in: Proceedings of the ICANN 1996, C. v.d. Malsburg,
W. v. Seelen, J.C. Vorbrüggen, B. Sendhoff (eds.),
Springer Verlag, Berlin, Heidelberg, New York, 1996.

Binding — a proposed experiment and a model

Jochen Triesch and Christoph von der Malsburg

Institut für Neuroinformatik, Ruhr-Universität Bochum, Germany

Abstract. The binding problem is regarded as one of today's key questions about brain function. Several solutions have been proposed, yet the issue is still controversial. The goal of this article is twofold. Firstly, we propose a new experimental paradigm requiring feature binding, the *"delayed binding response task"*. Secondly, we propose a binding mechanism employing fast reversible synaptic plasticity to express the binding between concepts. We discuss the experimental predictions of our model for the delayed binding response task.

1 Introduction

What is the binding problem? If several objects have to be represented simultaneously in our brains, superposition of the activity patterns may lead to confusion. This effect is called the "*superposition catastrophe*" and the general problem behind it is called the "*binding problem*". Several solutions have been proposed, only two of which shall be mentioned here. For a recent review of theoretical ideas the reader may refer to [vdM95].

Combination coding cells. The seemingly simplest approach is to introduce *combination coding cells*, which only react to combinations of features like e.g. an object of a particular shape and color at a particular retinal position. The problem with this approach is that it quickly leads to a combinatorial explosion of the number of needed cells. Furthermore, how should cells with these specificities be created? If they were to be learned, many examples of objects of all shapes and colors in all locations would have to be given before the system would work. If they were prewired, most of the connectivity pattern of the network would have to be stored in the genes, an unlikely proposition.

Binding by synchronous activity. This idea has received much attention recently. The signals of neurons representing features of the same object are mutually correlated in time, the signals of neurons representing features of different objects are not correlated or anti-correlated [vdM81]. The binding between cells referring to the same object is expressed in the signal correlations. For a review of experimental data pointing in this direction see [Sin93]. One special version of temporal binding is based on attentional mechanisms [TG80]: The responses to all but one stimulus are suppressed — only the units referring to one object are active. One way of implementing this uses inhibiton between units, although according to this proposal disambiguation is paid for by the restraint that only one stimulus may be represented at any time.

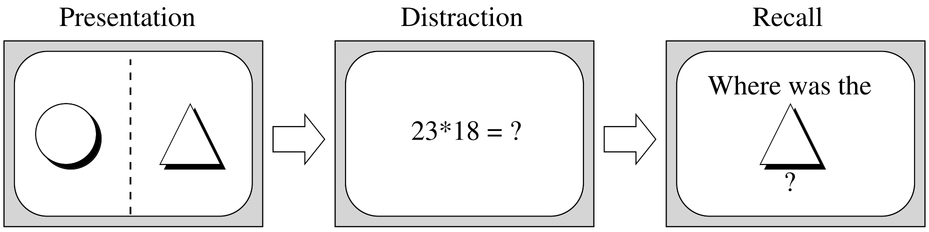

Fig. 1. The delayed binding response task. During the presentation phase the subject is asked to memorize binding relations between concepts, here place on the screen and shape. During a distraction phase the subject has to keep the binding relations (circle–left, triangle–right) in memory. Finally, the subject is asked to report a binding, by complementing a concept given as a cue (in the example the triangular shape is to be complemented by positional information).

2 The delayed binding response task

What would an experiment have to look like in order to test possible binding mechanisms? We propose a *delayed binding response task*, where subjects have to flexibly bind concepts and have to keep the bindings in memory during a distraction task before reporting them.

Behavioural task. The subject (human subject or non-human primate) is located in front of a computer screen, where visual objects are displayed in two different locations (see figure 1). For example the two objects may be a circle in the left half of the display and a triangle in the right half. In this example we have four concepts which have to be bound together: *circle* and *left* as well as *triangle* and *right*. After the subject has memorized the concepts and their binding relations, the stimuli are switched off and the subject performs a distraction task, which prevents him from rehearsing the concepts and their binding relations. After a few seconds one of the (four) concepts is given as a cue, and the subject has to indicate the concept that was bound to it during the presentation phase. If for example, the subject is cued with the triangle he is to indicate that it had been on the right-hand side of the stimulus display (e.g., by moving a joystick to the right).

It is important that the animal is trained (or the human subject instructed) for flexibility in the nature of the concepts and bindings implicit in the stimuli, using pairings such as shape-color, shape-motion, shape-shape, or even auditory or tactile stimuli instead of just shape-position as in our example — the common aspect of all tasks being that a binding is probed by just one of the concepts and the subject is to indicate the other. This is to keep the subject from over-learning the task and forming combination-coding cells specific to it. In the electrophysiological experiments proposed below this flexibility furthermore gives the possibility to select as current stimuli the particular response properties of neurons that the electrodes happen to have found. Our proposed paradigm is similar in many respects to the classical delayed response tasks [FBGR89] [Fus90], although requiring flexibility.

3 A model of binding

An overview of our model is given in figure 2. The model is composed of "units," pools of neurons that respond to one of the stimulus features, or "concepts." Individual neurons in a unit could be concentrated in cortical columns or could be dispersed over several areae, and presumably they are integrated by internal excitatory connections. The units can be activated by sensory stimulation or by queries sent from a center organizing behavior during performance of the experiment (not shown), the center being formed during training or verbal instruction. Units are permanently connected by synaptic links that reflect their logical relationship (mutually exclusive or potentially combined in a single stimulus).

When several stimuli are presented simultaneously, unit activity fluctuates. We interpret this fluctuation as the expression of attention that sequentially focuses on one stimulus at a time. Due to this mechanism, all concepts belonging to a single object or stimulus are activated simultaneously, while concepts belonging to other objects are suppressed. This assumption is supported by many neurophysiological and neuropsychological studies (see, e.g., [MD85] or [Luc94] for a recent review). We do not want to settle on a definite implementation of the attentional mechanism but it seems likely that lateral inhibition in cortical and subcortical structures might be important [Cri84]. At least part of this lateral inhibition may correspond to the inhibitory connections, shown in figure 2, which implement mutual exclusiveness of concepts.

The shifting activity of selective attention realizes temporal binding in our model. In the absence of eye movements we expect for these signals typical switching times in the range of a few tens of milliseconds (for a discussion of time spans typical for covert attention see [TG80]. Other binding paradigms, such as figure-ground separation, may be based on shorter time constants.)

The correlations in activity fluctuations control rapid reversible synaptic plasticity. The existence of fast synaptic plasticity mechanisms has been known for more than thirty years, for a review see [Zuc89]. Theoretical reasons for the importance of fast synaptic mechanisms for information processing in the brain have been emphasized in [vdM81], where a control mechanism is postulated according to which the efficacy of a synapse is rapidly increased if the pre- and postsynaptic neurons fire simultaneously and is rapidly decreased if activity is present but uncorrelated, whereas once signals cease the synapse slowly decays back to a base level, with a time-constant of the order of magnitude of the time span of short-term memory.

In our model for the delayed binding response task the coactivated units belonging to one object quickly strengthen connections between each other during stimulus presentation, which is the mechanism for short-term memory of binding. Later, when (part of) the concepts connected with an object are reactivated the still enhanced connections can reactivate all the concepts belonging to the object, hence giving access to all the information that was bound together.

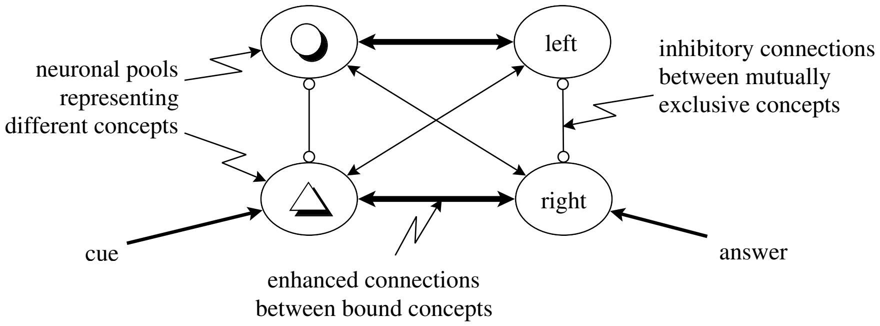

Fig. 2. Binding by enhanced connections between units. The unit pairs circle–left and triangle–right are bound by temporarily enhanced connections between them. Units for mutually exclusive concepts are connected by inhibitory connections, which prevent them from becoming active simultaneously.

Here is how the model would behave in the concrete delayed binding response task illustrated in figure 2. During *presentation* the attentional mechanism would activate either "circle" and "left" or "triangle" and "right" at a time, expressing binding between the concepts. During the times when, e.g., the units representing "circle" and "left" are highly active the synapses between them are strengthened, storing the binding between those concepts. In the *distraction* period the stimuli are switched off and the subject is prevented from rehearsing the concepts and their binding relations. During this time there is no activity in the four neural units (unless the distraction task makes use of their concepts) but the synapses between bound units do remain enhanced.

During *recall*, finally, a triangle is displayed in a central position as a cue and the subject now has to indicate that the triangle had been on the right during the stimulus presentation. With the help of the enhanced connection, the "triangle" unit will now activate the "right" pool, which additionally surpresses the "left" unit. Firing of the "right" unit can easily be imagined to trigger the correct (verbal or other) response to the task.

Note that the presentation of the triangle in a central position will generally also trigger the activiation of "middle" cells, which compete with the desired "right cells". Thus, current stimuli and recalled memories may interfere. We assume that during training of the task or instruction the subject learns to actively surpress such interferences.

4 Experimental Predictions

The model predicts *signal correlations.* Those units that correspond to different concepts present in a given stimulus element fire in a correlated way, whereas units that correspond to concepts belonging to different elements (if present in the stimulus) fire with no correlation or with anticorrelation.

What would be an appropriate recording technique to find and monitor relevant unit activity? As the temporal resolution characterizing the signal correlations to be found is unknown (although we guessed it would be dozens of milliseconds), it would be desirable to record activity with a resolution of a few milliseconds. Candidate techniques could be microelectrode recordings, optical imaging using voltage sensitive dyes, EEG or MEG. We suggest that the spatial resolution should be good enough to pick up signals specific for stimulus features (concepts), perhaps of the order of the diameter of a functional cortical column, between 0.3 and 0.6 mm. Thus the spatial resolution of non-invasive techniques might be inadequate, excluding human subjects from the proposed study.

With microelectrode recordings as an example, a concrete experiment could look like this: While the subject performs the task, simultaneous recordings are made from two neurons selective for different object shapes and from two neurons responsive to different object locations. Suitable shape specific neurons of sufficient complexity may be found in the inferior temporal cortex (IT), and their stimulus specificity may be discovered using a simplified version of the techniques described by K. Tanaka [Tan93]. Neurons selective for different locations in space but relatively unsensitive to stimulus shape could be found in parietal cortex [UH94]. Of course, this is only an example and there are many other ways for choosing stimulus classes and corresponding neurons. However, it is important that the concept pairs (e.g., shape and location) should be chosen such that they are unlikely to be represented by combination coding cells (if they existed then there would be no binding problem). A. Treisman's [TG80] paradigm could be used to test for the existence of combination-coding cells, by showing that the bindings in question are not subject to pop-out but require time proportional to the number of distractors.

A second prediction of the model is *rapid reversible synaptic plasticity.* After memorization during the presentation phase EPSP size of at least some synaptic connections between bound neural units should be increased, while it should remain small or even be diminished between neural units not being bound. To validate this prediction, intracellular recordings from target neurons would be required and individual EPSPs should be discernable (increased compound EPSPs could also be the result of an increased number of active presynaptic neurons). The relevant EPSPs should be identifiable by the correlation of their timing with signals in another unit involved in the task. We fear, however, that a direct test of the prediction rapid reversible synaptic plasticity in the delayed binding response task may be beyond current experimental techniques.

5 Discussion

The temporal binding hypothesis fails to be universally accepted, although the binding problem is widely discussed and recognized as such, and although relevant signal correlations have been found experimentally (as reviewed in [Sin93]). The lingering doubt is fed from two sources, the suspicion that those signal correlations are an insignificant epiphenomenon, and the suspicion that perhaps the brain manages to circumvent the binding problem altogether with the help of combination-coding cells, in spite of all the conceptual difficulties mentioned.

The model we are proposing as a neural basis of the experimental paradigm is minimal and in essence just consists in hypothesizing the presence of appropriate signal fluctuations and rapid reversible synaptic plasticity. Only on the basis of this utmost simplicity and ubiquitous availability of the binding machinery the flexibility can be reached that the experimental paradigm requires. Our model posits rapid reversible synaptic plasticity as basis for short-term memory. This is to be seen in contrast to the idea of reverberating neural activity, as demonstrated in the classical delayed response task [FBGR89] [Fus90]. We feel that neural reverberation is not a viable short-term memory mechanism in situations where flexibility is required.

We recognize that the experiment we propose pushes the state of the art to its limits and that it would constitute an enormous investment from the side of the experimenter, but we feel that an equally enormous scientific reward could make the effort worth while.

References

[Cri84] F. Crick. Function of the thalamic reticular complex: the searchlight hypothesis. *Proc. of the National Academy of Sciences*, 81:4586–4590, 1984.

[FBGR89] S. Funahashi, C. J. Bruce, and P. S. Goldman-Rakic. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61(2):331–349, 1989.

[Fus90] J. M. Fuster. Inferotemporal units in selective visual attention and short-term memory. *Journal of Neurophysiology*, 64(3):681–697, 1990.

[Luc94] S. J. Luck. Cognitive and neural mechanisms of visual search. *Current Opinion in Neurobiology*, 4:183–188, 1994.

[MD85] J. Moran and R. Desimone. Selective attention gates visual processing in the extrastriate cortex. *Science*, 229:782–784, 1985.

[Sin93] W. Singer. Synchronization of cortical activity and its putative role in information processing and learning. *Annu. Rev. Physiol.*, 55:349–374, 1993.

[Tan93] K. Tanaka. Neural mechanisms of object recognition. *Science*, 262:685–688, 1993.

[TG80] A.M. Treisman and G. Gelade. A feature integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.

[UH94] L. G. Ungerleider and J. V. Haxby. 'What' and 'where' in the human brain. *Current Opinion in Neurobiology*, 4:157–165, 1994.

[vdM81] C. von der Malsburg. The correlation theory of brain function. Internal Report 81-2, Dept. of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany, 1981. Reprinted in: Models of Neural networks II, edited by E. Domany, J.L. van Hemmen, and K. Schulten (Springer, Berlin, 1994) Chapter 2, pp. 95–119.

[vdM95] C. von der Malsburg. Binding in models of perception and brain function. *Current Opinion in Neurobiology*, 5:520–526, 1995.

[Zuc89] R. S. Zucker. Short-term synaptic plasticity. *Ann. Rev. Neurosci.*, 12:13–31, 1989.