Chaos and Complexity Letters
Volume 2, Number 2/3, pp. 435–452
ISSN 1555-3995

THE CORTICAL IMPLEMENTATION OF COMPLEX ATTRIBUTE AND SUBSTANCE CONCEPTS: SYNCHRONY, FRAMES, AND HIERARCHICAL BINDING

Markus Werning[1]****and Alexander Maye***[†]
[1]Department of Philosophy
Heinrich-Heine University Düsseldorf
Universitätsstraße 1, 40225 Düsseldorf, Germany
[2]Department of Neurophysiology and Pathophysiology
University Medical Center Hamburg-Eppendorf
Martinistraße 52, 20246 Hamburg, Germany

Abstract

Combining frame theory with the theory of neural synchronization, the paper proposes oscillatory networks as a model for the realization of complex attribute and substance concepts in the cortex. The network has both perceptual and semantic capabilities. Coherency chains and hierarchical binding mechanisms are postulated.

1. Introduction

Frame theory (Minsky, 1975; Barsalou, 1992) provides us with a universal account not only for the psychology of categorization, but also for the decomposition of concepts. A frame is defined for a large domain of things and contains a fixed set of attribute dimensions (e.g., color, form) each of which allows for a number of different attribute values (*red*, *green*, ...; *round*, *square*, ...). Frames can be nested hierarchically and mutual constraints between attribute dimensions (e.g., between size and weight) can be incorporated. The frame-theoretic approach is especially interesting for the decomposition of substance concepts into attributive concepts. Whereas attributive concepts represent features that are volatile in the sense that one and the same thing can fall under different attributive concepts at different times (an object may, e.g., change its color, its size or its speed), substance concepts represent permanent features that govern the identity conditions of objects (a tomato, e.g., ceases to exist when it no longer falls under the substance concept [*tomato*], say, because it has been mashed). If one now assumes that the attribute values of a frame predominantly are attributive concepts, each substance concept that corresponds to a category captured by a frame is likely to be decomposable into the respective attributive concepts.

[*] E-mail address: werning@phil.uni-duesseldorf.de
[†] E-mail address: a.maye@uke.uni-hamburg.de

When it comes to the neuronal implementation of frames, however, few biologically realistic models have been proposed yet.[1] This paper attempts to compensate this disadvantage and tries to combine frame theory with the theory of neural synchronization (Engel & Singer, 2001), which has been proposed as a solution to the binding problem. According to this theory, neurons synchronize their activity if they code properties of the same object and de-synchronize if they code for different objects. Synchronization, thus, is considered as a vital mechanism for the binding of property representations into object representations.

In the first half of the paper we will consider the composition of complex attributive concepts and their neural correlates according to the theory of neural synchronization. An example is the complex attributive concept [*red vertical*], which is composed from the concepts [*red*] and [*vertical*]. Complex concepts of this kind are usually not lexicalized: There is, e.g., no syntactically primitive expression for the concept [*red vertical*].

In the second half of the paper we will then turn to complex substance concepts. According to frame theory, a substance concept like [*elephant*] is complex because it can be decomposed into less complex concepts. A potential decomposition of the concept [*elephant*] might not only involve concepts of prototypical properties of elephants, e.g., [*animate*], [*big*], [*gray*], but also concepts of prototypical parts of elephants, such as [*trunk*], [*tusks*], [*big ears*], [*short tail*]. Many substance concepts are lexicalized.

A main working hypothesis of ours is that the distinction between primitivity and complexity on the conceptual level corresponds to a distinction between locality and distributedness on the neural level. We thus expect the neural correlates of primitive attributive concepts (e.g., for color, orientation) to be relatively local. In fact, cortical areas for more than 30 attribute dimensions could be anatomically located in the monkey (Felleman & van Essen, 1991). In contrast, the correlates of substance concepts, we assume, are highly distributed neural states. Substance concepts are thus not expected to be realized by single cells – so-called 'grandmother cells' (Gross, 2002) – as, e.g., Quian Quiroga, Reddy, Kreiman, Koch, and Fried (2005) argue, but by cell assemblies that may pertain to highly distinct parts of the cortex (Hebb, 1949; Wennekers & Palm, 2000).

Another working hypothesis assumes that the mechanism that binds together the distributed neural activities in the case of complex attributive concepts is the intra-cortical synchronization of the electrical discharges of neurons (Engel, König, Gray, & Singer, 1990) and that this mechanism can be modeled by oscillatory networks (Schillen & König, 1994) in a biologically realistic way.

What we envisage is that the mechanism of neural synchronization is general enough to bind together also the representations of the attribute values for a complex substance concept within a frame. For this purpose neuronal coherency chains are postulated. In the case of nested frames – as is relevant, e.g., if the representations of the parts of an object are integrated in the representation of the object with its parts – a mechanism of hierarchical

[1] Werning (2005b) explicitly discusses Smolensky's Integrated Connectionist/Symbolic Architecture (Smolensky, 1991/1995a, 1995b; Smolensky & Legendre, 2005a, 2005b). Together with Shastri and Ajjanagadde's (1993) it shares some of the advantages of the model proposed in the present paper. However, the main disadvantage of Smolensky's as well as Shastri and Ajjanagadde's models is that they cannot explain how internal semantic states get empirical content. There is no co-variation relation between internal states and external content. (see also Fodor & McLaughlin, 1990, and Fodor, 1997, for this line of criticism). This failure is due to the fact that unlike our model – for a detailed vindication of co-variation see Werning (2005a) – their models have no perceptual capabilities.

binding is postulated.

2. Oscillatory Networks

Von der Malsburg's (1981) supposition that the synchronous oscillation of neuronal responses constitutes a mechanism that binds the responses of feature specific neurons when these features are instantiated by the same object has been frequently applied to explain the integration of distributed responses. Object-related neuronal synchrony has been observed in numerous cell recording experiments and experiments related to attention, perception and expectation (reviewed by Singer, 1999). Those data suggest the hypothesis that the neural basis of object concepts are oscillation functions and that the neural basis of predicate concepts are clusters of feature-specific neurons.

From Gestalt psychology the principles governing object concepts are well known. According to some of the Gestalt principles, spatially proximal elements with similar features (similar color/similar orientation) are likely to be perceived as one object or, in other words, represented by one and the same object concept. 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, while 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 exhibit synchronous oscillation functions 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.

In our model a single oscillator (Fig. 1a) consists of two mutually coupled excitatory and inhibitory neurons. They are assigned the variables x and y, which statistically represent the electrical discharge behavior of a population of 100 to 200 biological cells. Populations of recurrently coupled excitatory and inhibitory neurons can be found in the primary visual cortex. Here, excitatory (pyramidal) cells in layers 2 and 3 are tightly coupled to local inhibitory neurons in layers 2 to 6. If the number of excitatory and inhibitory biological cells is large enough, the dynamics of each oscillator can be statistically described by the temporal evolution of the variables x and y according to the following differential equations:

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

Here, τ_ξ ($\xi \in \{x,y\}$) are constants that can be chosen to match refractory times of biological cells. g_ξ are semi-linear transfer functions chosen to prevent neurons from reaching saturation:

$$g_\xi(x) = \begin{cases} m_\xi(x - \theta_\xi) & \text{if } x > \theta_\xi \\ 0 & \text{else.} \end{cases} \tag{2}$$

L_0^{xx} describes self-excitation of the excitatory cell population, I_ξ are external inputs and N_ξ white noise, which models fluctuation within the cell populations. With I_ξ above threshold, the solutions of (1) are limit-cycle oscillations (Maye, 2003).

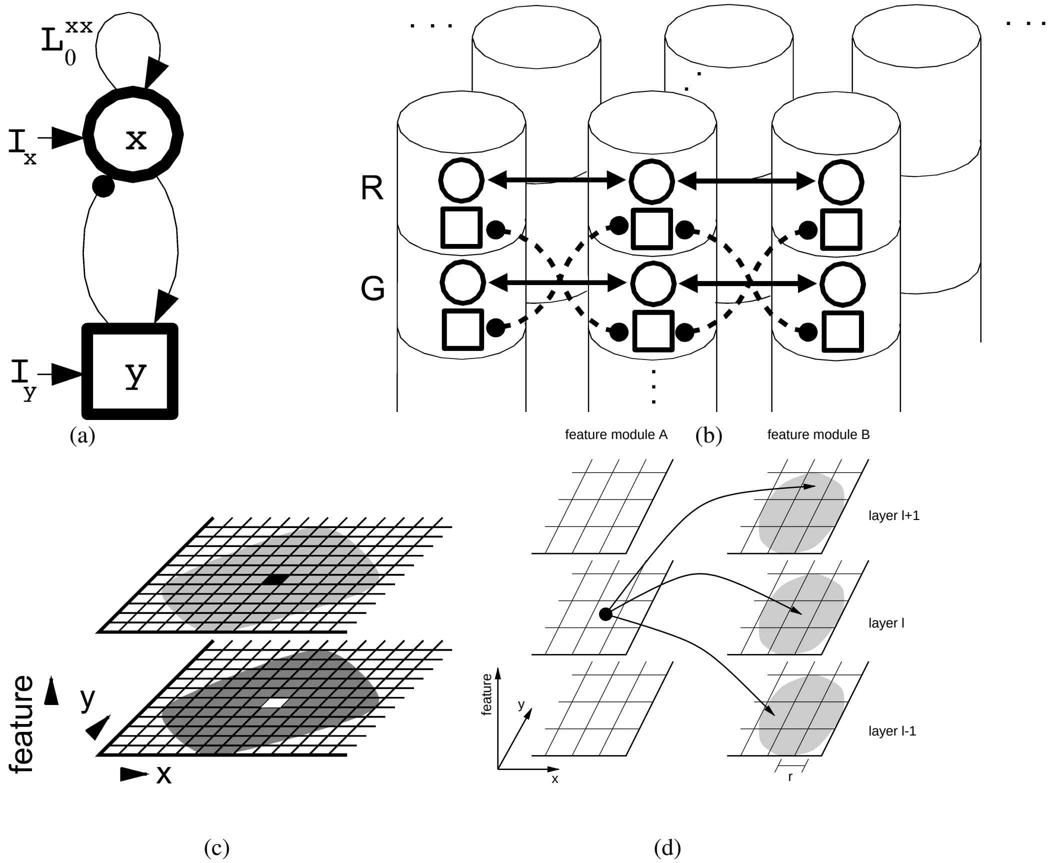

Figure 1. Oscillatory network. a) A single oscillator consists of an excitatory (x) and an inhibitory (y) neuron. Each neuron represents the average activity of a cluster of biological cells. L_0^{xx} describes the self-excitation of the excitatory neuron. I_x and I_y amounts to external input. b) Synchronizing 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 & Werning (2004).

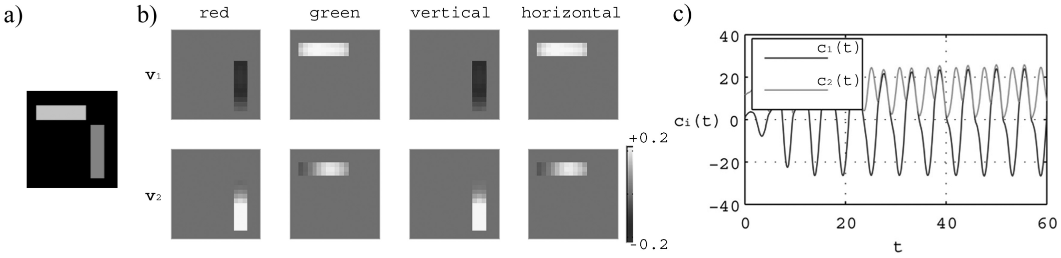

Figure 2. a) Stimulus: one vertical red bar and one horizontal green bar. It was presented to a network with $32 \times 32 \times 4$ oscillators. b) The two stable eigenmodes. The eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are shown each in one line. The four columns correspond to the four feature layers. Dark shading signifies negative, gray zero and light shading positive components. c) The characteristic functions for the two eigenmodes.

Oscillators are arranged on a three-dimensional grid forming a feature module (Fig. 1c). 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. Synchronizing (resp. de-synchronizing) connections between two oscillators are realized by mutually excitatory (inhibitory) connections between the excitatory (inhibitory) neurons of both oscillators (Fig. 1b). 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 (Fig. 1d).

Stimulated oscillatory networks (e.g., by stimulus of Fig. 2a, for other stimuli see Maye & Werning, 2004), 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 the properties of the object in question.

3. The Neural Correlate of Complex Attributive Concepts

An oscillation function $x(t)$ of an oscillator is the activity of its excitatory neuron as a function of time during a time window $[0,T]$. Mathematically speaking, activity functions can be conceived of as vectors in the Hilbert space $L_2[0,T]$ of functions that are square-integrable in the interval $[0,T]$. Thus, a precise measure of synchrony can be established and a powerful algebraic framework for the semantic interpretation of the network is provided. The Hilbert space $L_2[0,T]$ has the inner product

$$\langle x(t)|x'(t)\rangle = \int_0^T x(t)\, x'(t)dt. \tag{3}$$

The degree of synchrony between two oscillations lies between -1 and $+1$ and can now be defined as

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

The dynamics of complex systems is often governed by a few dominating states, the eigenmodes. The corresponding eigenvalues designate how much of the dynamics is accounted for by that mode. The two stable eigenmodes of a stimulated network are shown in Fig. 2b.[2] The overall dynamics of the network is given by the Cartesian vector $\mathbf{x}(t) = (x_1(t),...,x_k(t))^T$ that contains the excitatory activities of all k oscillators as components. The network state at any instant is considered as a superposition of the temporally constant, but spatially variant eigenvectors $\mathbf{v}_i$ weighted by the corresponding spatially invariant, but temporally evolving characteristic functions $c_i(t)$ of Fig. 2c:

$$\mathbf{x}(t) = \sum c_i(t)\mathbf{v}_i. \tag{5}$$

The eigenmodes, for any stimulus, can be ordered along their eigenvalues so that each eigenmode can be signified by a natural number i beginning with 1 for the strongest ($\mathbf{v}_i$ is the corresponding eigenvector).

The Hilbert space analysis allows us to interpret the dynamics of oscillatory networks in semantic terms. Since oscillation functions reliably co-vary with objects, they may be assigned to some of the individual terms $\mathrm{a},\mathrm{b},...,\mathrm{x},\mathrm{y},... \in \mathrm{Ind}$ of a predicate language by the partial function

$$\alpha : \mathrm{Ind} \to L_2[0,T]. \tag{6}$$

The sentence $\mathrm{a} = \mathrm{b}$ expresses a representational state of the system (i.e., the representation of the identity of the objects denoted by the individual terms a and b) to the degree the oscillation functions $\alpha(\mathrm{a})$ and $\alpha(\mathrm{b})$ of the system are synchronous. The degree to which a sentence ϕ expresses a representational state of the system, for any eigenmode i, can be measured by the value $d_i(\phi) \in [-1,+1]$. In case of identity sentences we have:

$$d_i(\mathrm{a} = \mathrm{b}) = \Delta(\alpha(\mathrm{a}),\alpha(\mathrm{b})). \tag{7}$$

When we take a closer look at the eigenvector of the first eigenmode in Fig. 2b, we see that most of the vector components are exactly zero (gray shading). However, few components in the greenness and the horizontality layers are positive (light shading) and few components in the redness and the verticality layers are negative (dark shading). 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 characteristic function $+c_1(t)$ and the representation of the second object is its mirror image $-c_1(t)$ (Because of the normalization of the Δ-function, only the signs of the eigenvector components matter). These considera-

[2] Only the stable eigenmodes, i.e. those eigenmodes whose characteristic function does not converge to zero, will be of concern to us when it comes to semantic interpretation. A related stability constraint has been proposed recently by Atmanspacher and beim Graben (2005).

tions justify the following evaluation of non-identity:[3]

$$d_i(\neg \mathrm{a} = \mathrm{b}) = \begin{cases} +1 \text{ if } d_i(\mathrm{a} = \mathrm{b}) = -1, \\ -1 \text{ if } d_i(\mathrm{a} = \mathrm{b}) > -1. \end{cases} \tag{8}$$

A great advantage of the eigenmode analysis is that object representations are no longer identified with the actual oscillatory behavior of neurons, but with the eigenmode-relative characteristic functions. In this approach the representation of objects does not require strict synchronization of neural activity over long cortical distances, but tolerates a traveling phase change as it has been observed experimentally (Eckhorn et al., 2001) as well as in our network simulation.

Feature layers function as representations of properties and thus can be expressed by predicates $\mathrm{F}_1, ..., \mathrm{F}_p$, i.e., to every predicate F a diagonal matrix $\beta(\mathrm{F}) \in \{0,1\}^{k\times k}$ can be assigned such that, by multiplication with any eigenvector $\mathbf{v}_i$, the matrix renders the subvector of those components that belong to the feature layer expressed by F. To determine to which degree an oscillation function assigned to an individual constant a pertains to the feature layer assigned to a predicate F, we have to compute how synchronous it maximally is with one of the oscillations in the feature layer. We are, in other words, justified to evaluate the degree to which a predicative sentence Fa (read: 'a is F', e.g., 'This object is red') expresses a representational state of our system, with respect to the eigenmode i, in the following way (the f_j are the components of the vector $\mathbf{f}$):

$$d_i(\mathrm{Fa}) = \max\{\Delta(\alpha(\mathrm{a}), f_j) | \mathbf{f} = c_i(t)\beta(\mathrm{F})\mathbf{v}_i\}. \tag{9}$$

If one, furthermore, evaluates the conjunction of two sentences $\phi \wedge \psi$ by the minimum of the value of each conjunct, we may regard the first eigenvector $\mathbf{v}_1$ of the network dynamics resulting from the stimulus in Fig. 2a as a representation expressible 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* $(= \mathrm{a})$ and *that* $(= \mathrm{b})$ to the oscillatory functions $-c_1(t)$ and $+c_1(t)$, respectively, and the predicates *red* $(= \mathrm{R})$, *green* $(= \mathrm{G})$, *vertical* $(= \mathrm{V})$ and *horizontal* $(= \mathrm{H})$ to the redness, greenness, verticality and horizontality layers as their neuronal meanings. Simple computation then reveals:

$$d_1(\mathrm{Ra} \wedge \mathrm{Va} \wedge \mathrm{Gb} \wedge \mathrm{Hb} \wedge \neg \mathrm{a} = \mathrm{b}) = 1. \tag{10}$$

Using further methods of formal semantics, the semantic evaluation of sentences has been extended to sentences comprising disjunction (∨), implication (→), and existential as well as universal quantifiers (∃,∀). The compositionality of meaning and the compositionality of content could be proven (Werning, 2005b). Co-variation with content can always be achieved if the individual assignment α and the predicate assignment β are chosen to match the network's perceptual capabilities. Werning (2003b) extends this approach from an ontology of objects to an ontology of events. Werning (2003a) integrates relational concepts

[3] The negation is sharp: it allows only -1 and $+1$ as values. In its definition we follow Gödel's (1932) intuitionist min-max-system. Double negation digitalizes the semantic value of the original sentence. The deeper reasons for this choice of the negation lie in systematic considerations concerning the envisaged calculus and the proof of completeness and compositionality theorems (Werning, 2005a).

like $[in]$. All complex attributive concepts that can be expressed by a first order predicate language with identity are in principle compositionally realizable in our network. We will now turn to semantically complex substance concepts.

4. Simple Frames and Coherency Chains

Among semanticists still some authors believe that lexical concepts are altogether not decomposable (e.g., Fodor & Lepore, 1992). According to those so-called atomist positions, only concepts that are linguistically expressible by syntactically combined expressions can be complex. In neuroscience, moreover, some authors still hold that substantial features like that of being an elephant, or even features as particular as that of being my grandmother are represented by highly specific single neurons (Logothetis & Sheinberg, 1996).

The position we advocate is contrary to those, in two respects: We think that (i) some lexical substance concepts are decomposable into less complex concepts, and that (ii) the neural realization of those concepts is distributed over assemblies of neurons and meta-assemblies thereof. In psychology, philosophy and linguistics various theories have been proposed to account for the decomposition of concepts. The most prominent ones are: prototype theory (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), lexical decomposition grammar (Hale & Keyser, 1993), and frame theory (Barsalou, 1992). For our purposes a modified version of frame theory seems to be most fruitful.

Frames are defined for a large reference class or domain D of objects and allow for a categorization therein. Lowest-level frames can be rendered by a matrix with one row per attribute dimension i and the attribute values F_{ij} of the respective dimension in the j-th column:

$$\text{dimensions}\left\{\begin{array}{c} \mathit{color}\\ \mathit{form}\\ \mathit{brightness}\\ \mathit{size}\end{array}\right. \qquad \underbrace{\begin{pmatrix} \mathit{red} & \mathit{green}\\ \mathit{round} & \mathit{square}\\ \mathit{light} & \mathit{dark}\\ \mathit{small} & \mathit{big}\end{pmatrix}}_{\text{values}} \tag{11}$$

Relative to a frame, a category or concept $\mathbf{C}$ is rendered by a matrix that results from an assignment of typicality values C_{ij} ($\in [0,1]$) to each attribute value. C_{ij} tells how typical the j-th attribute value, regarding the i-th attribute dimension, is for an object that satisfies the concept $\mathbf{C}$. Relative to the frame shown above, the matrix for the concept $[cherry]$ might, e.g., look as follows:[4]

$$\begin{pmatrix} 0.8 & 0.2\\ 1 & 0\\ 0.1 & 0.9\\ 0.9 & 0.1\end{pmatrix} \tag{12}$$

[4] Alternatively structured typicality matrices have been proposed, e.g., by Smith, Osherson, Rips, and Keane (1988). It has been extensively debated whether representations that are based on typicality measures are compatible with compositionality constraints (Osherson & Smith, 1981; Kamp & Partee, 1995; Fodor & Lepore, 1996). The authors, with Robbins (2002) and others, believe that they are, but for time and space reasons cannot enter into this discussion here.

If we assume that the list of attribute values in each dimension are a partitioning of domain D, and if we choose an appropriate system of many-valued logic, we can prove the following inequality for the truth degree $d(\in [-1,+1])$ of the existential claim $\exists \mathrm{x}\, \mathrm{Cx}$ (read: 'There are C's', e.g., 'There are cherries'):

$$d(\exists \mathrm{x}\, \mathrm{Cx}) \geq \max_{\substack{\pi \in \Pi \\ \mathrm{x} \in D}} \min_{i=1}^{m} (C_{i\pi(i)} d(\mathrm{F}_{i\pi(i)}\mathrm{x})). \tag{13}$$

Here, C is the predicate expressing the concept **C**, Π is the set of all variations of the index $j \in \{1,...,n\}$ at i-th position in a sequence of length m, where m is the number of attribute dimensions in the reference frame. For a proof of the inequality see the appendix.

On the right side of the inequality (13) we can directly apply the neuronal interpretation of the value $d(\mathrm{F}_{ij}\mathrm{x})$, which has already been developed in equation (9). From these considerations we may infer the following hypothesis about coherency chains:

Hypothesis 1 (Coherency chains)**.** *Provided that a concept, relative to a frame, is completely decomposable into primitive attributive concepts according to its matrix* **C***, the lowest boundary of the degree to which the network represents an object as an instance of the concept is given by the strength of the strongest (*max*) weighted coherency chain (weights: C_{ij}) that pertains to layers of neurons selective for the attribute values (F_{ij}) of the frame in question. Here, any coherency chain is regarded just as strong as the weakest (*min*) weighted coherence among the activities of the feature-selective neurons involved (see Fig. 3).*

5. Nested Frames and Hierarchical Binding

So far all substance concepts that are analyzable by simple frames can in principle be dealt with. Their neural correlate can be conceived of as characteristic chains of synchronized activity across various feature modules. Long-range synchronization between distributed neuronal populations has been experimentally demonstrated (Engel, König, Kreiter, & Singer, 1991). In fact it has been shown that oscillatory activity may be instrumental for establishing long-range synchronization (König, Engel, & Singer, 1995). Some complex concepts, however, are not directly decomposable into primitive attributive concepts because their analysis requires nested frames. This might, e.g., be because the objects of the category in question have characteristic parts. A frame for animals may, e.g., contain attribute dimensions for head, legs and body. The attribute values for the dimension *legs* might then be captured by another frame with attribute dimensions of its own (*number*, *length*, *orientation*, etc.).

This hierarchy of frames poses a dilemma for the theory of neural synchronization. If one, on the one hand, were to bind the attribute values of daughter frames to the attribute values of the mother frame by the overall synchronization of the corresponding feature-selective neurons, the information to which frame an attribute value belongs would be lost. If one, on the other hand, were to refrain from binding the attribute values together by some synchronization mechanism, the information that those attribute values are features of the

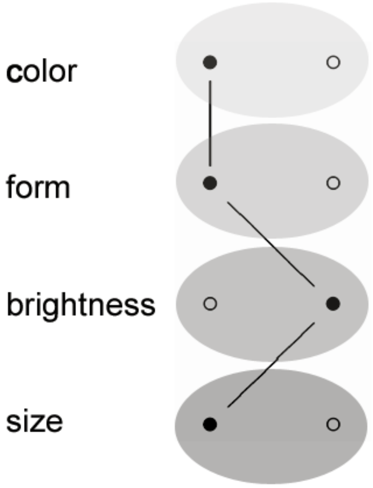

Figure 3. Schematic view of a network that represents an object as a cherry relative to the frame (11). Each oval represents an attribute dimension that is modeled by a feature module. Each of the small circles represents an attribute value that corresponds to a layer of feature(F_{ij})-selective neurons. The lines mark the strongest chain of temporally coherent activity.

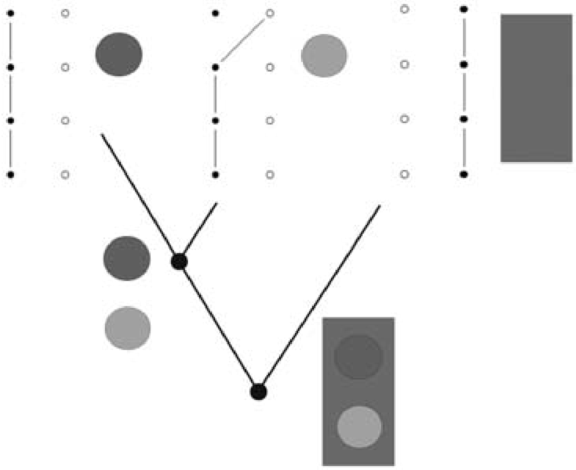

Figure 4. Schematic view of the successive composition of the complex concept [*traffic-light*] relative to the frame (11). The first row shows the characteristic patterns of temporally coherent network activity for the concepts of the object parts. A higher-order binding mechanism stepwise composes the concept of the whole object from the concepts of its parts.

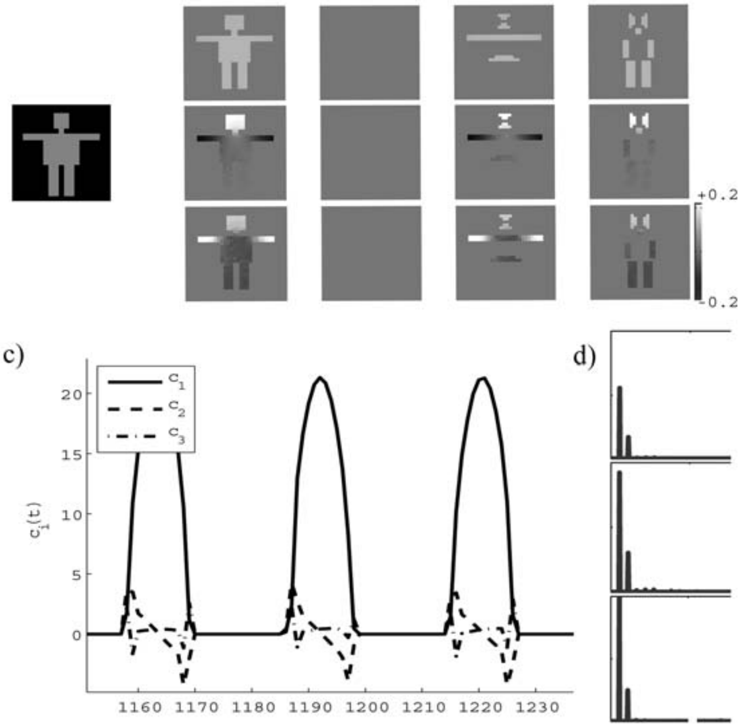

Figure 5. a) Stimulus: a red manikin with horizontally spread arms. b) The eigenvectors of the first three eigenmodes. $\mathbf{v}_1$ represents the manikin's body as a single object. $\mathbf{v}_2$ represents two parts of the body as distinct objects: the head (white) and the pair of arms (black). $\mathbf{v}_3$ distinguishes between the body's upper (head plus arms) and lower part (belly plus legs). c) The characteristic functions of the three eigenmodes. $c_1(t)$, which represents the whole body, is an envelope of the representations of the body parts $\pm c_{2/3}(t)$. d) Power spectra of the three characteristic functions.

parts of one and the same object would not be preserved. To illustrate this dilemma, take, e.g., an animal with long, vertically oriented legs and a small horizontally oriented body, say a flamingo. You can't simply represent the flamingo by synchronizing neurons for the features *long*, *vertical*, *small* and *horizontal*. But you can't either assume that there is no binding mechanism between the features of the legs and those of the body. For, how do you then represent the fact that legs and body belong to the same object?

What is in need to escape the dilemma obviously is a mechanism of hierarchical binding as schematically shown in Fig. 4. To explore this option, we stimulated our network with a complex figure: a red manikin (Fig. 5a). The first eigenmode represents the stimulus as one object, while the second and third eigenmode represent parts of the object (head, arms, belly, legs) as distinct from each other (Fig. 5b). A look at the characteristic functions of Fig. 5c reveals an interesting fact: The representation of the whole object, i.e., the first characteristic function, contains the second and third characteristic function in an envelope-like way. In the frequency spectra of the characteristic functions (Fig. 5d) one recognizes that they are made up of the same frequencies, although their temporal correlation measured by the Δ-function is close to zero. One might hence hypothesize that the envelope-like relation as well as the exploitation of the same frequencies might provide a hierarchical binding mechanism consistent with the theory of neural synchronization. Further investigation is needed here.

6. Conclusion

Both, complex attributive concepts as well as substance concepts that can be directly composed from attributive concepts can be modeled by oscillatory networks in a biologically realistic way. With regard to the latter, coherency chains are required. Further research needs to be done to account for substance concepts that decompose into concepts of parts. Here, mechanisms of hierarchical binding are postulated. Frame theory and the theory of neural synchronization may be combined to identify the neural correlate of complex concepts.

Acknowledgements

We would like to thank two anonymous referees for very helpful comments and are especially grateful to the organizers and participants of the Tandem Workshop *Advanced Methods of Electrophysiological Signal Analysis (Part A)* and *Symbol Grounding? Dynamical Systems Approaches to Language (Part B)* in Potsdam. This study has been supported by the grant *FOR 600* of the *Deutsche Forschungsgemeinschaft* and by the *Volkswagenstiftung*.

Appendix

To prove the inequality (13), we will first turn to the case of classical first order logic and will afterwards give a semantic interpretation in a many-valued system. We make the following assumptions:

1. On a given domain D there are $i = 1,...,m$ attribute dimensions (color, size, form, etc.), such that the content of a substance concept is uniquely determined by the values in those attribute dimensions.
2. For each attribute dimension i, there is a symmetric and reflexive (but not necessarily transitive) relation $\simeq_i$ of comparison with the interpretation: $x \simeq_i y$ if and only if x is the same as y with respect to the attribute dimension i (Example: x has the same color/size/form as y).
3. For each attribute dimension i, we have a class of values $F_{i1},..,F_{in(i)}$ (e.g., *red*, *green*, *blue*, *yellow*, *white*, *black* in the dimension of color) whose extensions are a partitioning of the domain.

We may now use assumption 2 for each attribute dimension i to define a functional concept B_i that maps any substance concept C to the property of being alike with the instances of C in the respective attribute dimension. Using the λ-operator for abstraction we get:

$$B_i(C) = \lambda x(\exists y\,(x \simeq_i y \wedge Cy)). \tag{14}$$

Example: If C denotes the set of cherries and i marks the attribute color, then $B_i(C)$ denotes the set of all things in the domain that have the color of some cherries.

Assumption 1 allows us to say that the content of a substance concept is fully determined by the values of all m functional concepts:

$$Cx \leftrightarrow \bigwedge_{i=1}^{m} B_i(C)x. \tag{15}$$

Example: Cherries are exactly those things that have the color, size, form, etc. of cherries.

Assumption 3, finally, justifies us to presuppose that every object x of the domain exemplifies some value in each attribute dimension. I.e., for each i, we have:

$$\bigvee_{j=1}^{n(i)} F_{ij}x. \tag{16}$$

Example: Everything has some color, i.e., it is *red*, *green*, *blue*, *yellow*, *white*, or *black* (if those are all possible values in the dimension of color; *colorless* and *varicolored* might be adopted as further values if one wishes to).

To simplify the notation, we will assume that the number of values $n(i)$ in all attribute dimensions i be the same. To achieve this, we may set

$$n = \max_{i=1}^{m} n(i) \tag{17}$$

and introduce empty predicates F_{ij} such that the denotation of F_{ij} is $\varnothing$ if $j > n(i)$.

We can now derive the following bi-conditional ('$\rightarrow$' by conjunction of premise (16); '$\leftarrow$' by simplification), for each i:

$$B_i(C)x \leftrightarrow (\bigvee_{j=1}^{n} F_{ij}x) \wedge B_i(C)x. \tag{18}$$

By insertion of (18) in (15) we get:

$$Cx \leftrightarrow \bigwedge_{i=1}^{m}((\bigvee_{j=1}^{n} F_{ij}x) \wedge B_i(C)x). \tag{19}$$

The distribution of the conjunction into the disjunction leads to:

$$Cx \leftrightarrow \bigwedge_{i=1}^{m} \bigvee_{j=1}^{n} (F_{ij}x \wedge B_i(C)x). \tag{20}$$

Since $p \wedge q$ is equivalent to $p \wedge (p \rightarrow q)$ we may derive:

$$Cx \leftrightarrow \bigwedge_{i=1}^{m} \bigvee_{j=1}^{n} (F_{ij}x \wedge (F_{ij}x \rightarrow B_i(C)x)). \tag{21}$$

Let Π now be the set of all functions $\pi : \{1,...,m\} \to \{1,...,n\}$, i.e., the set of all n^m variations of m-ary sequences from the index set $\{1,...,n\}$. Then the law of distributivity (out-multiplication of the disjunction) allows us to infer:

$$Cx \leftrightarrow \bigvee_{\pi \in \Pi} \bigwedge_{i=1}^{m} (F_{i\pi(i)}x \wedge (F_{i\pi(i)}x \to B_i(C)x)). \tag{22}$$

Existential generalization on both sides of the bi-conditional yields:

$$\exists x Cx \leftrightarrow \exists x \bigvee_{\pi \in \Pi} \bigwedge_{i=1}^{m} (F_{i\pi(i)}x \wedge (F_{i\pi(i)}x \to B_i(C)x)). \tag{23}$$

The first order sentence (23) can be semantically evaluated in Gödel's (1932) many-valued intuitionist min-max system that we have already chosen to interpret the first order sentences (7) to (10) of the main text. In the Gödel system the disjunction is evaluated as the maximum of the truth-degrees of the disjuncts. The conjunction comes to the minimum of the truth-degrees of the conjuncts. The existential quantifier, finally, is evaluated as the supremum over the domain (Gottwald, 2001):

$$d(\exists x Cx) = \sup_{x \in D} d(Cx). \tag{24}$$

The semantic interpretation of sentence (23) thus is:

$$d(\exists x Cx) = \sup_{x \in D} \max_{\pi \in \Pi} \min_{i=1}^{m} \min\{d(F_{i\pi(i)}x), d(F_{i\pi(i)}x \to B_i(C)x)\}. \tag{25}$$

According to the main text, a substance concept expressed by the predicate C is rendered by a matrix of typicality values C_{ij}. Those values tell how typical the j-th attribute value is for instances of C with respect to the i-th attribute dimension. It is possible to conceive of typicality as the truth degree of some fuzzy-valued implication, e.g.: Red is typical for tomatoes to the degree it is true that red things have the color of tomatoes. We will henceforth define typicality in precisely this manner:[5]

$$C_{ij} = \begin{cases} d(F_{ij}x \to B_i(C)x) & \text{if } d(F_{ij}x \to B_i(C)x) > 0 \\ 0 & \text{else.} \end{cases} \tag{26}$$

We assume that the truth-degree of the conditional is independent of the choice of x. Inserting the typicality value in (25) we arrive at:[6]

$$d(\exists x Cx) = \sup_{x \in D} \max_{\pi \in \Pi} \min_{i=1}^{m} \min\{d(F_{i\pi(i)}x), C_{i\pi(i)}\}. \tag{27}$$

[5] We cut the range of values at 0 to avoid negative typicality values. Other typicality measures are discussed, e.g., by Smith et al. (1988).

[6] Cutting off potential negative truth degress for $F_{ij}x \to B_i(C)x$ does not imperil the equality because the least possible value for any predicative sentence $F_{i\pi(i)}x$ in the semantics of the main paper is 0, itself.

Since the supremum of a set is always greater or equal to the maximum of the set and the minimum of two numbers in the interval $[0,1]$ is always greater or equal to their product, we can finally derive the inequality of the main text:

$$d(\exists x C x) \geq \max_{\substack{\pi \in \Pi \\ \mathrm{x} \in D}} \min_{i=1}^{m} C_{i\pi(i)} d(F_{i\pi(i)} x). \tag{28}$$

Remark: Instead of the disjunction (16), one may prefer to capture the partioning condition of assumption 3 in a many valued system semantically by a sum:

$$\sum_{j=1}^{n(i)} d(F_{ij} x) = 1. \tag{29}$$

Provided that the truth-degrees are transposed to range from 0 to 1,

$$\sum_{j=1}^{n(i)} d(F_{ij} x) \geq 1 \tag{30}$$

may be taken to express the exhaustivity of a class of fuzzy sets, whereas

$$\sum_{j=1}^{n(i)} d(F_{ij} x) \leq 1 \tag{31}$$

may replace the classical exclusivity condition.

With equation (29), the proof must be slightly altered, but leads to the same result. Instead of sentence (19), one then semantically has:

$$d(Cx) = \min_{i=1}^{m} \min\{(\sum_{j=1}^{n} d(F_{ij} x)), d(B_i(C)x)\}. \tag{32}$$

Now the sum of some non-negative real numbers is always greater or equal to their maximum. We thus may derive the inequality:

$$d(Cx) \geq \min_{i=1}^{m} \min\{(\max_{j=1}^{n} d(F_{ij} x)), d(B_i(C)x)\}. \tag{33}$$

And distributivity gives us:

$$d(Cx) \geq \min_{i=1}^{m} \max_{j=1}^{n} \min\{d(F_{ij} x), d(B_i(C)x)\}. \tag{34}$$

The right side of this inequality, however, is just the semantic value of the right side of sentence (20). Due to the parallelism of the lattice properties of the min-max-algebra and the $\wedge$-$\vee$-algebra, one may from here on proceed as shown above, being aware that one has turned from equation to inequality a couple of steps earlier.

References

Atmanspacher, H., & beim Graben, P. (2005). Contextual emergence of mental states from neurodynamics. *Chaos and Complexity Letters*. (Forthcoming)

Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In A. Lehrer & E. F. Kittay (Eds.), *Frames, fields, and contrasts* (pp. 21–74.). Hillsdale, NJ: Erlbaum.

Eckhorn, R., Bruns, A., Saam, M., Gail, A., Gabriel, A., & Brinksmeyer, H. J. (2001). Flexible cortical gamma-band correlations suggest neural principles of visual processing. *Visual Cognition*, *8*(3–5), 519–30.

Engel, A. K., König, P., Gray, C., & Singer, W. (1990). Stimulus-dependent neuronal oscillations in cat visual cortex: inter-columnar interaction as determined by cross-correlation analysis. *European Journal of Neuroscience*, *2*, 588–606.

Engel, A. K., König, P., Kreiter, A. K., & Singer, W. (1991). Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex. *Science*, *252*, 1177–9.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, *5*, 16–25.

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

Fodor, J. (1997). Connectionism and the problem of systematicity (continued): Why Smolensky's solution still doesn't work. *Cognition*, *62*, 109–119.

Fodor, J., & Lepore, E. (1992). *Holism: A shopper's guide*. Oxford: Blackwell.

Fodor, J., & Lepore, E. (1996, Feb). The red herring and the pet fish: why concepts still can't be prototypes. *Cognition*, *58*(2), 253–70.

Fodor, J., & McLaughlin, B. (1990). Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. *Cognition*, *35*, 183–204.

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.

Gross, C. G. (2002, Oct). Genealogy of the 'grandmother cell'. *Neuroscientist*, *8*(5), 512–8.

Hale, K., & Keyser, S. (1993). On argument structure and the lexical expression of syntactic relations. In K. Hale & S. Keyser (Eds.), *The view from building 20* (pp. 53–109). Cambridge, MA: MIT Press.

Hebb, D. O. (1949). *The organization of behaviour*. London: Wiley.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195*, 215–43.

Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition*, *57*, 129–91.

König, P., Engel, A., & Singer, W. (1995). Relation between oscillatory activity and long-range synchronization in cat visual cortex. *Proceedings of the National Academy of Science (USA)*, *92*(1), 290–4.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621.

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.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision.* McGraw-Hill.

Osherson, D., & Smith, E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, *9*, 35–58.

Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single-neurons in the human brain. *Nature*, *435*, 1102–7.

Robbins, P. (2002). How to blunt the sword of compositionality. *Noûs*, *36*, 313–34.

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

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

Smith, R., Osherson, D., Rips, L., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science*, *12*, 485–527.

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

Smolensky, P. (1995b). Reply: Constituent structure and explanation in an integrated connectionist/symbolic cognitive architecture. In C. Macdonald & G. Macdonald (Eds.), *Connectionism* (pp. 223–90). Cambridge, MA: Blackwell.

Smolensky, P., & Legendre, G. (2005a). *The harmonic mind: From neural computation to optimality-theoretic grammar* (Vol. 1: Cognitive Architecture). Cambridge, MA: MIT Press.

Smolensky, P., & Legendre, G. (2005b). *The harmonic mind: From neural computation to optimality-theoretic grammar* (Vol. 2: Linguistic and Philosophical Implications). Cambridge, MA: MIT Press.

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

Wennekers, T., & Palm, G. (2000). Cell assemblies, associative memory and temporal structure in brain signals. In R. Miller (Ed.), *Time and the brain* (pp. 251–73). Brighton: Harwood Academic Publishers.

Werning, M. (2003a). 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. (2003b). Ventral vs. dorsal pathway: the source of the semantic object/event and the syntactic noun/verb distinction. *Behavioral and Brain Sciences*, *26*(3), 299–300.

Werning, M. (2005a). Neuronal synchronization, covariation, and compositional representation. In E. Machery, M. Werning, & G. Schurz (Eds.), *The compositionality of meaning and content* (Vols. II: Applications to Linguistics, Philosophy and Neuroscience, pp. 283–312). Frankfurt: Ontos Verlag.

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