The Neural Basis of the Object Concept in Ambiguous and Illusionary Perception

Markus Werning (werning@phil.uni-duesseldorf.de)
Department of Philosophy, Heinrich-Heine University Düsseldorf
Universitätsstr. 1, D-40225 Düsseldorf, Germany

Alexander Maye (a.maye@uke.uni-hamburg.de)
Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf
Martinistraße 52, D-20246 Hamburg, Germany

Abstract

To test the hypothesis that synchronous neural oscillation constitutes the cortical representation of objects, an oscillatory network is designed and stimulated with non-uniform, ambiguous and illusionary objects. Alternative perceptive possibilities correspond to a multitude of eigenmodes of the network dynamics. A semantic interpretation of the network is developed. The data support the view that oscillation functions are the neural basis of object concepts, while clusters of feature responsive neurons constitute the basis of predicate concepts.

Introduction

Cognition is defined over conceptual structures, (i) which have content and (ii) are in principle (not necessarily by the subject itself) expressible by languages with object and predicate terms. The first condition derives from the fact that cognitive processes are epistemic in the sense that the criterion of truth-conduciveness, which is reserved for bearers of content, applies. The second condition grounds in the assumption that cognition presupposes categorization. Truth-conducive processes would be practically useless and without any evolutionary benefit if they did not subsume objects under categories. To do so the cognitive system must dispose over object and predicate concepts.

The role of the object concept in cognition and perception has been of particular interest not only in the developmental literature (reviewed by Scholl & Leslie, 1999), but also in neural modeling. Von der Malsburg’s (1981) supposition that the synchronous oscillation of neural 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 neural synchrony has been observed in numerous cell recording experiments (reviewed by Singer, 1999) and experiments related to attention (Steinmetz et al., 2000), perception (Fries, Roelfsema, Engel, König, & Singer, 1997), expectation (Riehle, Grün, & Aertsen, 1997) and mental representation (Werning, 2003a). 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 (Werning, 2005b).

From Gestalt psychology the principles governing object concepts are well known. According to two 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 word, represented by one and the same object concept. Most real-world objects, however, are non-uniform in one or more of their feature dimensions, e.g., within one object illumination, edge orientation and/or color can vary. Aside from non-uniformity there are cases of ambiguous stimuli: two distinct objects that overlap with each other and are alike in one or more feature dimensions, can generate the same retinal activation pattern as a single object with non-uniform properties. Furthermore, some stimuli as a matter of illusion arouse the perception of an object where no object really exists. Cases of non-uniformity, ambiguity and illusionary perception not only challenge the applicability of the Gestalt principles, but also provide an interesting test for our hypothesis about the neural basis of object and predicate concepts.

Using structural principles well known from the neurophysiology of the visual cortex, we designed an oscillator network for multidimensional feature binding and presented non-uniform, ambiguous and illusionary stimuli as input. To confirm our hypothesis, it should be expected that, even under these challenging condition, (i) the network assigns exactly one oscillation function (i.e., one object concept) to each object normally perceived by human subjects and (ii) the clusters of neurons (i.e., the predicative concepts) responsive for properties of the object show activity that is rendered by the oscillation function representing the object in question. We will, furthermore, take a look at the way the network manages to

1In some experiments, e.g., on the perception of plaid stimuli (Thiele & Stoner, 2003) synchronization was apparently not relevant.
represent the various, more or less likely, perceptive possibilities. To account for the representational capabilities of the network, a formally explicit semantic interpretation will be given.

**Multidimensional feature binding**

Schillen and König (1994) investigate the synchronization properties of an oscillator network for a stimulus that is uniform in one feature dimension (orientation), but differs in two others (features dimensions chosen are orientation, disparity and color). It turned out that the oscillators receiving input from the same object synchronized with each other, while the oscillatory functions of oscillators receiving input from two distinct objects differed by a phase shift. This corresponds to the perception of two distinct objects.

For an even number of feature dimensions or varying feature values in the object, the binding task can become ambiguous. A possible solution is the simultaneous representation of multiple representational candidates. Maye (2003) shows that the dynamics of an oscillator network can simultaneously represent multiple binding solutions.

We used a network of coupled oscillators to implement Gestalt-based feature binding in the temporal domain. The subnetwork for binding a single feature has been detailed by Maye (2003), but the general structure will be shortly reproduced: A single oscillator consists of an excitatory and inhibitory neuron with recurrent synaptic connections (Fig. 1a). Each model neuron is considered a representative of a larger group (100 to 200) of spatially proximal and physiologically similar biological neurons. Oscillators are arranged on a three-dimensional grid. Two dimensions represent the retinotopic mapping of the spatial domain, while the third dimension represents discrete values of a single feature. If a specific feature value is present in the receptive field, the corresponding oscillator will be activated by an input signal. The oscillators are locally connected by synchronizing and desynchronizing connections (Fig. 1b). This network implements binding within a single feature dimension and will be called a feature module. A mathematical analysis of a single oscillator as well as of the network has been carried out by Maye (2002).

The current work extends this model to multiple features. Here, the network consists of several feature modules, one for each feature dimension. For the qualitative design of the coupling between the feature modules, two criteria were relevant: (i) The distinctive features of a single object should synchronize the activity of oscillators activated by this object in the respective feature modules. For this reason, feature modules are coupled by synchronizing connections that preserve the topology. (ii) No particular entailments (e.g., ‘green things are vertical’) are specified between features of different feature layers. Therefore, couplings between feature modules are unspecific across feature layers. Fig. 2 illustrates a subset of the network.

In order to synchronize different feature modules, the excitatory neurons of oscillators were coupled. Quantitatively, the coupling strength \( L_{AB}(i,j) \) between oscillator \( i \) in feature module \( A \) and oscillator \( j \) in feature module \( B \) is given by:

\[
L_{AB}(i,j) = \begin{cases} \frac{L_0}{\sqrt{\pi \sigma^2}} e^{-\left(\frac{d(i,j)^2}{2\sigma^2}\right)} & \text{if } d(i,j) < r \\ 0 & \text{else} \end{cases}
\]

The distance in geometric space between the receptive fields of both oscillators is denoted by \( d(i,j) \) and the weight parameter is \( L_0 \). Connections emanating from oscillator \( i \) are allowed to contact oscillators in a surround of size \( r \) from the target oscillator \( j \).
Non-uniform and ambiguous stimuli

For the first series of experiments two feature dimensions were used: color and orientation. In order to investigate the binding capabilities of the network, two types of stimuli were tested (Fig. 3a–d). The first contained a horizontal and a vertical bar that overlap in the center. When both bars share the same color, this is usually perceived as a cross. This stimulus is uniform in the color dimension, but non-uniform in the orientation dimension. If the bars have different colors, they are non-uniform in both feature dimensions. In the other type of stimuli two non-overlapping bars were shown. In one case they had the same color and were parallel so that they might be perceived either as one object (a grating) or as two objects. In the other case both bars were different in color and orientation. The input to the network was computed from these stimuli. Since the feature values are binary in each dimension (two colors, two orientations), at most two layers of each feature module received input. The dynamic equations (Maye, 2003) were then solved numerically by a fourth order Runge-Kutta method. The activity of the \( j \)-th oscillator is characterized mathematically by the activity function \( x_j(t) \) during a time window \([0, T] \). Activity functions are vectors in the Hilbert space \( L^2[0, T] \), which comprises all functions square-integrable in the interval \([0, T] \). In case of real-valued functions, this space has the inner product

\[
\langle x(t)|y(t) \rangle = \int_0^T x(t) y(t) dt. \tag{2}
\]

The degree of synchrony between two functions is defined as

\[
\Delta(x, y) = \langle x|y \rangle / \sqrt{\langle x|x \rangle \langle y|y \rangle} \tag{3}
\]

and lies between \(-1\) and \(+1\). The degree of synchrony corresponds to the cosine of the angle between the Hilbert vectors \( x \) and \( y \). The vectors are parallel (synchronous), anti-parallel (anti-synchronous) and orthogonal (uncorrelated) depending on whether \( \Delta(x, y) \) is \(+1\), \(-1\) or \(0\). The overall dynamics of the network is given by the Cartesian product of the eigenvalues to the right. Below, the eigenmodes dominating states. These states are the eigenmodes of the system. The corresponding eigenvalues designate how much of the variance is accounted for by that mode. The eigenmodes \( e \) of the network dynamics are computed as the eigenvectors of the auto-covariance matrix \( C \), where its components \( C^{ij} \) are given as

\[
C^{ij} = \langle x_i|x_j \rangle.
\]

The network state at any instant is considered as a superposition of the eigenmodes \( e_i \):

\[
x(t) = \sum_i \alpha_i(t) e_i,
\]

where the \( \alpha_i(t) \) are the temporally evolving superposition coefficients determined by projecting the activity \( x(t) \) into the eigenspace. \( \alpha_i(t) \) will be called the characteristic function of the \( i \)-th eigenmode.

---

3Coupling parameters: \( L_0 = 0.1, r = 2 \); module parameters as defined by Maye (2003): \( \tau_x = \tau_y = 1, m_x = m_y = 2, \theta_x = 2, \theta_y = 1, L_0 = 2, L_0^x = 0.6, L_0^y = 0.5, W_0 = 0.1, r_x = r_y = 4 \). Parameters apply to Fig. 4 as well.

4To compute the components of the auto-covariance matrix, the integral was approximated by a sum over discrete unitary time steps

\[
\langle x|y \rangle \approx \sum_{0 \leq t \leq T} x(t) y(t).
\]
Maye (2002) demonstrates that the eigenvectors approximate the eigenmodes of the solution of the system of ODEs describing the network dynamics if the time course of the superposition coefficients is sinusoidal and if there are strong differences in the variances of each principal direction (expressed by the magnitude of the corresponding eigenvalues). Under these conditions the superposition coefficients associated with each eigenvector correspond to the order parameters associated with each eigenmode. The temporally evolving order parameters, i.e., the characteristic functions, are shown in Fig. 3e–f and eigenvalues in Fig. 3a–d.

For display purposes the eigenvectors were split into the contributions from each activated layer and reshaped to a matrix. To analyze the eigenmodes, the signs of the components of each mode (visualized by light and dark shades of grey) are taken into account. Regions with the same sign are activated together, while regions with opposite signs are activated alternately. This shows which parts of the stimulus are bound by synchronous activity of the oscillators.

Considering the eigenmode with the highest eigenvalue (the first row in Fig. 3a–d), a number of interesting observations can be made. To begin with, only oscillators that are stimulated by input in their receptive field become activated. The activity of spatially proximal oscillators is synchronized. The activity of oscillators in different feature modules activated by the same bar is synchronized as well. Finally, if the bars have different colors, the oscillations are desynchronized.

Applying the hypothesis that an object is represented by synchronous oscillation, the patterns of synchrony in the first eigenmodes of the four stimulations are to be read as representing: (3a) two objects, i.e., one R-colored V-oriented and one G-colored H-oriented object, (3b) one R-colored and partially V- and H-oriented object, (3c) one R-colored and V-oriented object, (3d) one R-colored V-oriented object and one G-colored H-oriented object.

When the same analysis is applied to the eigenmodes with the second largest eigenvalue (second row), alternative representations are to be read from the distribution of synchrony. This is most obvious in the stimulus in Fig. 3b. The two, differently oriented components of the cross are now represented as distinct objects. The stimulus of 3c in the second eigenmode is no longer represented as one, but as two objects. The second eigenmode of stimulus 3d represents an alternative possibility to the one in the first mode: now the two bars are bound together as one object. This perceptive possibility is also displayed by the third eigenmode for stimulus 3a.

The eigenvalues and the time-course of the order parameters show that the conditions under which eigenvectors of the covariance matrix can be considered an approximation of the eigenmodes of the underlying system of differential equations are fulfilled.

**Illusionary stimuli**

In the second series of experiments, a Kanisza triangle was presented as stimulus (Fig. 4a, top). It generates the percept of a white triangle in front of three black circles at the corners. The perception of this stimulus involves integration of contours that are induced by collinear line fragments at the corners of the illusionary triangle. This can be viewed as an instance of the Gestalt law of good continuation. The current model, however, does not account for contour integration. Hence, edge information was omitted altogether. Extending the model to exhibit contour integration is possible by means of an anisotropic connection schema of the oscillators in the orientation module as suggested by Li (1998).

Analyzing again the signs of the components of the eigenmodes, the most prominent eigenmode (first row in Fig. 4a) shows no distinction between the illusory figure and the background. This corresponds to a possible perception that only groups the three circle segments, on the one hand, and the background, on the other hand, to objects. The second eigen-
mode, however, clearly distinguishes the triangular object from the background. Due to missing edge information the shape of the triangle is not perfectly rendered (black, corresponding to the representation of the triangular object, seems to “flow out” on all three sides). Despite the vague information about contours, the negative pattern of synchrony (black) for the triangle in the left color layer is clearly distinguishable from the positive pattern of synchrony (white) for the rest of the stimulus, which is distributed over both color layers.

In order to figure out how far the eigenmodes are due to the illusory figure a control stimulus was tested. It had a similar structure but did not generate visual illusions (Fig. 4b top). For this stimulus none of the eigenmodes exhibits a difference between a foreground object in between the circles and the background. Subsequent eigenmodes distinguish between individual circles (data not shown).

Semantic interpretation

The dynamics of the network can be understood in semantic terms. We are allowed to regard oscillation functions as internal representations of individual objects, i.e., as object concepts. They may be assigned as meanings of some of the individual terms of a predicate language. Let Ind be the set of individual terms, then the partial function

$$\alpha : Ind \rightarrow \mathbb{L}_2[0, T]$$

(4)

is a constant individual assignment of the language into the set of activity functions \(\mathbb{L}_2[0, T]\). The sentence \(a = b (a, b \in Ind)\) – read, e.g., ‘this object is identical with that object’ – expresses a representational state of the system to the degree the oscillation functions \(\alpha(a)\) and \(\alpha(b)\) of the system are synchronous. Provided that \(Cls\) is the set of sentences, the degree to which a sentence expresses a representational state of the system, for any eigenmode \(e\), can be measured by the function

$$v_e : Cls \rightarrow [-1, +1].$$

(5)

In case of identity sentences we have:

$$v_e(a = b) = \Delta(\alpha(a), \alpha(b)).$$

(6)

Most vector components of a given eigenmode are exactly zero (illustrated by middle gray), while few in some cases are positive (light grey) and few in some cases are negative (dark grey). Since the contribution of an eigenmode \(e\) to the entire network state temporally evolves according to the characteristic function \(o(t)\), any positive eigenmode component \(e^i = +|e^i|\) contributes to the activity of the \(i\)-th oscillator with \(+|e^i|o(t)\), while any negative component \(e^j = -|e^j|\) contributes with \(-|e^j|o(t)\) to the activity of the \(j\)-th oscillator. Since the \(\Delta\)-function is normalized, only the signs of the constants matter to determine that the activities of the \(i\)-th and the \(j\)-th oscillator, contributed by an eigenmode, are exactly anti-parallel, while any two, with \(o(t)\) temporally evolving components of equal signs contribute mutually parallel activity. We may interpret this by saying that each eigenmode represents maximally two objects as different from one another. The representation of the first object is the positive function \(+o(t)\) and the representation of the second object is the negative function \(-o(t)\). Both, the positive and the negative function can be assigned to individual constants, \(a\) and \(b\), respectively, and thus play the role of object concepts. These considerations, for every eigenmode \(e\), justify the following evaluation of non-identity:

$$v_e(a = b) = \begin{cases} +1 & \text{if } v_e(a = b) = -1, \\ -1 & \text{if } v_e(a = b) > -1. \end{cases}$$

(7)

Feature clusters function as representations of properties. They can be expressed by monadic predicates. We will assume that our language has a set of monadic predicates \(Pred\) (containing, e.g., ‘red’, ‘green’, ‘vertical’, ‘horizontal’) such that each predicate denotes a property featured by some neural feature cluster. To every predicate \(F \in Pred\) we now assign a diagonal matrix \(\beta(F) \in R^{k \times k}\) that, by multiplication with any eigenmode \(e\), renders the sub-vector of the \(F\)-components, i.e., those vector components that belong to the feature cluster expressed by \(F\). The components of the matrix \(\beta(F)\) are defined as follows:

$$(\beta(F))_{ij} = \begin{cases} 1 & \text{if } i = j \text{ and } i \text{ indexes an } F\text{-component}, \\ 0 & \text{else}. \end{cases}$$

(8)

We are, hence, justified to call \(\beta(F)\) the neural intension of the predicate \(F\), or in other words, the (neural basis of the) predicate concept expressed by \(F\).

The neural intension of a predicate, for every eigenmode, determines its neural extension, i.e., the set of those oscillations that the neurons on the assigned feature layer, per eigenmode, contribute to the dynamics of the network. Hence, for every predicate \(F\) its neural extension in the eigenmode \(e\) comes to the set of activity functions

$$\{f_j | f = \beta(F)e\alpha(t)\},$$

where the \(f_j\) are the vector components of \(f\). To determine to which degree an oscillation function assigned to an individual constant \(a\) is in the neural extension of a predicate \(F\), we have to compute how synchronous it maximally is with one of the oscillation functions in the neural extension. This value then gives us the degree to which the sentence \(Fa\) (‘the object \(a\) satisfies the predicate \(F\)’) expresses a representational state of the system:

$$v_e(Fa) = \max\{\Delta(\alpha(a), f_j) | f = \beta(F)e\alpha(t)\}.$$
Table 1: Object concepts and the representational states of the network expressible by a sentence φ, per stimulus and eigenmode e. R, G: color predicates; H, V: predicates for orientation; a, b, c: individual terms.

<table>
<thead>
<tr>
<th>i</th>
<th>φ such that ν_e(φ) = 1</th>
<th>obj concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3a) 1</td>
<td>Ra ∨ V a ∧ G b ∧ H b ∧ ¬a = b</td>
<td>α(a) = −α(t)</td>
</tr>
<tr>
<td>2</td>
<td>Re ∧ V c ∧ G c ∧ H c</td>
<td>α(b) = +α(t)</td>
</tr>
<tr>
<td>3</td>
<td>Ra ∨ V a ∧ H a</td>
<td>α(c) = +α(t)</td>
</tr>
<tr>
<td>(3b) 1</td>
<td>Ra ∧ V a ∧ H a</td>
<td>α(a) = +α(t)</td>
</tr>
<tr>
<td>2</td>
<td>Rb ∧ Rc ∧ V b ∧ H c ∧ ¬b = c</td>
<td>α(b) = −α(t)</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>α(c) = −α(t)</td>
</tr>
<tr>
<td>(3c) 1</td>
<td>Ra ∧ V a</td>
<td>α(a) = +α(t)</td>
</tr>
<tr>
<td>2</td>
<td>Rb ∧ Rc ∧ V b ∧ V c ∧ ¬b = c</td>
<td>α(b) = +α(t)</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>α(c) = +α(t)</td>
</tr>
<tr>
<td>(3d) 1</td>
<td>Ra ∧ V a ∧ G b ∧ H b ∧ ¬a = b</td>
<td>α(a) = −α(t)</td>
</tr>
<tr>
<td>2</td>
<td>Re ∧ V c ∧ G c ∧ H c</td>
<td>α(b) = +α(t)</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>α(c) = +α(t)</td>
</tr>
</tbody>
</table>

Werning (2005a) extends this semantics to all logical constants of a first order predicate language and proves that it is compositional with respect to meaning and content. The conjunction, in particular, is evaluated by the minimum of the values of the conjuncts. Let φ, ψ be sentences of such a language, then, for any eigenmode e, we have:

$$
ν_e(φ ∧ ψ) = \min\{ν_e(φ), ν_e(ψ)\}.
$$

For each stimulus the network activity is governed by a number of eigenmodes specific for that stimulus. Each eigenmode represents different perceptive possibilities. The semantic interpretation of the network states now allows us to provide a precise analysis of the network’s representations and the object concepts involved therein (see table 1).

### Conclusion

The view on the neural basis of the object concept we presented in this paper competes, i.e., with a view recently proposed by Hurford (2003). He argues that object concepts and predicate concepts are processed separately, viz. in the dorsal and the ventral stream of the visual system, respectively (for discussion see Werning, 2003b). We, in contrast, hold that predicate and objects concepts are processed at the same location, at the same time and by the same mechanism. Since the generation of an object concept is governed by the Gestalt principles, which are formulated in terms of feature similarity, the processing of the object concept is inseparably intertwined with the generation of property representations. The theory of neural synchrony combined with the model of oscillatory networks takes this interdependence into account. The simulations reported here confirm our hypothesis that object concepts are to be identified with neural oscillations. Our hypothesis leads to successful predictions and explanations even under such ambitious conditions as non-uniform, ambiguous and illusionary stimuli.

### References


