Cortical Algorithms for Perceptual Grouping

Pieter R. Roelfsema

The Netherlands Ophthalmic Research Institute, Meibergdreef 47, 1105 BA
Amsterdam, The Netherlands; and Department of Experimental Neurophysiology,
Center for Neurogenomics and Cognitive Research, Vrije Universiteit, de Boelelaan
1085, 1081 HV Amsterdam, The Netherlands; email: p.roelfsema@nin.knaw.nl

Key Words

visual cortex, binding, visual attention, contour grouping,
contextual modulation, grandmother cell

Abstract

A fundamental task of vision is to group the image elements that
belong to one object and to segregate them from other objects and
the background. This review provides a conceptual framework of
how perceptual grouping may be implemented in the visual cor-
tex. According to this framework, two mechanisms are responsi-
ble for perceptual grouping: base-grouping and incremental group-
ing. Base-groupings are coded by single neurons tuned to multiple
features, like the combination of a color and an orientation. They
are computed rapidly because they reflect the selectivity of feedfor-
ward connections. However, not all conceivable feature combina-
tions are coded by dedicated neurons. Therefore, a second, flexible
form of grouping is required called incremental grouping. Incre-
mental grouping enhances the responses of neurons coding features
that are bound in perception, but it takes more time than does base-
grouping because it relies also on horizontal and feedback connec-
tions. The modulation of neuronal response strength during incre-
mental grouping has a correlate in psychology because attention is
directed to those features that are labeled by the enhanced neuronal
response.

Contents

INTRODUCTION

We live in a complex visual world. The typical visual scene that hits the retina is composed of large numbers of image elements that vary in luminance, color, shape, and motion. Our visual system's task is to impose structure on all this information: It must group together image elements that belong to a single object and segregate them from elements that belong to different objects and the background. In most instances, we immediately know where one object ends and another begins. Only occasionally do we experience difficulty in segmenting a scene (one example is given in **Figure 1***A*). The effortlessness of image segmentation could mean either that this task is easy or that the efficiency of our visual system prevents a full appreciation of the computational complexity. Computer science supports the latter view. The complexity of image segmentation was fully realized only when computer scientists tried to de-

velop algorithms for grouping and segmentation (Marr 1982) and, in spite of the striking developments in this field, the best computer algorithms are still no match for the human visual system (Barrett & Myers 2003).

Human vision uses many cues for image segmentation. One the one hand, it uses low-level grouping cues described by the Gestalt psychologists many years ago (Wertheimer 1923, Koffka 1935; reviewed by Rock & Palmer 1990). For instance, the Gestalt rule of similarity states that similar elements tend to be bound as an object in perception. On the other hand, human vision uses high-level grouping cues. In **Figure 1***A* we group the legs of the bird with the rest of the animal, because we know from our previous visual experience that this is where the legs should be. This grouping is not supported by low-level cues, suggesting that the legs, with their different color and texture, should rather be segregated from the rest of the animal. Familiarity with the shape of objects has a profound influence on image segmentation (e.g., Peterson et al. 1991, Ullman 1996, Vecera & Farah 1997).

The neuronal mechanisms for grouping and segmentation are, at best, partially understood. A satisfactory theory must explain how subjects combine different cues to converge on the correct segmentation of an image (Palmer 1992). This is not easy because image elements interact in complex ways in perception. Some of these complexities are illustrated in **Figure 2**. The first example illustrates the Gestalt law of proximity (**Figure 2***A*). On the left side of this figure, the blue dot forms a perceptual group with the yellow dot but not with the red dot. This grouping occurs because the blue and the yellow dots are close to each other and the red dot is farther away. By contrast, on the right side of the figure, the blue dot is grouped with the red dot rather than with the yellow dot even though the distances remain the same. The reason for this grouping is that the blue dot is close to a black dot, which in turn groups with the other black dots, which then group

with the red dot on the other side. These local groupings are combined in our perception, resulting in the blue dot being grouped with the red dot. This example demonstrates the transitivity of perceptual grouping: If A is grouped with B, and B with C, then A is also grouped with C.

Transitivity implies that grouping image elements depends on the context of the scene provided by other elements located at nearby or remote locations. **Figure 2B** illustrates the role of context through the Gestalt law of connectedness. On the left hand side, the two red dots are positioned on the same object, but on the right hand side, they are not. These two images differ at a location far from the two red dots. Connectedness is a transitive property, and the grouping of connected image elements can therefore depend on long chains of more local groupings (see also Minsky & Papert 1990). The third example, illustrated in **Figure 2C**, shows both transitivity and context sensitivity for the Gestalt law of similarity. Nearby line elements in this figure are grouped together when they have a similar orientation. However, the law does not necessarily apply to image elements that are farther apart. Line elements at locations 1 and 2, for example, have the same orientation but do not group because they are separated by boundaries where the orientation changes abruptly. In contrast, line elements at locations 2 and 3 have an orthogonal orientation, but they are grouped because the orientation changes gradually at intermediate locations, which allows transitive grouping. Taken together, these examples demonstrate that our visual system evaluates features across many image locations to arrive at a correct global segmentation of the image (see also the review by Albright & Stoner 2002).

Here, I review the cortical mechanisms responsible for perceptual grouping in the context of an incremental grouping theory (Roelfsema et al. 2000). The theory proposes a distinction between two types of grouping: base-grouping and incremental grouping (see also Ullman 1984). Base grouping relies on
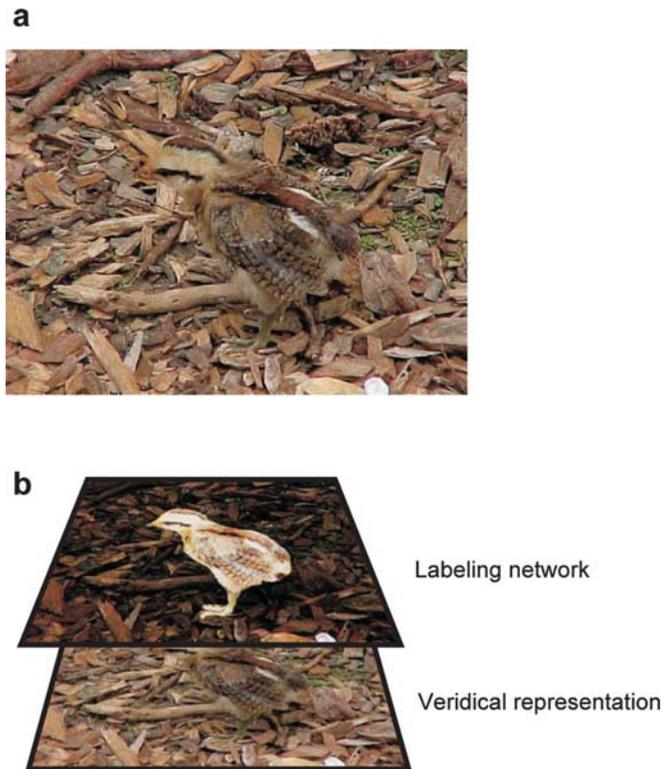
Figure 1

Difficult perceptual grouping task. (*A*) The bird is hard to see on a background that has a similar texture and color (photograph made by Dr. David Jefferies, and reproduced with permission). (*B*) Incremental grouping. The region occupied by the bird is shown at a higher luminance to illustrate how the enhanced neuronal responses label image elements that are grouped perceptually in areas of the visual cortex.

neurons that are tuned to feature conjunctions. These groupings are rapidly computed after the presentation of a new visual image by a cascade of feedforward connections that shape the neurons' receptive field (RF) properties in lower and higher areas of the visual cortex (purple in **Figure 3A,C**) (Felleman & Van Essen 1991). The feedforward processing phase roughly corresponds to what is often referred to in psychology as preattentive processing (Treisman & Gelade 1980). Incremental grouping, by comparison, requires recurrent processing: the exchange of information between neurons in the same area by horizontal connections (yellow in **Figure 3A,C**), and between neurons in higher and lower areas by

Incremental grouping: grouping by labeling neurons with an enhanced activity

Base grouping: grouping by neurons tuned to feature conjunctions

RF: receptive field

FEEDFORWARD AND RECURRENT PROCESSING

It is possible to distinguish between feedforward and recurrent processing on the basis of the following criteria: (*a*) feedforward connections drive the neurons' responses and determine the location and spatial extent of the classical receptive field. Modulatory influences on a neuron's response by information outside the RF can be attributed to recurrent loops through the same and higher visual areas (see **Figure 3C**). (*b*) A neuron's initial response is dominated by feedforward connections that provide the shortest route from the retina to the neuron. Effects mediated by horizontal and feedback connections are associated with additional synaptic delays. (*c*) In the macaque monkey a new stimulus arrives in the highest areas of visual cortex within ~120 ms (Nowak & Bullier 1997, Lamme & Roelfsema 2000). Tasks that take more time suggest recurrent processing.

feedforward and feedback connections (green in **Figure 3A,C**) (see Feedforward and Recurrent Processing). This time-consuming, attentive process labels a population of neurons that encodes the features to be grouped perceptually by enhancing the strength of their responses (highlighted region in **Figure 1B**). Incremental grouping is transitive, and it is therefore sensitive to the context set by other elements in the display. This review begins by focusing on grouping of contour elements. Later sections generalize the ideas to other feature domains and then return to the distinction between preattentive and attentive processing.

BASE-GROUPING: GROUPING BY SINGLE NEURONS

Images projected onto the retina are initially parceled into tiny fragments represented by neurons with small RFs. The information is relayed through the lateral geniculate nucleus (LGN) to the visual cortex, which has the difficult task of putting the objects back together again. Over 30 years of research has provided important insight into how object
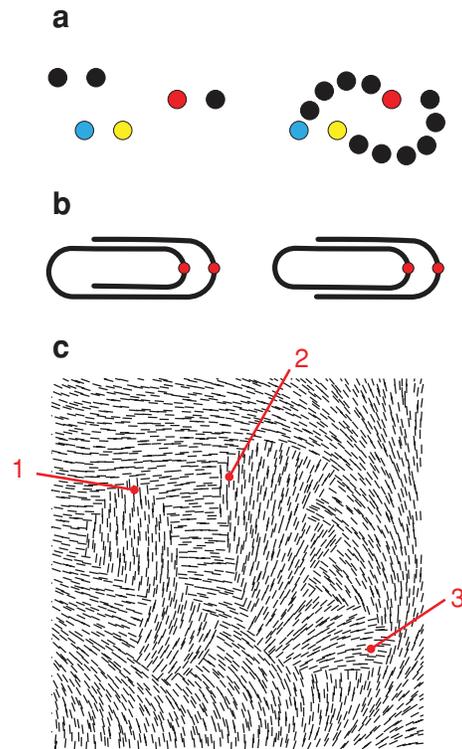
Figure 2

Context sensitivity of perceptual grouping. (*A*) Grouping by proximity. In the left image, the nearby blue and yellow circles are grouped together. In the right image, the blue dot is close to a string of black dots that, in turn, is close to the red dot. Grouping by proximity is a transitive process, and now the blue and red dots belong to the same perceptual group. (*B*) Grouping by connectedness. In the left image, the two red dots are located on a single, connected object, and in the right image they are not. This example illustrates that one can influence grouping by changing the stimulus at remote locations. (*C*) Grouping by similarity. Nearby image elements with a similar orientation are grouped in perception, whereas boundaries form at positions where the orientation changes abruptly. Image locations with similar orientations (locations 1 and 2) do not group if there is a boundary between them. In contrast, separate image locations with dissimilar orientations (locations 2 and 3) are grouped if the orientation changes gradually at intervening locations (modeled after Nothdurft 1992).
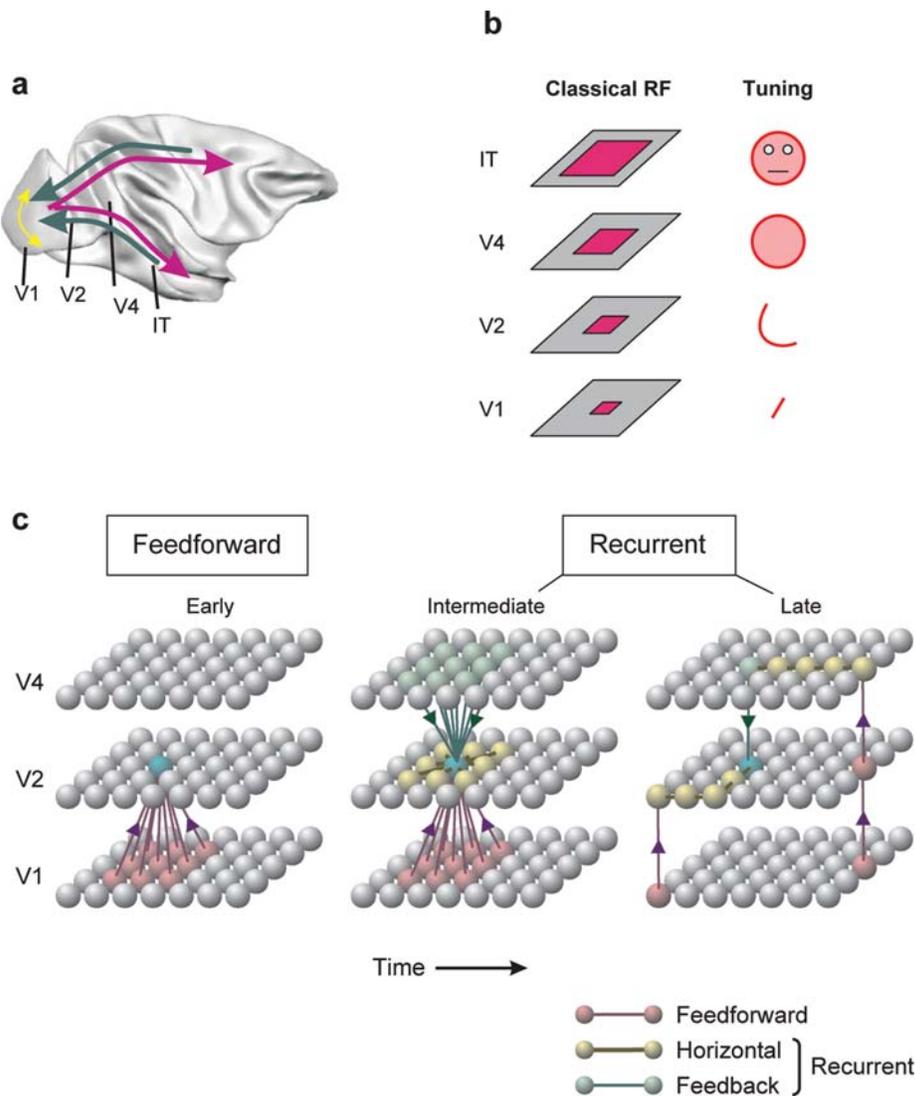
Connections between neurons in the visual cortex. (*A*) Schematic representation of feedforward connections (*purple*), horizontal connections (*yellow*), and feedback connections (*green*) between areas of the visual cortex. (*B*) In higher areas the size of the RFs increases and tuning becomes more complex. (*C*) *Left*: When a new image is presented, neuronal responses are initially dominated by the feedforward connections (*purple*). *Middle*: Horizontal (*yellow*) and feedback connections (*green*) take effect after a delay. *Right*: As time progresses longer recurrent routes start to influence the neuronal responses.

representations are gradually reconstructed from lower to higher regions of the visual cortex (Felleman & Van Essen 1991, Kobatake & Tanaka 1994) (**Figure 3**). Neurons in the primary visual cortex (area V1) have small RFs that are tuned to simple features, such as the orientation of a contour (**Figure 3***B*). Orientation tuning is a simple form of perceptual grouping because the detection of an oriented contour implies a set of smaller elements aligned in a specific oriented configuration. In addition, many neurons in V1 are tuned to combinations of features, such as colors, orientations, disparities, and motion directions (Leventhal et al. 1995, Sincich & Horton 2005). A neuron tuned to a red and vertical bar encodes a conjunction, i.e., a grouping of these features. Therefore, the distinction between a feature (unique and not divisible into smaller elements) and a conjunction (a combination of two or more features) is blurred at the neuronal level—a fact that has not been appreciated by many psychological theories (e.g., Treisman & Gelade 1980). It is

BIASED COMPETITION

The biased competition model, proposed by Desimone & Duncan (1995), suggests that different objects in the visual field compete for representation in the visual cortex. The competition is particularly strong in higher areas, where multiple objects are likely to fall into one RF. These inhibitory interactions occur on a fast timescale so that the competition already has an effect during the initial feedforward processing phase (Knierim & Van Essen 1992, Miller et al. 1993). The biased competition model proposes that bottom-up saliency as well as top-down factors that reflect the stimulus' behavioral relevance influence the outcome of the competition (for recent reviews of experimental evidence, see Kastner & Ungerleider 2000, Reynolds & Chelazzi 2004). The incremental grouping theory is compatible with biased competition between the features that belong to different objects. However, it also proposes excitatory interactions between neurons that code features of the same object that thereby tend to be coselected for perception and action.

convenient to use the term base-grouping for a feature conjunction that is coded by a single neuron, although it is just another word for the neuron's tuning.

The complexity of tuning (and thus of the base-groupings) increases in higher visual areas where RFs are larger. Many neurons in area V4 are selective for the shape and curvature of contour elements (Pasupathy & Connor 2001). In the inferotemporal cortex (area IT) the complexity of tuning increases further. **Figure 4*A*** illustrates a neuron that was recorded by Brincat & Connor (2004) in area IT of a monkey. They presented a large set of simple objects to the monkey and systematically investigated how the neuron's response depended on the object's contour elements. The tuning of the example cell could be explained by the relative position of four contour elements. The neuron was excited by two of the elements (A or B) and responded strongly to stimuli with three of the elements (A, B, and C all present); a fourth contour element (D) had a suppressive effect. Thus, the tuning could be described as the sum of terms

for the individual elements together with an interaction term. Joint tuning to multiple contour elements was observed in the large majority of IT neurons (Brincat & Connor 2004, see also Kobatake & Tanaka 1994). A simple feedforward model can explain this tuning; V4 provides information about the individual contour elements that are integrated into shapes at the level of area IT (insert in **Figure 4*A***). In such a simple model, tuning of V4 would depend on the input it receives from area V2, which in turn would depend on the input it receives from area V1 (e.g., Fukushima 1988, Riesenhuber & Poggio 1999).

In such a model, the selectivity of feedforward connection is responsible for the tuning of the neurons. Feedforward connections provide the shortest route from the retina to a cortical neuron, and in such a model, tuning is therefore observed from the first spike onward (van Rullen et al. 1998). Consistent with this idea, many V1 neurons are tuned to orientation during their initial visual response (Celebrini et al. 1993). The initial neuronal responses in higher visual areas are also selective for complex shapes (Oram & Perrett 1992, 1994; Sugase et al. 1999). This is direct support for the feedforward model because computational schemes that implicate horizontal and feedback connections are associated with additional synaptic and axonal propagation delays and do not give rise to the tuning of initial neuronal responses. Although base-grouping is fast (Lamme & Roelfsema 2000), the feedforward sweep can be curtailed by other objects in the scene. Competitive interactions that occur between the representations of different objects reduce the depth of processing and thereby decrease the number of computed base-groupings at higher levels in the visual hierarchy (as originally proposed by the biased competition model of Desimone & Duncan 1995; see Biased Competition).

The complexity of tuning increases even further in areas beyond area IT. Highly specific neurons have been observed in the medial temporal lobe of humans who were fitted

$$\text{Response} = 6.1A + 5.2B + 35.2ABC - 21.4D + 0.2$$

Figure 4

Base-grouping in the temporal lobe. (*A*) *Left*: Responses of an IT neuron to a large set of stimuli. Background gray levels indicate response strength. A simple equation based on four contour elements (A, B, C, and D) accounted for a large fraction of the variability in response strength. *Inset*: Simple feedforward model that could implement this equation. *Right*: Comparison of observed (*black bars*) and predicted response magnitude (*red bars*) for four of the stimuli (*red stars on the left*). From Brincat & Connor (2004). (*B*) Neuron that responds to pictures of Bill Clinton. The cell does not respond to a large variety of other pictures (only three of them are shown here). Adapted from Kreiman et al. (2002).

with electrodes to investigate the source of their epilepsy. **Figure 4B** shows an example of a cell in the amygdala that responded reliably and selectively to pictures of Bill Clinton but not to other pictures (Kreiman et al. 2002; see Quiroga et al. 2005 for more examples). The existence of these highly selective neurons was predicted by Barlow (1972) who called them "cardinal cells" (their nickname is grandmother cells). It is useful to have cardinal cells in an associative neuronal network because knowledge about the object or individual can be retrieved irrespective of the low-level features. However, there are limitations to using cardinal cells to achieve perceptual grouping (Singer & Gray 1995, von der Malsburg 1999). There are more feature conjunctions possible than neurons available in the brain. Moreover, many groupings never occur in a lifetime, and it is unrealistic to expect that there would be a neuron waiting for them.

DISTRIBUTED REPRESENTATIONS AND THE BINDING PROBLEM

Any single object always activates a pattern, or what Hebb (1949) called an assembly of neurons distributed across cortical areas. Distributed representations have several virtues. For example, they can represent the features of an object at a level of detail that is difficult to achieve with one or a few cardinal cells. A set of neurons in early visual areas can collectively represent the contour elements of an object in detail, while neurons in higher areas code aspects of its global shape. Moreover, a distributed representation would allow new objects to be represented as new patterns of activity across existing neurons.

A limitation of distributed representations is the binding problem, which is the tendency to confuse features that belong to different objects (Treisman & Schmidt 1982, von der Malsburg 1999). Imagine the simple problem of the simultaneous appearance of two objects composed of a number of contour ele-

ments. The objects activate two sets of neurons in early and higher visual areas, but this representation does not reveal which contour elements belong together as part of a single object. Similar problems can occur with the integration of features from different domains, like colors, motions, and shapes. The assembly code signals which features are present in the display, but it does not reveal which features belong to the same object (see also Treisman & Gelade 1980).

The binding problem can be solved by labeling neurons that encode the features of one object in a manner that distinguishes them from the neurons representing features of other objects. Two neurophysiologically plausible labels have been proposed. The first is synchrony. According to the binding-by-synchrony theory, neurons that respond to features that belong to the same object fire their action potentials at approximately the same time (von der Malsburg 1981). This theory initially received much support (reviewed by Singer & Gray 1995). However, recent studies that measured neuronal synchrony in monkeys that carried out perceptual grouping tasks observed no direct relationship between synchrony and perceptual grouping (Lamme & Spekreijse 1998, Thiele & Stoner 2003, Palanca & DeAngelis 2005). In some instances, grouping is even associated with a reduction in synchrony (Roelfsema et al. 2004), which suggests that synchrony is not the universal code for binding.

The incremental grouping theory, proposed here, claims that neurons encoding the features of one object enhance their firing rate compared with neurons encoding the features of other objects (**Figure 1B**). For ease of description, the theory refers to this differential activity as a response enhancement, although these effects are usually observed on a background of inhibitory interactions between the representations of different objects (see Biased Competition). The theory proposes that the enhancement of neuronal responses spreads across the network of horizontal and interareal connections that link

the neurons representing features of the same object (**Figure 3C**). These recurrent connections form chains of varying lengths so that the effects of short chains become evident at an earlier point in time than do the effects of longer chains. The next section reviews the interactions between neighboring contour elements that form local groupings. This sets the stage for the following section on a form of incremental contour grouping that is serial and time-consuming because it requires the combination of many local groupings in a transitive way.

THE LOCAL ASSOCIATION FIELD

A powerful tool to study contour grouping was introduced with the so-called pathfinder displays (Field et al. 1993). These displays contain Gabor patches (or contour elements), some of which are aligned colinearly to form a path (**Figure 5A**). Human observers efficiently detect the presence of such a path if the elements are close together and well aligned. The performance of observers decreases if the elements are not aligned (e.g., when the elements are rotated by 90°). On the basis of these results, Field et al. (1993) proposed that oriented elements interact with each other through a "local association field": Colinearly aligned contour elements mutually excite each other and increase each other's saliency, whereas noncolinear elements inhibit each other (Ullman 1992). The local association field (shown in **Figure 5B**) formalizes the Gestalt rule of good continuation, which states that well-aligned contour elements tend to be grouped together (Wertheimer 1923, Koffka 1935; see also Kellman & Shipley 1991).

Several studies have demonstrated that interactions between the neuronal representations of contour elements take place according to the local association field, even at the level of the primary visual cortex (area V1) (Kapadia et al. 1995, Polat et al. 1998, Bauer & Heinze 2002). **Figure 5C** shows the response

of a V1 neuron that was monitored in a monkey by Kapadia et al. (1995). The response of this neuron to the appearance of a short line in its RF was augmented when a colinear line segment was placed outside the neuron's RF. The flanking line did not affect the neuron's activity when it was presented alone. As shown in **Figure 5C**, the colinear line elements had stronger effects when a background texture was added that consisted of line elements with a random orientation. In this case, the neuron's response increased further with the addition of more colinear elements. These effects, which are observed in a large fraction of neurons in area V1 (Polat et al. 1998, Kapadia et al. 2000), indicate that neurons represent the local grouping of colinearly aligned contour elements by enhancing their response.

One important feature of these and other studies is that neuronal responses can be influenced only by information outside the RF when the cell is driven by a stimulus inside the RF (Maffei & Fiorentini 1976, Allman et al. 1985, Nelson & Frost 1985, Knierim & Van Essen 1992, Kapadia et al. 1995, Zipser et al. 1996). These contextual effects are gated by appropriate RF stimulation.

Neuroanatomy

Under the assumption that feedforward connections from the LGN determine the RF of a V1 neuron, the local association field, which is a zone outside the RF, must be the result of recurrent processing. These recurrent effects can be mediated by horizontal connections within area V1, by feedback from higher visual areas, or by both (see **Figure 3C**). The topology of horizontal connections in area V1 can account for the shape of the local association field. This is illustrated in **Figure 5D** with data from a study by Bosking et al. (1997) in the visual cortex of the tree shrew. In this study, the layout of orientation columns in area V1 was first established with optical imaging and then a tracer was injected to investigate the termination pattern of the horizontal connections. Connections originating from a cortical

a
Pathfinder display

b
Local association field

c

1   2   3   RF

4   5

d

V1   V2   VM

VM

Lateral
Rostral
1 mm
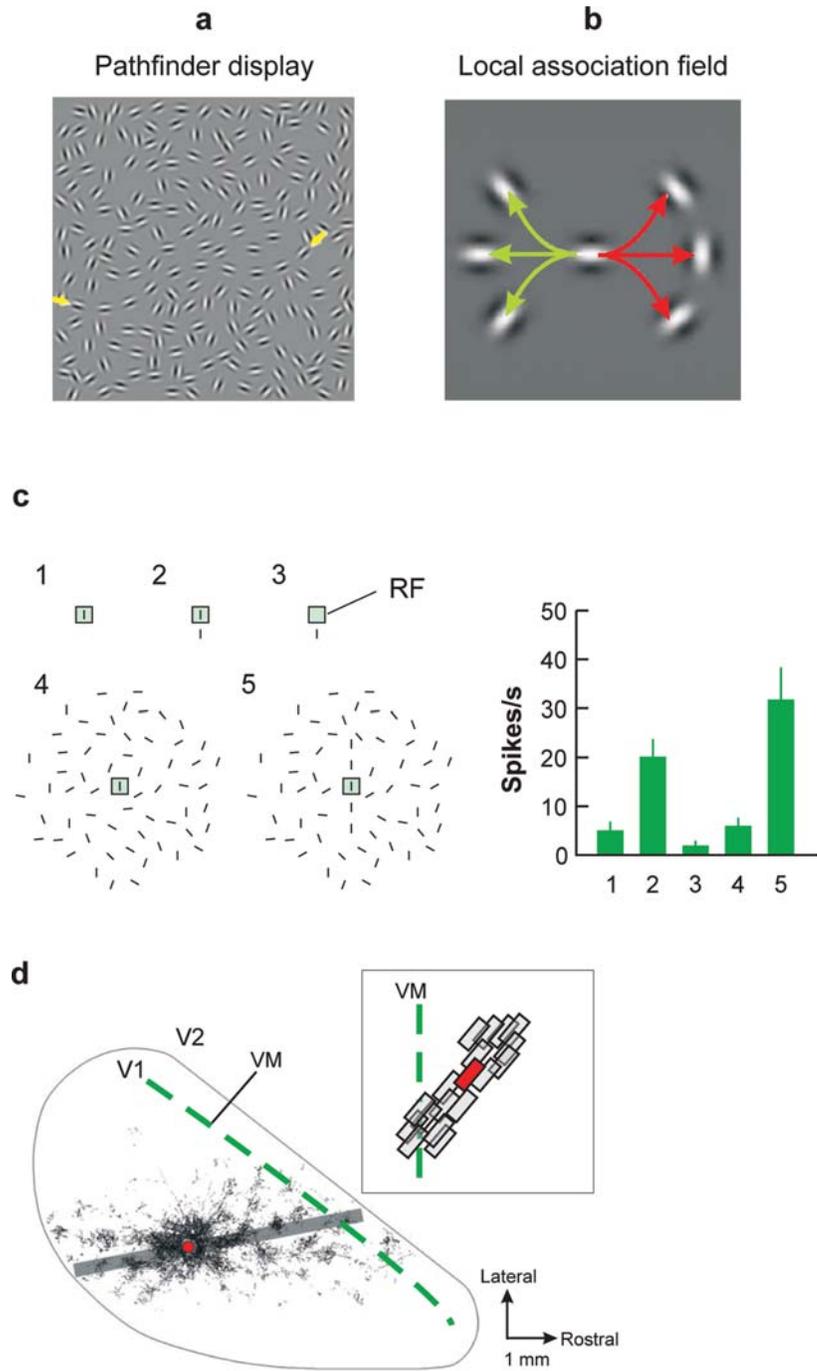
column with a particular orientation preference predominantly terminated in columns of neurons tuned to the same orientation (see also Gilbert & Wiesel 1989, Malach et al. 1993). This termination pattern may be responsible for the tendency of similar elements to group together: the Gestalt law of grouping by similarity. Moreover, most connections were made along the direction of the cortical representation of the preferred orientation (thick gray line in **Figure 5D**). Thus, horizontal connections tended to interconnect neurons that were tuned to colinear line elements, i.e., elements in each other's good continuation, in accordance with the shape of the local association field (this effect is less pronounced in the macaque monkey; see Angelucci et al. 2002). This remarkable convergence of results strongly suggests that the Gestalt laws of good continuation and similarity are caused by the selectivity of horizontal connections (Schmidt et al. 1997).

These findings do not, however, exclude a role for feedback connections in shaping the local association field. Neurons in areas V2, V4, and IT that encode the shape of an elongated curve could feed back to V1 to raise the responses evoked by contour elements that are part of the curve. For example, Li et al. (2004) demonstrated that the degree of colinear facilitation in area V1 depends on the

behavioral task. Because horizontal connections would not be affected by the monkey's task, feedback connections from higher visual areas may also play a role in shaping the local association field. Future studies could explore the relative involvement of horizontal and feedback connections in producing these effects.

SERIAL CONTOUR GROUPING

One feature of the pathfinder task is that it can be solved very efficiently; observers can rapidly distinguish between the presence and the absence of a string of colinearly aligned elements, which suggests that these groupings are formed in parallel. Not all contour grouping tasks can be solved in parallel, however. **Figure 6A** shows a curve-tracing task that was introduced by Jolicoeur et al. (1986). In this task, observers had to maintain gaze on a central fixation marker (red in **Figure 6A**) and determine whether a second dot (shown in green) appeared on the same or on a different curve. The observers' reaction time increased approximately linearly with the distance between the dots, if they fell on the same curve (**Figure 6B**), demonstrating that this grouping task requires serial processing.

Why do curve-tracing tasks require serial processing, whereas pathfinder tasks can

---

Figure 5

Local grouping of contour elements. (*A*) Pathfinder display. The string of colinearly aligned Gabor elements pops out (*yellow arrows*). (*B*) Local association field: Neurons with RFs that are well aligned mutually excite each other (*green arrows*), whereas neurons that respond to contour elements that are unlikely to belong to a single continuous curve inhibit each other (*red arrows*). From Hess & Field (1999). (*C*) Colinear facilitation in area V1. The RF of a V1 neuron is shown as a green rectangle. The cell gave a moderate response to a single contour element (stimulus 1). The response was enhanced by a colinear flanker (stimulus 2). The flanker did not evoke a response if presented alone (stimulus 3). The effect of colinear flankers was particularly pronounced if a background of randomly orientated contour elements was added (stimulus 4 and 5). Adapted from Kapadia et al. (1995). (*D*) Anatomy of horizontal connections in area V1 of the tree shrew. A tracer was injected at the red dot, and synaptic boutons were plotted. The distribution of labeled boutons was elongated along an axis that corresponds to the preferred orientation of the injection site (*gray line*). The inset shows a schematic representation of the RF at the location of tracer injection (*red*) and the RFs of cells at densely labeled cortical locations (*gray*) that have a similar orientation preference as the neurons at the injection site. Green dashed line indicates the vertical meridian (VM), which, in tree shrew, runs at a distance of ∼1 mm from the V1/V2 border (Bosking et al. 2000). Adapted from Bosking et al. (1997).
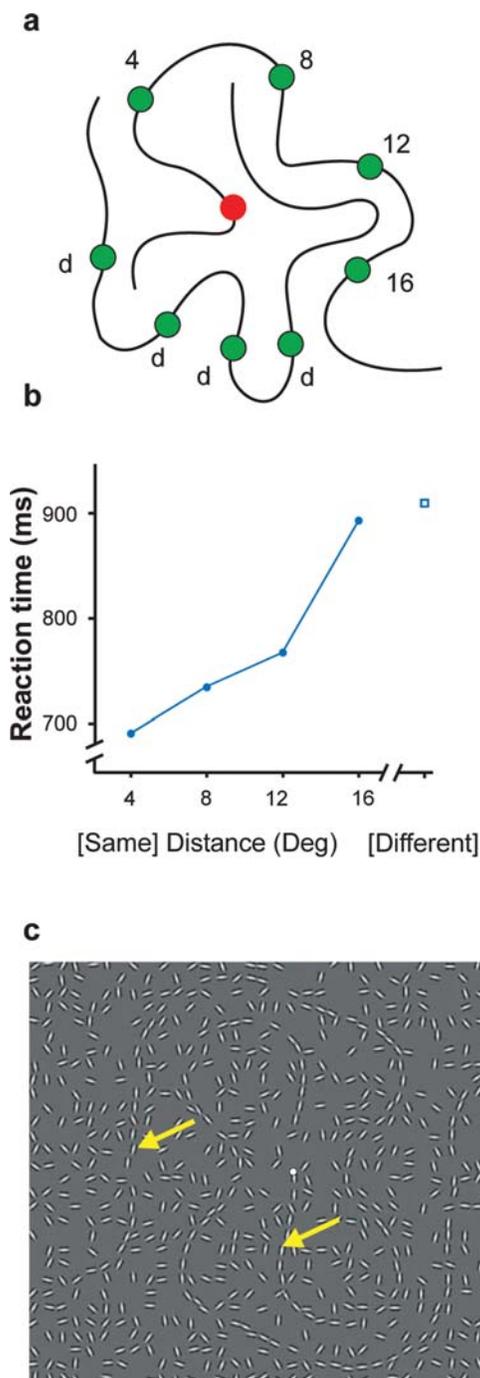
be performed in parallel? **Figure 6C** illustrates the critical difference between these two tasks (R. Houtkamp & P. Roelfsema, submitted manuscript). A parallel process that is sensitive to the local degree of colinearity can determine that the two elements indicated by arrows belong to a path. Indeed, configurations of a few colinear contour elements can be detected either as base-groupings in higher visual areas (e.g., Gigus & Malik 1991) or by local recurrent interactions in lower areas, as discussed above (**Figure 3C**, left and middle panel). However, when the task is to indicate whether the elements are on the same path in the presence of an equally colinear distractor path, many local groupings must be combined in a transitive way. This extra step apparently requires serial processing.

NEUROPHYSIOLOGY

To gain insight into the neurophysiological mechanisms underlying serial contour grouping, Roelfsema et al. (1998) trained monkeys to perform the curve-tracing task illustrated in **Figure 7A**. The monkeys had to report which of two circles was connected to a fixation marker by a target curve (T in **Figure 7A**) while they ignored a distractor curve (D). Neuronal activity was recorded

←———————————————————

Figure 6

Serial contour grouping. (*A*) Curve-tracing task. Subjects fixated the central red dot and indicated whether a second green dot was presented on the same curve or on a different curve. Only one of the green dots was presented on a given trial. (*B*) Reaction time on trials where both dots fell on the same curve increased linearly with the distance between the dots measured along the curve. Average reaction time on trials during which the dots fell on different curves is shown on the right (*open square*). From Jolicoeur et al. (1986). (*C*) Pathfinder display with two curves. A parallel process can segregate colinear elements from background elements, but the segregation of one curve from another one requires serial processing. From R. Houtkamp & P. Roelfsema, submitted manuscript.

in area V1 using chronically implanted electrodes that pick up the activity of multiple neurons in the vicinity of the electrode tip. **Figure 7A** shows the RFs of five of these multiunit recording sites relative to four stimuli. Neuronal responses were compared between complementary stimuli that differed only in their connection to the fixation point so that the target and distractor curves are switched (stimulus I is complementary to II, and III to IV). The RFs of neurons with an enhanced response are shown in yellow. Note that, at each of the recording sites, the neuronal responses evoked by the target curve were stronger than those evoked by the distractor curve. Thus, all contour elements of the target curve were labeled with an enhanced response, an outcome that is consistent with the idea that the label for incremental grouping is an enhancement of neuronal firing rates.

In **Figure 7**, the contour element in the RF of the V1 neurons is held constant, and the difference between complementary stimuli is far from the RF. This finding suggests that the modulation of neuronal activity is caused by recurrent processing. Moreover, the enhancement of responses does not occur during the initial transient response but later in time (**Figure 7B**). There appears to be a clear temporal separation between an early feedforward processing phase and a later recurrent processing phase, when the neurons are informed about whether their RF is on the target or distractor curve. The modulation of the neuronal responses in the recurrent phase is a reliable signal. Across the whole population of V1 neurons, the target curve evokes a 20%–50% stronger response than does the distractor curve. The delayed modulation of the response of some of the V1 neurons is so strong that it can be used to infer on single trials whether the RF is on the target or distractor curve. Moreover, these neurons change their activity if the monkey groups the wrong contour elements together so that they give insight in the monkey's interpretation (Roelfsema & Spekreijse 2001). A different population of V1 neurons does not discrimi-

A VERIDICAL AND AN INTERPRETATION NETWORK

In the curve-tracing task, neurons at ∼60% of the V1 recording sites reflect the attention shifts associated with recurrent processing (A-sites), and the other 40% do not (N-sites) (Roelfsema et al. 2004). This suggests that neurons at N-sites are sensitive only to information in their RF, whereas neurons at A-sites are also sensitive to the effects of recurrent connections. Neurons at A- and N-sites also differ in how they correlate with other cells. The firing rates of neurons at A-sites with nonoverlapping RFs are correlated with an average correlation coefficient of 0.4, which is further evidence that these neurons are linked by recurrent connections. In contrast, the average correlation coefficient between the firing rates of neurons at N-sites is only 0.02 (Roelfsema et al. 2004).

One tentative conclusion is that two networks coexist in area V1 (see also **Figure1B**). The first is a veridical network formed by neurons at N-sites that reliably represent the stimulus. The second is a labeling network of neurons at A-sites that take part in incremental grouping. Such a division of labor would have clear advantages. Labeling need not change the perception of low-level features that are always coded reliably at N-sites. At the same time, the labeling operation can be isolated from changes in firing rates induced by variations in low-level features. Labeled image elements can always be identified by comparing the activity of neurons at N- and A-sites if they have a similar tuning to low-level features.

nate between the target and distractor curve. These neurons may always convey a reliable representation of the low-level features in the stimulus, irrespective of task demands (see A Veridical and an Interpretation Network).

Algorithms

To understand the selectivity of the labeling process, it is useful to consider a simple algorithm, illustrated in **Figure 8**. It uses a single retinotopic area of neurons that receive feedforward connections from the retina, as well as horizontal connections from their nearest neighbors that can propagate an enhancement of the response. Suppose that the task is to group all squares connected to the red square.
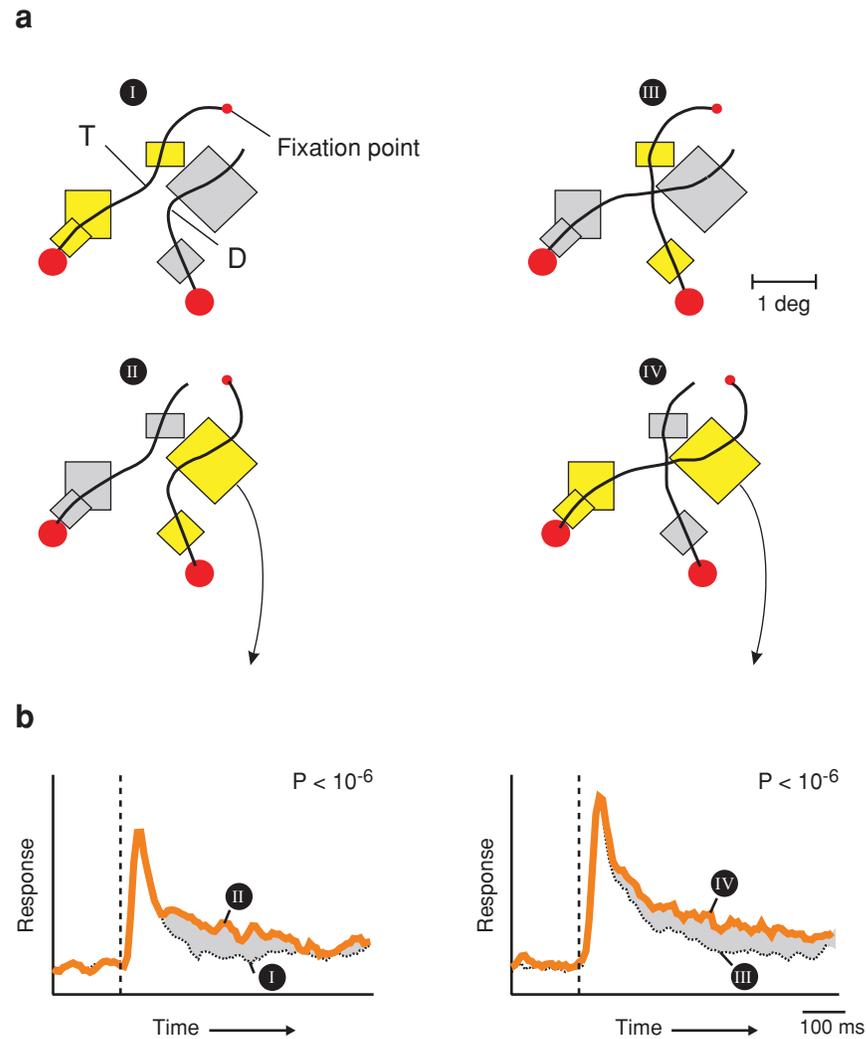
Figure 7

Neuronal activity in area V1 during serial contour grouping. (*A*) Four stimuli that contain a target curve (T) that was connected to the fixation point and a distractor curve (D) that was not. The monkey had to make an eye movement to the larger red circle at the end of the target curve. Neuronal responses were compared between complementary stimuli, which differed in their connection with the fixation point. Stimulus I is complementary to II, and III to IV. RFs are shown in yellow (*gray*) if the strength of the response was stronger (weaker) than that evoked by the complementary stimulus. (*B*) Time course of the responses at one of the recording sites. The initial transient response did not differ between complementary stimuli, but the sustained response was strongest if the RF fell on the target curve.

In the model, neurons with a red or black square in their RF are activated by feedforward connections (gray circles in **Figure 8***B*); the other neurons have only spontaneous activity. The algorithm requires that neurons are sensitive to the horizontal input only if they also receive input from feedforward connections. This type of gating is in accordance with neurophysiological data, as discussed above. Gating subdivides the connections into two
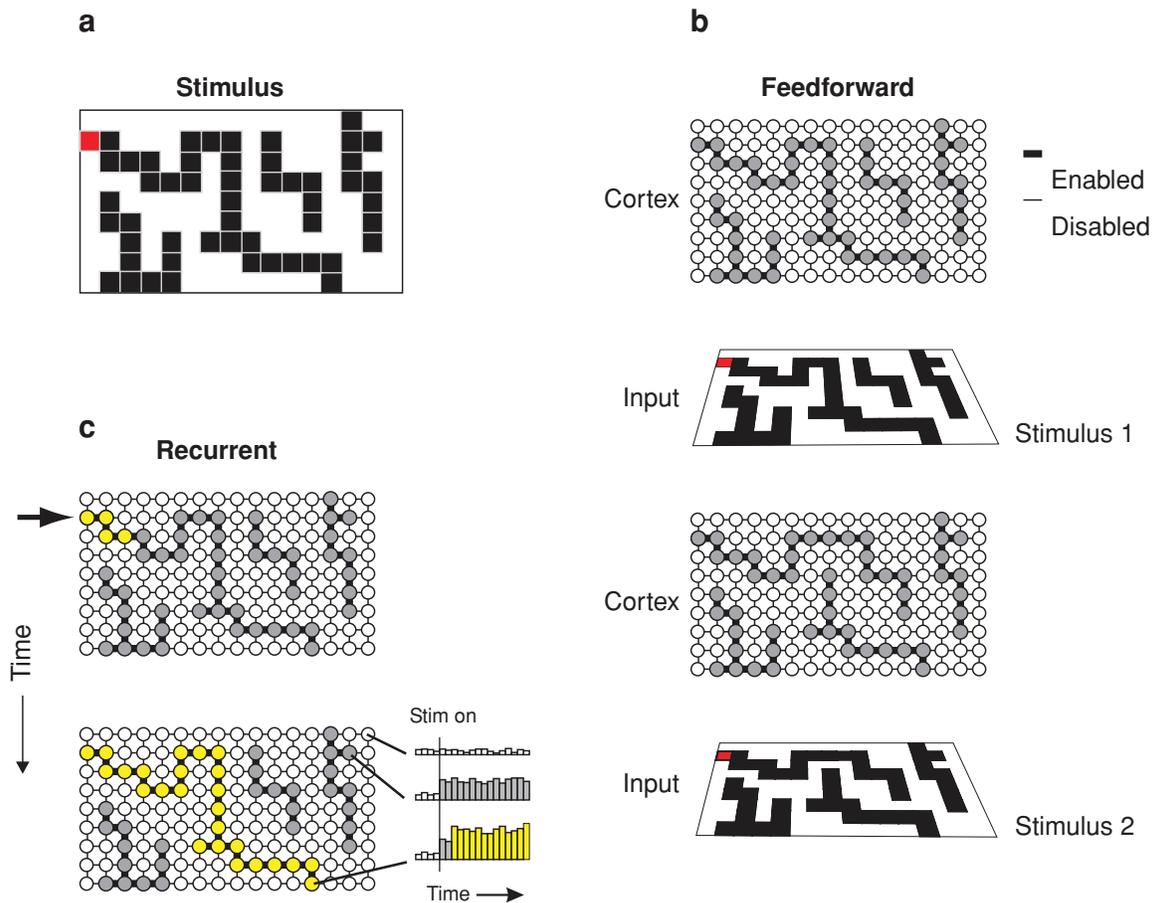
Figure 8

Algorithm for incremental grouping of connected image regions. (*A*) The task is to identify all squares that are connected to the red one. (*B*) Feedforward connections activate a subset of neurons in a retinotopic area (*grey circles*). Neurons also receive recurrent connections from their nearest neighbors; connections between active neurons are enabled (*thick lines*); the others are disabled (*thin lines*). If the input changes, different connections are enabled so that neurons that respond to connected squares are linked. (*C*) An enhanced response propagates gradually through the network of enabled connections (interaction skeleton) to make the incremental groupings explicit.

classes. The first class has an inactive neuron on one or both sides and is disabled (thin lines in **Figure 8B,C**). The second class is enabled because there is an active neuron on both sides (thick lines). I refer to this class as the "interaction skeleton." The interaction skeleton links the neurons that respond to squares that are connected to each other in the image.

Neurons in other cortical areas do not have access to the pattern of enabled con-

nections. To make the groupings explicit and accessible to these other areas, an enhanced response must spread starting at the representation of the red square until it reaches all neurons that represent other, connected squares (**Figure 8C**). The algorithm illustrates many of the key features of incremental grouping. First, it is serial because the amount of time that the label takes to reach a particular square increases linearly with its distance from

Interaction skeleton: set of enabled connections

Enabled connection: recurrent connection between two active neurons

the red square. The delay corresponds to the number of synapses that must be crossed. Second, the algorithm is transitive. When the interaction skeleton connects square 1 to square to 2 and it connects square 2 to square 3, then square 1 and square 3 are also connected. Third, it is sensitive to small changes in the input, as can be seen by comparing the two stimuli of **Figure 8B** that differ in the position of a single square. The shifted square activates a different cell and therefore enables a slightly different set of horizontal connections. One can see that the resulting change in the set of neurons linked by the interaction skeleton is much larger. The algorithm therefore naturally combines many local groupings in a transitive way to obtain a correct global grouping.

The model of **Figure 8** groups adjacent squares, but it is not sensitive to other grouping cues. It will not, for example, handle stimuli in which one curve intersects another one (see, e.g., **Figure 7A**). It is possible to accommodate other grouping cues, however, without fundamentally changing the model. The inclusion of orientation-selective neurons that are connected according to the local association field would make it sensitive to colinearity (Grossberg & Raizada 2000) and promote grouping of contour elements on both sides of an intersection. Another limitation of the simple algorithm of **Figure 8** is that it uses a single spatial scale. The processing time increases when the same image is presented at higher magnification. This effect is not observed in the performance of human observers (e.g., if we move closer to **Figure 6A**), where reaction time is approximately constant across different magnifications (Jolicoeur & Ingleton 1991). To account for this scale invariance, the model of **Figure 8** must be extended to include neurons at higher levels in the cortical hierarchy that have larger RFs (Edelman 1987, Roelfsema et al. 2000). At these higher levels, the horizontal connections link neurons with RFs that are farther apart, and this linking would allow the labeling of contour

elements to occur at higher speeds. If we go back to a low image magnification, however, the distance between the target and distractor curves decreases so that both curves fall in the same RF in higher areas, and this prohibits the selective labeling of one of the curves. In these situations, the response enhancement has to be propagated in lower areas with smaller RFs, at the cost of a reduced grouping speed.

GENERALIZATION TO OTHER FEATURE DOMAINS

So far, this review has focused on the mechanisms responsible for the grouping of contour elements into spatially extended objects. One important question is whether these ideas generalize to other feature domains and to other tasks that require perceptual grouping. Texture segregation provides one good example. In this task, a figure is segregated from a background on the basis of differences in a low-level feature, such as motion, orientation, or color (**Figure 2C**) (see e.g., Beck 1966, Julesz 1981). Image elements of the figure are grouped on the basis of the Gestalt law of similarity, while they are segregated from a dissimilar background.

Previous studies that explored the neurophysiological correlates of texture segregation showed that neurons in area V1 enhance their response when the RF lies within the figure compared with when it lies on the background (Lamme 1995, Zipser et al. 1996). The modulation of neuronal responses by figure-ground segregation has been observed for figures that were defined by various features, including orientation, motion, color, and depth. The figural response enhancement does not occur during the initial transient response in area V1 but instead at an additional delay, which suggests the involvement of recurrent connections. The response enhancement occurs first at the boundary between the figure and the background, and it then propagates toward the interior of the figure until all figural elements are labeled by an enhanced response (Lamme et al. 1999). These results indicate

that incremental grouping also occurs in texture segregation and that the proposed mechanisms generalize to perceptual grouping on the basis of a large variety of visual cues.

Binding Features and Locations

Many tasks require grouping of features that are represented in different areas of the visual cortex. Although much remains to be learned about interareal interactions, I briefly indicate how studies on spatial cuing and visual search in monkeys fit into the present scheme. The typical search task used in monkeys first presents a shape or color cue. Then a search array appears, and the monkey indicates the location of a matching item in the array. Thus, the target object is cued by a feature, and the monkey reports the object's location. The logic is reversed in many spatial cuing tasks. In this case, the monkey is first presented with a spatial cue, then an array of two or more objects appears and the animal must report a feature of the object at the cued location. Search and cuing tasks require the monkey to link the features of a target object to its location in the presence of distractors.

The incremental grouping theory proposes that the binding of features across different brain areas occurs by the enabling of interareal connections (for previous models that proposed similar ideas, see, e.g., Phaf et al. 1990, van der Velde & de Kamps 2001, Hamker 2005). **Figure 9** illustrates this idea with a highly simplified model that consists of three types of visual cortical areas: (*a*) two early visual areas (EV) where neurons are tuned to colors and shapes at a specific position; (*b*) two areas where neurons are tuned to the same features but have larger RFs (TI, translational invariance); and (*c*) an area of cortex that represents the location of salient and behaviorally relevant items, irrespective of their features, as a saliency map (Koch & Ullman 1985). Gray circles in **Figure 9** represent neurons that are activated by feedforward connections. The model uses recurrent connections between neurons that are tuned to

the same shape, color (grouping law of similarity), or location (grouping law of proximity). If only recurrent connections between active cells are enabled (thick lines), then this produces a selective linkage of neurons that respond to features of the same object. A switch in the color of the two objects activates different neurons, and changes the set of enabled connections, so that neurons that represent features of the same object are also linked for this stimulus (compare the two networks in **Figure 9**).

Now suppose that the task is a visual search and that the model has been cued to look for a red item. To implement the search, the model would first enhance the activity of neurons in TI that code red. This response enhancement then propagates through the enabled connections to neurons in EV and SaM. The neurons in SaM that enhance their response represent the location of the red object, and the search task is thereby solved. Conversely, if the task were to report the color of an object at a cued location, the model would first enhance the response of neurons in SaM that represent this location. The response enhancement would then spread in the opposite direction, through EV, to neurons in TI that code the target object's color.

Results from neurophysiological studies in monkeys are consistent with the key aspects of this scheme. During visual search, for example, neurons in area IT that code the object that the monkey is looking for in a translational invariant way enhance their response (Chelazzi et al. 1993). A similar response enhancement is observed in early visual areas (Motter 1994, Chelazzi et al. 2001, Roelfsema et al. 2003), as well as in areas of the parietal and frontal cortex, which act as saliency maps during visual search (Schall & Hanes 1993, Gottlieb et al. 1998, Schall & Thompson 1999). Similarly, the neuronal correlates of spatial cuing have been observed in many cortical areas. The representations of cued objects are enhanced in early visual areas (Treue & Maunsell 1996, Luck et al. 1997), in area IT (Moran & Desimone 1985), as well
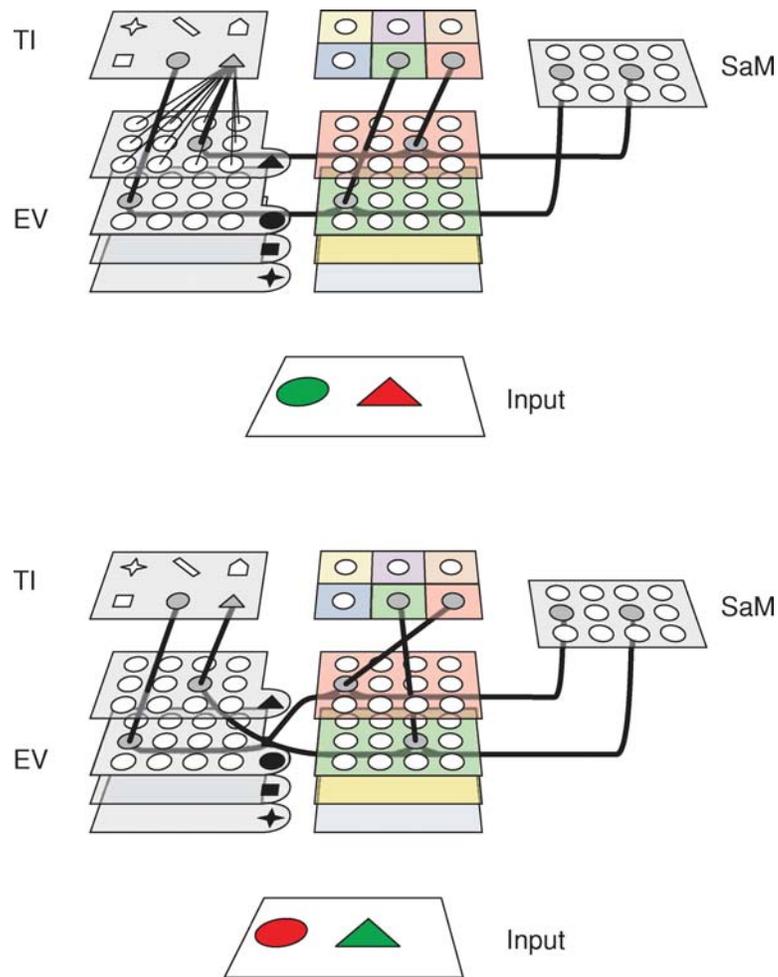
Figure 9

Incremental grouping of features represented in different areas of the visual cortex. Neurons in the two early visual areas (EV) are tuned to color and shape and have small RFs; neurons in two higher areas are tuned to the same features but have larger RFs (TI, translational invariance); and neurons in an area that functions as a saliency map (SaM) are tuned to the spatial location of objects irrespective of their features. Neurons tuned to the same location, shape, or color are linked with recurrent connections. Out of 384 connections, only 10 are enabled (*thick lines*; note that three enabled connections between cells with coinciding RFs are represented by a single curve). A few disabled connections are shown as thin lines. If the colors of the objects are switched, the pattern of active cells in EV changes, and this change enables a different set of connections. Note that for both images the interaction skeleton links neurons that respond to features of the same object.

as in areas of the parietal and frontal cortex (Everling et al. 2002, Bisley & Goldberg 2003). Thus, if the task requires the monkey to link features to locations, this linkage is achieved by labeling the respective neuronal representations with an enhanced response.

Another aspect of the model of **Figure 9** that is supported by neurophysiological data is the gating of recurrent connections by feedforward activation. A recent study by Bichot et al. (2005), for example, investigated the neuronal correlates of visual search in area V4.

When the monkeys searched for a particular target item, the V4 neurons with a matching feature preference enhanced their response. The response enhancement occurred only for neurons that were activated by the stimulus, which suggests that the effects of feedback connections, which transmit information about the target feature to area V4, are gated by the feedforward activation. An equivalent form of gating occurs in spatial cuing tasks. Neurons in early visual areas with a RF at the cued location enhance their response, but only if they are driven by the stimulus (McAdams & Maunsell 1999, Treue & Martínez Trujillo 1999). These findings concur with the idea that only a subset of the recurrent connections is enabled: those connections that link neurons that are activated by bottom-up (see also Moore & Armstrong 2003).

Taken together, the mechanisms that bind features represented in different cortical areas appear to be analogous to those involved in the grouping of image elements represented within a single retinotopic area. A single visual object is represented by an assembly of neurons that is linked by an interaction skeleton consisting of enabled connections. When the object becomes relevant for behavior, an enhanced response propagates through the interaction skeleton to make additional, incremental groupings explicit.

THE ROLE OF ATTENTION IN PERCEPTUAL GROUPING

The incremental grouping theory is a neurophysiological theory, and some caution is required when translating it into psychological terms. It is nevertheless useful to make contact with the large body of experimental work on the psychology of perceptual grouping and the role of attention therein. Base-grouping maps onto preattentive processing because it occurs rapidly and in parallel across the visual scene (Lamme & Roelfsema 2000, Hochstein & Ahissar 2002) (**Figure 3C**,

THE FEATURE INTEGRATION THEORY

In 1980, Treisman & Gelade proposed the feature integration theory (FIT). This theory claims that feature conjunctions are formed by an attentional spotlight that binds all the features at one location. The data reviewed here demonstrate that grouped features are labeled by an enhanced response in the visual cortex. These findings are consistent with the FIT because the response enhancements are correlates of visual attention (Reynolds & Desimone 1999, Kastner & Ungerleider 2000). Nevertheless, the incremental grouping theory also incorporates a number of significant deviations from the FIT:

- Many complex feature conjunctions are established without attention as base groupings.
- The FIT uses a spotlight of attention to group features at one location in space. In the incremental grouping theory, attention can also group spatially separate image elements. Attention is object-based, and it can even be directed selectively to one of two overlapping objects (Duncan 1984), which is impossible with a spotlight.

The FIT proposes that Gestalt cues are evaluated preattentively. In the incremental grouping theory, the enabling of recurrent connections also occurs in parallel during the feedforward sweep. However, to make the groupings explicit, an enhanced response must spread through the network of enabled connections, and this spread corresponds to a time-consuming spread of attention.

left). The fast extraction of elaborate feature conjunctions by the feedforward sweep may explain, for example, why subjects can rapidly determine whether a complex visual scene contains an animal (Thorpe et al. 1996, Fabre-Thorpe et al. 2001). As a result of the feedforward sweep, a subset of the recurrent connections is enabled, but they take effect only in the subsequent attentive-processing phase.

Recurrent interactions initially occur at a local scale (**Figure 3C**, middle), and this suffices for some tasks such as the pathfinder task. However, tasks that require the evaluation of transitive chains of groupings require

more time because the response enhancement must be propagated across multiple synapses (**Figure 3C**, right). Houtkamp et al. (2003) obtained direct evidence for a role of attention in Gestalt grouping. The spatial distribution of attention was investigated in subjects that had to trace a target curve connected to a fixation point. They found that the subjects initially direct their attention to the first segments of the target curve and that attention then gradually spreads across the entire curve until all contour elements are attended. The close correspondence with the neurophysiological data in monkeys suggests that the spread of attention is a correlate of the spread of an enhanced firing rate in the visual cortex. Thus, image elements that are grouped incrementally are labeled by attention. Additional support for this conclusion comes from studies demonstrating that Gestalt grouping does not occur under conditions of inattentional blindness (Mack & Rock 1998) or when attention is directed to other items in the display (Ben-Av et al. 1992). The idea that features are grouped by attention is in line with the feature integration theory of Treisman & Gelade (1980). However, the present incremental grouping theory also incorporates significant deviations from that theory (see The Feature Integration Theory).

FUTURE RESEARCH DIRECTIONS

The incremental grouping theory assigns specific roles to feedforward, horizontal, and feedback connections that can be tested in future experiments. A promising approach has been taken by studies that combine the activation or inactivation of neurons in one area with the recording of neuronal activity in another (Vanduffel et al. 1997, Hupé et al. 1998, Galuske et al. 2003, Moore & Armstrong 2003). These methods give direct insight into how cortical neurons influence other neurons in lower or higher cortical areas. These techniques could be used during perceptual grouping tasks so that the effects of activation and inactivation can also be monitored at the behavioral level (e.g., Moore & Fallah 2001). Another important avenue for future research is to assess how recurrent connections are gated by feedforward activation. Gating of recurrent connections ensures the selective linkage of neurons that represent features of the same object. This gating is an essential ingredient of perceptual grouping theories that combine computational, neurophysiological, and psychological insights into one single theoretical framework (Grossberg & Mingolla 1985, van der Velde & de Kamps 2001, Hamker 2005).

LITERATURE CITED

Albright TD, Stoner GR. 2002. Contextual influences on visual processing. *Annu. Rev. Neurosci.* 25:339–77

Allman J, Miezin F, McGuinness E. 1985. Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu. Rev. Neurosci.* 8:407–30

Angelucci A, Levitt JB, Walton EJS, Hupé JM, Bullier J, Lund JS. 2002. Circuits for local and global signal integration in primary visual cortex. *J. Neurosci.* 22:8633–46

Barlow HB. 1972. Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1:371–94

Barrett HH, Myers KJ. 2003. *Foundations of Image Science*. New York: Wiley

Bauer R, Heinze S. 2002. Contour integration in striate cortex. *Exp. Brain Res.* 147:145–52

Beck J. 1966. Effect of orientation and of shape similarity on perceptual grouping. *Percept. Psychophys.* 1:300–2

Ben-Av MB, Sagi D, Braun J. 1992. Visual attention and perceptual grouping. *Percept. Psychophys.* 52:277–94

Bichot NP, Rossi AF, Desimone R. 2005. Parallel and serial neural mechanisms for visual search in macaque area V4. *Science* 308:529–34

Bisley JW, Goldberg ME. 2003. Neuronal activity in the lateral intraparietal area and spatial attention. *Science* 299:81–86

Bosking WH, Kretz R, Pucak ML, Fitzpatrick D. 2000. Functional specificity of callosal connections in tree shrew striate cortex. *J. Neurosci.* 20:2346–59

Bosking WH, Zhang Y, Schofield B, Fitzpatrick D. 1997. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J. Neurosci.* 17:2112–27

Brincat SL, Connor CE. 2004. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat. Neurosci.* 7:880–86

Celebrini S, Thorpe S, Trotter Y, Imbert M. 1993. Dynamics of orientation coding in area V1 of the awake monkey. *Vis. Neurosci.* 10:811–25

Chelazzi L, Miller EK, Duncan J, Desimone R. 1993. A neural basis for visual search in inferior temporal cortex. *Nature* 363:345–47

Chelazzi L, Miller EK, Duncan J, Desimone R. 2001. Responses of neurons in macaque area V4 during memory-guided visual search. *Cereb. Cortex* 11:761–72

Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18:193–222

Duncan J. 1984. Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.* 113:501–17

Edelman S. 1987. Line connectivity algorithms for an asynchronous pyramid computer. *Comp. Vision Graphics Image Process.* 40:169–87

Everling S, Tinsley CJ, Duncan J. 2002. Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nat. Neurosci.* 5:671–76

Fabre-Thorpe M, Delorme A, Marlot C, Thorpe S. 2001. A limit to the speed of processing in ultrarapid visual catgorization of novel natural scenes. *J. Cognit. Neurosci.* 13:171–80

Felleman DJ, Van Essen DC. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1:1–47

Field DJ, Hayes A, Hess RF. 1993. Contour integration by the human visual system: evidence for a local "association field." *Vision Res.* 33:173–93

Fukushima K. 1988. A neural network for visual pattern recognition. *IEEE Comp.* 21:65–75

Galuske RAW, Schmidt KE, Goebel R, Lomber SG, Payne BR. 2003. The role of feedback in shaping neural representations in cat visual cortex. *Proc. Natl. Acad. Sci. USA* 99:17083–88

Gigus Z, Malik J. 1991. Detecting curvilinear structure in images. *Tech. Rep. 91/619*, Univ. Calif., Berkeley

Gilbert CD, Wiesel TN. 1989. Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J. Neurosci.* 9:2432–42

Gottlieb JP, Kusunoki M, Goldberg ME. 1998. The representation of visual salience in monkey parietal cortex. *Nature* 391:481–84

Grossberg S, Mingolla E. 1985. Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychol. Rev.* 92:173–211

Grossberg S, Raizada RDS. 2000. Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Res.* 40:1413–32

Hamker FH. 2005. The reentry hypothesis: the putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cereb. Cortex* 15:431–47

Hebb DO. 1949. *The Organization of Behavior. A Neuropsychological Theory*. New York: Wiley

Hess RF, Field DJ. 1999. Integration of contours: new insights. *Trends Cogn. Sci.* 3:480–86

Hochstein S, Ahissar M. 2002. Views from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36:791–804

Houtkamp R, Spekreijse H, Roelfsema PR. 2003. A gradual spread of attention during mental curve tracing. *Percept. Psychophys.* 65:1136–44

Hupé JM, James AC, Payne BR, Lomber SG, Girard P, Bullier J. 1998. Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394:784–87

Jolicoeur P, Ingleton M. 1991. Size invariance in curve tracing. *Mem. Cognit.* 19:21–36

Jolicoeur P, Ullman S, MacKay M. 1986. Curve tracing: a possible basic operation in the perception of spatial relations. *Mem. Cognit.* 14:129–40

Julesz B. 1981. Textons, the elements of texture perception, and their interactions. *Nature* 290:91–97

Kapadia MK, Ito M, Gilbert CD, Westheimer G. 1995. Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron* 15:843–56

Kapadia MK, Westheimer G, Gilbert CD. 2000. Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *J. Neurophysiol.* 84:2048–62

Kastner S, Ungerleider LG. 2000. Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* 23:315–41

Kellman PJ, Shipley TF. 1991. A theory of visual interpolation in object perception. *Cognit. Psychol.* 23:141–21

Knierim JJ, Van Essen DC. 1992. Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol.* 67:961–80

Kobatake E, Tanaka K. 1994. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* 71:856–67

Koch C, Ullman S. 1985. Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* 4:219–27

Koffka K. 1935. *Principles of Gestalt Psychology*. New York: Harcourt Brace

Kreiman G, Fried I, Koch C. 2002. Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proc. Natl. Acad. Sci. USA* 99:8378–83

Lamme VAF. 1995. The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.* 15:1605–15

Lamme VAF, Rodriguez-Rodriguez V, Spekreijse H. 1999. Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cereb. Cortex* 9:406–13

Lamme VAF, Roelfsema PR. 2000. The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23:571–79

Lamme VAF, Spekreijse H. 1998. Neuronal synchrony does not represent texture segregation. *Nature* 396:362–66

Leventhal AG, Thompson KG, Liu D, Zhou Y, Ault SJ. 1995. Concomitant sensitivity to orientation, direction, and color of cells in layers 2, 3, and 4 of monkey striate cortex. *J. Neurosci.* 15:1808–18

Li W, Piëch V, Gilbert CD. 2004. Perceptual learning and top-down influences in primary visual cortex. *Nat. Neurosci.* 7:651–57

Luck SJ, Chelazzi L, Hillyard SA, Desimone R. 1997. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* 77:24–42

Mack A, Rock I. 1998. *Inattentional Blindness*. Cambridge, MA: MIT Press

Maffei L, Fiorentini A. 1976. The unresponsive regions of visual cortical receptive fields. *Vision Res.* 16:1131–39

Malach R, Amir Y, Harel M, Grinvald A. 1993. Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. *Proc. Natl. Acad. Sci. USA* 90:10469–73

Marr D. 1982. *Vision*. New York: Freeman

McAdams CJ, Maunsell JHR. 1999. Effects of attention on orientation-tuning functions of single neurons in macaque area V4. *J. Neurosci.* 19:431–41

Miller EK, Gochin PM, Gross CG. 1993. Suppression of visual responses of neurons in inferior temporal cortex of the awake monkey by addition of a second stimulus. *Brain Res.* 616:25–29

Minsky ML, Papert SA. 1990. *Perceptrons*. Cambridge, MA: MIT Press

Moore T, Armstrong KM. 2003. Selective gating of visual signals by microstimulation of frontal cortex. *Nature* 421:370–73

Moore T, Fallah M. 2001. Control of eye movements and spatial attention. *Proc. Natl. Acad. Sci. USA* 98:1273–76

Moran J, Desimone R. 1985. Selective attention gates visual processing in the extrastriate cortex. *Science* 229:782–84

Motter BC. 1994. Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J. Neurosci.* 14:2178–89

Nelson JI, Frost BJ. 1985. Intracortical facilitation among co-oriented, coaxially aligned simple cells in cat striate cortex. *Exp. Brain Res.* 61:54–61

Nothdurft HC. 1992. Feature analysis and the role of similarity in preattentive vision. *Percept. Psychophys.* 52:355–75

Nowak LG, Bullier J. 1997. The timing of information transfer in the visual system. In *Cerebral Cortex*, Vol. 12, ed. KS Rockland, JH Kaas, A Peters, pp. 205–41. New York: Plenum

Oram MW, Perrett DI. 1992. Time course of neural responses discriminating different views of the face and head. *J. Neurophysiol.* 68:70–84

Oram MW, Perrett DI. 1994. Modeling visual recognition from neurobiological constrains. *Neural Netw.* 7:945–72

Palanca BJA, DeAngelis GC. 2005. Does neuronal synchrony underlie visual feature grouping? *Neuron* 46:333–46

Palmer SE. 1992. Common region: a new principle of perceptual grouping. *Cognit. Psychol.* 24:436–47

Pasupathy A, Connor CE. 2001. Shape representation in area V4: position-specific tuning for boundary conformation. *J. Neurophysiol.* 86:2505–19

Peterson MA, Harvey EM, Weidenbacher HJ. 1991. Shape recognition contributions to figure-ground reversal: Which route counts? *J. Exp. Psychol. Hum. Percept. Perform.* 17:1075–89

Phaf RH, van der Heijden AHC, Hudson PTW. 1990. SLAM: a connectionist network for attention in visual selection tasks. *Cognit. Psychol.* 22:273–341

Polat U, Mizobe K, Pettet MW, Kasamatsu T, Norcia AM. 1998. Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature* 391:580–84

Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I. 2005. Invariant visual representation by single neurons in the human brain. *Nature* 435:1102–7

Reynolds JH, Chelazzi L. 2004. Attentional modulation of visual processing. *Annu. Rev. Neurosci.* 27:611–47

Reynolds JH, Desimone R. 1999. The role of neural mechanisms of attention in solving the binding problem. *Neuron* 24:19–29

Riesenhuber M, Poggio T. 1999. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2:1019–25

Rock I, Palmer S. 1990. The legacy of Gestalt psychology. *Sci. Am.* 263(6):48–61

Roelfsema PR, Khayat PS, Spekreijse H. 2003. Subtask sequencing in the primary visual cortex. *Proc. Natl. Acad. Sci. USA* 100:5467–72

Roelfsema PR, Lamme VAF, Spekreijse H. 1998. Object-based attention in the primary visual cortex of the macaque monkey. *Nature* 395:376–81

Roelfsema PR, Lamme VAF, Spekreijse H. 2000. The implementation of visual routines. *Vision Res.* 40:1385–411

Roelfsema PR, Lamme VAF, Spekreijse H. 2004. Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nat. Neurosci.* 7:982–91

Roelfsema PR, Spekreijse H. 2001. The representation of erroneously perceived stimuli in the primary visual cortex. *Neuron* 31:853–63

Schall JD, Hanes DP. 1993. Neural basis of saccade target selection in frontal eye field during visual search. *Nature* 366:467–69

Schall JD, Thompson KG. 1999. Neural selection and control of visually guided eye movements. *Annu. Rev. Neurosci.* 22:241–59

Schmidt KE, Goebel R, Löwel S, Singer W. 1997. The perceptual grouping criterion of colinearity is reflected by anisotropies of connections in the primary visual cortex. *Eur. J. Neurosci.* 9:1083–89

Sincich LC, Horton JC. 2005. The circuitry of V1 and V2: integration of color, form, and motion. *Annu. Rev. Neurosci.* 28:303–26

Singer W, Gray CM. 1995. Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* 18:555–86

Sugase Y, Yamane S, Ueno S, Kawano K. 1999. Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400:869–73

Thiele A, Stoner GR. 2003. Neuronal synchrony does not correlate with motion coherence in cortical area MT. *Nature* 421:366–70

Thorpe S, Fize D, Marlot C. 1996. Speed of processing in the human visual system. *Nature* 381:520–22

Treisman AM, Gelade G. 1980. A feature-integration theory of attention. *Cognit. Psychol.* 12:97–136

Treisman AM, Schmidt H. 1982. Illusory conjunctions in the perception of objects. *Cognit. Psychol.* 14:107–41

Treue S, Martínez Trujillo JC. 1999. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399:575–79

Treue S, Maunsell JHR. 1996. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382:539–41

Ullman S. 1984. Visual routines. *Cognition* 18:97–159

Ullman S. 1992. Low-level aspects of segmentation and recognition. *Phil. Trans. R. Soc. London B* 337:371–79

Ullman S. 1996. *High-Level Vision*. Cambridge, MA: MIT Press

van der Velde F, de Kamps M. 2001. From knowing what to knowing where: modeling object-based attention with feedback disinhibition of activation. *J. Cognit. Neurosci.* 13:479–91

Vanduffel W, Payne BR, Lomber SG, Orban GA. 1997. Functional impact of cerebral connections. *Proc. Natl. Acad. Sci. USA* 94:7617–20

van Rullen R, Gautrais J, Delorme A, Thorpe S. 1998. Face processing using one spike per neuron. *BioSystems* 48:229–39

Vecera SP, Farah MJ. 1997. Is visual image segmentation a bottom-up or an interactive process? *Percept. Psychophys.* 59:1280–96

von der Malsburg C. 1981. The correlation theory of brain function. In *Internal Report 81–2*, pp. 1–38. Göttingen: Max-Planck-Inst. Biophys. Chem.

von der Malsburg C. 1999. The what and why of binding: the modeler's perspective. *Neuron* 24:95–104

Wertheimer M. 1923. Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung* 4:301–50

Zipser K, Lamme VAF, Schiller PH. 1996. Contextual modulation in primary visual cortex. *J. Neurosci.* 16:7376–89