# The Binding Problem

This lecture is based on following articles:
Adina L. Roskies: *The Binding Problem*; Neuron 1999 24: 7
Charles M. Gray: *The Temporal Correlation Hypothesis of Visual Feature Integration: Still Alive and Well*; Neuron 1999 24: 31
Wolf Singer: Neuronal Synchrony: *A Versatile Code for the Definition of Relations?*; Neuron 1999 24: 49

## Background
There are few, if any, places in the nervous system where all the information necessary to carry out a particular task is localized. This means that sensory, cognitive and motor processes result from parallel interactions among large populations o neurons in different regions of the brain.

## What is the "Binding Problem"?
Binding can be spatial as well as temporal. It can be rapid and preattentive or may require focused attention. It can occur within a single modality or may require cross modal or sensory-motor integration. One could say that binding is about "combining stimulus features to form an object representation". The problem is how does information variously distributed in patterns of neural firing result in coherent representations?

## The Term "binding problem" refers to a whole class of problems.

**perceptual binding**
*visual binding*:
binding information across visual space, across types of features and binding of neural signals across cortical space.

*auditory binding*:
may be needed to discriminate the sound of a single voice in a crowd.

*binding across time*:
required for interpreting object motion

*cross-modal binding*:
required to associate the sound of a ball striking a bat with the visual percept of it (aspects of a single event

**cognitive binding**
*relating a concept to a percept*:
linking the visual representation of an apple to all the semantic knowledge stored about it

*cross modal identification*:
being able to identify an item that has previously only been seen by how it feels

*sensory-motor binding*:
association between the sensory representation of objects and the motor commands to act in relation to those object

1

*memory reconstruction*:
linking of encoded information to form a unified representati

Because these processes are so diverse it is possible that they depend on differen mechanisms. In the following we will focus on *visual binding*.
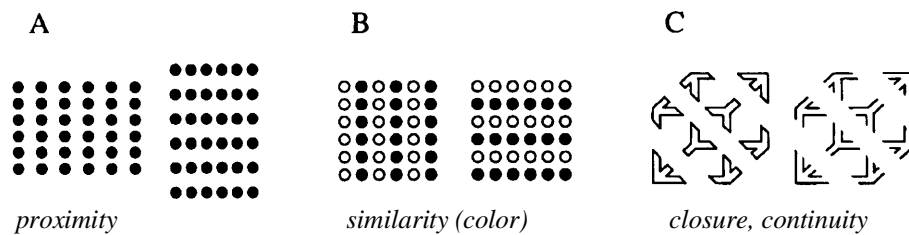
## Visual Perceptual Grouping



*Figure 1: Simple patterns illustrating several Gestalt principles of perceptual grouping.*

Perceptual Grouping is a form of binding. Every time we inspect a visual scene, we are confronted with a complex pattern of features, some static and some moving in various directions and at various velocities. In spite of this complexity, our visual system is able to rapidly identify the features that belong together as objects. In the first half of the last century, the field of Gestalt psychology formulated a set of principles specifying the properties of visual features that observers tend to group together. These were features of high visual salience. Several objects tend to be seen as one, if they are arranged according to the following Gestalt principles: proximity, similarity, continuity, closure, and common fate.

Features with these common characteristics tend to be automatically bound together as coherent units. If features conflict, the weakest feature is disregarded.
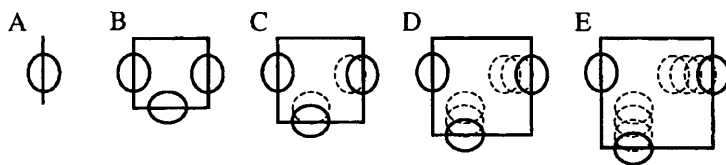
## Properties of Perceptual Grouping



*Figure 2: Schematic diagram illustrating the requirements for flexibility in the grouping of visual features.*

P.G. is a *flexible* process. A given feature element, at a given location in the visual field, may be grouped with one set of elements in one scene and with another, completely different, set of elements in another scene. This means the activity signalling the vertical contour must be recombined in a flexible manner so that a large number of other feature relationships can be perceptually grouped. Because visual scenes exist in near infinite variety, the perceptua grouping process appears to have infinite capacity.

Another feature of perceptual grouping is its *parallel* organization. Visual scenes usually consist of many objects often overlapping one another. Grouping must take place simultaneously at different locations of the visual field.
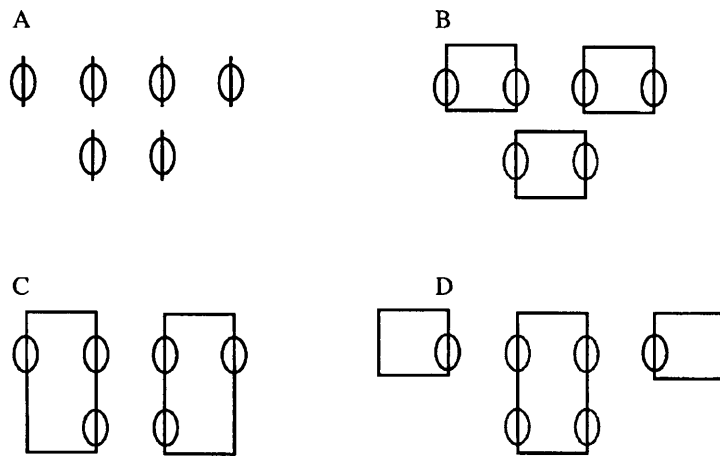
*Figure 3*

As the visual scene changes, features that are grouped together in one image may be segregated from one another in a different image. This may occur even if the local features themselves do not change (i.e. two vertical contours are grouped when the form opposite sides of a square but are segregated from one another when they form components of different squares).

Another important aspect of PG is its *speed*. When observers scan visual scenes they usuall make saccadic eye movements anywhere from one to four times a second, with the durati of visual fixation occasionally being as brief as 100 ms. Because the retinal image changes with each new visual fixation, perceptual grouping should take place at least as quickly as the interval between saccadic eye movements.

Together, these considerations suggest that the neuronal mechanisms responsible for perceptual grouping must operate within specific constraints. The mechanism must be capable of evaluating the relations among visual features and of identifying those combinations that exhibit Gestalt properties. It should be fast, flexible, have a large combinatorial capacity, and be parallel in its organization so that many feature combinations can be grouped simultaneously.
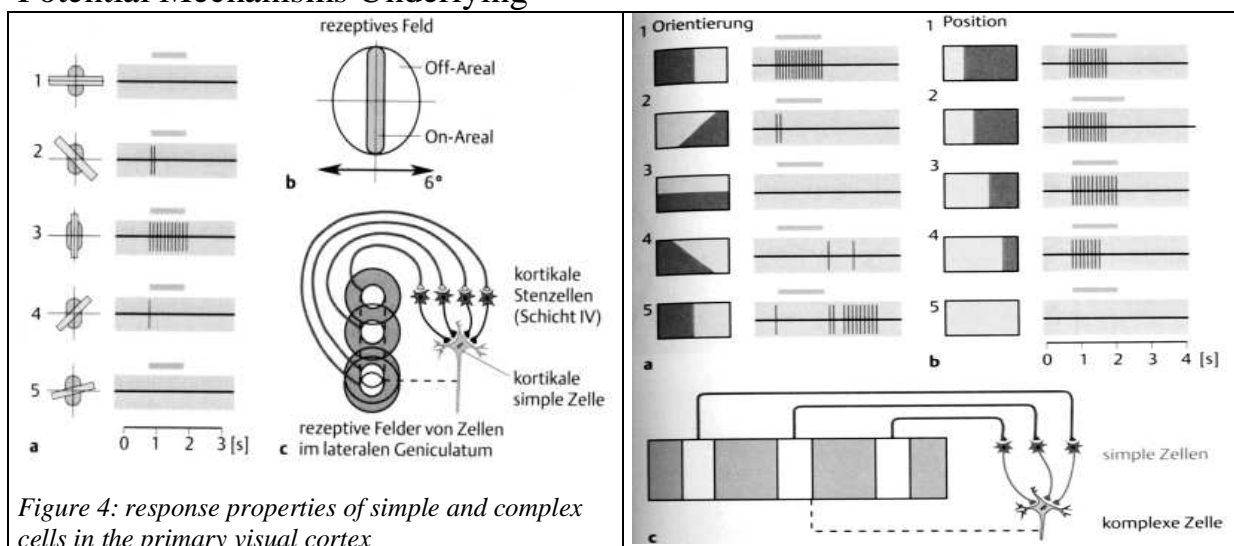
## Potential Mechanisms Underlying



*Figure 4: response properties of simple and complex cells in the primary visual cortex*

## Convergent Hierarchical Coding

The visual cortex is organized hierarchically. Each level in the hierarchy carries out operations that depend on the input from earlier levels. This hierarchical processing leads to a an increase in the complexity of neuronal representations. Increasingly complex stimul features are represented by higher levels in the hierarchy.

Evidence:
1. neuroanatomical, cytoarchitectonical: converging feed-forward connections
2. sensory discrimination tasks: At highest levels, neurons exhibit selective responses to object-like stimul
3. fast

This organization has specific implications . It suggests that grouping operations may be expressed by the activity of a small group o neurons or even a single neuron. Features and constellations of features are represented by the activity of neurons that receive convergent inputs from populations of neurons at lower levels in the hierarchy

**Combinatorial Problem**
The number of objects capable of sharing a common contour is very large. This means tha the higher-level neurons, whose receptive fields code for objects sharing a common contour, would all have to receive input from the neurons activated by the common contour (*figure 2*). One can see that this type of mecha nism quickly becomes untenable, when it is realized tha the number of objects capable of sharing a common contour is very large. Each neuron woul have to send an equally large number of connections to the next successive stage of the hierarchy, leading to an unacceptably large number of connections. It would require as man binding units as there are distinguishable objects, if all possible conjunctions were to be represented explicitly by conjunction neurons.

## Population Coding

The combinatorial problem can be overcome by a simple modification of the convergent coding. Rather than represent the integration of features by the activity of a few or even single neurons at specific cortical location, complex feature combinations could be represented b the activity of a population o neurons distributed within and across levels of the cortica hierarchy. In this scheme, each stimulus pattern could be represented by a distinct pattern o firing in a distributed population of cells. Such a coding scheme would greatly increase the representational capacity of the cortical network, because the number of distinct patterns o activity far outstrips the number of neurons available to represent the stimuli. Thus, the combinatorial complexity of the sensory world would no longer pose a problem.

Evidence:
1. The anatomical connectivity within and between different levels of the cortical hierarchy is both divergent and convergent. Neurons in a given cortical area send divergent outputs to other areas, enabling the cells in these areas to receive convergent input from many sources. This type of organization provides neuronal populations at the highest levels of the hierarchy with broad access to the entire visual field and the full range of visual feature properties, clearly marking them as candidate areas for feature integration and perceptual grouping. Similarly, there is substantial physiological evidence from both sensory and nonsensory cortical areas that stimuli, as well as motor actions, are represented by distributed populations of cellular activity.

2. Population coding is more economic in neuron numbers.

Attractive though it may be, this model of population coding is again incomplete. What happens if there is more than one stimulus? Then there would be two populations of neurons firing. How could these be distinguished? If features and objects are represented solely by the firing of distributed populations of cells, how are the vast numbers of active neurons in the cortical network disambiguated from one another? This is the so called "superpos tion problem."

## Temporal Correlation

Given that the activity evoked by the features comprising an object is distributed, that means, there is no single conjunction unit representing the feature, some mechanism is necessary to identify the members of a representation as belonging together and to distinguish them from other representations that may be present at the same time. What internal dynamic structure of a cell assembly would distinguish it from all other neuronal activity present simultaneously in the cortical network

To solve this problem was to define cell assemblies on the basis of the temporal correlation or synchronization of neuronal firing. Because neurons are sensitive to the timing of thei synaptic inputs, it was argued that synchronization provides many important advantages.

Synchronization reinforces the interactions among the members of the same assembly, provides a highly effective neuronal signal to systems responding to the output of an assembly. Therefore synchronicity serves as a signature of relatedness.

Different assemblies are distinguished from one another by the independence of their firing patterns. Thus, multiple, highly salient distributed signals can coexist in the same network of cortical areas at the same time, each providing an independent representation of grouped features.

Whenever conjunctions have to be signalled for which there are no specific conjunction units, synchrony can accomplish a *dynamic grouping of responses*.

**Use-Dependent Plasticity**
If synchronization of responses serves as a mechanism to group responses, synchronizing connections must be susceptible to use-dependent modifications. Indeed the synchronization probability increases between groups o neurons if these engage repeatedly in synchronous firing. This probability can be reduced again by having the same cells engage repeatedly in oscillatory firing patterns that are decorrelated.

Evidence:
1. neurons engage in synchronous firing on a time scale of 1ms
2. anatomical: horizontal fibre system
3. synchronous activity occurs preferentially when cells are activated by features having similar properties (Gestalt principles).

## Conclusion

Both Convergent Hierarchical Coding and Population Coding can be found in the visual system. Stereotyped, frequently occurring conjunctions are represented by specific binding units, because this strategy is faster and less susceptible to binding errors. However, because there are not enough neurons to exhaust the whole combinatorial space, population codes provide a possibility to represen  meta-conjunctions for which there are no binding units. Synchronicity serves as a tag of relatedness between neurons of a population by increasing the saliency of their synchronized responses, which in turn favours their jointly evaluation (binding) at subsequent processing stages.

It was suggested that this dual coding strategy (binding by convergence & dynamic binding) is iterated during the processing hierarchy, whereby some feature conjunctions ge represented explicitly by individual neurons and others by population codes.