ where $\gamma' = \gamma \cup \{\langle z, x\rangle\}$ and $x \in L_2[-\frac{T}{2}, +\frac{T}{2}]\}$.

Universal Quantifier: Given any individual variable $z \in Ind \setminus \mathrm{dom}(\gamma)$ and any sentential formula $\phi \in SF$, then $i \models^{d}_{\gamma} \forall z \phi$ iff $d = \inf\{d' \,|\, i \models^{d'}_{\gamma'} \phi$ where $\gamma' = \gamma \cup \{\langle z, x\rangle\}$ and $x \in L_2[-\frac{T}{2}, +\frac{T}{2}]\}$.

Let me briefly comment on these definitions: Most of them should be familiar from previous sections. The degree d, however, is no longer treated as a function, but as a relatum in the relation $\models$.

The semantic evaluation of negation has previously only been defined for negated identity sentences. The generalized definition, here, is a straightforward application of the Gödel system.[9] An interesting feature of negation in the Gödel system is that its duplication digitalizes the values of d into $+1$ and -1.

The evaluation of implication, too, follows the Gödel system. The deeper rationale behind this definition is the adjointness condition, which relates the evaluation of implication to the t-norm ($=$ min, by our choice).[10] Calculi for our semantics have been developed in the literature (cf. Gottwald, 2001). As far as propositional logic is concerned, the calculi are in principle those of intuitionist logic.[11] Fig. 5 gives a calculus of the Gödel system for the propositional case.

To evaluate existentially quantified formulae, the well-known method of cylindrification (Kreisel & Krivine, 1976, p. 17) is adjusted to the many-valued case. The supremum (sup) takes over the role of existential quantification in the meta-language and can be regarded as the limit case of the maximum-function in an infinite domain. This is analogous to the common idea of regarding the existential quantifier as the limit case of disjunction over an infinity of domain elements. It should be noted that the value of an existentially quantified sentence of the form

$$(\exists z)(Fz)$$

measures whether the neurons in the feature cluster expressed by F oscillate.

For the evaluation of universally quantified formulae, the method of cylindrification is used and adjusted again. This time the infimum (inf) assumes the role of universal quantification in the meta-language. It can be regarded as the limit case of the minimum for infinite domains in the same way as one might think of the universal quantifier as the limit case for infinite conjunction. To mention a

[9] In t-norm based many-valued logics a function $\mathbf{n} : [-1, +1] \rightarrow [-1, +1]$ is generally said to be a negation function if and only if $\mathbf{n}$ is non-increasing, $\mathbf{n}(-1) = 1$ and $\mathbf{n}(1) = -1$ (cf. Gottwald, 2001, p. 85).

[10] The adjointness condition relates the evaluation of implication, the function $\mathbf{i} : [-1, +1]^2 \rightarrow [-1, +1]$, to the t-norm $\mathbf{t}$ by the following bi-conditional (cf. Gottwald, 2001, p. 92): $d' \leq \mathbf{i}(d'', d''') \Leftrightarrow \mathbf{t}(d', d'') \leq d'''$.

[11] Gödel (1932) developed his min-max-system under the title ‘Zum intuitionistischen Aussagenkalkül’.

The following system of axiom schemata provides a propositional calculus for an infinitely many-valued Gödel system $\mathbf{G}_\infty$ as chosen in this paper. Its completeness is proven by Gottwald (2001, p. 297).

$$H_1 \to (H_1 \wedge H_1) \tag{LC1}$$
$$(H_1 \wedge H_2) \to (H_2 \wedge H_1) \tag{LC2}$$
$$(H_1 \to H_2) \to (H_1 \wedge H_3 \to H_2 \wedge H_3) \tag{LC3}$$
$$((H_1 \to H_2) \wedge (H_2 \to H_3)) \to (H_1 \to H_3) \tag{LC4}$$
$$H_1 \to (H_2 \to H_1) \tag{LC5}$$
$$H_1 \wedge (H_1 \to H_2) \to H_2 \tag{LC6}$$
$$H_1 \to H_1 \vee H_2 \tag{LC7}$$
$$H_1 \vee H_2 \to H_2 \vee H_1 \tag{LC8}$$
$$(H_1 \to H_3) \wedge (H_2 \to H_3) \to (H_1 \vee H_2 \to H_3) \tag{LC9}$$
$$\neg H_1 \to (H_1 \to H_2) \tag{LC10}$$
$$(H_1 \to H_2) \wedge (H_1 \to \neg H_2) \to \neg H_1 \tag{LC11}$$
$$(H_1 \to H_2) \vee (H_2 \to H_1) \tag{LC12}$$

The only rule of inference for the calculus is *modus ponens*:

$$H_1, H_1 \to H_2 \,/\, H_2.$$

Figure 5: Propositional calculus for the Gödel system.

concrete example, the value of a universally quantified implication of the form

$$(\forall z)(Fz \to F'z)$$

can be viewed as providing a measure for the overall synchronization between feature clusters expressed by the predicates F and F'.

The propositional calculus for the Gödel system can be extended to capture the first order predicate case. See Fig. 6. With respect to the identity relation one should keep in mind that identity is not absolute but graded. To capture the identity relation, we can nevertheless supplement the first order predicate calculus of Fig. 6 by the axiom schemata of Fig. 7.

10 Compositionality Ratified

In this section I will finally prove that the adequacy conditions for internal representation are fulfilled for oscillatory networks. The work done so far leads us

For the infinitely valued Gödel system $\mathbf{G}_\infty$, the propositional calculus of Fig. 5 can be extended to capture the first order predicate case. This is achieved if one adds as a rule of inference the rule of generalization

$$H\,/\forall xH$$

and if one supplements LC1–LC12 with the following axiom schemata, where the variable x must not occur free in G (cf. Gottwald, 2001, p. 284–5, unfortunately no completeness proof is provided):

$$\forall x(H_1 \to H_2) \to (\forall xH_1 \to H_2) \tag{GPL1}$$

$$\forall x(G \to H) \to (G \to \forall xH) \tag{GPL2}$$

$$\forall x(H \to G) \to (\exists xH \to G) \tag{GPL3}$$

$$\forall x(H_1 \to H_2) \to (H_1 \to \forall xH_2) \tag{GPL4}$$

$\forall xH(x) \to H(t|x)$ for all terms t which are substitutable for x in H (GPL5)

$H(t|x) \to \exists xH(x)$ for all terms t which are substitutable for x in H (GPL6)

Figure 6: First order predicate calculus for the Gödel system.

For the infinitely valued semantics presented in this paper, the propositional calculus of Fig. 5 plus the first order predicate extension of Fig. 6 can be extended to capture sentences involving the identity relation. Since identity is evaluated by the Δ-function, identity in our case is gradual, but still reflexive (ID1) and symmetric (ID2), however, not transitive. Due to our evaluation of predication, one direction of the Leibniz law, i.e., ID3, also holds. One may thus add the following axiom schemata:

$$\forall x(x = x) \tag{ID1}$$

$$\forall x\forall y(x = y \to y = x) \tag{ID2}$$

$$\forall x\forall y((x = y \wedge F(x)) \to F(y)) \text{ for every predicate } F \tag{ID3}$$

Figure 7: Axioms of identity.

directly to the following theorem:

Theorem 1 (Compositional Meanings in Oscillatory Networks). *Let L be the set of terms of a PL$^=$-language, SF the set of sentential formulae and $\models$ the neuronal model relation. The function μ with domain L is a compositional meaning function of the language if μ, for every $t \in L$, is defined in the following way:*

$$\mu(t) = \begin{cases} \{\langle\gamma,\gamma(t)\rangle | \gamma \textit{ is a constant assignment}\} \textit{ if } t \notin SF, \\ \{\langle\gamma,i,d\rangle | i \models^d_\gamma \phi\} \textit{ if } t \in SF. \end{cases}$$

To simply notation, we may stipulate for any γ-grammatical term t:

$$\mu_\gamma(t) = \begin{cases} \gamma(t) \text{ if } t \text{ is not a sentence,} \\ \{\langle i,d\rangle | \langle\gamma,i,d\rangle \in \mu(t)\} \text{ if } t \text{ is a sentence.} \end{cases} \tag{29}$$

Proof. To prove the theorem, one has to show that for any of the syntactic operations σ in (28), there is a semantic operation μ_σ that satisfies the equation:

$$\mu(\sigma(t_1,...,t_n)) = \mu_\sigma(\mu(t_1),...,\mu(t_n)). \tag{30}$$

To do this for the first six operations, one simply reads the bi-conditionals in the definition of $\models$ as the prescriptions of functions:

$$\mu_= : (\mu(a),\mu(b)) \mapsto \{\langle\gamma,i,d\rangle | d = \Delta(\mu_\gamma(a),\mu_\gamma(b))\};$$

$$\mu_{pred} : (\mu(a),\mu(F)) \mapsto \\ \{\langle\gamma,i,d\rangle \quad | d = \max\{\Delta(\mu_\gamma(a), f_j) | \mathbf{f} = \mu_\gamma(F)\mathbf{v}^i c_i(t)\}\};$$

$$\mu_\wedge : (\mu(\phi),\mu(\psi)) \mapsto \\ \{\langle\gamma,i,d\rangle \quad | d = \min\{d',d'' | \langle\gamma,i,d'\rangle \in \mu(\phi), \langle\gamma,i,d''\rangle \in \mu(\psi)\}\};$$

etc.

To attain semantic counterpart operations for $\sigma_\exists$ and $\sigma_\forall$, we have to apply the method of cylindrification:

$$\mu_\exists : \mu(\phi(z)) \mapsto \\ \{\langle\gamma,i,d\rangle \quad | \exists\gamma' : \mathrm{dom}(\gamma') = \mathrm{dom}(\gamma) \cup \{z\} \text{ and } \langle\gamma',i,d\rangle \in \mu(\phi(z))\};$$

$$\mu_\forall : \mu(\phi(z)) \mapsto \\ \{\langle\gamma,i,d\rangle \quad | \forall\gamma' : \mathrm{dom}(\gamma') = \mathrm{dom}(\gamma) \cup \{z\} \Rightarrow \langle\gamma',i,d\rangle \in \mu(\phi(z))\}.$$

One easily verifies that equation (30) is satisfied. □

Theorem 1 proves that condition (a) of Principle 1 is satisfied. If one holds fix the constant assignment γ and consequently the grammaticality of the terms of the language $PL^{=}$, and if one regards L_γ as the set of grammatical terms of $PL^{=}$ under assignment γ, one may say that the neuronal structure

$$\mathscr{N} = \langle \{\gamma\} \times \mu_\gamma[L_\gamma], \{\mu_=, \mu_{pred}, \mu_\neg, \mu_\wedge, \mu_\vee, \mu_\rightarrow, \mu_\exists, \mu_\forall\}\rangle$$

compositionally evaluates the language

$$\langle L_\gamma, \{\sigma_=, \sigma_{pred}, \sigma_\neg, \sigma_\wedge, \sigma_\vee, \sigma_\rightarrow, \sigma_\exists, \sigma_\forall\}\rangle$$

with respect to meaning.

The *ideal meaning* of a term t under assignment γ, $\mu^1_\gamma(t)$, can be identified with the subset of $\mu_\gamma(t)$, for which all values d are 1. The formula

$$\langle i, d\rangle \in \mu_\gamma(\phi)$$

can then be read as: The eigenmode i, to degree d, realizes the ideal neuronal meaning of ϕ under assignment γ. The ideal meaning $\mu^1_\gamma(a)$ of an individual constant a is henceforth identified with an object concept. Recall that the object concept $\mu^1_\gamma(a)$ just is the oscillation $\alpha(a)$. The ideal meaning $\mu^1_\gamma(F)$ of a predicate F is identified with a predicate concept. Notice that $\mu^1_\gamma(F)$ just is the matrix $\beta(F)$, which we have called neuronal intension earlier and which identifies a specific cluster of feature-selective neurons.

To comply with the condition of co-variation, i.e., condition (c) of Principle 1, we can choose the assignment γ in a way so that the oscillation function $\gamma(a)$ tracks the object designated by any individual term a. We can, furthermore, make sure that $\gamma(F)$ is just the cluster of neurons representing the property expressed by the predicate F. In this case, the assignment will be called *natural*. As we have shown in our simulations, the network dynamics warrants that the neuronal meanings of terms with respect to the natural assignment reliably co-vary with the terms' denotations:

Fact 1 (Covariation with Content for Oscillatory Networks). *Let Γ be an intended external constant assignment for a language $PL^{=}$ with a set of terms L such that Γ maps individual terms and predicates to their intended denotations. Let ν_Γ, then, be a function that maps each element of the set of Γ-grammatical terms L_Γ to its denotation. The architecture of oscillatory networks now warrants that there is a natural neuronal assignment γ of the individual constants and predicates of L_Γ ($= L_\gamma$) into the set of neuronal states N of the network and, consequently, a meaning function μ_γ from L_γ into N, such that meanings co-vary with their contents, or, formally speaking: There is a content function*

$$\kappa : \mu_\gamma[L_\gamma] \rightarrow \nu_\Gamma[L_\Gamma].$$

and

$$\nu_\Gamma(t) = \kappa(\mu_\gamma(t))$$

for every $t \in L_\Gamma$.

Condition (c) of the adequacy conditions in Principle 1 thus turns out to be fulfilled by the network architecture. This is the central result of our simulations and has an explanation in the construction plan of the network. Recall that the construction scheme was chosen not only to match up with neurobiological data, but also to implement the *Gestalt* principles for object perception.

If the co-variation with content is warranted according to Fact 1, the compositionality of content can also be proven:

Theorem 2 (Compositional Contents of Oscillatory Networks). *Let γ be the natural neuronal, and Γ the intended external assignment of the language $PL^=$ with the set of terms L. Let L_γ and L_Γ be the set of grammatical terms of $PL^=$ with respect to the natural neuronal, respectively, the intended external assignment. Let, furthermore, be*

$$L_\gamma = L_\Gamma.$$

We assume that $L_\gamma (= L_\Gamma)$ have a compositional function of denotation ν and μ be a compositional neuronal meaning function with the same domain. Then, in the case of co-variation, the natural neuronal structure

$$\mathscr{N} = \langle \{\gamma\} \times \mu_\gamma[L_\gamma], \{\mu_=, \mu_{pred}, \mu_\neg, \mu_\wedge, \mu_\vee, \mu_\rightarrow, \mu_\exists, \mu_\forall\}\rangle$$

can be compositionally evaluated with respect to content.

Proof. Since co-variation is assumed to be the case in the antecedent of the theorem, we have

$$\nu = \kappa \circ \mu.$$

Since Γ is the intended external and γ the natural neuronal assignment, we may set $\nu := \nu_\Gamma$ and $\mu := \mu_\gamma$. Now, the theorem's antecedent tells us that the language can be compositionally evaluated with respect to denotation, i.e., there is a function $f (= \nu_\sigma)$ for every n-ary syntactic operation σ of the language such that

$$\nu(\sigma(t_1, ..., t_n)) = f(\nu(t_1), ..., \nu(t_n)),$$

which, in the case of co-variation, is equivalent to

$$(\kappa \circ \mu)(\sigma(t_1, ..., t_n)) = f((\kappa \circ \mu)(t_1), ..., (\kappa \circ \mu)(t_n)).$$

Since the language is compositional with respect to the meaning function μ, there is a function μ_σ in $\mathscr{N}$ for each and every σ of the language such that

$$\mu_\sigma(\mu(t_1), ..., \mu(t_n)) = \mu(\sigma(t_1, ..., t_n)).$$

From the former two equations we derive:

$$\kappa(\mu_\sigma(\mu(t_1),...,\mu(t_n))) = f(\kappa(\mu(t_1)),...,\kappa(\mu(t_n))).$$

Since the surjectivity of the meaning function μ warrants that, for every element m of the carrier set of $\mathscr{N}$, there is at least one grammatical term t in L_γ such that

$$m = \mu(t),$$

we may finally conclude that, for every n-ary operation $h (= \mu_\sigma)$ of $\mathscr{N}$, and every sequence $m_1,...,m_n$ in the domain of h, there is a function $\kappa_h (= f = \nu_\sigma)$ such that

$$\kappa(h(m_1,...,m_n)) = \kappa_h(\kappa(m_1),...,\kappa(m_n)).$$

The content function is hence proven to be compositional and the condition (b) of the adequacy conditions is fulfilled. □

11 Perceptual Necessities

It is sometimes useful to talk not only about what is represented by one eigenmode of the network dynamics, but to talk about what the network dynamics as a whole represents. This must take into account all stable eigenmodes of the network. Each eigenmode, as I have argued earlier, stands for one perceptual or, more generally speaking, one epistemic possibility. If we take the identification of eigenmodes with possibilities – 'possibility' always to be read in an epistemic sense – at face value, we can apply Leibniz's idea that necessity is truth in all possible worlds.

We can then say that what the network dynamics represents as a whole is what is represented as necessarily being true by the network dynamics when the network is stimulated with a certain stimulus. If we want to capture what the network dynamics represents as a whole and identify epistemic possibilities with eigenmodes, we thus have to express what is represented as true in all eigenmodes of the network dynamics. Formally, this can be done by use of the necessity operator $\Box$ of modal logic. With ϕ being a grammatical sentence, we hence write

$$\Box\phi$$

to express that the network dynamics represents ϕ to be necessarily true, given the current epistemic situation, i.e., the current stimulus input.

If we hold fix the assignment γ to be the natural neuronal assignment, the four place relation $\models$ reduces to a three place relation. Epistemic necessity with respect to a network dynamics $\mathbf{x}(t)$ is now defined as follows:

Definition 1 (Epistemic Necessity in the Network). *Given a network dynamics* $\mathbf{x}$ *with the set of stable eigenmodes* $E \subseteq \mathbb{N}$*, then, for every sentence* ϕ *of the respective language,*

$$\Box\phi$$

is true in $\mathbf{x}$ *if and only if, for all stable eigenmodes* $i \in E$*, the following holds:*

$$i \models^1 \phi.$$

Likewise epistemic possibility can be defined by the existence of an eigenmode. We write

$$\Diamond\phi$$

just in case there is an eigenmode of the networks dynamics that models ϕ:

Definition 2 (Epistemic Possibility in the Network). *Given a network dynamics* $\mathbf{x}$ *with the set of stable eigenmodes* $E \subseteq \mathbb{N}$*, then, for every sentence* ϕ *of the respective language,*

$$\Diamond\phi$$

is true in $\mathbf{x}$ *if and only if, there is a stable eigenmode* $i \in E$*, such that the following holds:*

$$i \models^1 \phi.$$

We can now apply our newly defined modal notions to describe what the network dynamics represents as a whole when the network is stimulated, e.g., with the stimulus of Fig. 4a. We may assume that $i = 1,2$ are the only two stable eigenmodes of the resulting network dynamics $\mathbf{x}$. As one sees in Fig. 4c the characteristic functions of the third and fourth eigenmode are decreasing over time and probably converge to zero. The eigenmodes higher than 2 thus are not stable. A little computation now reveals that

$$\Box\left(\begin{array}{c}(\exists x\forall y)\\ (Rx \wedge Vx \wedge ((Ry \wedge Vy) \rightarrow y = x))\\ \vee\\ (\exists x\exists y\forall z)\\ (\neg x = y \wedge Rx \wedge Vx \wedge Ry \wedge Vy \wedge ((Rz \wedge Vz) \rightarrow (z = x \vee z = y)))\end{array}\right)$$

is true in $\mathbf{x}$.[12] However, the sentence after the necessity operator just expresses what we are forced to perceive when we look at the ambiguous stimulus of

[12] The first eigenmode makes the first disjunct true while the second eigenmode makes the second disjunct true. If we look at the first disjunct, the existential quantifier requires us to search for the oscillation function (the value of x) that makes the evaluation of the subsequent formula supremal, namely 1. This must be an oscillation function parallel to $+c_1(t)$. Only then $Rx \wedge Vx$ becomes 1. Looking at the universal quantifier, we have

Fig. 4a, namely, that there is exactly one red vertical object or there are exactly two red vertical objects. The semantics developed thus pretty well accommodates the phenomenological facts.

12 Conclusion

Oscillatory networks show how a structure of the cortex can be analyzed so that elements of this structure can be identified with mental concepts. These cortical states can be regarded as the neuronal meanings of predicative expressions. As meanings they form a compositional semantics for a language. As concepts they can themselves be evaluated compositionally with respect to external content. The approach formulated in this paper is biologically rather well-founded. It is supported by a rich number of neurophysiological and psycho-physical data and is underpinned by various computer simulations.

Compared to connectionist alternatives (Smolensky, 1991/1995; Shastri & Ajjanagadde, 1993; Plate, 1995; van der Velde & de Kamps, 2006), the architecture proposed for large parts of the cortex in this paper is advantageous in that it not only implements a compositional semantics of meanings, but shows how internal representations can co-vary with external contents. As a consequence, the internal conceptual structure can itself be externally evaluated in a compositional way. It thus becomes transparent how concepts can have content and how they thereby mediate between utterances and their denotations.

Oscillatory networks and their biological correlates may be assigned a central role at the interface between language and mind, and between mind and world. This is due to the quasi-perceptual capabilities of oscillatory networks, which alternative connectionist models for semantic implementations lack completely. Linking oscillatory networks to mechanisms for the production of phonological sequences remains a challenge for future investigations.

The theory developed here amounts to a new mathematical description of the time-structure the cortex is believed to exhibit. Neuronal synchronization plays an essential role not only for binding, but, generally, for the generation of compositional representations in the brain.

to ask whether any oscillation function (evaluating y) other than one parallel to $+c_1(t)$ could make the antecedent of the value of the implication $Ry \wedge Vy$ greater than the value of the consequent. Only then the value of the implication would be less than 1. The answer is no because all non-zero components v_j^1 of the eigenvector are positive and pertain to the redness or verticality layer. Their contributions to the network dynamics $v_j^1 c_1(t)$ are hence parallel to $+c_1(t)$ such that $d(x=y,1)$ would be 1. Assigning y to the constant zero-function would also leave the value of the implication at 1. For, in that case the values of the antecedent and the consequent would be equally 0. The evaluation of the second disjunct follows similar considerations.

References

Felleman, D. J., & van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1,* 1–47.

Fodor, J. (1992). *A theory of content and other essays.* Cambridge, MA: MIT Press.

Gödel, K. (1932). Zum intuitionistischen Aussagenkalkül. *Anzeiger Akademie der Wissenschaften Wien, 69*(Math.-nat. Klasse), 65–66.

Gottwald, S. (2001). *A treatise on many-valued logics.* Baldock: Research Studies Press.

Gray, C., König, P., Engel, A. K., & Singer, W. (1989). Oscilliatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature, 338,* 334–7.

Hodges, W. (2001). Formal features of compositionality. *Journal of Logic, Language and Information, 10,* 7–28.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195,* 215–243.

Kreisel, G., & Krivine, J. L. (1976). *Elements of mathematical logic. Model theory* (Vol. 2). Amsterdam: North-Holland.

Kripke, S. (1980). *Naming and necessity.* Cambridge, MA: Harvard.

Lewis, D. (1986). *On the plurality of worlds.* Oxford: Blackwell.

Maye, A. (2003). Correlated neuronal activity can represent multiple binding solutions. *Neurocomputing, 52–54,* 73–77.

Maye, A., & Werning, M. (2004). Temporal binding of non-uniform objects. *Neurocomputing, 58–60,* 941–8.

Obermayer, K., & Blasdel, G. G. (1993). Geometry of orientation and ocular dominance columns in monkey striate cortex. *Journal of Neuroscience, 13,* 4114–29.

Plate, T. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks, 6*(3), 623–41.

Schillen, T. B., & König, P. (1994). Binding by temporal structure in multiple feature domains of an oscillatory neuronal network. *Biological Cybernetics, 70,* 397–405.

Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, *16*, 417–94.

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, *24*, 49–65.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, *18*, 555–86.

Smolensky, P. (1995). Connectionism, constituency and the language of thought. In C. Macdonald & G. Macdonald (Eds.), *Connectionism* (pp. 164–198). Cambridge, MA: Blackwell. (Original work published 1991)

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171–8.

van der Velde, F., & de Kamps, M. (2006). Neural blackboard architectures of combinatorial structures in cognition. *Behavioral and Brain Sciences*. (In press)

von der Malsburg, C. (1981). *The correlation theory of brain function* (Internal Report Nos. 81–2). Göttingen: MPI for Biophysical Chemistry.

Werning, M. (2001). How to solve the problem of compositionality by oscillatory networks. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the twenty-third annual conference of the cognitive science society* (pp. 1094–1099). London: Lawrence Erlbaum Associates.

Werning, M. (2003). Synchrony and composition: Toward a cognitive architecture between classicism and connectionism. In B. Löwe, W. Malzkorn, & T. Raesch (Eds.), *Applications of mathematical logic in philosophy and linguistics* (pp. 261–78). Dordrecht: Kluwer.

Werning, M. (2004). Compositionaltity, context, categories and the indeterminacy of translation. *Erkenntnis*, *60*, 145–78.

Werning, M. (2005a). Right and wrong reasons for compositionality. In M. Werning, E. Machery, & G. Schurz (Eds.), *The compositionality of meaning and content* (Vol. I: Foundational Issues, pp. 285–309). Frankfurt: Ontos Verlag.

Werning, M. (2005b). The temporal dimension of thought: Cortical foundations of predicative representation. *Synthese*, *146*(1/2), 203–24.