B2 Neuroframes: A Neuro-Cognitive Model of Situated Conceptualization

1 General information

1.1 Applicant

SCHURZ, Gerhard, Univ.-Prof. Dr.

1.2 Topic

Neuroframes: A Neuro-Cognitive Model of Situated Conceptualization

1.3 Scientific discipline and field of work

Philosophy, Cognitive Science, Neuro-semantics

1.4 Scheduled total duration

Six years

1.5 Application period

Three years

1.6 Summary

Frames, regarded as recursive attribute-value structures, allow for the semantic decomposition of nominal concepts and thus the categorization of objects. The project strives for an empirically plausible and philosophically motivated realization of frames on the neural level. Incorporating neurobiological data, primarily on visual object perception, mathematical methods of universal algebra, logic, model theory and topology as well as computational simulations in oscillatory networks are used to identify parts of the cortical reality with frame-based concept structures. Value spaces for attributes can be neurally traced to topologically structured “feature maps”. The simultaneous assignment of values from different attributes to the same object corresponds to the synchronization of neural activity. On this basis the neural correlates of frames with their attributes, values and object instantiations seem to be given. The project will increasingly also incorporate neurobiological findings on the premotor and motor cortex and thereby extend the frame-based approach to action concepts expressible by verbs. This extended notion of a neuroframe provides the foundation for a philosophically justified and empirically motivated theory that maintains a representationalist view on concepts and at the same time perceives them as situated, i.e., largely based on sensori-motor schemata. Mental concepts, here, fulfil a twofold function: they constitute the meanings of linguistic expressions – words, phrases and sentences – and they provide the contents of intentional states. The project conjoins a number of philosophical issues – the problem of lexical decomposition, modu-

larity, semantic compositionality, and intentionality – with linguistic topics: that is, verb semantics and its relation to affordance aspects in the meanings of nouns. It will attempt to develop a biologically and philosophically plausible model for the realization of situated concepts in the brain.

2 State of the art, preliminary work

2.1 State of the art

According to widely held views on the architecture of the mind (Fodor, 1998; Prinz, 2004; Murphy 2002), the primary role for concepts is the integration of perception and action control. In order to survive in a world with a multitude of things, subjects must subsume them under concepts. Categorization allows the subject to recognize objects and events in the world as well as states of the body, to generate generalizations, and to preserve this information over time. Only thus goal directed interaction between one's body and the world is possible to the degree observed in many species.

With regard to non-human animals, it is often held that all higher primates possess concepts (Call, 2006). It has, moreover, been argued that other mammals and even other vertebrates like birds can be ascribed concepts (Stephan, 1999, for review). With regard to humans, concepts are assigned a twofold explanatory role: (i) as content providers and (ii) as meaning providers. In their first role concepts provide contents to intentional states. In their second role concepts are identified with the meanings of linguistic expressions. Concept are apt to fulfill the two roles because they are individuated as internal states of the system that essentially bear a causal-informational relation of co-variation to external contents (Fodor, 1994; Werning, 2002). In this vein, concepts may explain why intentional states are *about* things and why the meanings of expressions in a given context determine the things being referred to.

Intentional states include such diverse modes as perception, belief, desire, memory, expectation, imagination, emotion, and volition. Concepts provide the satisfaction conditions of intentional states, enter into inferential relations, and play a role in the causation of action. The twofold role of concepts suggests a view that intimately links meaning to intentionality. A unified approach of meaning and intentional content holds that the meaning of the sentence *water is in the bath tub*, the perception of water being colorless, the belief that water boils at 100°C, and the desire to drink a glass of water have one thing in common: they involve the concept [water]. This follows if one assumes (i) the compositionality of linguistic meaning or internal compositionality, and (ii) the compositionality of intentional content or external compositionality (For the formal notion of compositionality see the forthcoming papers by Hodges, Fernando, and Sandu scheduled to appear in *The Oxford Handbook of Compositionality*, Werning, et al., Eds., 2009). The compositionality of meaning is the principle that the meaning of a complex expression is in a syntax-dependent way determined by the meanings of its parts. It explains how the concept [water] contributes to the meaning of the sentence *water is in the bath tub* (For compositionality in language see the forthcoming papers by Löbner, Jacobson, Pelletier, Recanati, Pietroski, and Higginbotham to appear in Werning, et al., 2009, Eds.). The compositionality of content says that the content of a complex intentional state is determined by the contents of its parts in a structure-dependent way. It explains how the content of the concept [water] determines the contents of the perceptual, doxastic, and volitional states mentioned above (a discussion of the reasons for compositionality is given in Werning, 2005e; on compositionality in general see also the forthcoming papers by Janssen, Westerstahl, Szabo, and Zimmermann as well as by Werning, Hinzen, & Machery to appear in Werning, et al., 2009, Eds.).

Since frames were first introduced by Minsky (1975), they have become a central paradigm in cognitive science and cognitive linguistics (Barsalou 1992) and have been used to account for the structure of concepts in a systematic way. Unlike older approaches, which are based on exemplars or prototypes (Rosch et al. 1976), frame theory relates the capacity of categorization to the possession of certain (potentially not yet lexicalized) functional concepts. According to frame theory, categories (e.g., [banana]) are generated within a system of attributes (captured by functional concepts like [color], [size], etc.) by assigning more or less sharp values or value intervals ([red], [green], etc.) to each relevant attribute as typical. A significant advantage of frame theory over older approaches is that values can themselves be attributes making recursive embedding possible. As will be shown in the project, there also are fewer problems with compositionality.

In modelling the psychological data on categorization two main accounts have been dominant over the last decades: prototype theory (in a narrow sense) and exemplar theory. The former (e.g., Mervis & Rosch, 1981) assumes that humans generate the representation of a prototype from the totality of experienced instances of the category in question as their mean (on a metric scale) or median (on an ordered scale). Category membership then is gradually determined by similarity to the prototype. In contrast, exemplar theory (Medin, 1975, see also Medin & Schaffer, 1978) and its generalization by Nosofsky (1984, 1987) assumes that humans store few specific exemplars in memory during the training phase and determine category membership by comparison to those exemplars. The controversy between both accounts has remained empirically undecided (e.g. Smith & Minda, 2000) to date and has led to a number of hybrid (Palmeri & Nosofsky, 1995, for review) or radically sceptical (Machery, 2005) conclusions. Osherson and Smith (1981), as well as Fodor and Lepore (1996), have argued against both approaches on the basis of compositionality arguments. Kamp and Partee (1995) have responded with a super-valuation theory. The problem of compositionality for prototype and exemplar theory is still virulent (Robbins, 2002, for review; see also the forthcoming papers by Hampton & Jonsson, Gleitman & Connolly, Wisnewski, Prinz, Machery and Schurz to appear in *The Oxford Handbook of Compositionality*, Werning, et al., Eds., 2009).

The psychological models have been supplemented by a number of connectionist network proposals for learning categories: Gluck's and Bower's (1988) adaptive network, which builds on networks of Hinton and Anderson (1981) as well as Rummelhart and McClelland (1986), uses a least-mean-square learning rule (Rescorla & Wagner, 1972; Kohonen, 1977). Yet again, compositionality arguments have been raised against connectionist models of that kind (Fodor & Pylyshyn, 1988; Fodor & McLaughlin, 1990; Fodor, 1997). The central problem is the composition of complex concepts from more primitive ones such that co-variation with external content is warranted. Smolensky's (1988, 1995a, 1995b, 2005) attempts to accommodate cognitive-conceptual structures in connectionist networks remain insufficient in many respects, as our own research has been able to show (Werning, 2003b, 2005a; see also the forthcoming papers by Horgan, Eliasmith and Werning in Werning, et al., Eds., 2009). Of special importance in this context is the so-called binding problem (Treismann, 1996): i.e., how can distributed property representations be bound together to form the representation of an object as having these properties?

Von der Malsburg (1981) theoretically postulated the synchronization of the electrical discharges of neurons, which may be distributed over one of the many “feature maps” or “feature zoos” of the cortex (Hubel & Wiesel, 1968), as a mechanism for object binding. This view has now been corroborated by numerous experimental findings (Eckhorn et al., 1988; Gray et al., 1989; Engel et al., 1997, 2001). Neuronal synchronization not only seems to correlate with perceptual and behavioral processes in awake animals (Fries et al., 1997, 2002) and humans (Tallon-Baudry et al., 1999; Rodriguez et al., 1999), but also in perceptual domains other than vision (e.g., Laurent, 1996). Its explanatory force extends to action (Schnitzler et al., 2006) and language comprehension (Weiss et al. 2000, 2005, see also the forthcoming papers by Engel & Maye and Lambalgen, Baggio, & Hagoort to appear in Werning, et al., Eds., 2009).

The discovery of the so-called mirror neuron system may provide a basis to link action control, perception, categorization, and eventually language understanding with one another (Rizzolatti. & Craighero, 2004, for review). Based on a review of neurobiological data, Pulvermüller (1999) suggests that neural assemblies pertaining to the sensori-motor cortices and bound by neural synchronization play an important role in understanding the meanings of words. FMRI studies (Pulvermüller, 2005) regarding the understanding of verbs, e.g., hint at a differential top-down activation of motor and pre-motors areas. These pieces of data fit nicely with the theoretical considerations in the context of situated cognition (Clancey, 1997) and are related to the emulation theory of representation (Grush, 2004).

2.2 Preliminary work

In our context, the most important distinction in the domain of concepts is that between attributive concepts and substance concepts. Attributive concepts represent features of objects that are volatile in the sense that one and the same object can fall under different attributive concepts at different times: an object may, e.g., change its color, size, or speed, yet still continues to exist. [Blue] thus is a paradigmatic attributive concept. Attributive concepts correspond to the values of attributes in frames.

Substance concepts, in contrast, are governed by the identity conditions of objects: a mug ceases to exist when it no longer falls under the substance concept [mug], say, because it has been shattered. Substance concepts serve to re-identify things over time in spite of their contingent changes of attributes and so allow us to gather, store and update information in a systematic and enduring way (Millikan, 1998). They are typically expressed by concrete nouns – in English by names of individuals like *mama*, names of kinds like *mouse* and names of stuffs like *milk*. Attributive concepts, in contrast, are typically expressed in English by adjectives or abstract nouns: *blue*(*-ness*), *warm*(*-th*), *lucid*(*-ity*).

The perspective to be developed in the project largely draws on the theory of neuro-frames (Werning & Maye, 2007). The theory of neuro-frames holds that (i) substance concepts are decomposable into less complex concepts, that (ii) the decompositional structure of a substance concept can be rendered by a recursive attribute-value structure, that (iii) the neural realization of a substance concept is distributed over assemblies of neurons and meta-assemblies thereof, that (iv) those neurons pertain to neural maps for various attributes in many afferent and efferent regions of the cortex, and that (v) an appropriate mechanism of binding together the distributed information together into the neural realization of the substance concept is the mechanism of neural synchronization.

Frame theory provides us with a universal account not only for categorization and its link to action-control, but also for the decomposition of concepts. Frames are recursive attribute-value structures. Attributes assign unique values to objects and thus describe functional relations. The values can be structured frames themselves. A frame is defined for a large domain of things and contains a fixed set of attributes, each of which allows for a number of different values. The attributes in question are not constrained to perceptual modalities, but may involve attributes of motor affordances as well. Frames can be nested hierarchically and mutual constraints between attributes (e.g., between states of an object and actions directed to it) and between larger frames can be incorporated (see Figure 1).

For many attributes involved in perceptual processing one can anatomically identify cortical correlates. Those areas often exhibit a twofold topological structure and justify the notion of a feature map: (i) a receptor topology (e.g., retinotopy in vision, somatotopy in touch): neighboring regions of neurons code for neighboring regions of the receptor; and (ii) a feature topology: neighboring regions of neurons code for similar features. With respect to the monkey, more than 30 cortical areas forming feature maps are experimentally known for vision alone (Felleman & van Essen, 1991).

Attributes of motor affordances may also be parts of frames and appear to have cortical correlates, predominantly in the premotor cortex (Werning, 2008). An affordance is a dispositional property of an object that allows a subject to perform an action upon the object in a way specific for the object (Gibson, 1977). Typical examples of objects with affordances are artifacts (e.g., tools: a scissors have the affordance for cutting; furniture: a chair has the affordance for sitting; etc.), but also fruits (a banana has the affordance for peeling) and plants (a tree may have the affordance for somebody to sit under it). The cortical organization of motor control with regard to the effectors follows similar topological principles as the cortical organization in perception with regard to the receptors. The discovery of the so-called mirror neuron system (Rizzolatti & Craighero, 2004, for review) may provide a basis to integrate affordances into frames. Figure 2 shows a number of neural maps that relate to attributes of frames.

The fact that values of different attributes may be instantiated by the same object, but are processed in distinct regions of cortex is a version of the binding problem: how is this information integrated in an object-specific way? How can the color and taste of a banana be represented in distinct regions of cortex, although they are part of the representation of one and the same object?

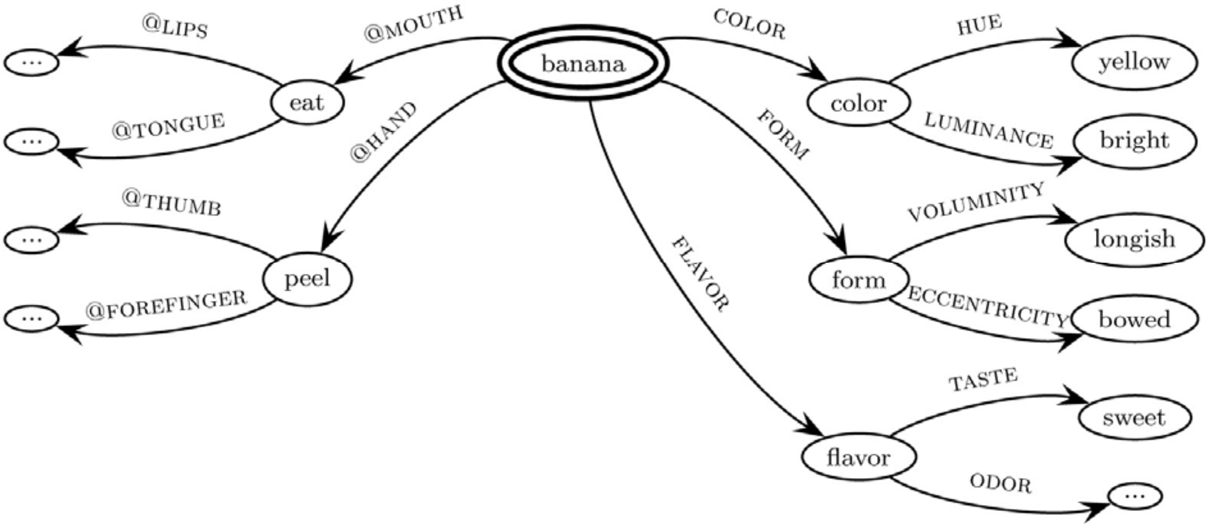

Figure 1. Hypothetical fragment of the frame for the concept [banana]. The substance concept to be decomposed is marked by a double-circle as the referring node of the frame. The labeled arrows denote attributes, the nodes their values. Nodes are themselves regarded as concepts and thus as conceptual parts of the central concept. In English, feature attributes (shown on the right) are frequently lexicalized – their arguments typically enter possessive constructions like *The color of the banana is yellow* or *The banana has the color yellow*. Based on linguistic and neurobiological evidence, we assume that affordances often relate to body parts and hence use the convention "@ + body part". Formally, attributes are mappings from domains of some type into domains of some other type. Petersen and Werning (2007) provide an explicit account of frames using a calculus of typed feature hierarchies and incorporating typicality effects.

A prominent and experimentally well supported solution postulates oscillatory neural synchronization as a mechanism of binding: Clusters of neurons that are indicative of different properties sometimes show synchronous oscillatory activity, but only when the properties indicated are instantiated by the same object in the perceptual field; otherwise they are firing asynchronously. Synchronous oscillation, thus, might be regarded as fulfilling the task of binding various property representations together to form the representation of an object having these properties (Singer, 1999). Using oscillatory networks as biologically motivated models, it could be demonstrated how the topological organization of information in the cortex by mechanisms of synchronization may yield a logically structured semantics of concepts (Maye & Werning, 2004, see Figure 3). Compositionality theorems have been provided (Werning, 2005a,d). Oscillation functions play the role of object concepts. Clusters of feature sensitive neurons play the role of attributive concepts (for illustration see Figure 4). The experimental findings by Schnitzler et al. (2006, Schnitzler is an external cooperation partner at the University of Düsseldorf) on the essential role of neural synchronization for action control may justify the extension of the synchrony-based neuro-frame approach from features to affordances.

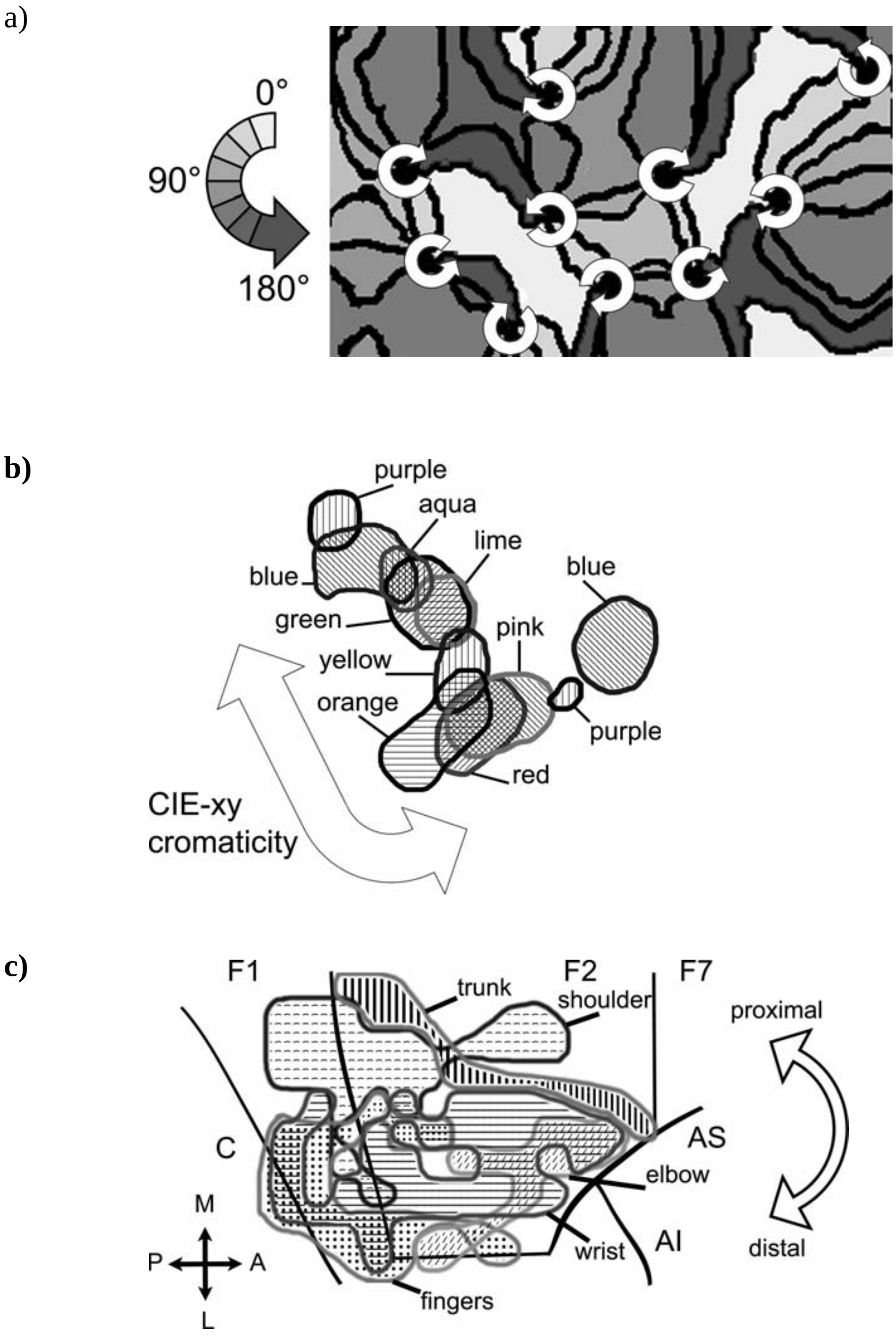

Figure 2. Cortical realizations of frame attributes. a) Fragment (ca. 4mm²) of the neural feature map for the attribute orientation of cat V1 (adapted from Crair et. al., 1997). The arrows indicate the polar topology of the orientation values represented within each hypercolumn. Hypercolumns are arranged in a retinotopic topology. b) Color band (ca. 1 mm²) from the thin stripes of macaque V2 (adapted from Xiao et. al., 2003). The values of the attribute color are arranged in a topology that follows the similarity of hue as defined by the Commission Internationale de l'Eclairages (xy-cromaticity). The topology among the various color bands of V2 is retinotopic. c) Neural map (ca. 250 mm²) of forelimb movement in macaque primary motor (F1) and dorsal premotor cortex (F2, F7) (adapted from Raos et al., 2003). The overarching topology is somatotopic from proximal to distal movement as shown by the arrow. Due to the size of the region one may expect it to comprise maps for more specific motor attributes. C: central sulcus, AS and AI: superior, respectively, inferior arcuate sulcus.

Support for the theory of neuro-frames also comes from a number of neuro-linguistic studies. Based on a review of neurobiological data, Pulvermüller (1999) suggests that neural assemblies that pertain to the sensori-motor cortices and are bound by neural synchronization play an important role in understanding the meanings of words. FMRI studies (Pulvermüller, 2005) regarding the understanding of verbs, e.g., hint at a differential top-down activation of motor and pre-motors areas. We know that the understanding of concrete nouns like *hammer,* for which not only features, but also affordances are salient, results in an activity distributed over the premotor and the visual cortex (Martin et. al. 1996, 2007). The hypothesis that words for substance concepts arouse more widely distributed activity than words for attributive concepts has also been supported by EEG studies (Rappelsberger et al., 2000).

The neuro-semantic approach used here is based on simulatory studies on perceptual oscillatory networks by Markus Werning and his co-operators (Werning, 2001a,b; Maye & Werning, 2007). Using an eigenmode analysis, it could be shown that the mechanism of oscillatory neural synchronization may subserve the realization of the semantics of a monadic first order predicative language (Werning, 2005a). The virtues of classicist and connectionist approaches could be attained, while the shortcomings of both paradigms were avoided (Werning, 2009; Maye, Werning, König, & Engel, 2005). As documented in Werning's doctoral dissertation (Werning 2004c) and in additional publications (Werning. 2005a,d), our theory has been related to the principle of semantic compositionality. The reasons for compositionality (Werning, 2005e) as well as the relation of compositionality to the context principle and the existence of categories (Werning, 2004a) have been discussed. Ambiguous und illusionary object perceptions like the Kanizsa-illusion (Werning & Maye, 2006) could be integrated into the approach as well as spatial (Werning, 2003b) and part-whole (Werning & Maye, 2005) relations and events (Werning, 2003a). The problem of lexical decomposition has also been discussed (Petersen & Werning, 2007). From a philosophical point of view, the topic of conscious perception (Werning, 2000) and the debate between pictorial and conceptual representations has been dealt with (Werning, 2005c). Issues like self-awareness (Werning, 2004b,c), the representation of higher-order intentionality (Abraham, Werning, Rakoczy, von Cramon, & Schubotz, 2008), metaphorization (Werning, Fleischhauer, & Beseoglu, 2006) as well as evolutionary and developmental issues (Werning, 2005b, 2008) have been approached as well. First attempts to integrate recursive embedding into neuro-frames by means of coherency chains and hierarchical binding were successful (Werning & Maye, 2004, 2005).

Gerhard Schurz has worked on normic conditionals and non-monotonic reasoning from the point of view of logic and semantics, philosophy of science, and evolutionary theory. Building on Schurz (1998, 2001, 2005a), Schurz (to appear 2009) has developed an evolutionist analysis and justification for prototype semantics. The advantages of prototype semantics lie not so much in a coherent classificatory framework, but in the highly efficient prognostic and diagnostic categorization of objects in our natural environment. The problem as to how the alleged non-compositionality of prototype combinations can be dealt with efficiently has been addressed (to be continued in co-operation with project B1). Schurz (2005b) has also shown that a compositional semantics for theoretical concepts and classification systems exists, although their content strongly depends on the background theory. Moreover, Gerhard Schurz is assigned a number of coordinating tasks in the SFB.

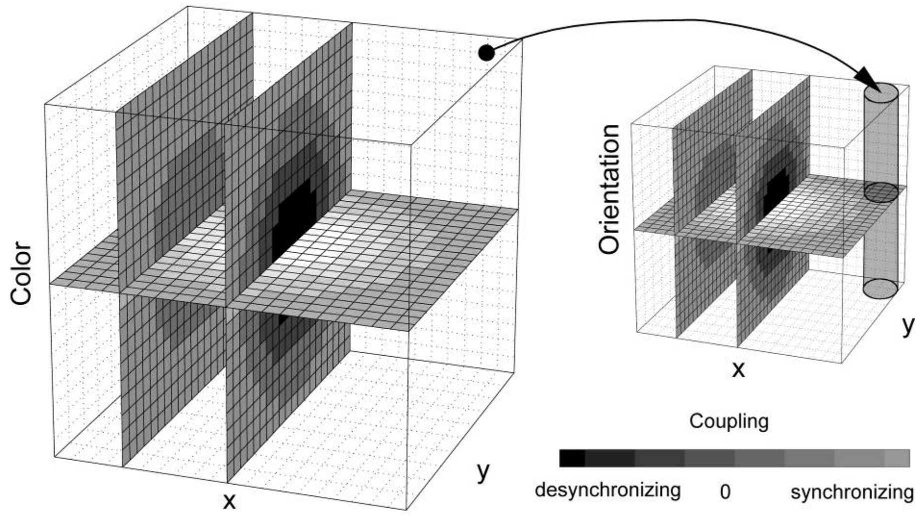

*Figure 3***.** Oscillatory network. The network topology reflects the receptor topology (xy-plane) and the feature topology (z-axis) of the neural maps. Each module realizes one attribute. The layers in each module realize the attribute values. Oscillators activated by neighboring stimulus elements with similar attribute values synchronize (light gray). Oscillators activated by neighboring stimulus elements with unlike attribute values de-synchronize (dark gray). The layers of different modules are connected in a synchronizing way that respects the common receptor topology. (From Maye & Werning, 2007).

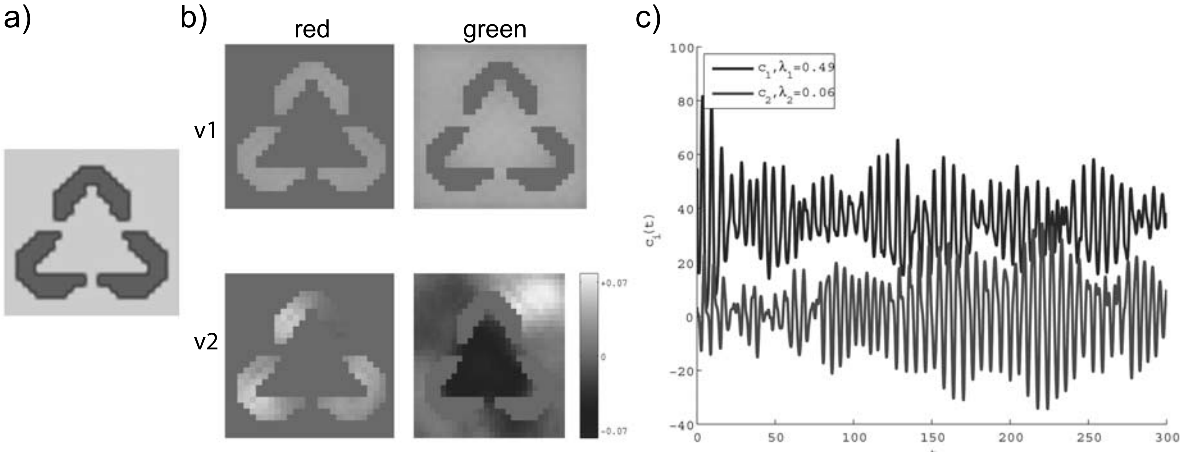

Figure 4. An oscillator network with a single module for color with layers for red and green is stimulated with the Kanizsa illusion. a) Stimulus: three red circle segments on a green ground. b) The two strongest eigenmodes of the network dynamics $\mathbf{v}_1$ and $\mathbf{v}_2$**,** each subdivided according to layers, are shown. The signs of the vector components are indicated by shades of gray: light gray: positive, middle gray: zero, dark gray: negative. c) Temporal evolution of the two eigenmodes are given by the characteristic oscillatory functions $c_1(t)$ and $c_2(t)$. The eigenvalues $\lambda_{1,2}$ yield the relative contribution of each eigenmode to the overall variability of the network dynamics. Semantic interpretation: The first eigenmode does not render figure ground segregation. The second eigenmode, however, renders a representation of the illusionary triangle (object concept: $-c_2$) as distinct from the background (mostly zero) and the united circle segments (object concept: $+c_2$).

3 Goals and work schedule

3.1 Goals

The primary goal of project B2 is to develop a biologically plausible and philosophically justified model for the cortical realization of frames and the functional concepts they contain so that frames may justly be regarded as a plausible model for conceptual decomposition from a neurobiologically informed point of view as well. If it can be shown how affordances can be integrated into frames, this in a natural way leads to a theory of situated conceptualization according to which conceptual representation is cortically grounded in sensori-motor schemata. The following hypotheses will be developed and evaluated in the project:

a) A general theory of neuroframes will be developed that covers not only substance concepts that decompose into flat frames, but also those that decompose into hierarchically embedded frames. A topic of increasing importance will be the role of action representations both in the decompositional analysis of substance concepts (e.g., concepts of tools) and that of event concepts. Provided that a concept is completely decomposable into a fully specified frame and provided that neural maps for each attribute can be identified in the cortex, the degree to which the cortex represents an object satisfying the concept is rendered by a general pattern of synchronizing neural activity distributed across the neural maps and over clusters of neurons that correspond to the basic values of the frame. This pattern may be called a concept's cortical fingerprint. To incorporate frames of unrestrained complexity, a mechanisms of hierarchical binding is postulated. To this end, the oscillatory networks, that have already been developed, will be expanded. Neuroframes are internally and externally compositional and can be used to model inferences in a cognitively realistic manner.

b) Conceptual representations are situated in sensori-motor schemata. The same concept tokens provide meanings to linguistic expressions and contents to intentional states. A strong modularity claim with respects to semantic processes is thus denied. However, a claim of moderate modularity might still be tenable. The issues of moderate modularity will be linked to the issues of hierarchical binding. Many substance concepts have affordance representations as conceptual parts. Neuroframes contribute to an explanation of processes of nominalizations and grammaticalizations for affordances. Linguistic, cognitive and neuronal evidence for affordance representations and their integration into frames will be exploited. Neuroframes also contribute to an explanation of non-compositional constructions in metaphors and metonymies.

3.2 Methods and work schedule

Ad a) The general theory of neuroframes (years 1-2)

One of the main virtue of frames is that they allow for the decomposition of substance concepts by means of attributes. This decomposition enables us to explain how a subject may subsume a perceived or otherwise given object under a substance concept. The framework of typed feature hierarchies (Petersen & Werning, 2007) has been developed in close cooperation with project B1 and allows us to translate graphical representations of frames like that of

Figure 1 into statement about types. The degree – between 0 and 1 – to which an object of the universe U instantiates a certain type of the set Type is given by the function:

$$d : \text{Type} \times U \rightarrow [0,\ 1].$$

In every frame the root node corresponds to the decomposed concept (‘banana’). The set of maximal paths MaxPath is well-defined for every frame. In a fully specified frame, end nodes, e.g, ‘yellow’ are atomic minimal types and are identified by maximal paths, i.e., [color:hue:] beginning at the root node. It is natural to assume that the cognitive subject is endowed with a detector system that for all atomic minimal types renders the degree d to which it is instantiated by a given object. These might be hue detectors, eccentricity detectors etc.

It is important to notice that many attributes shift the referential object when applied to an object. One might say that the eccentricity of the form of a banana is still a property of the banana and an eccentricity detector may well be directed to the banana in order to assign a value. However, the eccentricity of the form of the stalk of a banana is not a property of the banana, and detecting its value requires a potential eccentricity detector to be directed to the stalk. It is hence useful to introduce a reference-shifting function

$$\sigma : U \times \Pi \rightarrow U$$

that maps every object of the universe relative to the path in question (Π: the set of all paths) onto the same or another object of the universe. In the classical bi-valued case, the values of d are restricted to 0 and 1. For a fully specified frame we can conclude that an object x is to be subsumed under the decomposed concept C if and only if all the types of the end nodes are properly instantiated:

$$d(C,x) = \min_{m \in \text{MaxPath}} d(\Theta(m),\ \sigma(x,m)),$$

where $\Theta(m)$ is the type of the path m. A cognitively more realistic picture, however, is attained if we specify how typical a certain minimal type is for instances of the concept. Yellow may, e.g., be more typical than green as the hue of the color of bananas. Nevertheless the hue of the color of some bananas is still green. We can achieve this by considering alternative types for each maximal path. For each maximal path m we then have a set alt(m) containing the minimal type $\Theta(m)$ and all its alternative types with regard to the path. For each of the types t of alt(m) we can then specify a typicality value relative to the maximal path m of the fully specified frame for a decomposed concept C. The typicality value $\tau(C,m,\ t)$ tells how typical the type t is for the object $\sigma(x,m)$ given that x instantiates C. With these conventions we can apply previous results documented by Werning and Maye (2005, 2007) and, on the basis of the detector outputs, estimate to which degree an object x instantiates the decomposed concept C:

$$d(C,\ x) \geq \min_{m \in \text{MaxPath}} \max_{t \in \text{alt}(m)} \tau(C,m,\ t)\ d(t,\ \sigma(x,m)).$$

This approach will be further developed in close co-operation with project B1. Of particular interest is the role of the reference shifting function as postulated by the theory. This closely relates to the integration of part-whole structures into the theory. On the linguistic level reference shifting, e.g., occurs with nominal compounds in possessive constructions like *John’s hair length*. In this respect functional concepts of parts might eventually behave differently from functional concepts of family relations (*John’s mother length* seems ungrammatical, even though *hair* and *mother* both express functional concepts.). With regard to typicality issues the problem of compositionality will attract special attention.

Neuronal implementation

For many attributes (hue, brightness, orientation, direction, size, etc.) involved in the course of visual processing – we call them qualitative attributes – one can anatomically identify neuronal feature map. We may assume that such neurons function as detectors and thus evaluate atomic minimal types for a given stimulus object. This validates the hypothesis that there may be neural correlates of attributes and their subtypes in the cortex. Synchronous oscillation might be regarded as fulfilling the task of binding various property representations together in order to form the representation of an object as having these properties Using oscillatory networks (Schillen & König, 1994; Maye & Werning, 2004) as models, the structure of object-related neural synchronization could be interpreted (Werning, 2005a) as providing a conceptual structure expressible in a first-order predicate language. To show this, an eigenmode analysis of the network dynamics has been computed. Per eigenmode, oscillation functions play the role of object representations or concepts. Clusters of feature sensitive neurons play the role of property representations or predicate concepts. The following theorem (Werning & Maye, 2007) nicely links the results of frame theory given above to previous results on the neural implementation of conceptual structure. The degree to which the object x is represented as instantiating the atomic type t by a network eigenmode is given by the equation:

$$d(t, x) = \max\{\mathit{\Delta}(\alpha(\mathrm{x}), f_j) | \mathbf{f} = \beta(t)c\mathbf{v}\}.$$

Here $\alpha(\mathrm{x})$ is the oscillation function representing the object x, and $\beta(t)$ is a matrix identifying the neural clusters which function as detectors for the type t. $\mathbf{v}$ and c are the results of the eigenmode analysis and account for the spatial, respectively, temporal variation of the network activity in that eigenmode. $\mathit{\Delta}$ is defined as the normalized inner product of two square-integrable time-dependent functions in a given temporal interval and measures the degree of synchrony between an object-related oscillation and the actual oscillatory activity in a neural cluster. $d(t, x)$ approaches 1 if the oscillation function $\alpha(\mathrm{x})$, which represents the object x, is highly synchronous with some component oscillatory activity f_j of $\mathbf{f}$ – i.e., the vector containing the eigenmode-relative temporal evolution of the type-related cluster of detector neurons. If we conjoin the estimation of $d(C, x)$ in terms of type-specific detector outputs $d(t, x)$ with the identification of the latter with particular oscillatory network activity, we may conclude with the following hypothesis: Provided that a concept is completely decomposable into a fully specified frame with detectors for each type of a maximal path, the degree to which the cortex represents an object as an instance of the concept can be estimated by a general pattern of synchronizing neural activity distributed over various feature-selective neural clusters corresponding to the atomic types of the frame. This pattern may be called the cortical fingerprint of the concept.

In our representationalist and situated approach the principles of internal and external compositionality (Fodor & Lepore, 2002; Werning, Machery, & Schurz, Eds., 2005; Machery, Werning, & Schurz, Eds., 2005; Machery, Schurz, & Werning, 2005; Werning, Machery, & Hinzen, to appear 2009) as well as the co-variation of internal concepts with external contents are crucial. Only if internal and external compositionality are warranted can we explain how the semantics of languages can be learned in finite time, how communication about the external world is possible and, more generally, how finite systems are able to grasp a potentially infinite manifold of situations in a systematic way. The aim is to provide a proof of the general internal and external compositionality of frames. The proof may build on previous compositionality proofs regarding the implementation of simpler first order structures in oscillatory networks (Werning, 2005a,d). Neurobiological results of external co-operation partners (Andreas Engel, Alexander Maye, Alfons Schnitzler, Rüdiger Seitz, Ricarda Schubotz) will be used to modify and extend the model.

Hierarchical binding

Some complex concepts 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 might, 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, on the one hand, one 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, on the other hand, one were to refrain from binding the attribute values together by some synchronization mechanism, the information that those attribute values are features of the parts of one and the same object would not be preserved. To illustrate this dilemma, take an animal with long, vertically oriented legs and a small horizontally oriented rump, say a flamingo. You cannot simply represent the flamingo by synchronizing neurons for the features *long*, *vertical*, *small,* and *horizontal*. Neither can you assume that a binding mechanism between the features of the legs and those of the rump does not exist. For, how do you then represent the fact that legs and rump belong to the same object?

What is needed to avoid the dilemma is obviously a mechanism of hierarchical binding. 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 could 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.

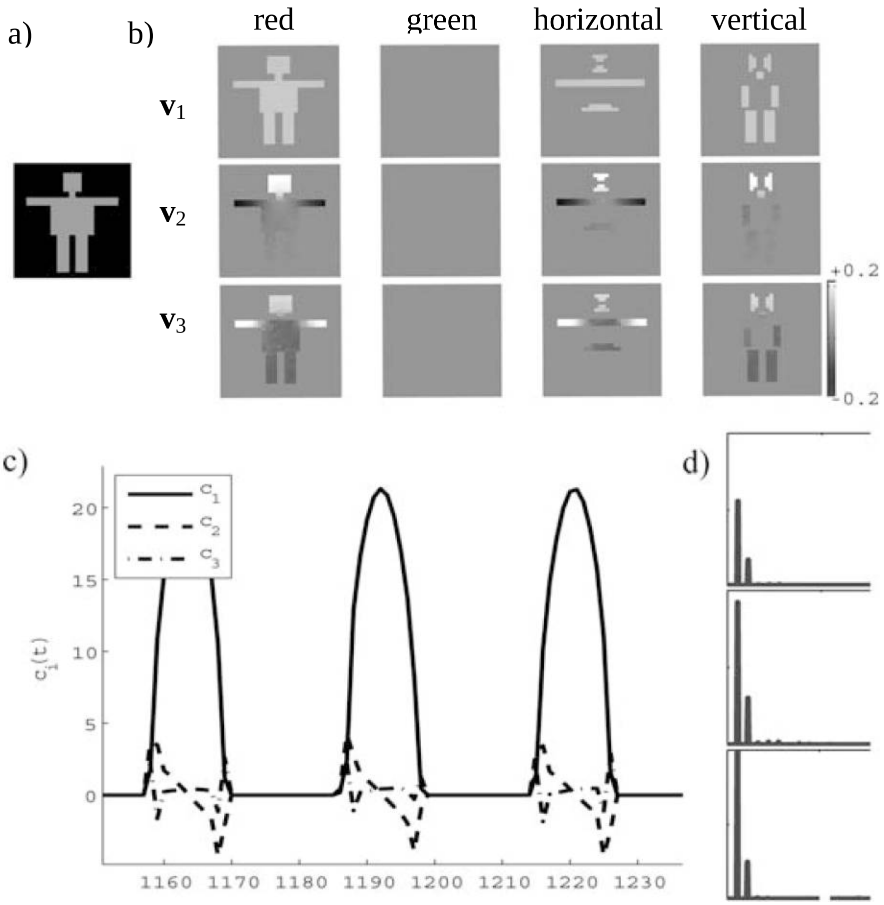

Figure 5. a) Stimulus: a red figure with horizontally spread arms. b) Eigenmodes of the stimulated oscillatory network with modules for the attributes color (values: *red*, *green*) and orientation (values: *horizontal*, *vertical*). The eigenvectors of the first three eigenmodes are shown. $\mathbf{v}_1$ 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.

Ad b) Situated conceptualization (years 2-3)

One of the main controversies regarding the processing and neuro-cognitive implementation of meaning is whether the semantics of language is processed in a modular or non-modular way. According to modular approaches, the meanings of words and sentences are processed in an informationally largely encapsulated, autonomous, and amodal way (Clifton & Ferreira, 1987). Candidates for cortical correlates of semantic processes are often supposed to be localized in left temporal and partially frontal regions (Friederici, 2002). Regions typically associated with either perceptual or motor processes in this paradigm are typically not regarded as contributing to semantics.

Modular approaches towards perception, in turn, argue for informationally encapsulated, domain-specific and cognitively impenetrable modules for various perceptual tasks (Fodor, 1983; Barrett & Kurzban, 2006, for review). Modularism with respect to semantics, perception, and perhaps other types of intentional states would thus be hardly compatible with the view that the same mental concept, respectively its neural correlate, is both a meaning provider for linguistic expressions and a content provider for various types of intentional states. A manifold of concept tokens with the content of water would be required: the concepts [water]-in-meaning, [water]-in-perception, [water]-in-desire, etc. – eventually even [water]-in-desires-to-drink, [water]-in-desires-to-swim, etc. It is easy to imagine that such a view would quickly lead to an ontological explosion of concepts, at least, if concepts are supposed to exist in a realist manner.

Much more compatible with a realist attitude towards concepts and the methodological goal of ontological parsimony is the anti-modularist view of situated conceptualization (Barsalou, 2005). Here concepts are regarded as situated, i.e., largely based on sensori-motor schemata (Arbib et al. 1987). This view, on the other extreme, also dissociates itself from radical proponents (Brooks, 1991) in the embodied cognition movement, who reject a representationalist model of the mind *tout court* and replace the explanatory role of mental concepts by some notion of simulation.

The controversy between semantic modularism and semantic anti-modularism relates to the question as to whether some lexical concepts – concepts listed in the lexicon and thus expressed by single words – decompose into conceptual parts. Some authors believe that lexical concepts are altogether not decomposable (Fodor & Lepore, 1992). According to those so-called atomist positions, only concepts that are linguistically expressible by syntactically explicitly combined expressions can be complex (see also the forthcoming papers by Wunderlich, Hinzen, and Harley to appear in Werning et al., Eds. 2009). In neuroscience some researchers maintain that substantial features like that of being an elephant or even features as specific as that of being Halle Berry are represented by highly specialized single neurons (Quian Quiroga et. al., 2005). Lexical atomism is a view semantic modularists can easily live with. For, if meanings are unstructured, it is completely unproblematic to conceive of them as localizable elements in an encapsulated module. Proponents of a situated view of meaning, in contrast, assume that at least some lexical meanings are structured such that parts of the meaning providing concepts may involve various sensori-motor schemata. Semantic anti-modularism seems to exclude lexical atomism.

In the project a philosophically grounded and empirically motivated theory will be developed that holds on to a representationalist view on concepts (Fodor, 1998; Werning, 2004c, 2005a) and at the same time views concepts as situated, i.e., largely based on sensori-motor schemata. This will take place in close cooperation with the project A4.

Prima facie conflicting evidence in favour of modularism are accounted for by a mechanism of hierarchical binding in our approach (Werning & Maye, 2004) that allows attribute values in more domain-specific subframes to be more tightly bound to each other than to other attribute values of the overall frame. Our anti-modularist, situated approach towards meaning is entirely compatible with the idea that, in the comprehension of sentences, syntax resolution as well as particular semantic subtasks such as the assignment of thematic roles may well be processed in a relatively modular, amodal way. We also consider the possibility that, at later stages in phylogeny and ontogeny, concepts may become more abstract and decouple from their sensori-motor basis. The decoupling mechanism may relate to processes of metonymy, metaphor, and analogy and will be investigated in the project.

Linguistic evidence for affordances

Many languages have developed lexical or grammatical means to express affordances. In Indo-European languages one often finds lexically explicit word-word compositions of the head noun with nouns or verbs that refer to affordances: English: *finger food, hand driller, football*; German: *Lesebuch* ‘read book’, *Trinkgefäß* ‘drink vessel’. To denote objects through their affordances, we here also have rather productive morphological means – English mixer; German: *Schläger* ‘racket’ (from *schlagen* ‘to hit’) – as well as fossilized forms like German *Griff* ‘handle’ (from *greifen* ‘to grasp’). In several languages affordances grammaticalize. In the Austronesian language Paamese, for a noun like *ani* ‘coconut’, the choice between the classifiers *ā*, *emo, ese, one*, which are obligatory in alienable possessor constructions and carry the possessive suffix (e.g., 3.Sg: *-n* ‘his/her’), indicates the affordance of the substance for the possessor: *ani ā-n/emo-n/ese-n/one-n* ‘his/her coconut with the affordance for him/her to eat (the flesh)/to drink (the milk)/to grow (on his/her land)/to use in any other way’ (Crowley, 1995). In the North-American language Dakota instrumental prefixes relating to body parts (*ya-* ‘with the mouth’, *yu-* ‘with the hand’, *na-* ‘with the foot or leg’) occur with verbs (*yu-ho'm.ni* ‘to turn with the hand (like a screw)’ vs. *nawa'hom.ni* ‘I turn it with the foot’), in denominalizations (*ya'ite* ‘to flatter’ from *'ite* ‘face’) and with locatives (*nao'hlat’e* ‘kicked underneath’ from *‘ohla't’e* ‘underneath’) (Boas & Deloria, 1939). In co-operation with the linguistic projects of the Forschergruppe (A1-A6) further evidences will be acquired.

Cooperative neuroscientific studies

In co-operation with external co-operation partners (Schnitzler and Seitz Neurology, Univ. Düsseldorf; Schubotz AG Prämotorik, MPI Köln), cognitive and neuronal evidences for affordance representations will be investigated. The long-standing cooperation with the Institute of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf (Engel and Maye) will be enhanced to integrate further evidence from EEG and MEG recordings on neural synchronization and hierarchical binding (e.g., in the context of Navon stimuli, in which small letters are parts of a bigger letter). With this group also new simulation methods will be approached.

A logic of frames

In a more long-term perspective a “logic for frames” will be developed in cooperation with project B1. The logic can in part build on the calculus provided by Werning (2005d). However, non-monotonic inferences and default values have to be integrated. This integration will be based on the study of non-monotonicity in Schurz (2005a, 2009). Processes relating to abstraction, metaphors and metonymies will also be investigated in cooperation with project A4.

Own publications

I. Refereed Publications in

a) journals

Abraham, A., Werning, M., Rakoczy, H., von Cramon, D. Y., & Schubotz, R. I. (2008). Minds, persons, and space: An fMRI investigation into the relational complexity of higher-order intentionality. *Consciousness and Cognition,* (In press).

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

Maye, A., & Werning, M. (2007). Neuronal synchronization: From dynamic feature binding to object representations. *Chaos and Complexity Letters, 2*(2/3), 315-325.

Schurz, G. (1998). Probabilistic semantics for Delgrande's conditional logic and a counterexample to his default logic. *Artificial Intelligence*, *102*(1), 81-95.

Schurz, G. (2001). What is 'normal'? An evolution-theoretic foundation of normic laws and their relation to statistical normality. *Philosophy of Science*, *28*, 476-497.

Schurz, G. (2005). Non-monotonic reasoning from an evolutionary viewpoint: Ontic, logical and cognitive foundations. *Synthese*, *146*(1–2), 37-51.

Werning, M. (2000). Synchronous oscillation and the transparency of consciousness. *Consciousness and Cognition*, *9*, 42-43.

Werning, M. (2001a). Eine Brücke gesucht, aber keine in Sicht: Warum sich Psychoanalyse und Konnektionismus nicht verbinden lassen. *Ethik und Sozialwissenschaften*, *12,* 590-592.

Werning, M. (2002). How to compose contents. *Psyche*, *8*(7).

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

Werning, M. (2004a). Compositionality, context, categories and the indeterminacy of translation. *Erkenntnis*, *60*, 145-178.

Werning, M. (2005a). The temporal dimension of thought: Cortical foundations of predicative representation. *Synthese*, *146*(1/2), 203-224.

Werning, M. (2005b). The 'complex first' paradox: What's needed to compose neurally distributed lexical meanings? *Behavioral and Brain Sciences*, *28* (Online suppl.).

Werning, M. (2008). The 'complex first' paradox: Why do semantically thick concepts so early lexicalize as nouns? *Interaction Studies, 9*(1). (In press)

Werning, M., & Maye, A. (2007). The cortical implementation of complex attribute and substance concepts: Synchrony, frames, and hierarchical binding. *Chaos and Complexity Letters, 2*(2/3), 435–52.

b) scientific congresses

Maye, A., Werning, M., König, P., & Engel, A. (2005). Advancing dynamic binding theory: Implementation of complex concepts. In H. Zimmermann, & K. Krieglstein (Eds.), *Proceedings of the 6th Meeting of the German Neuroscience Society / 30th Göttingen Neurobiology Conference* (p. 36B). Heidelberg: Elsevier.

Schurz, G. (2005). Semantic holism and (non-)compositionality of scientific theories. In M. Werning, E. Machery, & G. Schurz (Eds.), *The compositionality of meaning and content*, *1* (pp. 271-284). Frankfurt: Ontos.

Schurz, G. (2007). Human conditional reasoning explained by non-monotonicity and probability: An evolutionary account. In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.), *Proceedings of the 2nd European Cognitive Science Conference* (pp. 628-633). New York: Lawrence Erlbaum Associates.

Werning, M. (2001b). How to solve the problem of compositionality by oscillatory networks. In J. D. Moore, & K. Stenning (Eds.), *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society* (pp. 1094-1099). London: Lawrence Erlbaum Associates.

Werning, M. (2004b). On the possibility of self-awareness. In R. Bluhm, & C. Nimtz (Eds.), *Selected papers from the Fifth international congress of the Society for Analytical Philosophy* (pp. 470-478). Paderborn: Mentis.

Werning, M. (2005c). Time structure and the structure of perception. In F. Stadler, & M. Stöltzner (Eds.), *Time and history: Contributions of the Austrian Ludwig Wittgenstein Society* (pp. 336-339). Kirchberg am Wechsel: Austrian Ludwig Wittgenstein Society.

Werning, M., & Maye, A. (2005). Frames, coherency chains and hierarchical binding: The cortical implementation of complex concepts. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the Twenty-seventh Annual Conference of the Cognitive Science Society* (pp. 2347-2352). London: Lawrence Erlbaum Associates.

Werning, M., & Maye, A. (2006). The neural basis of the object concept in ambiguous and illusionary perception. In R. Sun, & N. Miyake (Eds.), *Proceedings of the Twenty-eighth Annual Conference of the Cognitive Science Society* (pp. 876-881). London: Lawrence Erlbaum Associates.

Werning, M., Fleischhauer, J., & Beseoglu, H. (2006). The cognitive accessibility of synaesthetic metaphors. In R. Sun, & N. Miyake (Eds.), *Proceedings of the Twenty-eighth Annual Conference of the Cognitive Science Society* (pp. 2365-2370). London: Lawrence Erlbaum Associates.

c) anthologies

Petersen, W., & Werning, M. (2007). Conceptual fingerprints: Lexical decomposition by means of frames – a neuro-cognitive model. In U. Priss, S. Polovina, & R. Hill (Eds.), *Conceptual structures: Knowledge architectures for smart applications* (pp. 415-428). Heidelberg: Springer-Verlag.

Schurz, G. (to appear 2009). Prototypes and their composition from an evolutionary point of view. In M. Werning, W. Hinzen, & E. Machery (Eds.), *The Oxford handbook of compositionality*. Oxford: Oxford University Press.

Werning, M. (2003b). 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. 261-278). Dordrecht: Kluwer.

Werning, M. (2004c). Self-awareness and imagination. In J. Sáàgua (Ed.), *A explicação da interpretação humana / The explanation of human interpretation* (pp. 369-377). Lisbon: Colibri.

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

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

Werning, M. (to appear 2009). The neural foundation of semantic composition. In M. Werning, W. Hinzen, & E. Machery (Eds.), *The Oxford handbook of compositionality*. Oxford: Oxford University Press.

Werning, M., & Maye, A. (2004). Implementing the (de-)composition of concepts: Oscillatory networks, coherency chains and hierarchical binding. In S. D. Levy, & R. Gayler (Eds.), *Compositional connectionism in cognitive science* (pp. 76-81). Menlo Park: AAAI Press.

II. Non-refereed publications in

a) journals

b) scientific congresses

c) anthologies

Machery, E., Schurz, G., & Werning, M. (2005). Preface. In M. Werning, E. Machery, & G. Schurz (Eds.), *The compositionality of meaning and content, Vol. I: Foundational Issues* (pp. 7-21). Frankfurt: Ontos Verlag.

Werning, M., Hinzen, W., & Machery, E. (to appear 2009). Compositionality: Reasons, problems and perspectives. In M. Werning, W. Hinzen, & E. Machery (Eds.), *The Oxford handbook of compositionality*. Oxford: Oxford University Press.

III. Anthologies and editions

Machery, E., Werning, M., & Schurz, G. (2005). *The compositionality of meaning and content.* Vol. II: Applications to Linguistics, Psychology and Neuroscience. Frankfurt: Ontos Verlag.

Werning, M. (1997). *Erkenntnis und Schlußfolgerung.* Berlin: Master thesis, Free University of Berlin.

Werning, M. (2004c). *The compositional brain: A unification of conceptual and neuronal perspectives.* Düsseldorf: Ph.D. Thesis, Heinrich-Heine University Düsseldorf.

Werning, M., Hinzen, W., & Machery, M. (Eds.). (ca. 2009). *The Oxford Handbook of Compositionality*. Oxford: Oxford University Press.

Werning, M., Machery, E., & Schurz, G. (Eds.). (2005). *The compositionality of meaning and content.* (Vol. I: Foundational Issues). Frankfurt: Ontos Verlag.

Relevant Publications

Arbib, M., Conklin, E., & Hill, J. (1987). *From schema theory to language*. Oxford: OUP.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review, 113*, 628-647.

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.

Barsalou, L. W. (2005). Situated conceptualization. In H. Cohen, & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 619-650). St. Louis: Elsevier.

Boas, F., & Deloria, E. (1939). Dakota grammar. *Memoirs of the National Academy of Sciences*, *23*(2).

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence, 47*, 139-159.

Clancey, W. J. (1997). *Situated cognition – on human knowledge and computer representations.* Cambridge University Press.

Call, J. (2006). Descartes' two errors: Reason and reflection in the great apes. In S. Hurley & M. Nudds (Eds.), *Rational animals?* (pp. 219–33). Oxford: Oxford University Press.

Clancey, W. J. (1997). Situated cognition – on human knowledge and computer representations. Cambridge University Press.

Clifton, C., & Ferreira, F. (1987). Modularity in sentence comprehension. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding* (pp. 277-290). MIT Press.

Crair, M. C., Ruthazer, E. S., Gillespie, D. C., & Stryker, M. P. (1997). Ocular dominance peaks at pinwheel center singularities of the orientation map in cat visual cortex. *Journal of Neurophysiology*, *77*(6), 3381-3385.

Crowley, T. (1995). Inalienable possession in Paamese grammar. In H. Chappell, & W. McGregor (Eds.), *The grammar of inalienability* (pp. 383-432). Berlin: de Gruyter.

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., et al. (1988). Coherent oscillations: a mechanism for feature linking in the visual cortex? *Biological Cybernetics*, *60*, 121-130.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, *2*, 704-716.

Engel, A. K., Roelfsema, P. R., Fries, P., Brecht, M., & Singer, W. (1997). Role of the temporal domain for response selection and perceptual binding. *Cerebral Cortex*, *7*, 571-582.

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

Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology.* Cambridge, MA: MIT Press.

Fodor, J. (1994). *The elm and the expert – mentalese and its semantics.* Cambridge, MA: MIT Press.

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

Fodor, J. (1998). *Concepts: Where cognitive science went wrong*. New York, NY: Oxford University Press.

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

Fodor, J., & Lepore, E. (1996). The red herring and the pet fish: why concepts still can't be prototypes. *Cognition, 58*(2), 253-270.

Fodor, J., & Lepore, E. (2002). *The compositionality papers*. Oxford: Oxford University Press.

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

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3-71.

Friederici, A. D. (2002, February). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78-84.

Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science, 291*, 1560-1563.

Fries, P., Roelfsema, P. R., Engel, A. K., König, P., & Singer, W. (1997). Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. *Proc. Natl. Acad. Sci. USA*, *94*, 12699-12704.

Fries, P., Schröder, J.-H., Roelfsema, P. R., Singer, W., & Engel, A. K. (2002). Oscillatory neuronal synchronization in primary visual cortex as a correlate of stimulus selection. *Journal of Neuroscience*, *22*, 3739-3754.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67-82). Hillsdale, N.J.: Erlbaum.

Gluck, M. A., & Bower, G. H. (1988, September). From conditioning to category learning: an adaptive network model. *Journal of Experimental Psychology: General*, *117*(3), 227-247.

Gray, C., König, P., Engel, A. K., & Singer, W. (1989). Oscilliatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, *338*, 334-337.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, *27*(3), 377-442.

Hinton, G. E., & Anderson, J. A. (Eds.). (1981). *Parallel models of associative memory*. Hillsdale, NJ: Erlbaum.

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*, 129-191.

Keijzer, F. (2002). Representation in dynamical and embodied cognition. *Cognitive Systems Research*, *3*(3), 275-288.

Kohonen, T. (1977). *Associative memory: A system-theoretic approach.* Berlin: Springer-Verlag.

Laurent, G. (1996). Dynamical representation of odors by oscillating and evolving neural assemblies. *Trends in Neuroscience*, *19*(11), 489-496.

Machery, E. (2005). Concepts are not a natural kind. *Philosophy of Science*, *72*, 444-467.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology, 58*(1), 25-45.

Martin, A., Wiggs, C. L., Ungerleider, L., & Haxby, J. V. (1996). Neural correlates of categoryspecific knowledge. *Nature*, *379*, 649-652.

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

Medin, D. L. (1975). A theory of context in discrimination learning. In G. H. Bower (Ed.), *The psychology of learning and motivation*, *Vol. 9* (pp. 269-315). New York: Academic Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207-238.

Mervis, C., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*, 89-115.

Millikan, R. G. (1998). A common structure for concepts of individuals, stuffs and real kinds: More Mama, more milk, and more mouse. *Behavioral and Brain Sciences*, *21*, 55-100.

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

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*(1), 104-114.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *13*(1), 87-108.

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

Palmeri, T. J., & Nosofsky, R. M. (1995, May). Recognition memory for exceptions to the category rule. Journal of Experimental Psychology: Learning, Memory and Cognition, 21(3), 548–-568.

Prinz, J. J. (2002). *Furnishing the mind*: Concepts and their perceptual basis. Cambridge, MA: MIT Press.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*, 253-279.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576-582.

Pylyshyn, Z. (1986). *Computation and cognition*. Cambridge, MA: MIT Press.

Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single-neurons in the human brain. *Nature*, *435*, 1102-1107.

Raos, V., Franchi, G., Gallese, V., & Fogassi, L. (2003). Somatotopic organization of the lateral part of area f2 (dorsal premotor cortex) of the macaque monkey. *Journal of Neurophysiology*, *89*(3), 1503-1518.

Rappelsberger, P., Weiss, S., & Schack, B. (2000). Coherence and phase relations between EEG traces recorded from different locations. In R. Muller (Ed.), *Time and the brain* (pp. 297-330). Harwood Academic Publishers.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H.Black, & W. F. Prokasy (Eds.), *Classical conditioning II :current research and theory* (pp. 64-98). New York: Appleton Century Crofts.

Rizzolatti. G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169-192.

Robbins, P. (2002). How to blunt the sword of compositionality. *Noûs*, *36*, 313-334.

Rodriguez, E., George, N., Lachaux, J., Martinerie, J., Renault, B., & Varela, F. (1999). Perception's shadow: Long-distance synchronization of human brain activity. *Nature*, *397*, 430-433.

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

Rumelhart, D. E., & McClelland, J. L. (1986). PDP models and general issues in cognitive science. In D. Rumelhart, J. L. McClelland, & PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1: Foundations*. Cambridge, MA: MIT Press.

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

Schnitzler, A., Timmermann, L., & Gross, J. (2006). Physiological and pathological oscillatory networks in the human motor system. *Journal of Physiology*, *99*(1), 3-7.

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

Smith, J. D., & Minda, J. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*(1), 3-27.

Smolensky, P. (1988). On the proper treatment of connectionism. *The Behavioral and Brain Sciences*, *11*, 1-74.

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. 223-290). Cambridge, MA: Blackwell.

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

Stephan, A. (1999). Are animals capable of concepts? *Erkenntnis*, *51*(1), 583-596.

Tallon, C., Bertrand, O., Bouchet, P., & Pernier, J. (1995). Gamma-range activity evoked by coherent visual stimuli in humans. *European Journal of Neuroscience*, *7*(6), 1285-1291.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171-178.

von der Malsburg, C. (1981). *The correlation theory of brain function*. (Internal Report Nos. 81–82.) Göttingen: MPI for Biophysical Chemistry.

Weiss, S., Müller, H. M., & Rappelsberger, P. (2000). Theta synchronization predicts efficient memory encoding of concrete and abstract nouns. *Neuroreport*, *11*, 2357-2361.

Weiss, S., Müller, H., Schack, B., King, J. W., Kutas, M., & Rappelsberger, P. (2005). Increased neuronal synchronization accompanying sentence comprehension. *International Journal of Psychophysiology*, *57*, 129-141.

Xiao, Y., Wang, Y., & Felleman, D. J. (2003). A spatially organized representation of colour in macaque cortical area V2. *Nature*, *421*(6922), 535-539.

3.3 Experiments involving humans or human materials

☐ **yes** ☒ **no**

3.4 Experiments with animals

☐ **yes** ☒ **no**

3.5 Experiments with recombinant DNA

☐ **yes** ☒ **no**