MARKUS WERNING

THE TEMPORAL DIMENSION OF THOUGHT

Cortical Foundations of Predicative Representation

ABSTRACT. The paper argues that cognitive states of biological systems are inherently temporal. Three adequacy conditions for neuronal models of representation are vindicated: the compositionality of meaning, the compositionality of content, and the co-variation with content. Classicist and connectionist approaches are discussed and rejected. Based on recent neurobiological data, oscillatory networks are introduced as a third alternative. A mathematical description in a Hilbert space framework is developed. The states of this structure can be regarded as conceptual representations satisfying the three conditions.

1. CONCEPTS, COMPOSITIONALITY AND CO-VARIATION

The view that cognition takes place in the cortex constitutes a common ground for most contemporary philosophers and cognitive scientists. Highly controversial, however, is the question *how* this can be. Cognition is not just any form of information processing. Only processes that are defined over conceptual structures, (i) which have content and (ii) which are expressible by predicate languages, are properly called cognition. The first condition derives from the fact that cognitive processes are essentially epistemic: The criterion of truth-conduciveness, which is exclusive to bearers of content, i.e., representations, applies to them. The second condition grounds in the assumption that cognition presupposes categorization. Truth-conducive processes would be practically useless and without any evolutionary benefit if they did not subsume objects under categories. Non-categorial processes would not be *about* anything. Categories, however, are just what concepts are and predicates express. While the neuronal structure of the cortex, to this day, has been perceived as radically different from conceptual structure, this paper, using the dimension of time, will show how it is nevertheless possible to reduce the latter to the former.

Cognition is *systematic* in the sense that there are systematic correlations between representational capacities: If a mind is capable of certain cognitive states, it most probably is also capable of other cognitive states with related contents. The capacity to think that a red square is in a green circle, e.g., is statistically highly correlated with the capacity to think that

Synthese (2005) 146: 203–224

a red circle is in a green square. To explain this correlation, compositional operations are postulated. They enable the system to build complex representations from primitive ones so that the content of the complex representation is structure-dependently determined by the content of its parts. Not only are cognitive states compositional with respect to content, also expressions of natural languages are compositional with respect to meaning: The meaning of a complex expression is a syntax-dependent function of the meanings of its syntactic parts. The reasons for compositionality of content and meaning have been extensively discussed in the literature (Janssen 1997; Hodges 2001; and Fodor and Lepore 2002).

To explain the compositionality of content and meaning, Fodor and Pylyshyn (1988) take recourse to a language of thought, which they link to the claim that the brain can be modelled by a Turing-style computer. A subject's having a cognitive state, they believe, consists in the subject's bearing a computational relation to a mental sentence; it is a relation analogous to the relation a Turing machine's control head bears to the tape. A subject's thought that there is a red square in a green circle, thus, is conceived of as a computational relation between the subject and the mental sentence: [*There is a red square in a green circle*]. Likewise, when a subject understands the utterance *John loves Mary*, this utterance reliably causes the subject to bear a computational relation to the concatenation of mental words: [*John loves Mary*].

The trouble with classical computer models is well known and ranges from the frame problem, the problem of graceful degradation, and the problem of learning from examples (cf. Horgan and Tienson 1996) to problems that arise from the content sensitivity of logical reasoning. To avoid the pitfalls of classicism, connectionist models have been developed. In connectionist models that try to implement the semantics of natural languages (e.g., Smolensky 1995; for a survey of related models see Werning 2001) the syntax of a language is mapped homomorphically onto an algebra of vectors and tensor operations. Each primitive expression of the language is assigned to a vector. Every vector renders a certain distribution of activity within the connectionist network. The syntactic operations of the language have as counterparts tensor operations that generate vectors, which implement the meanings of complex expressions, from vectors which implement the meaning of the syntactic constituent expressions. As far as the compositionality of meaning is concerned, the semantics of languages with some, though limited combinatorial potential can, indeed, be implemented by a connectionist network.

To make the notion of compositionality explicit, one usually defines the syntax of a representational (linguistic, cognitive, or neuronal) structure

$\mathcal{R}$ as a pair $\mathcal{R} = \langle R, \Sigma \rangle$, where R is the set of representations and Σ is the set of syntactic operations $\sigma_1, \ldots, \sigma_j$. Each syntactic operation σ of some arity n is a partial function $\sigma : R^n \rightarrow R$ (not necessarily a concatenation). The set R is the closure of a fixed set of atomic representations with regard to recursive application of the syntactic operations. Given any representations $t, t' \in R$, t' is called an *immediate $\mathcal{R}$-syntactic part* (or constituent) of t just in case there are an n-ary syntactic operation $\sigma \in \Sigma$ and some representations $t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n \in R$ such that $t = \sigma(t_1, \ldots, t_{i-1}, t', t_{i+1}, \ldots, t_n)$. Any representation t' is said to be a *$\mathcal{R}$-syntactic part* of a representation t just in case t' is either an immediate $\mathcal{R}$-syntactic part of t or an immediate $\mathcal{R}$-syntactic part of some $\mathcal{R}$-syntactic part of t. (I will often omit the relativization of syntactic constituency to a certain syntax.)

Representational structures are characterized not only by a syntax, but also by the fact that they can be evaluated semantically. Expressions of some linguistic structure are semantically evaluated either with respect to their meaning or with respect to their denotation, while cognitive states (thoughts, concepts, etc.) are evaluated with regard to their content. If one entertains a mentalist (or cortical) view on meanings and identifies the meanings of expressions with the cognitive states the expressions express, the denotation of an expression may be identified with the content of its meaning. This is the view I will assume in this paper, being aware of the fact that non-actual entities will then probably have to be allowed as elements of denotations. I will assume that expressions can be disambiguated (by terms) and that cognitive states naturally are unambiguous such that both can be evaluated semantically by a function. The notion of compositionality is now defined for any function that semantically evaluates a representational structure:

DEFINITION 1 (Compositionality). Let $\mathcal{R} = \langle R, \Sigma \rangle$ be a representational structure with $\Sigma = \{\sigma_1, \ldots, \sigma_j\}$ and let μ be a function of semantic evaluation with domain R. Suppose that every $\mathcal{R}$-syntactic part of a μ-evaluable representation is μ-evaluable. Then μ is called compositional if and only if, for every syntactic operation $\sigma \in \Sigma$, there is a function μ_σ such that for every non-atomic μ-evaluable representation $\sigma(t_1, \ldots, t_n) \in R$ the following equation holds:

$$(1) \qquad \mu(\sigma(t_1, \ldots, t_n)) = \mu_\sigma(\mu(t_1), \ldots, \mu(t_n)).$$

A representational structure is called *compositional* just in case it has a total compositional function of semantic evaluation. In case (and just in case)

of compositionality, the representational structure has the homomorphic semantics $\langle \mu[R], \{\mu_{\sigma_1}, \ldots, \mu_{\sigma_j}\}\rangle$.

In order to show that a connectionist system of the kind mentioned above provides a compositional semantics of meaning for natural languages, it suffices to show that the tensor algebra (or the connectionist system, in general) is a homomorphic image of the syntax of natural language. There is no principled reason why this should not be possible. The problem with connectionist approaches to semantics lies elsewhere, viz. in the compositionality of content: The network structure (of vectors and tensor operations) is now itself regarded as a syntax whose semantics is an algebra of external contents, where most semantic theories explain the semantic properties of internal representations in terms of co-variation. They, e.g., hold that a certain internal state is a representation of redness because the state co-varies with nearby instances of redness.[1] This co-variation relation is backed by the intrinsic and extrinsic causal properties of the internal state that makes up the *redness* representation. Consequently an internal representation has its semantic value because it has a certain causal role within the world. The question of how the semantic value of an internal representation is determined by the semantic values of its syntactic parts leads to the question of how the causal properties of an internal representation are determined by the causal properties of the syntactic parts. From chemistry and other sciences we know that atoms determine the causal properties of molecules because atoms are *mereological* constituents of molecules. A state X is commonly regarded to be a mereological constituent of a state Y if and only if it is true that, if Y occurs at a certain region of space at a certain time, then X occurs at the same region at the same time. Independently from sciences, one can even make it a hard metaphysical point: If the causal properties of a state B are determined by the causal properties of the states $A_1, \ldots, A_n$ and their relations to each other, then $A_1, \ldots, A_n$ are mereological constituents of B. Its justification starts off with Kim's (1993) well established (although not everywhere accepted) principle of explanatory exclusion, which says that no two independent phenomena each completely determine one and the same phenomenon. Given the truism that the causal properties of a whole B are determined by the causal properties of an exhaustive sample $C_1, \ldots, C_m$ of mereological constituents of B (plus structure), it follows that the causal properties of the states $A_1, \ldots, A_n$ (plus structure) determine the causal properties of B only if $A_1, \ldots, A_n$ are not independent from $C_1, \ldots, C_m$. Since there is a limited repertoire of relevant metaphysical dependency relations, viz. identity, reduction, supervenience and mereological constituency, one may conclude that each A_i is either

(i) identical with, (ii) reducible to, (iii) supervenient on, (iv) a mereological constituent of, (v) or – the reverse – mereologically composed of one or more of the C_j. In all five cases every A_i would be a mereological constituent of B. In the first case, this follows from the reflexivity of mereological constituency. In the second and the third case, if A_i reduces to, or is supervenient on, one or more of the C_j, A_i co-occurs with the C_j in question. Since the latter, as mereological constituents of B, occur whenever and wherever B does, also A_i occurs whenever and wherever B does and is, thus, a mereological constituent of B. In the fourth case, it follows from the transitivity of mereological constituency. The fifth case holds because every mereological composition of mereological constituents of a whole is itself a mereological constituent of the whole.

We may conclude that the semantic values, i.e., the contents, of the syntactic constituents of an internal representation determine the content of the internal representation just in case the syntactic constituents are mereological constituents of the internal representation. Two remarks should be added: First, syntactic parts aren't mereological constituents *per se*. Syntactic constituency is the relation the arguments of a syntactic operation bear to the values thereof, while mereological constituency is a relation of spatio-temporal co-occurrence. Since many natural languages have deletion rules – in English exemplified by the mapping $(can, not) \mapsto can't$ – syntactic constituency does not correlate with mereological constituency. Second, the requirement that syntactic constituents of internal representations be mereological constituents of the latter does not follow from the constraint of compositionality alone. There may well be compositional representational structures for which syntactic constituents aren't mereological constituents, e.g., languages with deletion rules. However, the requirement that the syntactic constituents of internal representations be mereological constituents follows from the principle of compositionality together with the premise that internal representations owe their semantic values to their causal properties. The requirement highlights a particularity of *internal* representation and does not generalize to other representational structures. The words and phrases of English owe their meanings mainly to the interpretation of English speakers. There may well be a language whose tokens have the same causal properties (sound, loudness, etc.) as those of English, but differ with respect to their meanings. For internal representations, in contrast, causal properties are determinant with regard to their semantics because internal representations represent autonomously, i.e., without being interpreted by any other system.

Previous connectionist attempts to implement cognitive states, we may now diagnose, fail. What is in need are two mappings, not one and both

have to be compositional. The first, unproblematic mapping $\mu : L \rightarrow N$ maps a syntax of some natural language $\mathcal{L}$ to the network structure $\mathcal{N}$ and treats the latter as a semantics. The second mapping $\kappa : N \rightarrow W$, however, treats the network structure itself as a syntax and maps the network states onto their external contents. The external contents form the worldly structure $\mathcal{W}$, e.g., a structure of individuals, properties and possible worlds. Moreover, the mapping κ need not only be a formal homomorphism, but needs to be supported by a causal relation of co-variation. As I argued, this requires any $\mathcal{N}$-syntactic part t' of some state $t \in N$ to be a mereological constituent of t. Smolensky (1995) an others have frequently conceded, that this is not the case in connectionist approaches because the products of tensor operations do typically not contain the vectors they have been applied to as vector components, i.e., as mereological constituents.

The argument can also be formulated in less abstract terms: Assume the English expressions *brown, cow*, and *brown cow* have been mapped onto vectors of a connectionist network by some compositional function μ. Now, although *brown* and *cow*, in English, are not only syntactic, but also mereological parts of *brown cow*, and although their network counterparts $\mu(brown)$ and $\mu(cow)$, with respect to the network structure, are syntactic parts of $\mu(\textit{brown cow})$, the states $\mu(brown)$ and $\mu(cow)$ aren't mereological constituents of the state $\mu(\textit{brown cow})$. This implies that even if $\mu(brown)$ co-varied with brown things and even if $\mu(cow)$ co-varied with cows, $\mu(\textit{brown cow})$ would not be necessitated to co-vary with its content, brown cows. If mereological constituency, on the other hand, had been correlated with syntactic constituency on the network level, any co-variation between $\mu(cow)$ and cows, respectively, $\mu(brown)$ and brown things, would have necessitated that $\mu(\textit{brown cow})$ co-varies with brown cows. We may conclude: The requirement that every function semantically evaluating the neuronal meanings of natural languages with respect to their contents should not only be compositional, but should also be backed by a relation of co-variation is violated, if, on the level of the neuronal structure, syntactic constituency does not correlate with mereological constituency. This is the reason why traditional connectionist approaches must fail, and indeed no semantically interpreted connectionist architecture, so far, has achieved co-variation between internal representations and external contents.

2. OSCILLATORY NETWORKS AND HILBERT SPACE

Mereological constituency is a synchronic relation, while causal connectedness is a diachronic relation. Whole and part co-exist in time, whereas causes and effects succeed in time. The reference to causal connections and

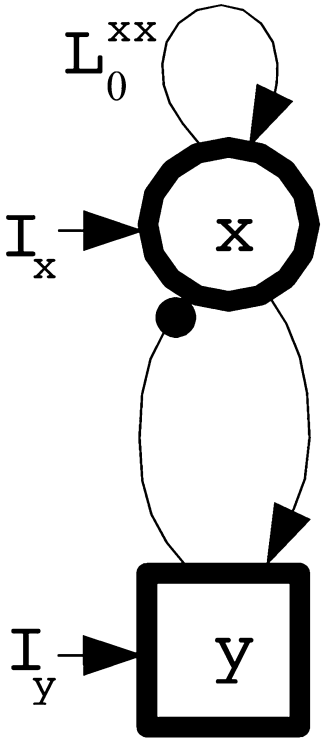

Figure 1a. A single oscillator consists of an excitatory (x) and an inhibitory (y) neuron. Each neuron represents the average activity of a cluster of biological cells. L_0^{xx}: self-excitation, I_x and I_y: input.

the flow of activation within the network will, therefore, not suffice to establish mereological constituent relations. What we, in addition, need is an adequate synchronic relation. Oscillatory networks provide a framework to define such a relation: the relation of synchrony between oscillations.

A single oscillator consists of two mutually excitatory and inhibitory neurons, each of which represents a population of biological cells (Figure 1a). If the number of excitatory and inhibitory biological cells is large enough, the dynamics of each oscillator can be described by two variables, i.e.:

$$\text{(2a)} \quad \dot{x} = -\tau_x x - g_y(y) + L_0^{xx} g_x(x) + I_x + N_x;$$

$$\text{(2b)} \quad \dot{y} = -\tau_y y + g_x(x) - I_y + N_y.$$

Here, τ_ξ ($\xi \in \{x, y\}$) are constants that can be chosen to match refractory times of biological cells, g_ξ are transfer functions, L_0^{xx} describes self-excitation of the excitatory cell population, I_ξ sums up the inputs from external stimuli and connected oscillators (minus a normalizing current).

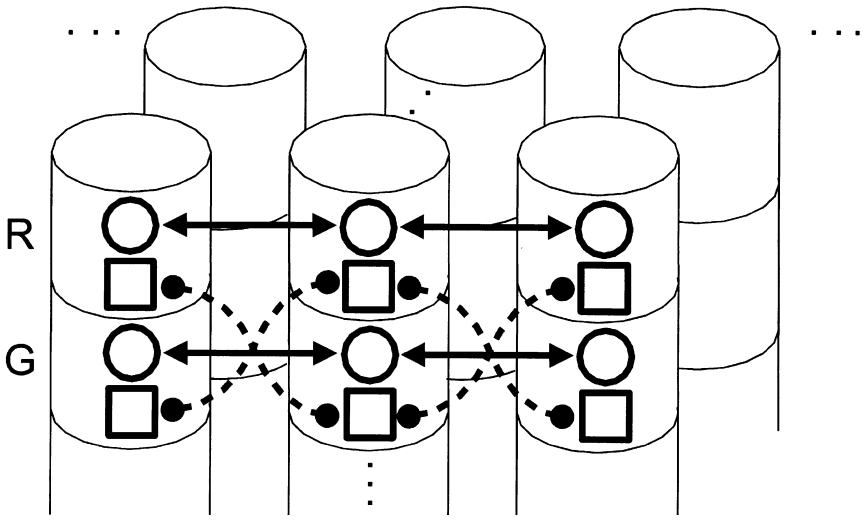

Figure 1b. Synchronizing connections (solid) are realized by mutually excitatory connections between the excitatory neurons and hold between oscillators within one layer. Desynchronizing connections (dotted) are realized by mutually inhibitory connections between the inhibitory neurons and hold between different layers. 'R' and 'G' denote the red and green channel.

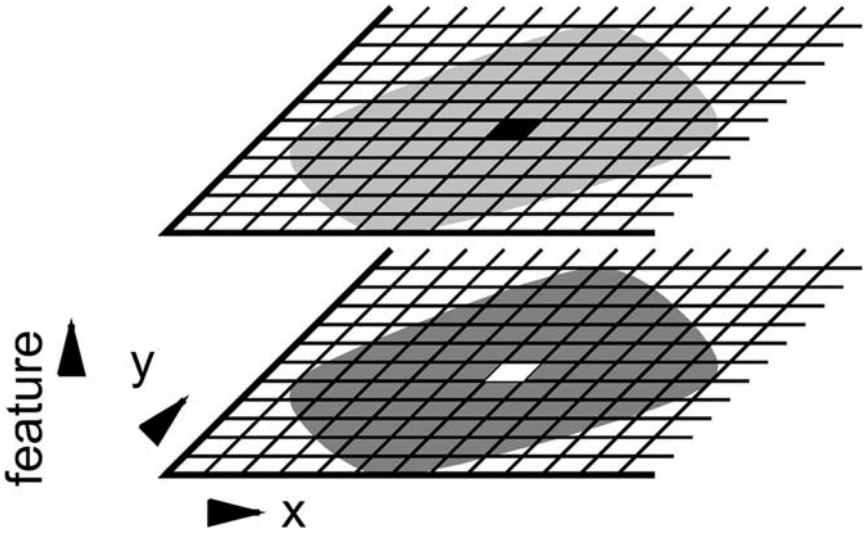

Figure 1c. Oscillators are arranged in a 3D-topology. The shaded circles visualize the range of synchronizing (light gray) and desynchronizing (dark gray) connections of a neuron in the top layer (black pixel).

Figure 2a. Stimulus: a green horizontal and red vertical bar.

The solutions of (2) are oscillations. For a more detailed description of the network see Maye (2002).

Oscillators are arranged on a three-dimensional grid forming a feature module (see Figures 1b and c). Two dimensions represent the spatial domain, while the feature is encoded by the third dimension. Spatially close oscillators that represent similar properties synchronize. The desynchronizing connections establish a phase lag between different groups of synchronously oscillating clusters. This can be viewed as an implementation of some of the well known Gestalt principles of perception. According to those principles, proximal elements in the stimulus tend to be perceived as belonging to one and the same object if they exhibit like properties. Feature modules for different feature dimensions, e.g., color and orientation, can be combined by establishing synchronizing connections between oscillators of different modules in case they code for the same stimulus region.

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. We observe that for each represented object a certain oscillation spreads through the network. The oscillation pertains only to oscillators that represent properties of the object in question.

A great number of neurobiological studies have by now corroborated the view that cortical neurons are rather plausibly modelled by oscillatory networks (Singer and Gray 1995; Schillen and König 1994; Werning 2001). Two hypotheses are supported:

HYPOTHESIS 1 (Indicativity). There are clusters of neurons that show activity only when an object in the receptive field instantiates a certain property (Hubel and Wiesel 1962). These clusters are called feature clusters (in the network: feature layers).

HYPOTHESIS 2 (Synchrony). Neurons of different feature clusters show synchronous oscillations only if the properties indicated by each feature

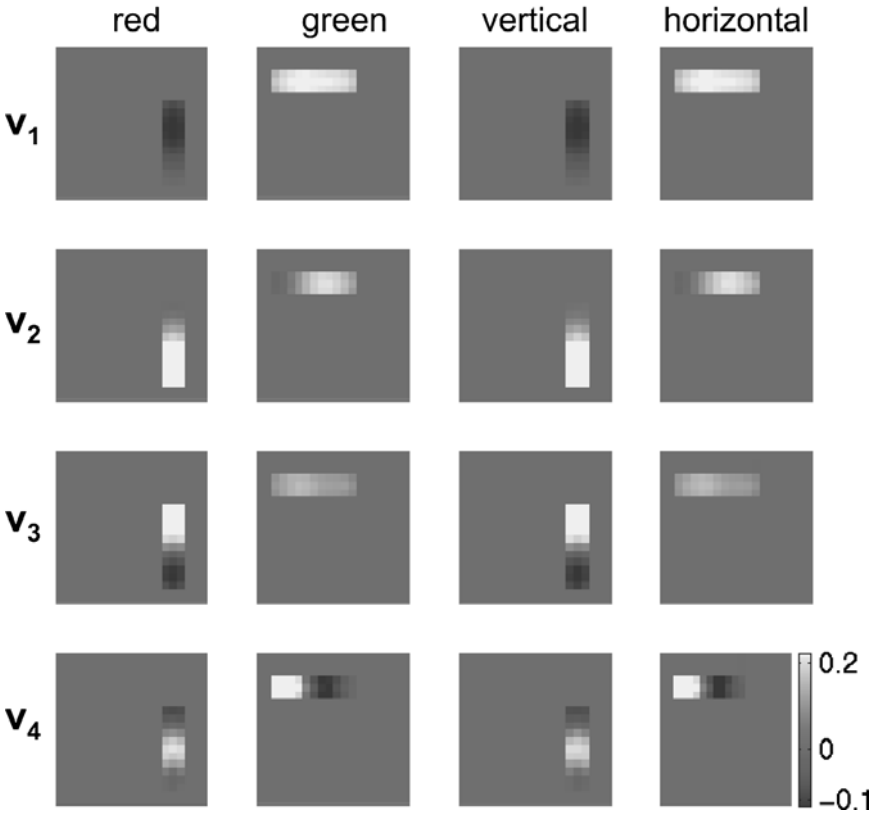

Figure 2b. Network state after stimulation with stimilus of Figure 2a. Each of the four eigenmodes $\vec{v_1}, \ldots, \vec{v_4}$ with the largest eigenvalues is shown in one line. The four columns correspond to the four feature layers.

cluster are instantiated by the same object in the receptive field (Gray and Singer 1989).

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 $[-\frac{T}{2}, +\frac{T}{2}]$. 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 countable basis $\left\{\frac{1}{\sqrt{T}} \exp(\frac{ni2\pi t}{T}) \mid n \in \mathbb{Z}\right\}$ and the inner product

$$\langle x(t)|x'(t)\rangle = \int_{-T/2}^{+T/2} \overline{x(t)}\, x'(t)dt, \tag{3}$$

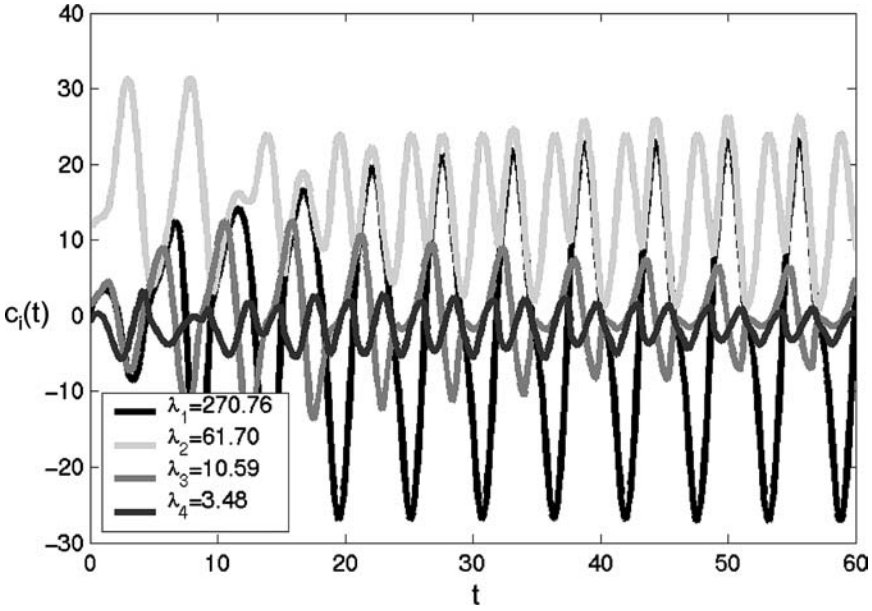

Figure 2c. The characteristic functions $c_i(t)$ show the temporal evolution of the four eigenmodes of Figure 2b.

where $\overline{x(t)}$ signifies the conjugate complex of $x(t)$. The degree of synchrony between two oscillations lies between -1 and $+1$ and is defined as

$$\Delta(x, x') = \langle x|x'\rangle / \sqrt{\langle x|x'\rangle\langle x|x'\rangle}. \tag{4}$$

The degree of synchrony corresponds to the cosine of the angle between the Hilbert vectors x and x'. The vectors are parallel, anti-parallel and orthogonal depending on whether $\Delta(x, x')$ is $+1, -1$ or 0. The overall dynamics of the network is given by the Cartesian vector $\vec{x}(t) = (x_1(t), \ldots, x_k(t))^T$. The vector comprises the activities of the excitatory neurons of all k oscillators of the network, each of which is determined by a solution of (2). From synergetics it is well known that the dynamics of complex systems is often governed by a few dominating states. These states are the eigenmodes of the system. The corresponding eigenvalues designate how much of the variance is accounted for by that mode. The eigenmodes $\vec{v}$ of the network dynamics are computed as the eigenvectors of the auto-covariance matrix $C \in \mathbb{R}^{k\times k}$, i.e., as solutions of the equation $C\vec{v} = C\lambda$, where the components $C_{jj'}$ of C are given as $C_{jj'} = \langle x_j|x_{j'}\rangle$. The temporal evolution of each eigenmode $\vec{v}_i$ (Figure 2b) is described by a characteristic function $c_i(t)$ (Figure 2c). The network state at any instant

is considered as a superposition of the eigenmodes $\vec{v}_i$ weighted by the corresponding characteristic functions $c_i(t)$:

$$\vec{x}(t) = \sum_i c_i(t)\vec{v}_i. \tag{5}$$

The eigenmodes, for any stimulus, can be ordered strictly along their (presumably non-degenerate) eigenvalues: $\lambda_i > \lambda_{i+1}$. This allows us to introduce the useful convention of signifying each eigenmode by the index $i \in \mathbb{N}$. For any stimulus we have the mapping: $i \mapsto \langle \vec{v}_i, c_i(t), \lambda_i \rangle$.

3. FIRST STEPS INTO SEMANTICS

In this section, I will develop a heuristics that allows us to interpret the dynamics of oscillatory networks in semantic terms. Later on, I will provide a more explicit and fully semantic account of the network dynamics. Oscillatory networks that implement the Hypotheses 1 and 2, I argue, realize a semantics of a monadic first order predicate language with identity $PL^{=}$.

Because of Hypothesis 2 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 $Ind = \{a_1, \ldots, a_m, z_1, \ldots, z_n\}$ be the set of individual terms of $PL^{=}$, then the partial function

$$\alpha : Ind \to L_2[-\frac{T}{2}, +\frac{T}{2}] \tag{6}$$

is a constant individual assignment of the language. By convention, I will assume that, unless indicated otherwise, $\mathrm{dom}(\alpha) = \{a_1, \ldots, a_m\}$ 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 $a_1, \ldots, a_m$. I will furthermore use bold print to signify the oscillation function assigned to an individual term: $\alpha(a) = \mathbf{a}$.

Following (4), the identity 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

$$d : Cls \times \mathbb{N} \to [-1, +1]. \tag{7}$$

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

(8) $\quad d(a = b, i) = \Delta(\mathbf{a}, \mathbf{b})$.

Most vector components of the first eigenmode of Figure 2b are exactly zero (marked middle grey), while few in the greenness and the horizontality layers are positive (marked light grey) and few in the redness and the verticality layers are negative (marked dark grey). Since the contribution of the eigenmode vector $\vec{v}_1$ to the entire network state temporally evolves according to its characteristic function $c_1(t)$, any positive eigenmode component $v_1^j = +|v_1^j|$ contributes to the activity of the j-th oscillator with $+|v_1^j|c_1(t)$, while any negative component $v_1^l = -|v_1^l|$ contributes with $-|v_1^l|c_1(t)$ to the activity of the l-th oscillator. Since the Δ-function is normalized, only the signs of the constants matter to determine that the activities of the j-th and the l-th oscillator, contributed by the first eigenmode, are exactly anti-parallel, while any two, with $c_1(t)$ temporally evolving components of equal signs contribute mutually parallel activity. We may interpret this by saying that the first eigenmode represents two objects as different 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. These considerations, for every eigenmode i, justify the following evaluation of non-identity (Notice that unlike identity, its negation is represented by the network as sharp, i.e., non-gradual):

$$(9) \qquad d(\neg a = b, i) = \begin{cases} +1 \text{ if } d(a = b, i) = -1, \\ -1 \text{ if } d(a = b, i) > -1. \end{cases}$$

Following hypothesis 1, feature clusters 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 $Pred = \{F_1, \ldots, F_r\}$ such that each predicate denotes a property featured 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 $\vec{v}_i$, renders the sub-vector of those components that belong to the feature cluster expressed by F:

(10) $\quad \beta : Pred \rightarrow \{0, 1\}^{k \times k}$.

With respect to our particular network, the matrix $\beta(red)$, e.g., is zero everywhere except for the first $\frac{k}{4}$ diagonal elements. Since $\beta(F)$ does not

vary from eigenmode to eigenmode, it is sensible to call it the *neuronal intension* of F. By convention, I will use bold print to signify the neuronal intension of predicates: $\beta(F) = \mathbf{F}$.

The neuronal intension of a predicate, for every eigenmode, determines its neuronal extension, i.e., the set of those oscillations that the neurons on the assigned feature layer, per eigenmode, contribute to the dynamics of the network. Hence, for every predicate F its *neuronal extension* in the eigenmode i comes to the set of activity functions $\{f_j | \vec{f} = \mathbf{F}\vec{v}_i c_i(t)\}$. 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 expresses a representational state of our system, with respect to the eigenmode i, in the following way:

$$(11) \qquad d(Fa, i) = \max\{\Delta(\mathbf{a}, f_j) | \vec{f} = \mathbf{F}\vec{v}_i c_i(t)\}.$$

Having now provided a sematic 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: A binary operation $\mathbf{t}$ in the real interval $[-1, +1]$ is a t-norm iff it is (i) associative, (ii) commutative, (iii) non-decreasing in the first element, i.e., satisfies $d \leq d' \Rightarrow \mathbf{t}(d, d'') \leq \mathbf{t}(d', d'')$ for all $d, d', d'' \in [-1, +1]$, and (iv) has 1 as neutral element. 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 connective can be derived by systematic considerations (cf. Gottwald 2001). 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:

$$(12) \qquad d(\phi \wedge \psi, i) = \min\{d(\phi, i), d(\psi, i)\}.$$

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 (Figure 2a), 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*

Figure 3a. Stimulus: two red vertical bars.

($= V$) and *horizontal* ($= H$) to the redness, greenness, verticality and horizontality layers as their neuronal intensions. Simple computation then reveals:

$$(13) \quad d(Ra \wedge Va \wedge Gb \wedge Hb \wedge \neg a = b, 1) = 1.$$

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 in Figure 3a 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 Figure 3b). Eigenmodes, thus, play a similar role for neuronal representation as possible worlds for semantics. They do not interfere with each other because eigenmodes are mutually orthogonal. Moreover, the identity of oscillation functions as well as the neuronal intensions of predicates apply across eigenmodes. It is also a nice feature that they can be ranked and re-ranked along their eigenvalues. The results of Spohn (1988), who provides a semantics of ranked models for a non-monotonic calculus, naturally apply. To spell this idea out would go far beyond the scope of this paper, though. I just want to mention that each of the two stable eigenmodes shown in Figure 3b can be expressed by a disjunctive sentence, if we semantically evaluate disjunction as follows:

$$(14) \quad d(\phi \vee \psi, i) = \max\{d(\phi, i), d(\psi, i)\}.$$

We are leaving the heuristic approach now and turn to a formally explicit description of the neuronal semantics realized by oscillatory networks.

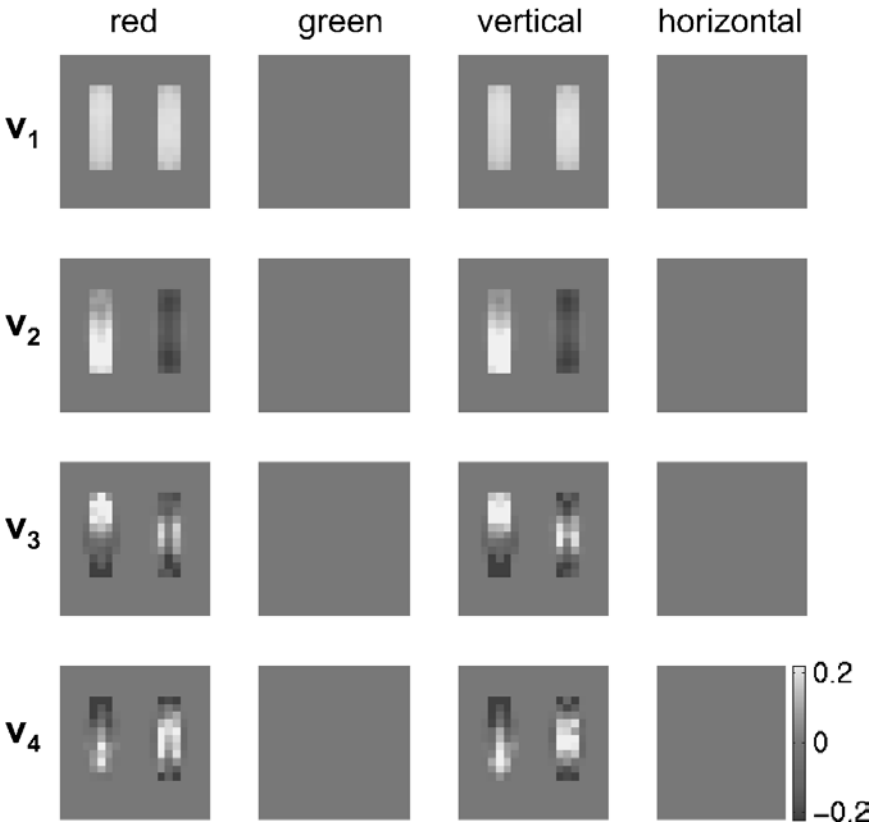

Figure 3b. The first eigenmode represents the stimulus of Figure 3a as one red vertical object, while the second mode represents it as two red vertical objects.

4. MAKING SYNTAX AND SEMANTICS EXPLICIT

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 eigenmode $i \in \mathbb{N}$, it renders a determinate eigenmode vector $\vec{v}_i$, 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 (6) and (10), the union $\gamma = \alpha \cup \beta$ 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

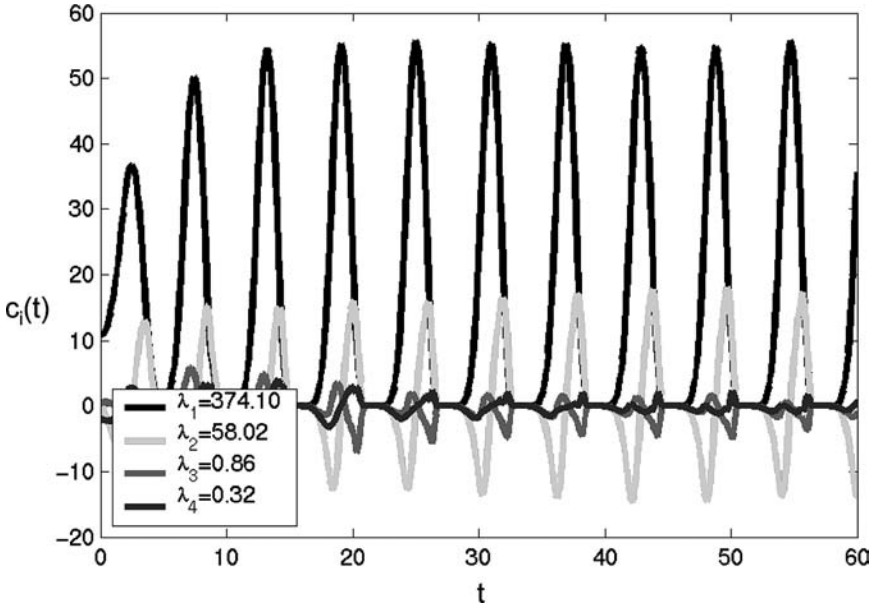

Figure 3c. The characteristic functions of the eigenmodes of Figure 3b. Only the first two characteristic functions are non-decreasing and thus belong to stable eigenmodes.

the set SF of sentential formulae as their recursive closure can be defined as follows, for arbitrary $a, b, z \in Ind$, $F \in Pred$, and $\phi, \psi \in SF$:

$$
\begin{array}{ll}
\sigma_{=} : (a, b) \mapsto a = b; & \sigma_{pred} : (a, F) \mapsto Fa; \\
\sigma_{\neg} : \phi \mapsto \neg\phi; & \sigma_{\wedge} : (\phi, \psi) \mapsto \phi \wedge \psi; \\
\sigma_{\vee} : (\phi, \psi) \mapsto \phi \vee \psi; & \sigma_{\rightarrow} : (\phi, \psi) \mapsto \phi \rightarrow \psi; \\
\sigma_{\exists} : (z, \phi) \mapsto \exists z\phi; & \sigma_{\forall} : (z, \phi) \mapsto \forall z\phi.
\end{array} \tag{15}
$$

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* iff, 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 universes), 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 $a, b \in Ind \cap \mathrm{dom}(\gamma)$ such that $\gamma(a) = \mathbf{a}$, $\gamma(b) = \mathbf{b}$, then $i \models^{d}_{\gamma} a = b$ iff $d = \Delta(\mathbf{a}, \mathbf{b})$.

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

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

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

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

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 $d = \sup\{d' \mid i \models^{d'}_{\gamma'} \phi \text{ where } \gamma' = \gamma \cup \{\langle z, \mathbf{z}\rangle\} \text{ and } \mathbf{z} \in L_2[-\frac{T}{2}, +\frac{T}{2}]\}$.

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 $d = \inf\{d' \mid i \models^{d'}_{\gamma'} \phi \text{ where } \gamma' = \gamma \cup \{\langle z, \mathbf{z}\rangle\} \text{ and } \mathbf{z} \in L_2[-\frac{T}{2}, +\frac{T}{2}]\}$.

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. An interesting feature of negation is that its duplication digitalizes the values of d into $+1$ and -1. The evaluation of implication, too, follows the Gödel system.[2] The evaluation of the quantifiers follows standard methods in semantics. Calculi for our semantics have been developed in the literature (cf. Gottwald 2001). The value of an universally quantified implication of the form $(\forall z)(Fz \rightarrow F'z)$ provides a measure for the overall synchronization between feature clusters expressed by the predicates F and F'. 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. The work done so far leads us directly to the following theorem:[3]

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, provided that μ, for every $t \in L$, is defined in the following way:

$$(16) \qquad \mu(t) = \begin{cases} \{\langle \gamma, \gamma(t) | \gamma \text{ is a constant assignment} \rangle\} \text{ if } t \notin SF, \\ \{\langle \gamma, i, d \rangle | i \models^{d}_{\gamma} \phi\} \text{ if } t \in SF. \end{cases}$$

Consequently, $\mu(t)$ can itself be regarded as a function on the domain of constant assignments. We stipulate for any γ-grammatical term t:

$$(17) \qquad \mu_{\gamma}(t) = \begin{cases} \gamma(t) \text{ if } t \text{ is not a sentence,} \\ \{\langle i, d \rangle | \langle \gamma, i, d \rangle \in \mu(t)\} \text{ if } t \text{ is a sentence.} \end{cases}$$

The *ideal meaning* of 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. To comply with the condition of co-variation, we can choose the assignment γ in a way so that the oscillation function $\gamma(a)$ tracks the object designated by some individual term a. We can, furthermore, make sure that $\gamma(F)$ is just the cluster of neurons featuring the property expressed by some predicate F. In this case, the assignment will be called *natural*. As we have seen earlier, the network dynamics warrants that the neuronal meanings of terms with respect to the natural assignment reliably co-varies with the terms' denotations.

The compositionality of content is achieved if co-variation is warranted and the content of a representational state is identified with the denotation of the term expressing it. The only additional assumptions we need are: (i) We have the intended external constant assignment Γ that maps individual constants to their designated objects and predicates to functions that determine their extension in every possible world. (ii) We have have a relation: a possible world w externally models ϕ to degree d by assignment Γ, in symbols $w \models^{d}_{\Gamma} \phi$, for any sentence ϕ of the language and any real number $d \in [-1, +1]$. (iii) $\models_{\Gamma}$ is defined in the same way $\models_{\gamma}$ is defined except that the set of oscillation functions is replaced by the set of objects, neuronal extensions are replaced by external extensions, and the Δ-function is interpreted as the (possibly digital) degree of identity between objects. (iv) We have a denotational function ν that, *mutatis mutandis*, is defined like μ.[4]

THEOREM 2 (Compositional Contents of Oscillatory Networks). Let L be the set of terms of a $PL^{=}$-language. We assume that L has a compositional function of denotation ν. Let μ be a neuronal meaning function with domain L and let γ be the natural neuronal, and Γ the intended external assignment of L. In the case of co-variation, the natural neuronal structure $\mathcal{N} = \langle \{\gamma\} \times \mu_\gamma[L], \{\mu_=, \mu_{pred}, \mu_\neg, \mu_\wedge, \mu_\vee, \mu_\rightarrow, \mu_\exists, \mu_\forall\}\rangle$ can be compositionally evaluated with respect to content.

5. CONCLUSION

Oscillatory networks show how a structure of the cortex can be analyzed in a way so that elements of this structure can be identified with the neuronal meanings of a full-fledged first order predicate language. These internal meanings form a compositional semantics and can themselves be evaluated compositionally with respect to their external contents. The approach formulated in this paper is biologically plausible and has been supported by a number of experimental neurobiological data. Compared to alternative connectionist approaches, the account presented here is superior in that it not only implements a compositional semantics of meanings, but shows how internal meanings can co-vary with external contents. The theory developed amounts to a new mathematical description of the temporal structure the cortex is known to exhibit. Cognition as realized by biological systems takes place inherently in the medium of time. The task of the neuronal hardware, only, is to keep this truly sublime structure alive.

ACKNOWLEDGEMENTS

Many data presented here have been attained in co-operation with the Neural Information Processing Group at TU Berlin. I am particularly thankful to Alexander Maye for the kind permission to print the diagrams in Figures 2 and 3.

NOTES

[1] For a defense of co-variationism see Fodor (1992). I favor the view that co-variation is an asymmetric and probabilistic dependency relation.

[2] The deeper rationale behind this definition is the adjointness condition, which relates the evaluation of implication $\mathbf{i}$ to the t-norm $\mathbf{t}$ (= min, by our choice) (cf. Gottwald 2001, p. 92): $d' \leq \mathbf{i}(d'', d''') \Leftrightarrow \mathbf{t}(d', d'') \leq d'''$.

[3] To prove the theorem, one has to show that for any of the syntactic operations in (15), there is a semantic operation that satisfies (1). To do this for the first six operations, one simply reads the bi-conditionals in the definition of $\models$ as the prescriptions of functions: $\mu_{=} : (\mu(a), \mu(b)) \mapsto \{\langle \gamma, i, d\rangle | d = \Delta(\mu_\gamma(a), \mu_\gamma(b)\}$; $\mu_{pred} : (\mu(a), \mu(F)) \mapsto \{\langle \gamma, i, d\rangle | d = \max\{\Delta(\mu_\gamma(a), f_j) | \vec{f} = \mu_\gamma(F)\vec{v_i}c_i(t)\}\}$; $\mu_\wedge : (\mu(\phi), \mu(\psi)) \mapsto \{\langle \gamma, i, d\rangle | d = \min\{d', d'' | \langle \gamma, i, d'\rangle \in \mu(\phi), \langle \gamma, i, d''\rangle \in \mu(\psi)\}\}$, etc. To attain semantic counterpart operations for $\sigma_\exists$ and $\sigma_\forall$, we have to apply the method of cylindrification: $\mu_\exists : \mu(\phi(z)) \mapsto \{\langle \gamma, i, d\rangle | \exists \gamma' : \mathrm{dom}(\gamma') = \mathrm{dom}(\gamma) \cup \{z\}$ and $\langle \gamma', i, d\rangle \in \mu(\phi(z))\}$; $\mu_\forall : \mu(\phi(z)) \mapsto \{\langle \gamma, i, d\rangle | \forall \gamma' : \mathrm{dom}(\gamma') = \mathrm{dom}(\gamma) \cup \{z\} \Rightarrow \langle \gamma', i, d\rangle \in \mu(\phi(z))\}$. One easily verifies that (1) is satisfied.

[4] Proof: Because of co-variation we have a function κ such that $\Gamma = \kappa \circ \gamma$. Furthermore, $\nu_\Gamma(t) = \mu_{\kappa\circ\gamma}(t)$ for every $t \in L$, provided the interpretation of the Δ-function is adjusted. The semantic operations ν_σ are the same as μ_σ except that the interpretation of the Δ-function is altered appropriately. The intended denotational structure $\mathcal{W} = \langle \{\Gamma\} \times \nu_\Gamma[L], \{\nu_{=}, \nu_{pred}, \nu_\neg, \nu_\wedge, \nu_\vee, \nu_\rightarrow, \nu_\exists, \nu_\forall\}\rangle$, hence, is a homomorphous image of $\mathcal{N}$.

REFERENCES

Fodor, J.: 1992, *A Theory of Content and Other Essays*, MIT Press, Cambridge, MA.

Fodor, J. and E. Lepore: 2002, *The Compositionality Papers*, Oxford University Press, Oxford.

Fodor, J. and Z. Pylyshyn: 1988, 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition* **28**, 3–71.

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*, Research Studies Press, Baldock.

Gray, C. M. and W. Singer: 1989, 'Stimulus-Specific Neuronal Oscillations in Orientation Columns of Cat Visual Cortex', *Proceedings of the National Academy of Sciences, USA* **86**, 1698–1702.

Hodges, W.: 2001, 'Formal Features of Compositionality', *Journal of Logic, Language and Information* **10**, 7–28.

Horgan, T. and J. Tienson: 1996, *Connectionism and the Philosophy of Psychology*, MIT Press, Cambridge, MA.

Hubel, D. H. and T. N. Wiesel: 1962, 'Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex', *Journal of Physiology* **160**, 106–154.

Janssen, T.: 1997, 'Compositionality', in J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Language*, Elsevier, Amsterdam, pp. 417–473.

Kim, J.: 1993, *Mechanism, Purpose and Explanatory Exclusion*, Cambridge University Press, Cambridge, MA.

Maye, A.: 2002, 'Neuronale Synchronität, zeitliche Bindung und Wahrnehmung', Ph.D. thesis, TU Berlin, Berlin.

Schillen, T. B. and P. König: 1994, 'Binding by Temporal Structure in Multiple Feature Domains of an Oscillatory Neuronal Network', *Biological Cybernetics* **70**, 397–405.

Singer, W. and C. M. Gray: 1995, 'Visual Feature Integration and the Temporal Correlation Hypothesis', *Annual Review of Neuroscience* **18**, 555–586.

Smolensky, P.: 1995, 'Connectionism, Constituency and the Language of Thought', in C. Macdonald and G. Macdonald (eds.), *Connectionism*, Blackwell, Cambridge, MA, pp. 164–198.

Spohn, W.: 1988, 'Ordinal Conditional Functions', in W. Harper and B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics*, Reidel, Dordrecht, pp. 105–134.

Werning, M.: 2001, 'How to Solve the Problem of Compositionality by Oscillatory Networks', in J. D. Moore and K. Stenning (eds.), *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum Associates, London, pp. 1094–1099.

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