Incremental grouping of image elements in vision

Pieter R. Roelfsema · Roos Houtkamp

Abstract One important task for the visual system is to group image elements that belong to an object and to segregate them from other objects and the background. We here present an incremental grouping theory (IGT) that addresses the role of object-based attention in perceptual grouping at a psychological level and, at the same time, outlines the mechanisms for grouping at the neurophysiological level. The IGT proposes that there are two processes for perceptual grouping. The first process is *base grouping* and relies on neurons that are tuned to feature conjunctions. Base grouping is fast and occurs in parallel across the visual scene, but not all possible feature conjunctions can be coded as base groupings. If there are no neurons tuned to the relevant feature conjunctions, a second process called *incremental grouping* comes into play. Incremental grouping is a time-consuming and capacity-limited process that requires the gradual spread of enhanced neuronal activity across the representation of an object in the visual cortex. The spread of enhanced neuronal activity corresponds to the labeling of image elements with object-based attention.

P. R. Roelfsema · R. Houtkamp
Department of Vision and Cognition,
The Netherlands Institute for Neuroscience,
Royal Academy of Arts and Sciences,
Amsterdam, the Netherlands

P. R. Roelfsema
Department of Integrative Neurophysiology, Center for
Neurogenomics and Cognitive Research, Vrije Universiteit,
Amsterdam, The Netherlands

R. Houtkamp
Department of Cognitive Biology, Otto-von-Guericke
Universitaet,
Leipzigerstrasse 44, Haus 91,
39120 Magdeburg, Germany

P. R. Roelfsema (✉)
Netherlands Institute for Neuroscience,
Meibergdreef 47,
1105 BA Amsterdam, the Netherlands
e-mail: p.roelfsema@nin.knaw.nl
URL: http://www.nin.knaw.nl/research_groups/roelfsema_group/

Introduction

Vision starts with a fragmentation of the visual scene. Neurons in low-level areas of the visual cortex extract the low-level features present in their small receptive fields, in parallel across the visual field. The representation of the visual scene in early visual areas therefore consists of a set of image fragments, like short contour elements and small texture patches. This is not how we perceive the visual world. The world with which we interact consists of coherent and unitary objects that comprise many features, rather than an unstructured collection of localized image fragments (Wertheimer, 1923). Thus, our visual system must have powerful mechanisms for grouping all the elements of an object together and for segregating them from other objects and the background. This process of perceptual grouping is important for object recognition and for the interaction with the objects that surround us. If we want to grasp an object, we have to know which parts belong to it and which parts do not. Somehow, our visual system synthesizes chairs, tables, trees, and animals from all the image fragments that are represented in early visual areas and assigns colors, motions, and depth structure to

these objects. Previous theoreticians convincingly pointed out that this is no small achievement, especially because we can even perceive new objects that consist of feature constellations that we never saw before (see also Treisman, 1996; von der Malsburg, 1999; cf. Singer & Gray, 1995).

The problem of perceptual grouping has intrigued psychologists for a long time, and the topic has an extended history. In the first half of the previous century, Gestalt psychologists described many of the rules that determine what groups with what in visual perception (Koffka, 1935; Wertheimer, 1923). A few of their Gestalt laws are illustrated in Fig. 1. One example is the law of *similarity* (Fig. 1a), stating that similar elements in a visual scene tend to be grouped in perception. Other Gestalt laws are the laws of *proximity* (Fig. 1b), implying that nearby elements are grouped; *connectedness* (Fig. 1c) suggesting that connected elements are grouped; *good continuation* (Fig. 1d), stating that well-aligned contours are assigned to the same perceptual object; and *common fate* (Fig. 1e), holding that image elements moving in the same direction tend to be grouped (Koffka, 1935; Kubovy, Holcombe, & Wagemans, 1998; Rock & Palmer, 1990; Wertheimer, 1923). This list was not final, since more recent work has added new grouping cues, including the *common region* rule stating that elements that are part of the same region tend to be grouped (Fig. 1f; Palmer, 1992).

Most previous theories have assumed that these Gestalt criteria are evaluated preattentively, by an unlimited capacity mechanism (Bergen & Julesz, 1983; Julesz, 1981; Neisser, 1967; Treisman & Gelade, 1980; Treisman & Gormican, 1988). There is experimental evidence to

support this claim, since grouping on the basis of Gestalt criteria indeed occurs, under some conditions, in parallel across the visual scene. In *pathfinder studies* (Fig. 2a), for example, observers see Gabor elements. Their task is to detect a string of elements that are aligned collinearly to form a curved path (Field, Hayes, & Hess, 1993; Kovács & Julesz, 1993). There are no simple cues to distinguish the elements of the path from the background elements. The percept of the path therefore derives from a process that integrates the relative orientation of the elements along the path. Nevertheless, the path appears to pop out (Field et al., 1993) and even influences perception of other image elements inside the region bounded by the contour if it is a closed contour (Kovács & Julesz, 1994), which suggests that grouping on the basis of good continuation occurs in parallel across the visual field. Palmer and Rock (1994) expressed a similar view about connectedness as a grouping cue, suggesting that the grouping of connected image elements occurs in parallel as one of the first steps in the analysis of the visual scene.

However, there are other conditions where the grouping of image elements on the basis of connectedness and good continuation is associated with substantial delays (Crundall, Dewhurst, & Underwood, 2008; Jolicoeur, Ullman, & MacKay, 1986, 1991; Pringle & Egeth, 1988; Roelfsema, Scholte, & Spekreijse, 1999). This occurs, for example, if stimuli consist of two curves and subjects have to judge whether contour elements belong to the same curve. The fishermen in Fig. 2b have to perceptually group the contour elements of their line to see who will catch the big fish. All the elements of the lines are related to each other by Gestalt grouping cues since they are locally collinear and connected—that is, they are in each other's good continuation. By studying a laboratory version of this task, Jolicoeur et al. (1986, 1991) demonstrated that the processing time increases linearly with the length of the lines. Grouping short lines requires tens of milliseconds
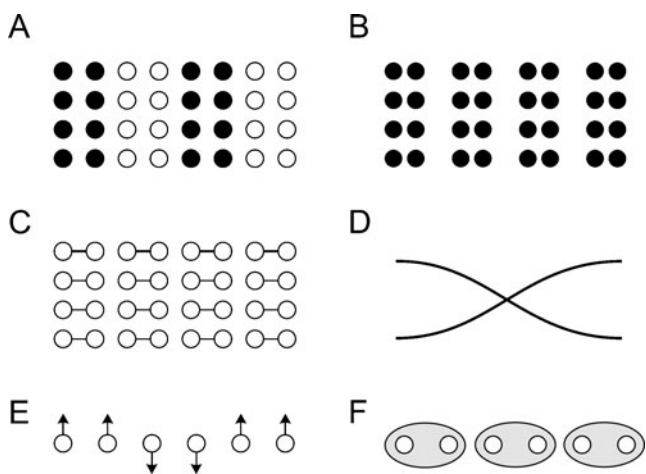

Fig. 1 Examples of Gestalt principles. **a** Grouping by similarity; note that circles with a similar color tend to form perceptual groups. **b** Proximity: Nearby image elements are grouped together. **c** Connectedness: Connected elements are grouped in our perception, **d** Good continuation: Line elements in each other's good continuation are grouped. **e** Common fate: Grouping of image elements that move in the same direction. **f** Grouping by common region. Panels a–d have been adapted from Rock and Palmer (1990)
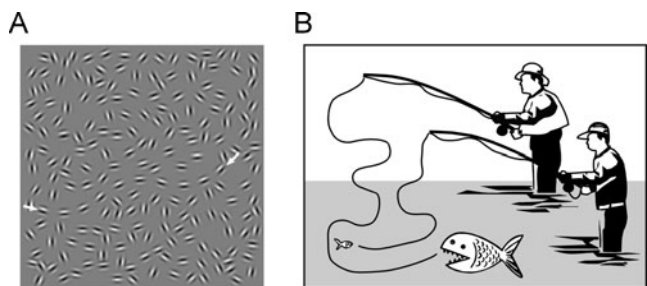

Fig. 2 Parallel and serial contour grouping. **a** Example of a pathfinder display. White arrows mark a subset of the Gabor patches that are collinearly aligned along a curved path (from Hess & Field, 1999). **b** A situation that requires serial contour grouping. To see who will catch the big fish, the visual system has to group together the contour elements that belong to one of the lines. Processing time in this task increases linearly with the length of the line

(Crundall, Cole, & Underwood, 2008; Pringle & Egeth, 1988), but the delays add up to hundreds of milliseconds for longer curves (Jolicoeur et al., 1986), implying that Gestalt criteria are not invariably evaluated by an unlimited capacity mechanism.

The IGT was inspired by these seemingly discrepant findings and aims to explain why perceptual grouping is sometimes a parallel process and why it requires serial processing in other situations. The theory relates neurophysiological mechanisms for perceptual grouping to preattentive and attentive psychological processes, addressing the time constraints and capacity limits that are observed in some but not all grouping tasks. As neuronal codes for grouping, we will consider, first, the roles of neurons that are selective for feature conjunctions and complex feature constellations (Ghose & Maunsell, 1999; Riesenhuber & Poggio, 1999a). These neurons compute what we call *base groupings* in parallel across the visual scene, and we will argue that this process maps onto preattentive vision (Fig. 3). Second, we will consider the propagation of an enhancement of neuronal firing rates (Roelfsema, 2006). This *incremental grouping* process has a limited capacity, is time consuming, and corresponds to the spread of object-based attention across the representation of a perceptual object. In the last part of the article, we will also touch upon the previously proposed role of neuronal synchrony in binding (e.g., Singer & Gray, 1995; Watt & Phillips, 2000), but we suggest that the experimental evidence does not strongly support the involvement of synchrony in perceptual grouping operations.

In what follows, we will use the term *grouping* for the processes that delineate the features that belong to the same object, such as its color and shape, and also for the processes that identify the various image elements that belong to an object. Previous authors have used the term *binding* for the same processes. Because the IGT links levels of description, we will have to switch back and forth between neurophysiology and perceptual psychology. The IGT also proposes a number of computational principles that underlie the specificity of perceptual grouping, which could inspire more specific implementations as neural network models or in machine vision. However, in its present form, the theory is not sufficiently detailed to address the complex computational problem of finding the best way to parse a visual image given a set of local consistency constraints (for computational approaches, see, e.g., Barbu & Zhu, 2003; Borenstein & Ullman, 2008; Sharon, Galun, Sharon, Basri, & Brandt, 2006).

The second section of the article will describe the core assumptions of the IGT that are inspired by neurophysiological findings. In a previous article (Roelfsema, 2006), we outlined some of these neurophysiological mechanisms, but the present article is the first to present the IGT compre-
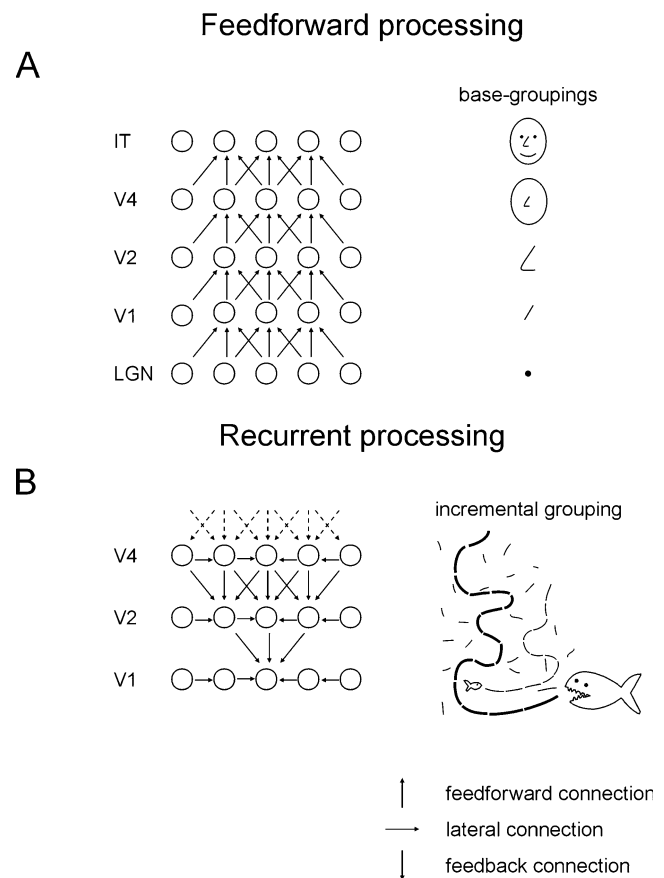

Fig. 3 Hierarchical organization of the visual cortex. **a** Neurons in lower areas of the visual cortex such as the primary visual cortex (area V1) code simple features. Feedforward connections propagate this information to higher areas V2, V4, and IT. Neurons in these higher areas are tuned to more complex features and feature conjunctions. These feature conjunctions are called *base groupings*. **b** Recurrent processing involves feedback connections that propagate information from higher to lower areas and lateral connections that interconnect neurons in the same area. The recurrent connections propagate an enhancement of neuronal firing rates to label all the neurons that respond to features of the same object (highlighted contour elements on the right). This labeling process is called *incremental grouping*

hensively, as a number of conjectures. In the third section, we then use the theory, for the first time, to reconcile apparently conflicting findings on the efficiency of grouping, thereby establishing new connections between the neurophysiology of perceptual grouping and the large psychological literature on this topic. We will discuss tasks that probe binding and object perception, but we will initially stay away from visual search and texture segregation, tasks that have been used to measure the efficiency of feature binding in previous work (e.g., Treisman & Gelade, 1980; Wolfe, 1994). We will postpone a discussion of these tasks to the fourth section, which specifies similarities and differences with previous theories of perceptual grouping. We will suggest that the neuronal mechanisms for search and texture segregation are more complex than may have

been anticipated in these earlier studies. Finally, the fifth section makes a number of new predictions that could be exploited to test the theory.

Neuronal mechanisms for perceptual grouping

The IGT proposes that there are two distinct mechanisms for perceptual grouping. The first is *base grouping*. In the visual cortex, base groupings are coded by specialized neurons. The computation of these groupings relies on the selectivity of feedforward connections that propagate activity from lower to higher areas of the visual cortex (Fig. 3a). The second type of grouping is called *incremental grouping*.[1] It is required for groupings that are not coded as base groupings. Incremental grouping relies on feedback connections that run from higher to lower visual areas, as well as on lateral connections between neurons in the same area. These *recurrent* connections propagate a neuronal response enhancement in order to label all the neurons that code image elements that belong to a single object (Fig. 3b). This labeling operation is associated with processing delays, and incremental grouping is, therefore, a serial process.

Conjecture 1: Two mechanisms for grouping There are two forms of grouping: (1) base grouping, the rapid activation of neurons tuned to feature conjunctions, and (2) incremental grouping, the time-consuming spread of an enhanced neuronal response.

Base grouping: activation of tuned neurons

Base grouping depends on the tuning of individual neurons to feature conjunctions. Neurons in early visual areas respond selectively to relatively simple features, such as the orientation of a line element (Fig. 3a). Orientation is usually considered to be a basic feature that is extracted before grouping operations come into play. However, a line element is also a grouping of simpler elements (e.g., pixels) that are aligned in a specific configuration. It is not easy to draw a line between what is a shape feature and what is a conjunction of shapes. If neurons are tuned to shapes, we call these shapes *base groupings*. In addition to their tuning to shapes, many neurons in early visual areas are also tuned to other features, such as colors and movement directions (Leventhal, Thompson, Liu, Zhou, & Ault, 1995). A neuron tuned to a red horizontal line represents a conjunction

---

[1] The terms *base grouping* and *incremental grouping* are based on a distinction initially made by Ullman (1984) between an early visual representation and a representation that can be modified by visual routines.

between these two features, and this is another example of base grouping. An important implication is that the representation of some feature conjunctions is not fundamentally different from the representation of single features.

Neurons in higher area neurons are tuned to more complex feature conjunctions (Fukushima, 1980; Riesenhuber & Poggio, 1999b). Neurons in the inferotemporal cortex, for example, code specific configurations of contour elements that form a shape (Brincat & Connor, 2004, 2006; Kayaert, Biederman, Op de Beeck, & Vogels, 2005; Tanaka, 1993). Consider the neurons in this brain region that are tuned to the shape of a face: Some of these cells are activated only if a number of face components, such as the mouth, eyes, and nose, are seen in their correct relative positions (Kobatake & Tanaka, 1994; Tsao, Freiwald, Tootell, & Livingstone, 2006). The neuron's activation implies detection of the face components as a perceptual group (analogous to a group of aligned pixels that are detected as a line element). Barlow (1972) called these highly selective neurons "cardinal cells." They are also known as grandmother cells, and these studies, together with recent findings of neurons tuned to specific individuals like Jennifer Aniston and Bill Clinton (Kreiman, Fried, & Koch, 2002; Quiroga, Reddy, Kreiman, Koch, & Fried, 2005) provide compelling evidence that cardinal cells exist.

Base groupings are extracted rapidly after the presentation of a visual image, in early visual areas (Celebrini, Thorpe, Trotter, & Imbert, 1993), as well as in higher visual areas (Hung, Kreiman, Poggio, & DiCarlo, 2005; Oram & Perrett, 1992; Sugase, Yamane, Ueno, & Kawano, 1999). The tuning to these feature conjunctions emerges as soon as the neurons are activated by a newly presented stimulus. The implication is that base grouping mainly reflects the selectivity of feedforward connections, because these connections provide the shortest route from the retina to any particular area of the visual cortex (Lamme & Roelfsema, 2000; Oram & Perrett, 1992; Thorpe, Fize, & Marlot, 1996). The fast emergence of tuning is incompatible with a major role for recurrent pathways that involve lateral connections and feedback connections, since these are associated with additional synaptic and axonal conduction delays. We will use the term *base representation* when we refer to the pattern of neuronal activity that is evoked by the selectivity of feedforward connections (Roelfsema, Lamme, & Spekreijse, 2000; Ullman, 1984). The set of features that are coded as base groupings is large because it corresponds to the features for which tuned neurons are found. It includes neurons tuned to properties of contours such as orientation (Livingstone & Hubel, 1987) and curvature (Brincat & Connor, 2004), surface properties such as color (Zeki, 1983) and texture (Komatsu & Ideura, 1993), and other features such as motion (Albright & Stoner, 2002) and shape (Tanaka, 1993).

And yet, there are limits to the number of groupings that can be coded in the base representation. In higher areas, receptive fields are larger, so that multiple objects fall into one receptive field. The representations of these objects compete with each other through mutual inhibition (Desimone & Duncan, 1995), and hence the depth of processing—that is, the number of computed base groupings—depends on the distance between an object and other objects in the surround. These inhibitory interactions take place on such a fast time scale that they curtail the initial wave of feedforward processing (Knierim & Van Essen, 1992; Miller, Gochin, & Gross, 1993; Sheinberg & Logothetis, 2001). Moreover, if multiple objects fall into these larger receptive fields, their features can become mingled, which is presumably the cause of crowding (Pelli, Palomares, & Majaj, 2004).

There also is a more fundamental limit on the number of base groupings, because there are more objects possible than there are neurons available in the visual cortex (Engel, König, Kreiter, Schillen, & Singer, 1992; von der Malsburg, 1999; von der Malsburg & Schneider, 1986). This problem can be solved by coding objects as patterns of activity across a number of neurons. An elongated curve, for example, can be coded as a collection of contour elements, even if there is no neuron that is tuned to the overall shape of the curve. Such a distributed representation has a number of virtues. First, it is efficient, because neurons can participate in the representation of many objects that share a particular feature. Second, objects that were not encountered previously can be coded as a new pattern of activity across the existing neurons (see also Singer & Gray, 1995). However, there is a disadvantage associated with distributed representations, called the *binding problem*: In the presence of multiple objects, information is lost about whether features belong to the same object or to different objects (Engel et al., 1992; Treisman & Gelade, 1980; von der Malsburg, 1999; von der Malsburg & Schneider, 1986). If there are multiple elongated curves, for example, neurons that code the contour elements will all be active, but this pattern of activity does not reveal which contour elements belong to the same curve (as in Fig. 2b). These situations therefore require incremental grouping.

Incremental grouping: grouping by labeling

Incremental grouping is necessary if feature conjunctions have to be established that are not coded as base groupings. The central idea is that neurons that respond to features that are grouped incrementally are *labeled* in the visual cortex with an enhanced neuronal response. This idea is illustrated in Fig. 3b, where a collection of contour elements that belong to one of the lines has been highlighted. The central hypothesis is that these contour elements are grouped together in perception when the neurons in the visual cortex that represent them enhance their firing rate.

Conjecture 2: Identity of the label Neurons coding features that are grouped incrementally enhance their response.

Why does the firing rate enhancement spread to neurons that represent image elements that belong to a single object, and not to image elements that belong to other objects? An important feature of perceptual grouping is that it is a *transitive* process. Transitivity means that if an image element 1 is grouped with element 2 and if 2 is grouped with 3, then 1 is also grouped with 3 (see Fig. 4a). Image elements can thus be grouped indirectly, through a chain of local groupings. To start with a simple example, we will first indicate how the IGT accounts for the grouping of image elements on the basis of connectedness. Figure 4a shows an image and a network of neurons that could be situated in an early area of the visual cortex. When the image is presented to the network, some neurons have an image element in their receptive field and are activated (gray circles) by feedforward connections. Other neurons do not have an image element in their receptive field and remain silent (white circles). At this stage (base representation), the pattern of activity represents a collection of image elements, but it does not reveal which elements belong to the same object. To compute the incremental groupings, a subset of the neurons will have to propagate an enhancement of their firing rate, but this propagation should occur only among neurons that respond to connected contour elements. The theory makes two assumptions to equip the labeling process with the required selectivity and transitivity. The first is a specific topology of the connections that propagate the neuronal response enhancement. Only neurons that are tuned to features that are likely to belong to the same object should be interconnected. Thus, to implement the detection of connectedness, lateral connections (short lines between the circles in Fig. 4a) interconnect neighboring neurons that would be activated by pixels that are directly connected to each other in the visual scene. The second assumption is an interaction between the base representation and the label-spreading process: Label spreading is permitted only between neurons that are activated by feedforward connections (i.e., between the gray circles), implying a multiplicative interaction between the feedforward connections and the recurrent lateral connections. According to the assumption, lateral connections will change a response of, for example, 40 spikes/s into one of 60 spikes/s but will not increase the response of inactive neurons. This multiplicative effect subdivides lateral connections into two classes. The first class of connections is *enabled* because there is an active neuron on both sides (thick lines on the right in Fig. 4a), and these connections could spread an increase in activity. The

second class of connections is *disabled* (thin lines) because neurons on one or both sides are silent. We refer to the set of enabled connections as the *interaction skeleton*. Linking is not an active process but, rather, the direct consequence of the activity pattern produced by feedforward processing. It can be seen that the interaction skeleton links only neurons that respond to pixels that are directly or indirectly connected to each other in the image. Linkage by the interaction skeleton thus ensures transitivity of the grouping process. It can be seen in Fig. 4 how the base representation constrains the label-spreading process by enabling some connections and disabling others. The two images in Fig. 4a differ in the location of only a single square (labeled with "2"). The change of this square between pictures disables two connections and enables two other connections, so that

squares 1 and 3 are either indirectly connected or disconnected. Due to the transitivity, small changes in the input can cause large changes in the set of neurons that are linked by the interaction skeleton.

Conjecture 3: Enabling of connections The base representation enables a set of connections: the connections between neurons that are activated by feedforward connections. Enabled connections can propagate the enhanced neuronal response.

An extra processing step is necessary to make incremental groupings explicit and determine which image elements belong to the same overall shape, because neurons

Fig. 4 A mechanism for incremental grouping on the basis of connectedness. **a** Left panels show two input patterns; right panels show the representation of these patterns in an early area of the visual cortex. Every circle denotes a neuron. Neurons that are activated by feedforward connections (the base representation) are shown in gray; neurons that remain silent are shown in white. Lines between the neurons indicate horizontal connections between neurons that respond to neighboring pixels. Connections between active neurons are enabled (thick lines); the other connections are disabled (thin lines). Note that neurons that respond to pixels that belong to the same object are linked by a chain of enabled connections. **b** During incremental grouping, an enhancement of neuronal firing rates (shown in black) spreads through the (enabled) recurrent connections between activated neurons to make the additional, incremental grouping explicit. **c** At a psychological level of description, incremental grouping corresponds to the selective spread of attention across elements that belong to the same object
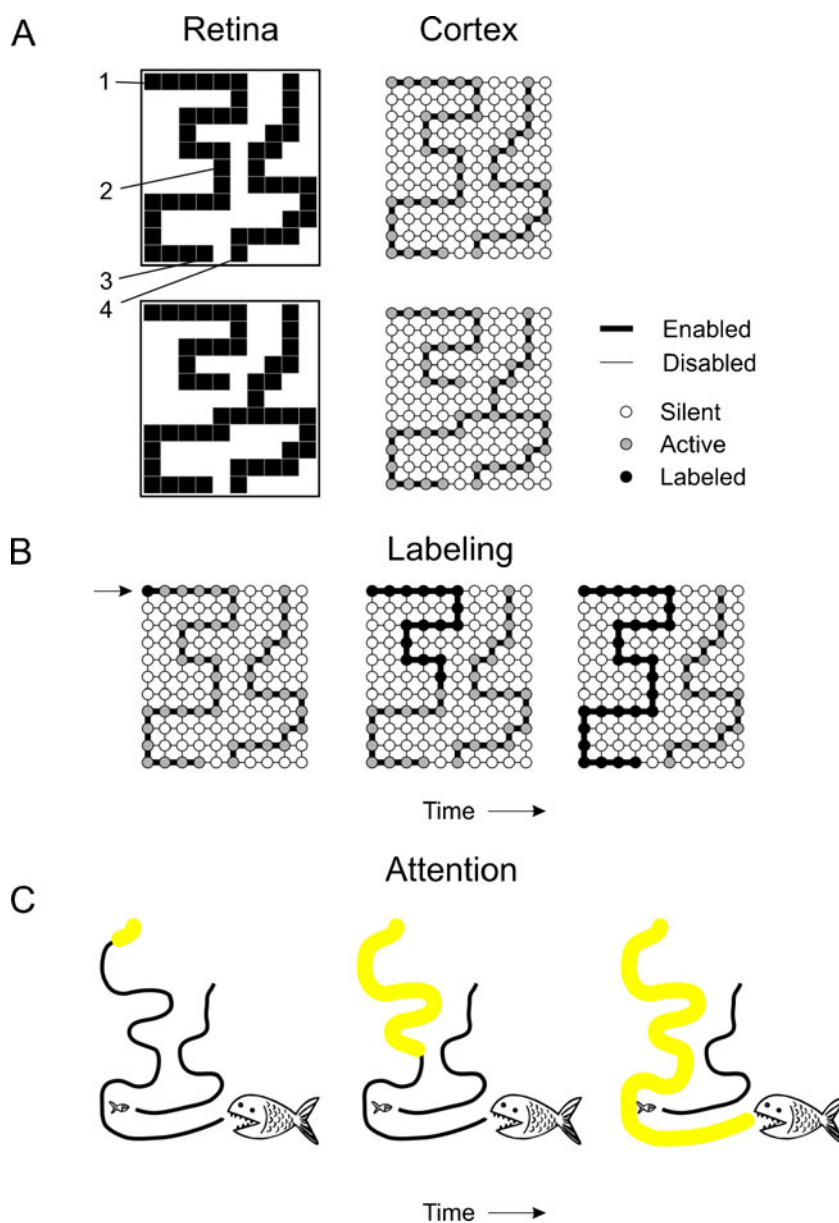
do not have access to the overall shape of the interaction skeleton. Neurons receive information only about the activity of other cells, but not about a (transitive) pattern of enabled connections. This distinction is important. The neurons that respond to connected pixels are initially only locally *linked*, but these pixels are not yet transitively *grouped*—that is, available for report. To make these latent groupings accessible to other neurons, an enhanced firing rate has to spread through the network of enabled connections. Figure 4b illustrates how, during incremental grouping, the enhanced firing rate starts to spread from one of the activated neurons so that it eventually highlights the representation of an entire connected image component. It follows that incremental grouping is a serial process. The buildup of processing delays during the spread of the rate enhancement produces a linear increase in reaction time with increases in the length of the curve and corresponds to the serial spread of attention at the psychological level (Fig. 4c), as is indeed observed experimentally (Houtkamp, Spekreijse, & Roelfsema, 2003; Jolicoeur et al., 1986).

The elements that are labeled by the enhanced neuronal response are segregated from the elements that are not labeled, so that grouping and segregation are two sides of the same coin. In some cases, incremental grouping can even override local cues for segregation. This can be seen in Fig. 4a, where squares 3 and 4 are separated by white squares that provide (local) evidence for segregation. However, in the lower panel of Fig. 4a, these squares are linked through a detour, so that incremental grouping could make their linkage explicit.

Conjecture 4: Linking and grouping The enabling of connections is called linking and occurs in parallel across the visual scene. An enhanced response has to be propagated across the enabled links to make incremental grouping explicit so that it is available for report. This propagation is a serial, time-consuming process.

The neurophysiology of incremental grouping

Neurophysiological data support the distinction between a feedforward (base grouping) and recurrent (incremental grouping) processing phase. Roelfsema, Lamme, and Spekreijse (1998) trained monkeys to carry out the contour grouping task illustrated in Fig. 5a. The animals had to look at a fixation point in a display with two curves and two larger circles. One of these curves was a target curve that connected the fixation point to one of the circles, while the other one was a distractor, not connected to the fixation point. The animal's task was to make an eye movement to the circle at the other end of the target curve—that is, the

circle that was connected to the fixation point. The initial V1 responses that were triggered at a latency of approximately 40 ms signaled the appearance of a contour element in the neurons' receptive field but did not convey information about the identity of the target curve (Fig. 5b, blue bar in Fig. 5c). However, after an additional delay, the V1 neurons started to enhance their response if their receptive field fell on the target curve (stimulus II in Fig. 5a), relative to when it fell on the distractor curve (stimulus I). Also note that the neurons with a receptive field at the beginning of the target curve (RF1) enhanced their response at an earlier point in time than did the neurons with a receptive field further along the curve, in accordance with a gradual spread of the enhanced response along the target curve (Fig. 5b). In a recent analysis, we found that it takes 50 ms, on average, for the enhanced activity to spread from one receptive field to the next abutting but nonoverlapping receptive field (Pooresmaeili & Roelfsema, manuscript in preparation).

As predicted by the IGT, the neuronal activity in the primary visual cortex (area V1) during the contour grouping task is characterized by two distinct processing phases. The initial responses caused by the feedforward connections code the contour elements and are selective for their orientation (Celebrini et al., 1993; Lamme, Rodriguez-Rodriguez, & Spekreijse, 1999) (blue bar in Fig. 5c). This is followed by a phase where lateral and feedback connections propagate an enhanced response along the target curve in order to compute the incremental groupings (yellow bar in Fig. 5c). The contour grouping task is solved when this enhanced response reaches the circle at the end of the target curve so that it can be selected for an eye movement.

A veridical and a labeling network

Roelfsema, Lamme, and Spekreijse (2004) observed that the response enhancement during the contour grouping task shown in Fig. 5 occurred at 60% of the V1 recording sites (A-sites, or attention sites; left panel in Fig. 5c). At the other 40% of the recording sites, the neuronal responses did not distinguish between target and distractor curve (N-sites, or nonattentional sites; right panel in Fig. 5c). These A- and N-sites did not form a clear dichotomy but should be considered as the extremes of a continuum in the strength of attentional signals across neurons. N-neurons are hardly influenced by attention, so that they can form a *veridical network* that codes features (and base groupings) reliably, whereas A-neurons can *label* image elements for incremental grouping. Such a division of labor between N- and A-neurons has a number of advantages. First, labeling can occur without changes in the perception of low-level features such as orientations or contrasts, which are always coded reliably at N-sites. This may explain why shifts of
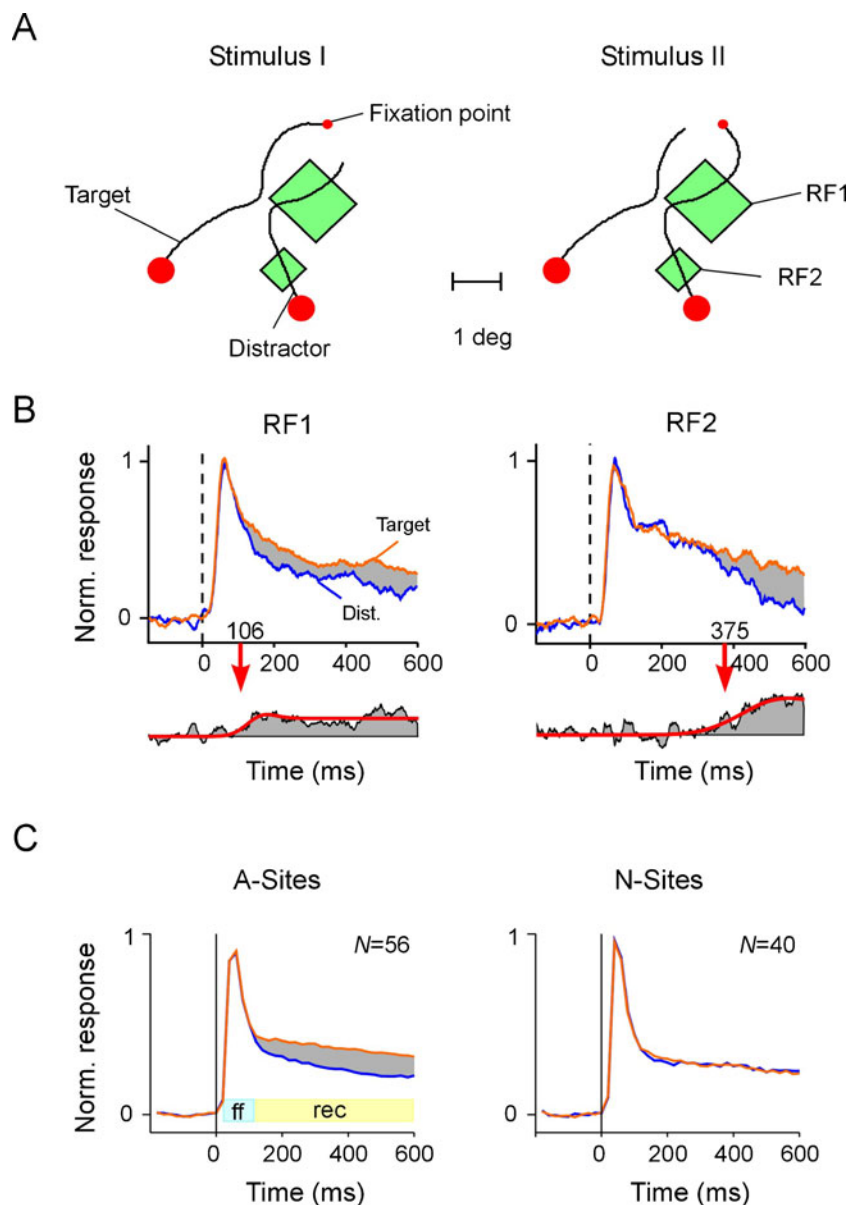
Fig. 5 Neuronal activity during incremental grouping. **a** Monkeys were tested in a task where they had to look at a fixation point while two curves and two circles appeared on the screen. One of the curves (target curve) was connected to the fixation point. The monkey had to group all the contour elements of this curve to locate a larger circle at the other end, which was the target for an eye movement. The second curve was a distractor. The receptive fields of two groups of V1 neurons (green rectangles) were on the distractor curve for stimulus I and on the target curve for stimulus II. **b** The initial responses did not distinguish between the target and the distractor curves. After about 140 ms, however, the response to the target curve (orange response) was enhanced. The gray region between responses indicates the response enhancement, caused by recurrent processing. Note that the onset of the enhanced response (indicated by the red arrow) occurred later for the neurons with a receptive field that was farther from the fixation point (RF2). The lower panels show the difference between the responses evoked by the target and distractor curves. The red curve shows a function that was fitted to the difference response to estimate the latency of the response enhancement (see Roelfsema, Khayat, & Spekreijse, 2003). **c** Some neurons in the primary visual cortex reflect incremental grouping in their response, while others do not. Left, neurons at so-called A-sites (56 out of 96 recording sites) enhanced their response if their receptive field fell on the target curve, while neurons at the so-called N-sites did not (average response of neurons at 40 recording sites). The blue bar shows the feedforward processing phase, and the yellow bar the recurrent processing phase

attention that influence the response of a large fraction of neurons in area V1 have a modest effect on perceived contrast. A shift of attention can cause a small increase in perceived contrast (Carrasco, Ling, & Read, 2004; Ling & Carrasco, 2006), but this effect has not been found consistently in all studies (Liston & Stone, 2008; Prinzmetal, Nwachuku, Bodanski, Blumenfeld, & Shimizu, 1997; Schneider, 2006). Figure 6a illustrates the responses of two simultaneously recorded V1 sites in a monkey performing a contour grouping task with curves of varying contrast from a

recent study by Pooresmaeili, Poort, Thiele, and Roelfsema (2010). The neurons at the A-site enhanced their activity if their receptive field fell on the target curve, as compared with when it fell on the distractor curve, irrespective of the contrast of the curve. The neurons at the N-site (blue responses), however, did not distinguish the target from the distractor curve. Their responses were determined mainly by the contrast of the curve. Shifts of attention need not interfere with contrast perception if it depends on the activity of neurons at N-sites.

The division of labor between A- and N-sites would also permit the incremental grouping of curves with varying contrast, as is illustrated in Fig. 6b. In this figure, the upper curve is attended so that the responses of A-neurons are stronger than the responses of N-neurons. Thus, in such a scheme, labeling with an enhanced response can even occur for curves with varying contrast. It is also possible to propagate a difference in activity between A- and N-

neurons, if A-neurons excite neighboring A-neurons with N-neurons providing inhibition. This idea is illustrated in Fig. 6c, where A-neurons in the lower row receive an equal amount of excitation and inhibition. In the upper row, however, the A-neurons propagate enhanced activity. When an A-neuron enhances its response, the neighboring A-neuron receives more excitation than inhibition, so that it will also enhance its response. Thus, the network can propagate the response enhancement in spite of the differences in activity caused by variation in stimulus contrast. The network uses two codes: The difference in response between A- and N-neurons labels contour elements for incremental grouping, while the response of N-neurons codes the contrast of the contour elements. In a recent study, we found that it is indeed possible to decode the contrast of a stimulus as well as the contour elements that have been labeled for grouping from the activity of a population of neurons in area V1 (Pooresmaeili et al., 2010).
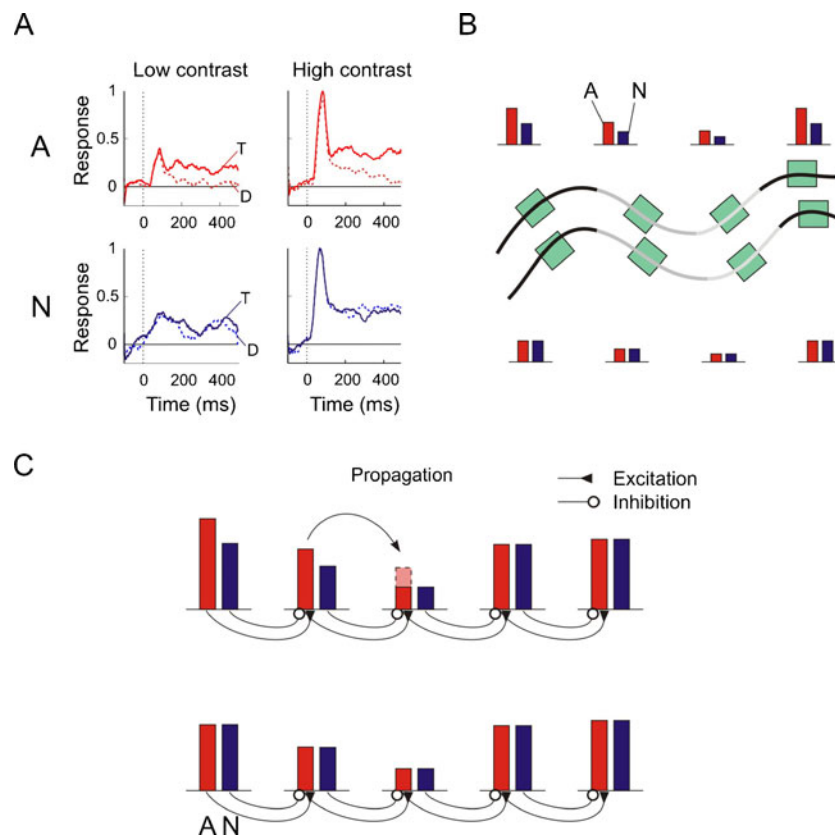

Fig. 6 Incremental grouping of contour elements with varying contrasts. **a** Simultaneous recordings from an A- and an N-site in the primary visual cortex of a monkey performing the curve-tracing task at different contrasts. For the A-site, the responses (shown in red) evoked by the target curve (continuous lines) were stronger than responses evoked by the distractor (broken lines) at low (4.3%; left panel) and high (19%) contrast. The target and distractor curves evoked responses of a similar magnitude at the N-sites (blue responses). Modified from Pooresmaeili, Poort, Thiele, and Roelfsema (2010). **b** Neurons at both A- and N-sites have stronger responses if their receptive field falls on a contour element with a higher contrast. Contour elements of the upper curve have been grouped, because neuronal responses evoked by this curve at A-sites (red bars) are stronger than those evoked at corresponding N-sites (blue bars). **c** A connection scheme where A-neurons receive excitation from other A-neurons and inhibition from N-neurons can propagate the enhanced response along a curve with varying contrast. In the lower row, where no incremental grouping takes place, the amount of excitation to the neighboring A-neurons is balanced by inhibition from N-neurons

Conjecture 5: A veridical and a labeling network There are N-neurons and A-neurons in the visual cortex. N-neurons form a veridical network that provides a reliable representation of the stimulus features, while A-neurons form a labeling network where features can be grouped incrementally.

Interactions between lower and higher visual areas

It is likely that the modulation of neuronal activity during perceptual grouping occurs in multiple visual areas. The IGT uses the propagation of enhanced neuronal activity for the formation of incremental groups, and the spread of enhanced activity between areas is therefore essential for the incremental grouping of low-level features and more complex features. Neurons in lower areas, like V1, label the individual contour elements of an object for grouping and neurons in higher areas could label the shape. Although no study has examined neuronal activity in visual cortical areas beyond V1 during perceptual grouping tasks (one exception is area FEF; see below), studies that have tested monkeys in visual search tasks have demonstrated that the selection of visual information increases the responses of neurons in multiple visual areas, including areas V1, V4, and the inferotemporal cortex (Bichot, Rossi, & Desimone, 2005; Chelazzi, Miller, Duncan, & Desimone, 1993, 2001; Motter, 1994; Roelfsema et al., 2003). Neurons with overlapping receptive fields in lower and higher visual areas are reciprocally connected by feedforward and feedback connections (Felleman & Van Essen, 1991; Salin & Bullier, 1995). Modeling studies have demonstrated that these connections could facilitate the coselection of the individual contour elements and high-level shape features of the same object (Borenstein & Ullman, 2008; Fukushima, 1988; Hamker, 2005; Sharon et al., 2006; Tsotsos, Rodríguez-Sánchez, Rothenstein, & Simine, 2008; van der Velde & de Kamps, 2001; see Fig. 7), and such a selective interaction between high-level and low-level visual representations has a number of advantages.

In the first place, shape representations in higher areas could facilitate the grouping of low-level features. Experimental support for such a facilitation has come from a study by Vecera and Farah (1997), who asked subjects to decide whether two markers were on the same or on different letters, a task that requires the grouping of contour elements of the same letter. This study compared the performance for letters in their upright position, which are presumably strongly represented in higher visual areas, with the performance for upside-down letters with weaker representations (Fig. 7b). Response times were shorter for the upright letter, implying that shape representations indeed facilitate the grouping of low-level features. The

benefit for upright letters is explained if feedforward processing activates shape representations in higher areas (base grouping). Neurons in the higher areas that represent the cued shape (the letter F) would enhance their activity over the noncued shape and would feed back to lower areas to enhance the representation of the individual contour elements of the F in lower areas. Inverted letters are presumably not represented or are more weakly represented in higher areas, so that the grouping of their contour elements has to rely on low-level grouping cues in lower areas like good continuation only.

Conversely, if a set of low-level features that form a perceptual group is labeled with enhanced activity in early visual areas, this enhanced activity can spread to higher areas to activate neurons that code the overall shape (see, e.g., Sáry, Vogels, & Orban, 1993). For example, subjects can group image elements that move coherently in one direction and can segregate them from background elements moving in different directions to identify the shape of this perceptual group (Large, Aldcroft, & Vilis, 2005).

Figure 7c gives another demonstration. In the left panel, some of the image elements are linked by good continuation and can be seen to form the letter F. Without the low-level grouping, it is more difficult to perceive the F (right panel in Fig. 7c). Thus, the interactions between lower and higher areas permit grouping of elementary features into a shape. These interactions between areas are important for many tasks; for example, it is crucial to know which parts do and which parts do not belong to an object if you plan to grasp it.

It is likely that the spread of enhanced activity in lower areas influences the competitive interactions between representations in higher areas. If objects are nearby or overlapping (as they are in Fig. 7b), they typically fall into the same receptive field of neurons in higher areas. In these situations, the neuronal activity is the average of the activity evoked by the individual objects (Armstrong, Fitzgerald, & Moore, 2006; Reynolds, Chelazzi, & Desimone, 1999). Thus, a neuron that responds well to the F in Fig. 7b but poorly to the A will have an intermediate response if both letters fall in its receptive field. When attention is directed to the strokes of the F, the neuron's response increases to the level evoked by the F (if presented alone). Vice versa, if attention is directed to the A, the activity of the neuron decreases to the level evoked by an individual A. Modeling studies have shown that the spread of enhanced neuronal activity in lower visual areas can account for these influences on neuronal activity in higher areas (Grossberg & Raizada, 2000).

Scale invariance of perceptual grouping

The spread of enhanced neuronal activity between lower and higher visual cortical areas for incremental grouping
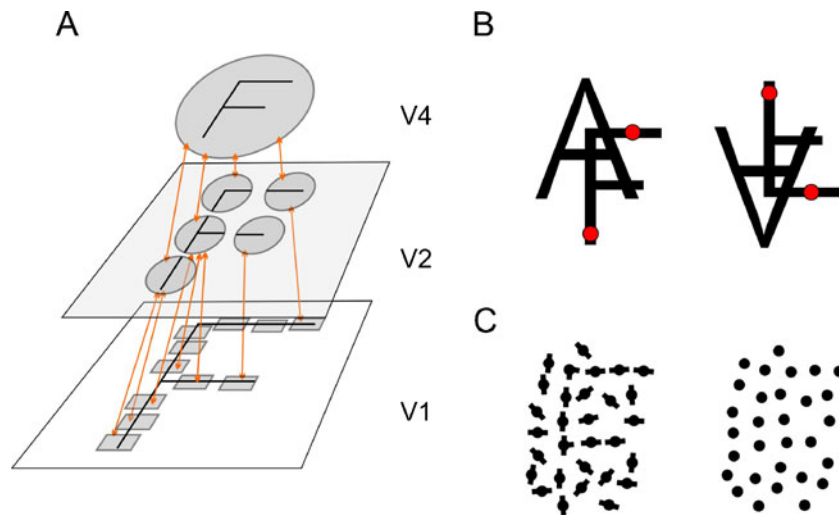
Fig. 7 Interactions between lower and higher areas of visual cortex. **a** The individual contour elements of a letter F are represented in area V1, whereas fragments of the shape and the shape itself are represented in higher areas. Interareal connections permit a bidirectional propagation of enhanced neuronal activity between areas. **b** If subjects have to indicate whether the two red dots are on the same or on different letters, their performance is more efficient for upright than for inverted letters (modified from Vecera & Farah, 1997). **c** Image elements that are related to each other by low-level grouping cues can define a shape that is represented in higher visual areas. Left, a subset of image elements is grouped by good continuation. Right, it is more difficult to see the letter F if the low-level grouping cues are lacking

can also explain the scale invariance that has been demonstrated for curve tracing (Jolicoeur & Ingleton, 1991). The time it takes to trace from the fishermen to the fish in Fig. 2b is largely independent of viewing distance. If we come close to the figure, the length of the curve that has to be traced (measured in degrees of visual angle) increases, but grouping speed increases accordingly, so that the overall response time remains the same. This result is incompatible with models that implement contour grouping at a single spatial scale (e.g., the model in Fig. 4). To explain this scale invariance, Roelfsema and Singer (1998) proposed that the propagation of the enhanced response occurs simultaneously in multiple areas of the visual cortex. If curves are far apart, tracing makes the fastest progress in the higher visual areas with larger receptive fields where many degrees of visual angle can be crossed by a few synapses (Fig. 8b). The response enhancement in higher areas is also fed back to the lower visual areas where labeling with the enhanced response proceeds faster than would have been possible without the higher areas. If curves come close together, however, the large receptive fields in higher areas fall on multiple curves, and the enhanced response might spill over to the distractor curve (dashed receptive field in Fig. 8a). This calls for a mechanism that blocks propagation in the higher areas whenever curves are nearby (Edelman, 1987; Jolicoeur et al., 1991). Propagation of the enhanced response is taken over by the lower visual areas where the smaller receptive fields fall on only a single curve. This higher spatial resolution comes at the cost of a decreased grouping speed because more synapses have to be crossed in the

visual cortex to bridge the same distance in the visual field. Thus, the scale invariance of contour grouping can be explained by the propagation in multiple areas. In such a model, tracing speed depends on the distance between the target curve and the nearest distractor, in accordance with response time data in human observers (Jolicoeur et al., 1991). Neural network modeling studies demonstrate that scale-invariant contour grouping is indeed possible in such a hierarchical network (Korjoukov & Roelfsema, manuscript in preparation).

Influence of the enhanced activity on areas involved in response selection

Theories that spread labels to group features into objects should also provide a mechanism for these labels to be read out by brain regions involved in response selection. In the curve-tracing task shown in Fig. 5, for example, the animals responded with an eye movement to the larger circle at the end of the target curve. Khayat, Pooresmaeli, and Roelfsema (2009) recorded from neurons in the frontal eye fields (area FEF), an area of the frontal cortex involved in the generation of eye movements, during this task. They configured the stimulus so that the receptive field of the FEF neurons fell on the circle at the end of either the target or the distractor curve (Fig. 9b). The lower panel of Fig. 9c illustrates the activity of a population of FEF neurons that are initially activated by the appearance of a stimulus in their receptive field, while it takes more time before the responses evoked by the target curve are enhanced over

Fig. 8 Scale invariance of contour grouping. **a** Incremental grouping of a curve that is near an adjacent curve relies on neurons with small receptive fields (RFs) in area V1 and proceeds slowly. Black circles in V2 and V4 denote neurons with multiple contour elements in their larger RFs that cannot participate in incremental grouping. **b** With a larger spacing of the curves, V2 neurons can contribute to incremental grouping, and the speed of incremental grouping increases. Orange lines denote enabled connections within and between areas. Gray dashed lines show disabled connections
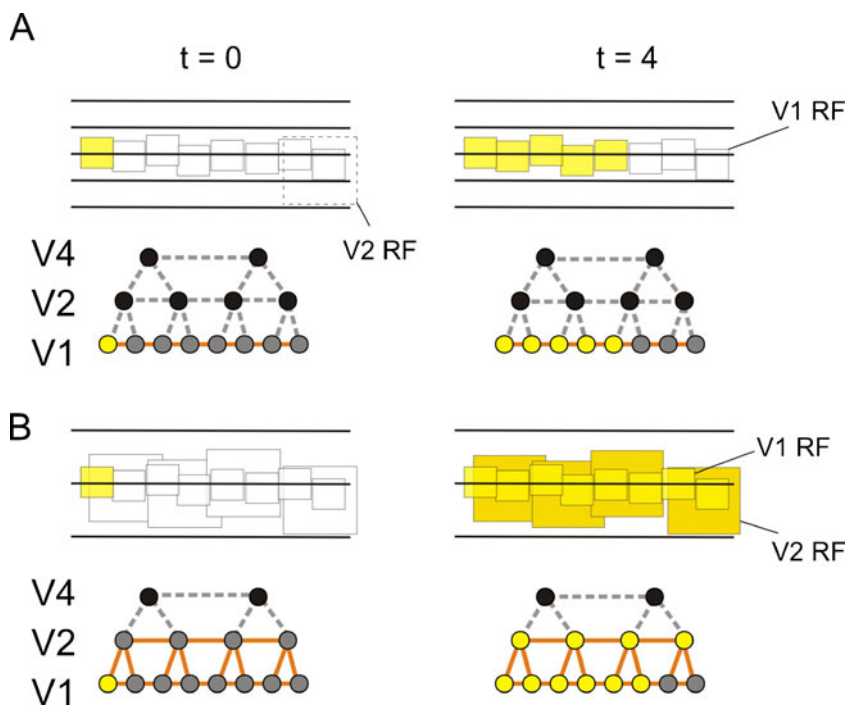

responses evoked by the distractor (for comparable observations during visual search, see Buschman & Miller, 2007; Schall & Thompson, 1999). The top panel of Fig. 9c shows the responses of neurons in the primary visual cortex for comparison. Neuronal activity in both areas is initially dominated by feedforward processing (blue bar in Fig. 9c), while the response enhancement caused by curve tracing takes more time to develop (yellow bar).

Incremental grouping is apparently associated with the enhancement of neuronal activity in the visual and frontal cortices. It is likely that the enhanced activity of V1

neurons also influences neuronal responses in higher visual areas (see the Interactions between lower and higher visual areas section). However, we do not know whether the enhanced activity in FEF originates from area V1. It is also possible that the response enhancement in V1 is caused by feedback from higher visual areas, including area FEF. Studies that have used microstimulation to artificially raise the activity of small populations of neurons have demonstrated that feedback from FEF enhances neuronal activity in the visual cortex (Moore & Armstrong, 2003), even in area V1 (Ekstrom, Roelfsema, Arsenault, Bonmassar, &

Fig. 9 Interactions between the visual and frontal cortex during contour grouping. **a** Lateral view of the cortex of the macaque monkey. Areas V1 and FEF are shown in red. **b** Curve-tracing stimuli where either the target or the distractor curve fell in the receptive fields of neurons in area V1 and FEF. **c** Responses evoked by the target (orange) and distractor curves (blue) in areas V1 (upper panel) and FEF (lower). In both areas, the initial response did not discriminate between target and distractor (blue bar, feedforward processing), but the later response was enhanced if the target curve fell in the receptive field (yellow bar, recurrent processing)
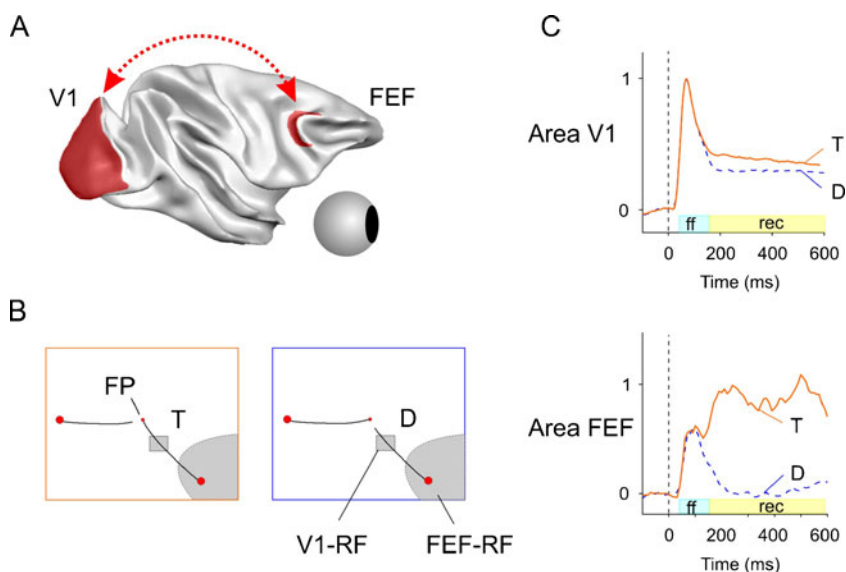
Vanduffel, 2008). Our working hypothesis is that neurons in the two areas engage in *reciprocal* interactions, causing the same curve to be selected in the visual and the frontal cortices (Duncan, Humphreys, & Ward, 1997). Interactions between area V1 and FEF involve intermediate processing stages as well (Ekstrom et al., 2008), since there are only a few direct connections between area V1 and FEF (Felleman & Van Essen, 1991). A recent study showed that the widespread neuronal correlates of the attentional selection of the target curve can also be measured as a sustained negativity in the human EEG (Lefebvre, Jolicoeur, & Dell'Acqua, 2010).

Conjecture 6: Consequences of incremental grouping Neurons that enhance their activity during incremental grouping have increased impact on other cortical areas and can thus provide the input to object recognition and response selection.

The role of attention

Incremental grouping requires the time-consuming spread of an enhanced firing rate. Many neurophysiological studies have indicated that these firing-rate modulations in the visual cortex are responsible for shifts of visual attention at a psychological level of description (reviewed by Desimone & Duncan, 1995; Roelfsema, 2006). Thus, while base grouping maps onto preattentive processing, incremental grouping maps onto attentive processing, and the spread of an enhanced neuronal activity in the visual cortex corresponds to the spread of object-based attention in psychology (Fig. 4c).

We obtained support for this idea by investigating the distribution of visual attention during contour grouping (Houtkamp et al., 2003). Subjects saw a target curve that started at a fixation point and a distractor curve. Their primary task was to indicate the location of a marker at the other end of the target curve. To probe the distribution of attention, colors were presented on different segments of the curves at various intervals during a trial, and the secondary task was to report one of these colors. The performance in the secondary task showed that, at the start of the trial, attention was directed to the initial contour elements of the target curve and that it subsequently spread across the entire curve until all contour elements were labeled by attention (schematically indicated in Fig. 4c). Thus, during incremental grouping, attention gradually adds elements to the evolving perceptual group by spreading from attended image elements to other elements that are related to them by Gestalt criteria, until the entire object has been labeled by attention.

Conjecture 7: Spread of attention The propagation of an enhanced neuronal response through the network of enabled connections corresponds to the spread of attention on the basis of Gestalt cues at a psychological level of description.

Incremental grouping on the basis of good continuation, similarity, and proximity

The IGT can explain the paradoxical finding that some contour grouping tasks are solved in parallel, while others require serial processing. In a pathfinder display (Fig. 2a), a single collinear figure is presented on an incoherent background consisting of contour elements with random orientations. Detection of these figures can occur in parallel if we assume that the visual system contains operators sensitive to the local degree of collinearity of contour, as can be implemented with a feedforward connectivity scheme (see Gigus & Malik, 1991, for a definition of these operators; see Kapadia, Ito, Gilbert, & Westheimer, 1995, for neurophysiological evidence). Thus, feedforward connections permit the parallel segregation of well-aligned contour elements from an incoherent background. However, such a feedforward process cannot determine whether the collinear elements indicated by the arrows in Fig. 10a belong to the same curve, because there is another curve with contour elements that are equally collinear. In this situation, the detection of local groups of aligned contour elements does not suffice. It becomes necessary to transitively combine a number of these local groupings, and this is achieved by the incremental-grouping process that serially labels contour elements of one curve with attention (Houtkamp & Roelfsema, 2010).

The connectivity scheme shown in Fig. 4 works well for the detection of connectedness but must be generalized to accommodate other Gestalt grouping laws, such as good continuation. For good continuation, we assume that there are connections that spread the enhanced neuronal activity between neurons tuned to well-aligned contour elements (Field et al., 1993; Grossberg & Raizada, 2000; Li, 1999). This assumption is in accordance with the anatomy of horizontal connections in the visual cortex, which interconnect neurons that code contour elements that are well aligned—that is, in each other's good continuation (Bosking, Zhang, Schofield, & Fitzpatrick, 1997; Schmidt, Goebel, Löwel, & Singer, 1997). Likewise, grouping by similarity can be implemented by connections between neurons tuned to similar features (e.g., Grossberg & Mingolla, 1985; Roelfsema, Lamme, Spekreijse, & Bosch, 2002), and proximity grouping by connections between neurons with nearby receptive fields.
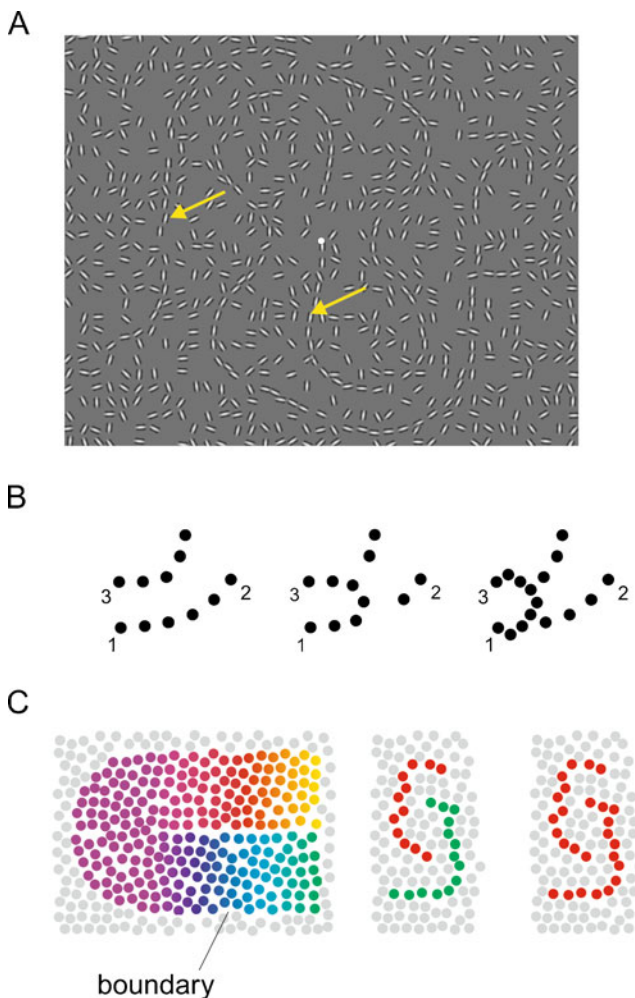
Fig. 10 Transitive grouping on the basis of good continuation, proximity, and similarity. **a** Transitive grouping on the basis of good continuation has been studied with a pathfinder display with two paths. A parallel process can segregate collinear elements from background elements and permits perception of the paths. However, a serial process is required to establish whether the two image elements indicated by arrows belong to the same or different paths. **b** Transitive proximity grouping. On the left, circle 1 groups with circle 2, although it is closer to circle 3. In the middle, the distances between circles 1–3 are the same, but other circles are displaced, and circle 1 now groups with circle 3. On the right, extra circles are added to the left picture, and circle 1 also groups with 3. **c** Left panel: transitive similarity grouping. Nearby elements with a similar color are grouped together, and boundaries form at locations with an abrupt change in color (e.g., between gray and blue circles). In this example, the yellow and green elements are on different sides of a boundary, but they are nonetheless grouped indirectly, through a detour with a gradual color gradient. The middle panel illustrates a situation where a string of red elements can be detected as base grouping during feedforward processing because they have a unique color. The right panel shows that the same elements require incremental grouping if there is a second string with the same color

Conjecture 8: Implementation of Gestalt grouping Gestalt grouping cues are implemented by connecting neurons tuned to image features that are likely to belong to the same perceptual object so that they spread enhanced activity—for example, similar features (similarity grouping), well-aligned features (good continuation), or nearby features (proximity grouping).

The IGT predicts that perceptual grouping becomes time consuming whenever groupings have to be formed transitively because they are not extracted as base groupings. In a recent study, we investigated whether it is possible to observe delays during perceptual grouping on the basis of good continuation, using displays similar to those shown in Fig. 10a (Houtkamp & Roelfsema, 2010). We found that the reaction time of subjects increased linearly with the number of elements that had to be grouped together (i.e., with the distance between the two arrows in Fig. 10a), just as had been observed for continuous contours by Jolicoeur et al. (1986, 1991).

Figure 10b shows that it is also straightforward to create stimuli where transitive grouping occurs on the basis of proximity. The circles in the left panel can be seen to form two strings. Circle 1 is close to a nearby circle, which is close to another one, and we eventually reach circle 2 through a chain of local groupings. Transitivity dictates that the entire chain is seen as a perceptual group, and circle 1 therefore groups with circle 2, although it is actually closer to circle 3. Thus, also in this example, the transitivity implies that grouping is sensitive to the context set by other elements in the display. In the middle panel of Fig. 10b, the distances between circles 1, 2, and 3 are the same, but other circles are displaced so that circle 1 groups with circle 3. According to the IGT, attention spreads from one circle to the next until the whole string is attended, and we observed that the reaction times of subjects indeed increased linearly with the number of items that had to be grouped using displays comparable to those shown in Fig. 10b (Houtkamp & Roelfsema, 2010). The IGT would model the spread of the enhanced response by assuming that neurons coding nearby items are linked by recurrent connections. It is well known that attention tends to spread from target elements to other items in their proximity (e.g., Eriksen & Eriksen, 1974). The IGT assigns a functional role to this effect: It promotes grouping of nearby items.

In the right panel of Fig. 10b, we added five elements to the left picture, and circle 1 now also groups with circle 3. This is remarkable because all circles that promoted grouping between circles 1 and 2 in the left panel have kept their position. How can a scheme that uses enabled connections to support grouping in the left panel fail to do so in the right panel? We propose that proximity grouping is implemented at multiple spatial scales. Horizontal connections in early visual areas link nearby locations in the visual field, while horizontal

connections in higher areas link locations that are farther apart. Higher areas could propagate an enhanced response between neurons that code image elements that are far apart, but propagation in the higher areas should be blocked if there are also image elements that are nearer, just as was proposed for scale-invariant contour grouping above. A psychophysical study demonstrated that proximity grouping is indeed largely scale invariant (Kubovy et al., 1998), so that the perceptual organization of the displays of Fig. 10b does not depend on viewing distance. The implementation of proximity grouping at multiple hierarchical levels of the visual cortex might account for this scale invariance, a hypothesis that could be tested by neural network studies.

Figure 10c illustrates transitive grouping by similarity. Nearby image regions with a similar color are grouped together. Boundaries form at locations where neighboring circles have a categorically different color so that elements on one side of a boundary do not group with elements on the other side (e.g., blue and gray circles). The left panel of Fig. 10c shows that a gradual change in color within a region permits transitive grouping between elements with a dissimilar color (e.g., green and yellow circles). Mumford, Kosslyn, Hillger, and Herrnstein (1987) and Wolfson and Landy (1998) demonstrated that subjects exploit gradual changes in feature values for the grouping of image regions, but they also evaluate abrupt feature changes for their segregation, in support of the idea that grouping and segregation are complementary processes. Note, however, that while an abrupt change in color may provide local evidence for the segregation of image regions, these regions may nevertheless be linked through a detour, as is illustrated for the yellow and green circles in the left panel of Fig. 10c. If subjects are tested with displays where image elements have to be grouped on the basis of their similarity, the reaction time increases linearly with the number of items that need to be grouped, indicating that incremental grouping also occurs when similarity defines the perceptual groups (Houtkamp & Roelfsema, 2010).

The middle and right panels of Fig. 10c illustrate that grouping on the basis of similarity can, in some situations, rely on base grouping, while it requires incremental grouping in others. Image elements with a color that differs from other elements can be detected by neurons in higher areas that are color selective (middle panel), rapidly and in parallel. However, incremental grouping has to come into play when there are multiple strings with the same color (right panel). In this situation, color selectivity does not suffice for grouping. To group the elements of one of the red strings and to segregate them from the other string, an enhanced response has to be propagated along a chain of more local groupings, based on both color and proximity relationships. We showed recently that grouping is serial in this situation, since subjects' response times increase linearly with the length of the string (Houtkamp & Roelfsema, 2010).

We recently tested whether the enhanced neuronal activity indeed spreads according to multiple Gestalt grouping rules in the visual cortex of monkeys (Wannig, Stanisor, & Roelfsema, 2011). The monkeys saw several stimuli, and we cued one of them as the target for a saccadic eye movement (Fig. 11). The appearance of the cue increased neuronal activity at the cue's location, and this enhanced response then spread to other stimuli that were in good continuation with the cued stimulus (Fig. 11a), had a similar color (Fig. 11b), or moved coherently (not shown). Control experiments demonstrated that the enhanced activity did not spread to stimuli that were unrelated to the target by any of these grouping cues. Thus, the enhancement of neuronal activity caused by an attention shift indeed spreads along several Gestalt grouping rules, including good continuation, color similarity, and common fate.

## Previous experiments on perceptual grouping

The key assumptions of the IGT have been summarized in conjectures 1–8. In summary, the presentation of an image triggers a parallel and resource-unlimited base-grouping process that depends on cascades of neurons tuned to features and feature conjunctions. This pattern of activity enables some of the links: activity-spreading connections between activated neurons. For conjunctions not coded as base groupings, there is a later serial and resource-limited incremental grouping process that relies on attention that is propagated along the enabled links to enhance the representation of a coherent, unified group. This pattern of enhanced activity can then be read out by other processes for object recognition (Fig. 7c) or for the programming of actions.

We will now explore whether and how the conjectures above, which are largely based on neurophysiology, can account for previous results in perceptual psychology. Let us, for the sake of the argument, step in the shoes of an outsider who tries to familiarize himself or herself with the literature on grouping in perception. It is likely that he or she will first be confused. Some workers have argued that perceptual grouping takes place in parallel across the visual scene, while others state that it requires serial processing. Some maintain that grouping depends on visual attention, while others claim that it largely happens at a preattentive stage. Many of these viewpoints have been supported by substantial experimental evidence. The IGT aims to provide a framework that resolves some of these discrepancies.

When grouping takes time and when it does not

Sometimes perceptual grouping can rely on base grouping, which is a rapid process that occurs in parallel across the visual scene. One example that we mentioned is the pathfinder task, where subjects detect a string of collinearly

Fig. 11 Enhanced neuronal activity spreads in the visual cortex along Gestalt grouping cues. **a** The monkey saw four lines, and one of these lines was cued as the target for a saccadic eye movement. The enhanced neuronal activity (yellow region) evoked by the cue spread to another line according to the Gestalt rule of good continuation. **b** In this task, the monkey had to make an eye movement to the circle that increased in size. The enhanced neuronal activity evoked by the cue spread according to the Gestalt rule of similarity

A

FP   Cue   Enhanced activity

Time

B

Cue   Enhanced activity
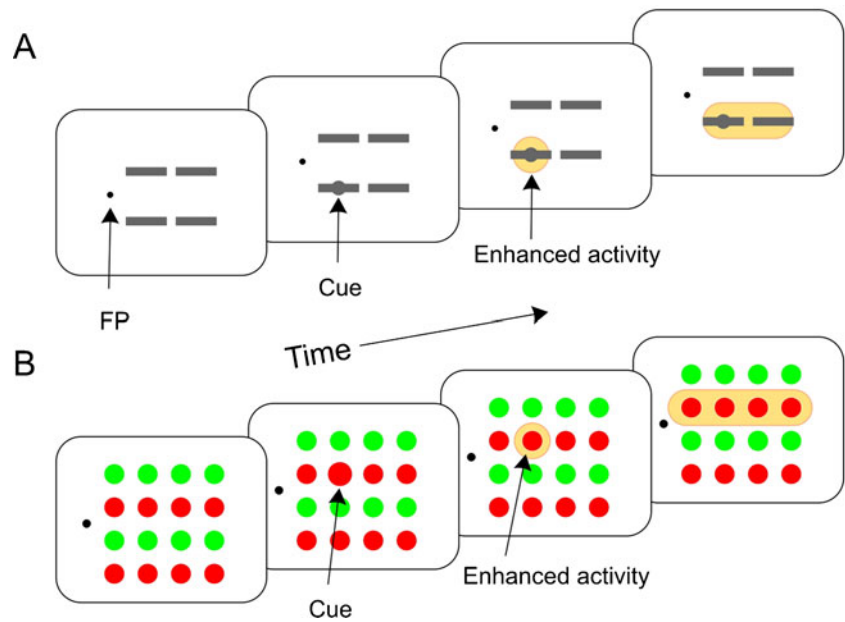
aligned contour elements (see the Introduction and Fig. 2a). Another example is the rapid detection of feature constellations that form familiar shapes, such as the letters of the alphabet, or familiar objects, such as, for example, faces and cars (Thorpe et al., 1996), that presumably relies on the category-selective neurons in the inferotemporal cortex and other cortical regions (Freedman, Riesenhuber, Poggio, & Miller, 2003; Hung et al., 2005; Oram & Perrett, 1992; Sugase et al., 1999). Incremental groupings will have to be formed when there is no neuron that codes the relevant constellation as a base grouping or if the representation of object identity does not suffice for the task—for example, if it is important to also group lower level image elements with the shape. This process is time consuming, especially when image elements are related only indirectly through a chain of local groupings. The flexibility of incremental grouping therefore comes at the cost of serial processing. A study by Holcombe and Cavanagh (2001) gives a beautiful illustration of the distinction between base and incremental grouping. Subjects saw a red leftward-tilted grating that alternated rapidly with a green rightward-tilted grating at the same location (Fig. 12a). The alternation rate was varied to determine the maximal rate at which the subjects could report the color of, say, the left-tilted grating. They could do this at the remarkably short exposure duration of 30 ms per stimulus. This rapid grouping worked, however, only if the color and orientation were present at the same location. If a leftward-tilted grating was juxtaposed to a homogeneous red region and this stimulus was alternated with a right-tilted grating next to a green region, correct grouping between color and orientation was possible only for exposure durations of approximately 200 ms (Fig. 12b). This suggests that the conjunction between an orientation and a color is a base

grouping coded by single neurons, but only if these features are present at a single location (i.e., if they fall in the same receptive field). Neurons tuned to both color and orientation are indeed abundant in visual cortical areas (Kobatake & Tanaka, 1994; Leventhal et al., 1995; Sincich & Horton, 2005). Conjunctions between a color at one location and an orientation at another location can, on the other hand, presumably form only as incremental groupings. They require the spread of enhanced neuronal activity between neurons coding the orientation at one retinal location and other neurons coding color at another location, and their formation is therefore associated with longer processing delays.

These considerations do not exclude that incremental grouping may, in some cases, also be required to establish conjunctions between features at the same location. To take an arbitrary example, imagine a rotating and shrinking elephant. Feature conjunctions between complex motion patterns and shapes are presumably not coded as base
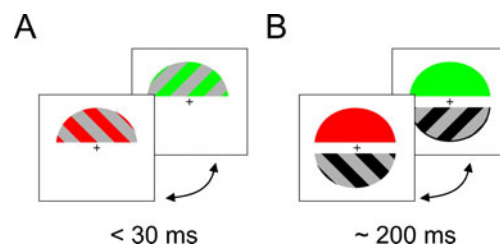
A   B

< 30 ms   ~ 200 ms

Fig. 12 Base and incremental grouping of nearby features. **a** If the two pictures are shown in alternation at the same location, subjects are able to report that the leftward-tilted grating is red and the rightward-tilted grating green when pictures are shown longer than 30 ms. **b** If the orientations and colors are shown at different locations, the subjects require at least 200 ms per picture before they can report the feature conjunctions. Adapted from Holcombe and Cavanagh (2001)

groupings and would require labeling the shape in one visual area and the motion pattern in another area with enhanced activity. Mechanisms to ensure that the enhanced activity spreads from one visual area to another one with sufficient specificity have been reviewed by Roelfsema (2006).

Grouping with and without attention

There are many discrepancies in the literature about the involvement of attention in perceptual grouping. Some studies have suggested that forms of grouping do not occur without attention, while other studies have suggested that Gestalt grouping takes place at a preattentive stage. Here, we will suggest how some of these discrepancies can be resolved by distinguishing between base and incremental grouping.

One example of attention-demanding grouping is the contour-grouping task, where object-based attention spreads over contour elements that have to be grouped into elongated curves (see Fig. 4c; Houtkamp et al., 2003; Roelfsema, Houtkamp, & Korjoukov, 2010; Scholte, Spekreijse, & Roelfsema, 2001). Other studies have found that some forms of grouping do not occur when attention is directed elsewhere. Ben-Av, Sagi, and Braun (1992), for example, investigated grouping of image elements surrounding a centrally displayed letter. When the subjects directed their attention to the letter, they were unable to report the perceptual organization of the other image elements, which were arranged in rows or columns on the basis of proximity or similarity, as if proximity and similarity groups do not form without attention.

Another line of evidence that seems to imply a role for attention in perceptual grouping comes from the *inattentional blindness* paradigm introduced by Mack, Tang, Tuma, Kahn, and Rock (1992; see also Mack & Rock, 1998). Their subjects had to report about the relative length of two arms of a central cross that was surrounded by a pattern of small elements. After a few trials with this task, the background elements were organized in columns or rows on the basis of proximity or similarity cues. The observers received a surprise question about the perceptual organization of the background elements and were usually unable to report about the grouping into rows or columns. Mack et al. concluded that Gestalt grouping does not take place without attention. However, their methodology was later criticized. First, it is conceivable that grouping took place outside awareness and that subjects therefore were not able to report about it (Driver, Davis, Russell, Turatto, & Freeman, 2001). Second, the observers may have forgotten their percept by the time of questioning (Wolfe, 1999).

Subsequent studies with similar arrays, but using more sensitive and implicit measures of grouping, have substantiated these criticisms and, instead, have obtained evidence for grouping without attention. C. M. Moore and Egeth (1997), for example, asked subjects to carry out a line length discrimination task on a background of black and white dots. On some of the trials, the black dots were configured to induce a line length illusion in the case of grouping. Remarkably, the dots influenced the line length judgments, even though subjects could not report about the groupings when asked (see also Chan & Chua, 2003). Studies by Kimchi and Razpurker-Apfeld (2004) and Russell and Driver (2005) extended these findings. They investigated the influence of perceptual grouping of background elements while subjects carried out a change detection task. Subjects had to compare stimuli in central vision to detect a change. Unbeknownst to the subjects, the image elements in the surround formed columns or rows on the basis of color similarity in some of the images. If both the central pattern and the grouping of background elements changed across displays, the subjects were more likely to report the change than when only the central stimulus changed. Again, the observers were unaware of the grouping of the background elements (see also Kimchi & Peterson, 2008).

These studies demonstrate that perceptual grouping can occur without attention and outside awareness. According to the IGT, this is possible only for feature constellations that are coded as base groupings. The arrays of black dots of C. M. Moore and Egeth (1997) looked like lines if observed through a low spatial frequency filter. It is therefore plausible that neurons in the visual cortex could detect these dot arrays as base groupings during feedforward processing, and this may have influenced perceived line length just as normal line inducers that are commonly used to produce the line length illusions. A similar explanation can be given for the isoluminant dot arrays used in the change detection task of Kimchi and Razpurker-Apfeld (2004) and Russell and Driver (2005). These patterns are likely to activate orientation-selective cells (Gegenfurtner, Kiper, & Levitt, 1997) and can be registered as base groupings without attention. Thus, although these results are indicative of grouping without attention, they are consistent with what is known about the tuning of visual cortical neurons to feature conjunctions.

Another line of evidence that, at least at first sight, seems to imply that Gestalt grouping occurs without attention has come from studies in patients with hemineglect. These patients often fail to perceive objects in the hemifield that is contralateral to a brain lesion that is often located in the parietal cortex (Halligan & Marshall, 1993; reviewed by Driver, 1995). Many of these patients suffer from extinction: If presented with two visual objects, one in each hemifield, they see only the object in the good hemifield and fail to see the one in the bad hemifield. This deficit occurs even though patients are able to see the same stimulus in the impaired hemifield if presented alone. The remarkable finding is that an item in the bad hemifield

can be rescued from extinction if it forms a perceptual group with an item in the good hemifield; that is, if the patients see two items that form a perceptual group, they perceive both. This relief from extinction has been observed for objects that are grouped on the basis of luminance similarity (Gilchrist, Humphreys, & Riddoch, 1996), connectedness (Driver, 1995; Humphreys & Riddoch, 1993), and good continuation (Gilchrist et al., 1996; Mattingly, Davis, & Driver, 1997; Pavlovskaya, Sagi, Soroker, & Ring, 1997).

Because the main impairment of neglect patients is to shift their attention to the bad hemifield, the results have been interpreted as evidence for grouping without attention (e.g., Driver, 1995). The IGT provides an alternative explanation that is based on linking—that is, the enabling of connections. The items in the good and bad hemifields of the patients are related to each other by Gestalt grouping cues, and neurons that represent them should, therefore, be linked by recurrent, attention-spreading connections in early visual areas that are usually spared by the lesion. The enabling of these connections occurs in parallel across the visual field but is without effect during the preattentive processing stage. However, when the patient attends to the item in the good hemifield, the enabled connections cause attention to spread to the item in the impaired hemifield and facilitate detection.

Linked but not grouped

The status of items linked by enabled attention-spreading connections is a subtle and potentially confusing issue. Enabling of connections in the base representation occurs in parallel across the image. One might, therefore, argue that these items are grouped in parallel and preattentively and that our theory is a variation on previous theories on perceptual grouping (Bergen & Julesz, 1983; Julesz, 1981; Neisser, 1967; Treisman & Gelade, 1980; Treisman & Gormican, 1988). However, the IGT suggests that only base groupings are computed outside the focus of attention as soon as an image like the one shown in Fig. 13a appears. Connections between neurons that code image elements of the same elongated object are enabled (Fig. 13a), but the base representation does not reveal whether element 1 belongs to the same object as element 2 or element 3. Transitive grouping of elements 1 and 2 does not take place until attention spreads across the enabled connections to make this additional, incremental grouping explicit. In other words, the elongated strings are perceived only once they are filled with attention, and this occurs for only one string at a time (Avrahami, 1999; Houtkamp & Roelfsema, 2010). Figure 13b illustrates incremental grouping in a more natural scene. The floor brush of the vacuum cleaner is only indirectly grouped with the plug at the end of the cable, and the IGT predicts that grouping should be associated with a substantial delays.

The distinction between *linking* and *grouping* also sheds light on intriguing results on the influence of grouping in the motion-induced blindness (MIB) paradigm (Bonneh, Cooperman, & Sagi, 2001). In this paradigm, a number of high-contrast stationary (or slowly moving) items are superimposed on a background of rapidly moving dots. Under these conditions, the stationary items disappear spontaneously from perception for a period of several seconds and then reappear. Bonneh et al. suggested that the moving dots increase the level of competition between the representations of visual objects, so that the stationary ones can completely disappear from perception. Importantly, Gestalt cues influence MIB. Visual objects disappear and reappear together if they form a perceptual group on the basis of collinearity and proximity cues. Ungrouped objects, on the other hand, disappear and reappear independently. Mitroff and Scholl (2005) extended these findings to other grouping cues, including connectedness. In their study, Gestalt grouping cues were added or removed between objects while they were invisible. Remarkably, the changes in the grouping cues occurring outside awareness nevertheless influenced the simultaneity of reappearance.

If grouping cues were removed outside awareness, the items tended to reappear independently, and vice versa, if grouping cues were added, items tended to reappear together.

The enabling of attention-spreading connections between items on the basis of Gestalt cues accounts for their simultaneous reappearance in MIB. Recall that the enabling process is the direct consequence of the pattern of activity in the base representation, and modifications in the base representation are therefore associated with changes in the set of enabled connections (as was discussed in relation to Fig. 4a). Thus, enabling and disabling of connections (i.e., linking) is independent of attention and can occur outside awareness. However, once attention is directed to one item, it will spread to other items that are linked so that they become visible at the same time.

An additional explanation is required to account for the simultaneous *disappearance* of grouped items. One possibility is that the mutual facilitation between grouped items causes them to remain visible for a longer time than they would if presented alone. Bonneh et al. (2001) indeed demonstrated that grouped items were less prone to disappear from awareness. However, as soon as one of the grouped items disappears from awareness, this decreases the facilitation of linked items, increasing the probability that they also become invisible.

The influence of grouping cues on the spread of attention

The IGT requires that attention flows within perceptual groups, a requirement that has been supported by many
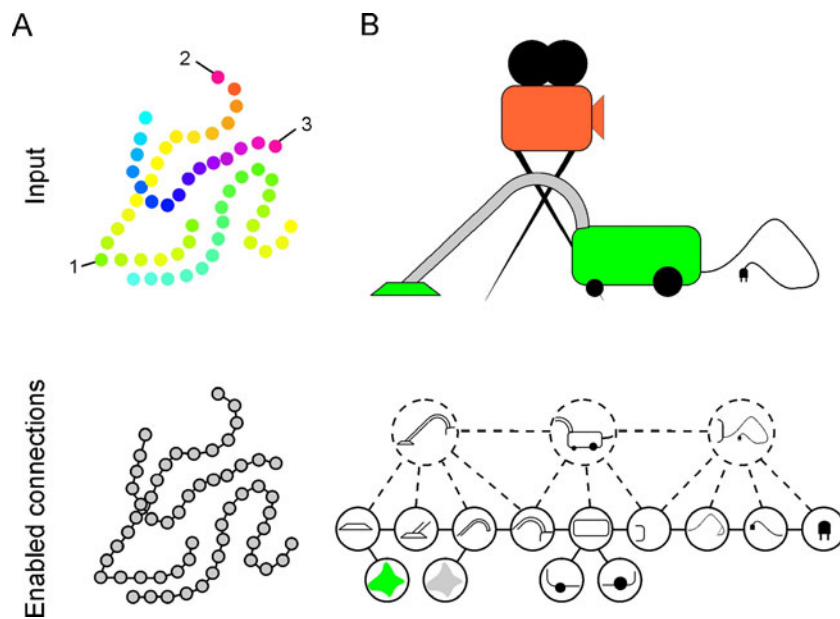
Fig. 13 Linking and grouping. **a** Circle 1 belongs to the same perceptual group as circle 2. Circles 2 and 3 are not linked, even though they have the same color. The lower panel illustrates that linking takes place in the base representation. When the image appears, connections between neurons that respond to nearby image elements with a similar color are enabled. Neurons that respond to elements 1 and 2 are linked indirectly, through a number of intermediate elements. This linkage is not accessible to neurons in higher visual areas, and the elements are therefore not yet grouped.

Grouping occurs only when a neuronal response enhancement spreads through the enabled connections. In psychological terms, incremental grouping depends on the spread of attention. **b** The grouping of many everyday objects may also involve serial processing—in particular, if they consist of multiple parts—because attention has to spread from one component to the next to make the grouping explicit. Dashed lines indicate additional base groupings and links that might be added by learning. The additional links shorten the route between parts of the same object and increase grouping speed

studies. Kahneman and Henik ([1981]) may have been the first to study the effect of perceptual grouping cues on the spread of attention. They investigated the effect of proximity and similarity cues in partial report tasks and found that perceptual groups act as units, because grouped items tend to be jointly reported or jointly missed, suggesting that they are coselected by attention. Later studies extended these findings with a great variety of techniques.

The flanker task is another powerful method to probe the influence of Gestalt grouping cues on the spread of attention. In this task, subjects map target objects onto arbitrary responses. The target is flanked by distractors that are response incompatible (they map onto the opposite response), neutral (not associated with a response), or compatible (associated with the same response as the target). The general finding is that response-incompatible flankers increase response time, while compatible flankers tend to reduce response time. Importantly, flankers that are linked to the target by Gestalt grouping cues cause more interference than do unlinked flankers. These effects can be explained if attention spreads from the target to the flankers if they are linked by Gestalt grouping cues. Eriksen and Eriksen ([1974]), for example, showed that nearby flankers generate

more interference than do flankers that are farther away, supporting the hypothesis that attention flows among items linked by proximity (see also Baylis & Driver, [1992]). Similar results have been obtained for similarity and motion. Flankers with a similar color or motion cause more interference than do flankers with a different color or motion (Baylis & Driver, [1992]; Driver & Baylis, [1989]; Harms & Bundesen, [1983]; Kramer & Jacobson, [1991]). Moreover, connectedness and good continuation modulate the flanker effect similarly (Baylis & Driver, [1992]; Kramer & Jacobson, [1991]; Richard, Lee, & Vecera, [2008]). These results, taken together, provide strong support for the hypothesis that attention spreads among items linked by Gestalt grouping cues, enhancing the impact of flankers linked to a target.

Cuing tasks provide a second line of evidence in support of the hypothesis that attention flows among linked items. In an elegant study, Egly, Driver, and Rafal ([1994]) presented two elongated objects and asked subjects to detect a probe item that was presented on one of these objects. The probe was preceded by a cue that could appear at another location on the same object or on a different object. Remarkably, a cue on the same object gave rise to shorter response times than did a cue on the other object, even if the distances between cues and probes were the

same. This result is explained if attention spreads across the entire representation of the object when it is cued on one end; that is, it spreads within a *linked array* of locations (Vecera, 1994).[2] This result was extended by Haimson and Behrmann (2001), who showed that attention spreads across the entire cued object even if parts of it are occluded, and by He and Nakayama (1995), who demonstrated that attention spreads among neighboring items that are linked by a disparity gradient defining a plane in the image. We recently discovered a neuronal correlate of this object-based cuing effect by showing that the enhanced neuronal activity evoked by a cue spreads to the representation of other image elements in the visual cortex that are linked to the cued element (see Fig. 11; Wannig et al., 2011).

Visual search experiments provide yet another line of evidence in support of the conjecture that perceptual groups act as units that can be selected by attention. Duncan and Humphreys (1989) demonstrated an important role for similarity grouping in visual search. They showed that visual search for a particular target item is most efficient if distractors are similar, but dissimilar from the shape of the target, and suggested that a set of similar distractors can be rejected efficiently as a perceptual group (see also Bundesen & Pedersen, 1983). Less interference during search has also been observed for distractors grouped by proximity (Banks & Prinzmetal, 1976), good continuation (Donnelly, Humphreys, & Riddoch, 1991), and connectedness (Wolfe & Bennett, 1997) and for those located on an image plane tilted in depth (He & Nakayama, 1995).

Finally, there are indirect effects of perceptual grouping on performance in other tasks, such as the repetition detection task. In this task, subjects see a string of image elements, and they have to detect the repetition of one element; that is, they have to detect whether there are two adjacent elements that are identical. Repetitions are easier to detect if they are part of the same perceptual group than if they are part of different groups. It is likely that attention has to be directed to the repeating elements for accurate detection. The spread of attention according to the grouping cues can therefore account for the better performance if the repeating elements are part of the same group.

Thus, there is converging evidence for the role of Gestalt grouping cues in attentional processing. Attention spreads among related items to form perceptual groups, and linked items are thereby either jointly selected or blocked from further processing. Previous studies have remarked that image segmentation influences attention (e.g., Driver et al., 2001; Kahneman & Henik, 1981), but by distinguishing between linking and grouping, the IGT introduces a notation that is more precise: The representation of the image in the visual cortex enables a set of recurrent connections (linking), but these connections do not come into effect before attention starts to flow among the linked items (grouping).

Perceptual learning

If a particular grouping of features is critical for performance and is required often, the visual brain may reserve a feedforward cascade of feature detectors to detect the relevant conjunction. It can use a base-grouping strategy, although this requires more dedicated and specialized neurons than does the labeling of a distributed representation with an enhanced response. There is substantial evidence that perceptual experience induces new base groupings. The experiment of Vecera and Farah (1997) described above (Fig. 7b) indicates that our lifelong experience with upright letters makes them easier to segregate from each other, and base groupings for letters presumably form during childhood. Figure 13b illustrates how new base groupings would increase grouping speed in the example with the vacuum cleaner. Grouping of the floor brush and plug can occur on the basis of low-level features (circles with solid lines), but this would require the spread of attention across a large number of local links. The addition of more complex base groupings (dashed lines) would reduce the number of links that need to be traversed, and it can thereby speed up incremental grouping. Effects of training have indeed been observed in contour grouping (DeSchepper & Treisman, 1996; Kourtzi, Betts, Sarkheil, & Welchman, 2005), where a few hours of training have long-lasting effects.

Neurophysiological experiments have also witnessed the emergence of new base groupings. Baker, Behrmann, and Olson (2002) trained monkeys to discriminate between "batons," elongated objects consisting of two distinct shapes joined by a straight line. Initially, many neurons in the monkey's inferotemporal cortex were tuned to the local shapes, but not to the overall configuration. The situation changed after training, because neurons became selective for conjunctions between the shapes. Thus, new base groupings were formed after experience with behaviorally relevant feature conjunctions.

These results, taken together, suggest that new base groupings can be formed as the result of perceptual experience. Tasks that initially require incremental grouping may later be solved by base grouping if sufficient training

---

[2] Vecera (1994) actually referred to a set of linked locations as a *grouped array*. Here, we avoided this terminology to prevent confusion. We use the word *linked* for a set of image elements that are related to each other by Gestalt grouping cues. At a neurophysiological level of description, these elements are linked by a chain of enabled connections. We use *incrementally grouped* for the set of elements that are labeled by an enhanced response—that is, attention. These groupings have been made explicit by the labeling process and, thereby, become available for the subject's report.

has taken place. Training thereby makes perceptual grouping more efficient, since the groupings can now be extracted rapidly and in parallel, replacing the slower, serial incremental-grouping process.

Introspective phenomenology and availability for report

Many studies on Gestalt grouping cues have investigated the phenomenology of visual perception. Observers looked at stimuli and reported whether some of the image elements appeared to be grouped or not (Koffka, 1935; Wertheimer, 1923), an approach that also yielded many valuable insights in more recent years (e.g., Kellman & Shipley, 1991), although there are new methods that permit probing the effects of Gestalt grouping cues through other tasks, such as the detection of a repeating element in a row of elements (Palmer & Beck, 2007), the estimation of numerosity (Franconeri, Bemis, & Alvarez, 2009) or distance (Coren & Girgus, 1980; Vickery & Chun, 2010). Here, we have approached the problem of perceptual grouping from a neurobiological perspective, and we have also not put emphasis on introspective phenomenology. However, we believe that the labeling process sets the stage for the ability to report about the groupings (see also Avrahami, 1999). In the curve-tracing task in Fig. 5a, for example, the monkeys reported their percept by making an eye movement to the larger red circle at the end of the relevant curve. In this task, the early visual areas propagate the enhanced response along the target curve, and when the enhanced neuronal activity reaches the circle at the end of this curve, other areas involved in the planning of eye movements can read out the enhanced activity to initiate a saccade to the appropriate location in the visual field. *Global workspace* theories of awareness propose that the main difference between sensory features that do and do not enter awareness is determined by their influence on processing in other brain areas (Baars, 2002; Dehaene, Sergent, & Changeux, 2003). In the IGT, features that are labeled with enhanced neuronal activity have more influence on higher areas (Fig. 7), which is consistent with global workspace theories. This correspondence is further supported by the results of Supèr, Spekreijse, and Lamme (2001), who showed that the enhancement of V1 responses is strong on trials where a monkey perceives a stimulus and is weaker on trials where it does not.

As was discussed above, the image elements coded by a set of active neurons (coding base groupings) linked by enabled connections is not yet available for report because neurons in other areas are insensitive to the enabling of connections (Fig. 13a, lower panel). The incremental-grouping process first has to make these linkages explicit by enhancing the responses of a subset of the neurons so that the groupings can be read out. In other words,

unconsciously established links can only set the stage for incremental grouping, while the propagation of the enhanced neuronal response causes the formation of perceptual groups that are reportable.

Base groupings, in contrast, can exist outside awareness. Dehaene et al. (1998), for example, demonstrated that digits that are followed by a mask can exert unconscious priming effects, implying that the activation of neurons that code digits is insufficient for awareness. It has been proposed that reciprocal interactions between higher and lower visual areas is necessary for awareness in these situations (Dehaene et al., 2003; Lamme, 2003; Lamme & Roelfsema, 2000).

Comparison with previous theories

The IGT aims to link neurophysiology to psychology. This section will compare the IGT with previous theories that were specified at the psychological level (the FIT) and the neurophysiological level (binding-by-synchrony). We will start by mentioning theoretical developments and experimental findings that inspired the IGT. An important inspiration was the work of Ullman (1984), who suggested that vision starts with an early *base representation* that is driven by the visual stimulus and a later *incremental representation* that is modified by elemental operators like, for example, visual search and contour grouping that, when applied sequentially, can form visual routines. The implementation of these elemental operators in the visual cortex has been discussed elsewhere (Roelfsema, 2005; Roelfsema et al., 2000).

A number of previous studies have demonstrated that image segmentation and perceptual grouping do not always precede object recognition and attentive object selection. Driver et al. (2001), for example, described how grouping cues determine the spread of attention but that the opposite relation also holds: Attentional processes can influence segmentation. This article clearly indicated that theories with a strict succession of preattentive processes responsible for segmentation followed by an attentive processing stage must be incomplete. Studies by Peterson, Harvey, and Weidenbacher (1991) and Vecera and Farah (1997) also provided evidence for "late" grouping processes by demonstrating, for the first time, that image segmentation sometimes depends on the results of object recognition, contrasting with the more popular view that image segmentation precedes object recognition. Moreover, Vecera and O'Reilly (1998) demonstrated the plausibility of reciprocal interactions between object recognition and perceptual grouping processes in artificial neural networks (see also Behrmann, Zemel, & Mozer, 1998; van der Velde & de Kamps, 2001), and these results inspired the proposed

interactions between higher and lower visual areas illustrated in Fig. 7. An interesting precursor of the distinction between base and incremental grouping can be found in the work of Zucker and Davis (1988), who noted that the grouping of nearby dots can be mediated by visual cortical neurons tuned to orientation, while grouping of dots with a larger spacing requires another mechanism.

The feature integration theory and related theories

The FIT (Treisman & Gelade, 1980) has been a very influential theory of the role of attention in perceptual grouping. This theory holds that features such as colors, motions, and shapes are initially registered in separate feature maps. A spotlight of attention has to be directed to the location of an image element to highlight all its features in the various feature maps so that they are bound in perception. The FIT was the first to propose that attention can be used to represent feature conjunctions that are not coded by dedicated neurons (Treisman & Gelade, 1980; Treisman & Schmidt, 1982), and this insight plays an important role in the IGT.

The distinction between a preattentive and an attentive processing stage also features in related theories of visual perception (Neisser, 1967). Many theories of visual search have adopted a preattentive processing stage that accounts for parallel search, followed by a serial stage that compares individual display items with the representation of the search target in memory (Egeth, Virzi, & Garbart, 1984; Hoffman, 1979; Wolfe, 1994). In the IGT, base and incremental grouping map onto preattentive and attentive processing, respectively, and yet there are important differences between the IGT and the FIT and these other, previous theories.

First, the FIT considers only spatial attention: A spotlight or zoom lens is directed to the spatial location of a target item to bind its features into a coherent representation. The grouping of information at a single location is a relatively easy problem if compared with the grouping of features at different locations of a spatially extended object (see Shadlen & Movshon, 1999). The incremental-grouping theory therefore holds that conjunctions between features at a single spatial location are often coded as base groupings that are extracted in parallel. Many examples of feature conjunctions that are coded by single neurons have been found in neurophysiology (Kobatake & Tanaka, 1994; Leventhal et al., 1995), and human observers extract these conjunctions rapidly (see Fig. 12; Holcombe & Cavanagh, 2001) and in parallel (McLeod, Driver, & Crisp, 1988), which is inconsistent with the FIT.

The hard problem that requires attention is the grouping of features of spatially extended objects. In this situation, attention does not act as a spotlight, however, but, rather, adopts the shape of the relevant object (Fig. 4c). There is compelling evidence that attention can be object based, which means that it can be directed selectively to an object that overlaps with another object (Behrmann et al., 1998; Blaser, Pylyshyn, & Holcombe, 2000; Duncan, 1984; O'Craven, Downing, & Kanwisher, 1999; Scholl, 2001; Watson & Kramer, 1999).

Another important difference between the IGT, on the one hand, and the FIT and other object-based theories of attention, on the other hand, is that most of the previous theories suggest that Gestalt grouping takes place preattentively (but see Driver et al., 2001), whereas the IGT claims that this is true only for the detection of local base groupings and *linking* (i.e., enabling of connections), while the transitive combination of local groupings on the basis of Gestalt criteria requires incremental grouping (i.e., the spread of object-based attention). Thus, while the FIT suggests that processing delays are caused by shifts of the attentional spotlight from one object to the next, the IGT holds that delays are also caused by the time-consuming spread of attention within the object representations (Houtkamp et al., 2003). Finally, the IGT proposes that the spread of attention at the psychological level corresponds to the spread of enhanced activity at the neurophysiological level and explains mechanistically why attention coselects image elements linked by Gestalt grouping cues.

Texture segregation The FIT and related theories have proposed that texture segregation can be used to distinguish between basic features and feature conjunctions (Bergen & Julesz, 1983; Treisman & Gelade, 1980). These theories hold that if a texture can be subdivided into regions with different features, these regions segregate effortlessly, while feature conjunctions do not support effortless texture segregation. If the base groupings of the IGT had the same status as the features of the FIT, image regions with different base groupings should segregate effortlessly (Bergen & Julesz, 1983; Julesz, 1981; Malik & Perona, 1990; see also Beck, Graham, & Sutter, 1991; Sutter, Beck, & Graham, 1989). However, the IGT does not propose that a difference between base groupings is *sufficient* for segregation, because theories that use texture segregation to distinguish between basic features and feature conjunctions may have oversimplified the texture segregation process. Texture segregation is actually a very complex process that relies on two complementary processes, boundary detection and region growing, which require different types of interactions between neurons and, hence, different types of connections (Grossberg & Mingolla, 1985; Mumford et al., 1987; Roelfsema et al., 2002).

The first process, boundary detection, is sensitive to abrupt changes in feature value. The processes for

boundary detection are related to *pop-out*, the effortless detection of image elements that differ from their neighbors. Neurophysiologically plausible algorithms for boundary detection and pop-out hold that neurons tuned to similar features with adjacent receptive fields *inhibit* each other (Grossberg & Mingolla, 1985; Itti & Koch, 2001; Li, 1999), unlike in incremental grouping, where these cells *excite* each other (conjecture 8). Neurons coding elements of a homogeneous region receive maximal inhibition from their neighbors, while neurons coding elements at a boundary receive less inhibition. Boundaries therefore evoke more activity than do homogeneous image regions, and this is why boundaries are more salient.

In contrast, region growing is related to incremental grouping (Ullman, 1984) and requires *excitatory* connections between neurons tuned to the same features. The selection of an image element should promote the coselection of similar image elements, and the IGT suggests that an enhanced response should spread between neurons representing image elements of the same figure. This incremental grouping process corresponds to the spread of object-based attention over a homogeneous region of the texture (Ben-Shahar, Scholl, & Zucker, 2007). Roelfsema et al. (2002) showed that boundary detection and region growing can occur within the same hierarchical neural network. The key elements of this model are illustrated in Fig. 14. In the feedforward pathway, neurons tuned to the same orientation inhibit each other for boundary detection and pop-out (red connections in Fig. 14a) at multiple spatial scales (Fig. 14b). The feedback pathway has the opposite connection scheme, because neurons tuned to the same orientation *excite* each other to propagate the enhanced response within the figural region (incremental grouping: green connections in Fig. 14c). In this model, boundary detection occurs early after the presentation of an image and in parallel across the visual scene (Bravo & Blake, 1990; Cavanagh, Arguin, & Treisman, 1990), while region growing occurs later (Fig. 14d). Neurophysiological studies have demonstrated that boundary detection indeed precedes region growing during texture segregation (Lamme et al., 1999; Roelfsema, Tolboom, & Khayat, 2007).

The complexity of the mechanisms for texture segregation implies that there are several reasons why not all base groupings may permit effortless texture segregation so that the IGT cannot separate conditions with effortless texture segregation from conditions where segregation is effortful. For example, the inhibition generated by dense displays during feedforward processing may curtail the computation of complex base groupings in higher areas (as was discussed in the Base grouping: activation of tuned neurons section). Moreover, neurons in higher visual areas that code these more complex base groupings have large receptive fields that may preclude the precise localization of

boundaries. It is also conceivable that the boundary signal is too weak if neurons coding a particular base grouping are not strongly linked with inhibitory connections. Similar arguments can be made for the process of region growing that depends on feedback. Neurons in higher areas that code the more complex base groupings may not uniquely identify the image locations that belong to a homogeneous region because of their large receptive fields, and these cells may not provide feedback to the appropriate neurons in lower areas.

Visual search The FIT and related theories (e.g., Wolfe, 1994) have proposed that visual search can be used to distinguish between elementary features and feature conjunctions, but again, the complexity of search may have been underestimated. Search for features should be parallel, while search for feature conjunctions should be serial (Treisman & Gelade, 1980). However, the neuronal implementation of visual search is complex, and the absence of base groupings is only one of a number of possible reasons for search to become serial. First, dense displays may curtail feedforward processing and, hence, the computation of complex base groupings in higher areas. Moreover, models of search use inhibitory connections between neurons with similar tuning for pop-out (Bravo & Nakayama, 1992; Wolfe, 1994) (just as for texture segregation; see Fig. 14a), and not all base groupings may support pop-out. Models of search also use feedback connections from higher to lower areas to "guide" search (Bravo & Nakayama, 1992; Egeth et al., 1984). These connections propagate activity from higher areas coding the target of search to enhance the activity of neurons in lower areas that respond to target features (van der Velde & de Kamps, 2001; Wolfe, 1994). These feedback connections may not have sufficient selectivity if the base groupings are complex and not represented in early visual areas. Finally, most models of search require an additional matching process that compares candidate display items with a representation of the search target stored in memory ("examination" in Wolfe, 1994). This putative matching process is only partially understood. We therefore conclude that there are many reasons why a feature conjunction coded as a base grouping cannot be searched for in parallel and that the IGT does not predict when search becomes serial. One illustrative example is the search for items with a particular orientation. There are combinations of target and distractor orientations that do not permit parallel search (Wolfe, 1994). This finding is remarkable because there are many neurons in early visual areas that are highly selective for stimulus orientation. Differences between the processes required for visual search and texture segregation may also explain why combinations of image elements that permit effortless texture segregation do not always give rise to parallel search and, vice versa, why image elements that can be searched in
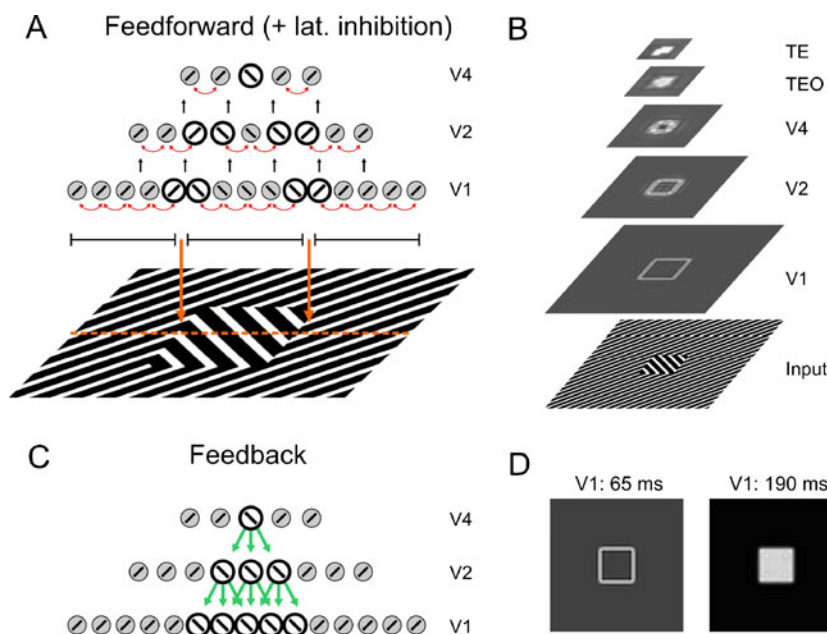
Fig. 14 Model for texture segregation. Neurons are tuned to the left and right orientations. **a** In the feedforward pathway, neurons tuned to the same orientation inhibit each other, a connection scheme opposite to that required for incremental grouping. Inhibition is weakest at orientation boundaries (larger cells). **b** Activity averaged across the two orientation maps. Note that the four boundaries merge in higher areas because receptive fields are larger. **c** Neurons in higher areas excite neurons tuned to the same orientation in lower areas to fill in the center of the figure with enhanced activity (incremental grouping). **d** Initially, V1 activity is enhanced at the boundaries, but at a later time, the response enhancement spreads to the figure center (modified from Roelfsema, Lamme, Spekreijse, & Bosch, 2002)

parallel do not always permit texture segregation (Wolfe, 1992). Another illustrative example is the demonstration that learning can change a condition where search is serial into one where it becomes parallel (Sireteanu & Rettenbach, 1995). One exciting possibility is that this learning relies on new feature conjunctions coded as base groupings, but it is also possible that other changes in the interactions between neurons are responsible, because the presence of base groupings is not sufficient for parallel search.

Illusory conjunctions In an influential study, Treisman and Schmidt (1982) demonstrated that subjects make conjunction errors: They sometimes perceive features in erroneous combinations. In their study, the subjects had to direct their attention to a large region in the display because their primary task was to report two briefly presented digits appearing at peripheral locations to the left and right of fixation. A number of colored letters were presented between the digits, and it was the subject's secondary task to report the identity and color of the letters. An illusory conjunction occurred if the subject reported a letter of the secondary task with a color that actually belonged to a different letter. Thus, the subject might, for example, report a brown T, although a blue T and a brown R had been presented. Treisman and Schmidt interpreted this result in the context of the FIT. They suggested that features within a large attentional focus are floating freely, because attention

has to constrict around one of the letters before the letter identity is correctly bound with the corresponding color.

Ashby, Prinzmetal, Ivry, and Maddox (1996) noted, however, that the subjects in this study (and other studies on illusory conjunctions) were much better at detecting genuine conjunctions than would be expected if features were completely free-floating and recombined randomly. This is consistent with the incremental grouping theory that proposes that base groupings provide hardwired conjunctions between shapes and colors at the same location. But why would subjects sometimes erroneously recombine features that are coded as base groupings? The answer may be that Treisman and Schmidt (1982) had to present the letters so briefly to observe conjunction errors that the subjects also made feature errors: That is, they reported a letter or a color not present in the display. The occurrence of conjunction errors is not incompatible with their coding as base groupings if the display duration is so short that even simple features are sometimes misperceived. It is conceivable that these short display durations occasionally cause the preserved representation of the identity of one letter and the color of another one while the feature conjunctions are lost, so that the subjects report the features that they did perceive and remember in erroneous combinations.

Conjunction errors occur more frequently between items that belong to the same perceptual group than between items of different groups (Prinzmetal, 1981). Illusory

conjunctions are, for example, more abundant between nearby items than between items at a larger separation (Cohen & Ivry, 1989). They also occur more frequently between items with a similar color, shape, or motion (Baylis, Driver, & McLeod, 1992; Ivry & Prinzmetal, 1991; Prinzmetal, 1981) and between items that are connected (Scholte et al., 2001). To explain the effects of grouping cues on illusory conjunctions, we note that the subjects may sometimes direct their attention to these items, even if they are part of a secondary task (see also Ashby et al., 1996; Treisman & Schmidt, 1982). The influence of grouping cues on the spread of attention section above reviewed the results of Kahneman and Henik (1981), who demonstrated that Gestalt grouping cues determine the features that are jointly reported in partial report tasks, presumably because these cues determine the spread of attention (for a similar view, see Prinzmetal, 1981). Thus, linked image elements are more likely to be coselected by attention, and their features are therefore more likely to be extracted together. It follows naturally that subjects are also more likely to report the features of different elements of a group in an erroneous combination than the features of items that belong to different groups.

Binding-by-synchrony

A second theory of binding that has been specified at the neurophysiological level is binding-by-synchrony, first proposed by von der Malsburg (1999; von der Malsburg & Schneider, 1986). This theory holds that neurons that code features of the same object are active synchronously; that is, they fire their action potentials at approximately the same time with a temporal resolution of a few milliseconds, while neurons coding features of different objects fire independently (Engel et al., 1992; Phillips & Singer, 1997; Singer & Gray, 1995; von der Malsburg, 1999; von der Malsburg & Schneider, 1986; Watt & Phillips, 2000). Thus, here a temporal tag, instead of an enhanced response, labels image elements to be grouped in perception. One of the hypothesized advantages of binding-by-synchrony is that multiple incremental groups may coexist in perception, because neurons coding the image elements of different groups would fire at different phases of an oscillation (Behrmann et al., 1998; Cowan, 2001; Singer & Gray, 1995). Neurons that participate in the representation of the same perceptual group can be synchronized, while there is no synchronization between groups. The IGT, on the other hand, has only a single label. If two groups of neurons are simultaneously labeled with the enhanced response, these two groups run the risk of merging into one large group. To test this difference between the theories, we recently devised a task where subjects had to group contour elements into two curves and found that only one incremental group could form at a time (Houtkamp & Roelfsema, 2010; in accordance with preliminary data by Jolicoeur, 1988). These results provide evidence compatible with the IGT and against binding-by-synchrony.

Clearly, the largest differences between binding-by-synchrony and the IGT are at the neurophysiological level. Experimental support for binding-by-synchrony was obtained in a series of neurophysiological studies demonstrating that the responses of visual cortical neurons evoked by features of the same object were better synchronized than the responses of neurons responding to features of different objects (Singer & Gray, 1995; reviewed by Eckhorn et al., 2001). Later studies did not support these findings, however, and cast doubt on the generality of binding-by-synchrony (Shadlen & Movshon, 1999; Thiele & Stoner, 2003). Some of the discrepancies between studies are presumably related to the use of anesthetized animals in some of them and the use of awake animals in others. Only a handful of studies have measured neuronal synchrony while the animals reported about their percepts, and these studies generally have not supported the binding-by-synchrony theory (Lamme & Spekreijse, 1998; Palanca & DeAngelis, 2005; Roelfsema et al., 2004; Thiele & Stoner, 2003). Roelfsema et al. (2004) even succeeded in dissociating grouping from synchrony by using the contour grouping task described above and created stimuli where the strength of synchrony between neuronal responses evoked by the same curve was weaker than the strength of synchrony evoked by different curves. These results imply that synchrony is not a universal code for binding. Instead, the contour elements that had to be grouped in perception were invariably labeled with enhanced neuronal activity, in accordance with the IGT.

Testing the incremental-grouping theory

We will close with a number of predictions of the incremental-grouping theory that can be tested experimentally. The first and major prediction is that grouping becomes serial whenever the following two conditions are met: (1) The task requires the transitive combination of local groupings, and (2) the overall configuration is unfamiliar, so that base groupings cannot have formed (Fig. 2b). We note that this prediction distinguishes the IGT from previous theories that did not envision serial grouping. We conjecture that serial grouping is required in many everyday scenes, which usually contain objects where multiple grouping cues have to be combined in a transitive manner. Figure 13b gives one example. The components of the vacuum cleaner are all indirectly connected, and the IGT predicts that grouping of all its components depends on a time-consuming incremental grouping process.

Fig. 15 Incremental grouping in natural images. **a** Picture with two zebras. **b** The picture activates representations of the texture elements in the visual cortex that are linked locally. Other neurons in the brain cannot "see" these links. **c** To represent all the image elements that belong to one of the zebras as a perceptual group, object-based attention spreads through the links to form a coherent representation of one of the zebras
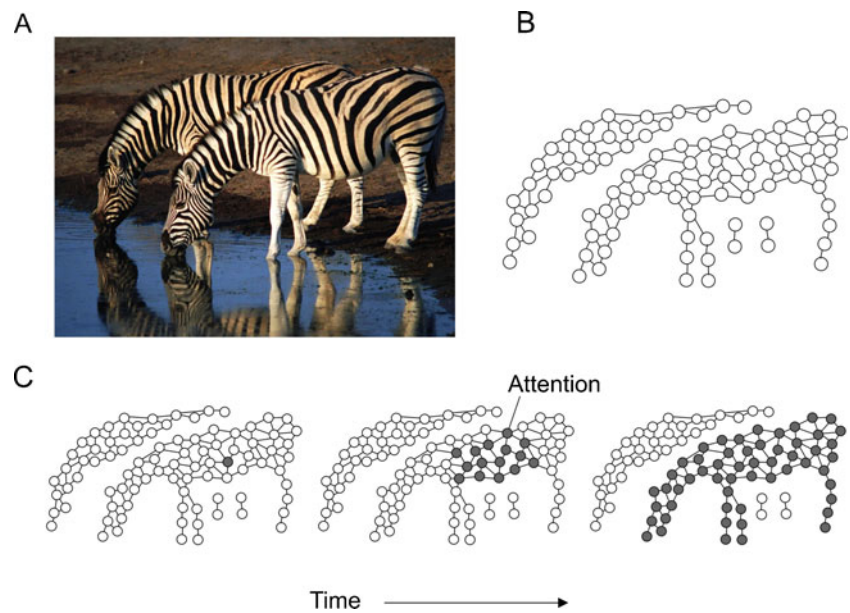
A second prediction is that attention has to spread over an extended object representation if grouping of its components is important. Consider the picture in Fig. 15a. Identification of these animals is straightforward and may depend on simple features like stripes, as well as more complex features. If the task is to group all image elements that belong to one of the zebras—for example, when checking whether it has four legs—attention would have to spread across these image elements. This is proposed to be a time-consuming process. Moreover, it should take more time to group image elements that are separated by a longer distance within the zebra. Psychophysical studies have confirmed this prediction for simple curves (Houtkamp et al., 2003; Scholte et al., 2001), but it has yet to be investigated for natural images. This incremental grouping process is predicted to be impossible if attention is directed elsewhere (Ben-Av et al., 1992). Future studies could also test the spread of attention with neurophysiological and neuro-imaging techniques. The theory predicts that the delays that occur during incremental grouping are the reflection of the gradual spread of an enhanced neuronal response.

Figure 16 illustrates that some tasks can rely on base grouping only, while other tasks with the same stimuli require incremental grouping. It is not too difficult to read the sentence in Fig. 16, although the letters are somewhat displaced and overlapping. Reading could rely on base groupings for these letters or even words, which are highly familiar objects. However, letter identification does not suffice if the task is to identify letters with two dots of the same color. This requires the precise localization of the dots, relative to the strokes, and has to happen in lower visual areas with a high spatial accuracy. In this situation, the enhanced response will have to propagate from higher visual areas that

represent letter identity to lower levels where neurons represent the location of the individual strokes (see the Interaction between lower and higher visual areas section).

A third prediction is the coexistence of two networks, a veridical network consisting of N-neurons and a labeling network composed of A-neurons. These networks should have ramifications in many visual areas that code different visual features such as motions, colors, shapes, and so forth. Previous neurophysiological studies on the effects of attention in these areas invariably observed neurons that were modulated by attention shifts and other neurons that were not, although the properties of the modulating neurons usually received the most emphasis (Luck, Chelazzi, Hillyard, & Desimone, 1997; Roelfsema et al., 1998; Treue & Maunsell, 1996). Future studies could examine whether the propagation of an enhanced response indeed depends on the specific interaction between A- and N-neurons proposed in Fig. 6c. A related aspect open to experimental testing is the proposed multiplicative interaction between the feedforward connections and recurrent connections. This specific interaction is responsible for the linking process—that is, the enabling of connections between neurons activated by feedforward connections and the disabling of connections between neurons that are not
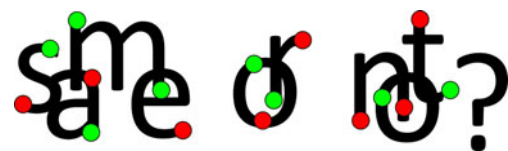
Fig. 16 Reading requires the identification of familiar letters, and this can rely on base grouping. If the task is to find letters with two dots of the same color, however, letter identification does not suffice. This task requires incremental grouping of low-level features (letter strokes), and the IGT holds that this will occur for only one letter at a time

activated by the stimulus. Previous studies have shown that attention mainly amplifies the response of well-driven neurons in the visual cortex and has comparatively little effect on neurons that are not well driven (McAdams & Maunsell, 1999; Treue & Martínez Trujillo, 1999), in support of such a multiplicative interaction. Future studies could test whether comparable interactions are responsible for the spread of the enhanced response among image elements related to each other by Gestalt grouping rules.

A fourth prediction derives from the availability of one label for incremental grouping. A strong prediction of the IGT is that it is not possible to simultaneously group two sets of image elements that are linked only transitively by a chain of local grouping cues. Our recent study (Houtkamp & Roelfsema, 2010) supported this prediction during a contour grouping task. We note for clarity that the theory does permit the coexistence of an incremental group with a number of base groupings that are extracted during feedforward processing.

The topic of perceptual grouping or *binding* has been a controversial issue for many years, with viewpoints ranging from the idea that binding is crucial for perception (Engel et al., 1992; Singer & Gray, 1995) to the view that most binding problems are solved during feedforward processing (Ghose & Maunsell, 1999; Riesenhuber & Poggio, 1999a). We believe that our consideration of a wide range of experimental findings in neurophysiology and psychology has provided a new conceptual framework for perceptual grouping that is able to reconcile many of the discrepancies. The predictions above are only a few of those made by the IGT, and it is exciting to anticipate the experiments that will put these predictions to the test.

References

Albright, T. D., & Stoner, G. R. (2002). Contextual influences on visual processing. *Annual Reviews in Neuroscience, 25,* 339–377.

Armstrong, K. M., Fitzgerald, J. K., & Moore, T. (2006). Changes in visual receptive fields with microstimulation of frontal cortex. *Neuron, 50,* 791–798.

Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review, 103,* 165–192.

Avrahami, J. (1999). Objects of attention, objects of perception. *Perception & Psychophysics, 61,* 1604–1612.

Baars, B. J. (2002). The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Sciences, 6,* 47–52.

Baker, C. I., Behrmann, M., & Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience, 5,* 1210–1216.

Banks, W. P., & Prinzmetal, W. (1976). Configurational effects in visual information processing. *Perception & Psychophysics, 19,* 361–367.

Barbu, A., & Zhu, S.-C. (2003). Graph partition by Swendsen–Wang cuts. In *Proceedings of the 9th International Conference on Computer Vision* (Vol. 1, p. 320). Washington, DC; IEEE Computer Society.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception, 1,* 371–394.

Baylis, G. C., & Driver, J. (1992). Visual parsing and response competition: The effect of grouping factors. *Perception & Psychophysics, 51,* 145–162.

Baylis, G. C., Driver, J., & McLeod, P. (1992). Movements and proximity constrain miscombinations of colour and form. *Perception, 21,* 201–218.

Beck, J., Graham, N., & Sutter, A. (1991). Lightness differences and the perceived segregation of regions and populations. *Perception & Psychophysics, 46,* 312–332.

Behrmann, M., Zemel, R. S., & Mozer, M. C. (1998). Object-based attention and occlusion: Evidence from normal participants and a computational model. *Journal of Experimental Psychology. Human Perception and Performance, 24,* 1011–1036.

Ben-Av, M. B., Sagi, D., & Braun, J. (1992). Visual attention and perceptual grouping. *Perception & Psychophysics, 52,* 277–294.

Ben-Shahar, O., Scholl, B. J., & Zucker, S. W. (2007). Attention, segregation, and textons: Bridging the gap between object-based attention and texton-based segregation. *Vision Research, 47,* 845–860.

Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature, 303,* 696–698.

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science, 308,* 529–534.

Blaser, E., Pylyshyn, Z., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature, 408,* 196–199.

Bonneh, Y. S., Cooperman, A., & Sagi, D. (2001). Motion-induces blindness in normal observers. *Nature, 411,* 798–801.

Borenstein, E., & Ullman, S. (2008). Combined top-down/bottom-up segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30,* 2109–2125.

Bosking, W. H., Zhang, Y., Schofield, B., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of Neuroscience, 17,* 2112–2127.

Bravo, M., & Blake, R. (1990). Pre-attentive vision and perceptual groups. *Perception, 19,* 515–522.

Bravo, M. J., & Nakayama, K. (1992). The role of attention in different visual-search tasks. *Perception & Psychophysics, 51,* 465–472.

Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience, 7,* 880–886.

Brincat, S. L., & Connor, C. E. (2006). Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron, 49,* 17–24.

Bundesen, C., & Pedersen, L. F. (1983). Color segregation and visual search. *Perception & Psychophysics, 33,* 487–493.

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prerontal and posterior parietal cortices. *Science, 315,* 1860–1862.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience, 7,* 308–313.

Cavanagh, P., Arguin, M., & Treisman, A. (1990). Effect of surface medium on visual search for orientation and size features. *Journal of Experimental Psychology. Human Perception and Performance, 16,* 479–491.

Celebrini, S., Thorpe, S., Trotter, Y., & Imbert, M. (1993). Dynamics of orientation coding in area V1 of the awake monkey. *Visual Neuroscience, 10,* 811–825.

Chan, W. Y., & Chua, F. K. (2003). Grouping with and without attention. *Psychonomic Bulletin & Review, 10,* 932–938.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature, 363,* 345–347.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. *Cerebral Cortex, 11,* 761–772.

Cohen, A., & Ivry, R. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology. Human Perception and Performance, 15,* 650–663.

Coren, S., & Girgus, J. S. (1980). Principles of perceptual organization and spatial distortion: The Gestalt illusions. *Journal of Experimental Psychology. Human Perception and Performance, 6,* 404–412.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences, 24,* 87–185.

Crundall, D., Cole, G. G., & Underwood, G. (2008a). Attentional and automatic processes in line tracing: Is tracing obligatory? *Perception & Psychophysics, 70,* 422–430.

Crundall, D., Dewhurst, R., & Underwood, G. (2008b). Does attention move or spread during mental curve tracing? *Perception & Psychophysics, 70,* 374–388.

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., et al. (1998). Imaging unconscious semantic priming. *Nature, 395,* 597–600.

Dehaene, S., Sergent, C., & Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences, 100,* 8520–8525.

DeSchepper, B., & Treisman, A. (1996). Visual memory for novel shapes: Implicit coding without attention. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 22,* 27–47.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Reviews in Neuroscience, 18,* 193–222.

Donnelly, N., Humphreys, G. W., & Riddoch, M. J. (1991). Parallel computation of primitive shape descriptions. *Journal of Experimental Psychology. Human Perception and Performance, 17,* 561–570.

Driver, J. (1995). Object segmentation and visual neglect. *The Behavioral and Brain Sciences, 71,* 135–146.

Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology. Human Perception and Performance, 15,* 448–456.

Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). Segmentation, attention and phenomenal visual objects. *Cognition, 80,* 61–95.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology. General, 113,* 501–517.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458.

Duncan, J., Humphreys, G. K., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology, 7,* 255–261.

Eckhorn, R., Bruns, A., Saam, M., Gail, A., Gabriel, A., & Brinksmeyer, H. J. (2001). Flexible cortical gamma-band correlations suggest neural principles of visual processing. *Visual Cognition, 8,* 519–530.

Edelman, S. (1987). Line connectivity algorithms for an asynchronous pyramid computer. *Computer Vision, Graphics, and Image Processing, 40,* 169–187.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology. Human Perception and Performance, 10,* 32–39.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology. General, 123,* 161–177.

Ekstrom, L. B., Roelfsema, P. R., Arsenault, J. T., Bonmassar, G., & Vanduffel, W. (2008). Bottom-up dependent gating of frontal signals in early visual cortex. *Science, 321,* 414–417.

Engel, A. K., König, P., Kreiter, A. K., Schillen, T. B., & Singer, W. (1992). Temporal coding in the visual cortex: New vistas on integration in the nervous system. *Trends in Neurosciences, 15,* 218–226.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16,* 143–149.

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

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field.". *Vision Research, 33,* 173–193.

Franconeri, S. L., Bemis, D. K., & Alvarez, G. A. (2009). Number estimation relies on a set of segmented objects. *Cognition, 113,* 1–13.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal cortex and inferior temporal cortices during visual categorization. *Journal of Neuroscience, 23,* 5235–5246.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics, 36,* 193–202.

Fukushima, K. (1988). A neural network for visual pattern recognition. *Computer, 21,* 65–75.

Gegenfurtner, K. R., Kiper, D. C., & Levitt, J. B. (1997). Functional properties of neurons in macaque area V3. *Journal of Neurophysiology, 77,* 1906–1923.

Ghose, G. M., & Maunsell, J. H. R. (1999). Specialized representations in visual cortex: A role for binding? *Neuron, 24,* 79–85.

Gigus, Z., & Malik, J. (1991). *Detecting curvilinear structure in images (Tech. Rep. 91/619).* Berkeley: University of California.

Gilchrist, I. D., Humphreys, G. W., & Riddoch, M. J. (1996). Grouping and extinction: Evidence for low-level modulation of visual selection. *Cognitive Neuropsychology, 13,* 1223–1249.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review, 92,* 173–211.

Grossberg, S., & Raizada, R. D. S. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research, 40,* 1413–1432.

Haimson, C., & Behrmann, M. (2001). Cued visual attention does not distinguish between occluded and occluding objects. *Psychonomic Bulletin & Review, 8,* 496–503.

Halligan, P. W., & Marshall, J. C. (1993). The history and clinical presentation of neglect. In H. Robertson & J. C. Marshall (Eds.), *Unilateral neglect: Clinical and experimental studies* (pp. 3–25). Hove, U.K.: Erlbaum.

Hamker, F. H. (2005). The reentry hypothesis: The putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cerebral Cortex, 15,* 431–447.

Harms, L., & Bundesen, C. (1983). Color segregation and selective attention in a nonsearch task. *Perception & Psychophysics, 33,* 11–19.

He, Z. J., & Nakayama, K. (1995). Visual attention to surfaces in three-dimensional space. *Proceedings of the National Academy of Sciences, 92,* 11155–11159.

Hess, R. F., & Field, D. J. (1999). Integration of contours: New insights. *Trends in Cognitive Sciences, 3,* 480–486.

Hoffman, J. E. (1979). A two-stage model of visual search. *Perception & Psychophysics, 25,* 319–327.

Holcombe, A. O., & Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience, 4,* 127–128.

Houtkamp, R., & Roelfsema, P. R. (2010). Parallel and serial grouping of image elements in visual perception. *Journal of Experimental Psychology. Human Perception and Performance, 36,* 1443–1459.

Houtkamp, R., Spekreijse, H., & Roelfsema, P. R. (2003). A gradual spread of attention during mental curve tracing. *Perception & Psychophysics, 65,* 1136–1144.

Humphreys, G. W., & Riddoch, M. J. (1993). Interactions between object and space systems revealed through neuropsychology. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 143–162). Cambridge, MA: MIT Press.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310,* 863–866.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2,* 194–203.

Ivry, R. B., & Prinzmetal, W. (1991). Effect of feature similarity on illusory conjunctions. *Perception & Psychophysics, 49,* 105–116.

Jolicoeur, P. (1988). Curve tracing operations and the perception of spatial relations. In Z. Pylyshyn (Ed.), *Computational processes in human vision: An interdisciplinary perspective* (pp. 133–168). Norwood NJ: Ablex.

Jolicoeur, P., & Ingleton, M. (1991). Size invariance in curve tracing. *Memory & Cognition, 19,* 21–36.

Jolicoeur, P., Ullman, S., & MacKay, M. (1986). Curve tracing: A possible basic operation in the perception of spatial relations. *Memory & Cognition, 14,* 129–140.

Jolicoeur, P., Ullman, S., & MacKay, M. (1991). Visual curve tracing properties. *Journal of Experimental Psychology. Human Perception and Performance, 17,* 997–1022.

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

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181–211). Hillsdale NJ: Erlbaum.

Kapadia, M. K., Ito, M., Gilbert, C. D., & 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,* 843–856.

Kayaert, G., Biederman, I., Op de Beeck, H. P., & Vogels, R. (2005). Tuning for shape dimensions in macaque inferior temporal cortex. *European Journal of Neuroscience, 22,* 212–224.

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology, 23,* 141–221.

Khayat, P. S., Pooresmaeili, A., & Roelfsema, P. R. (2009). Time course of attentional modulation in the frontal eye field during curve tracing. *Journal of Neurophysiology, 101,* 1813–1822.

Kimchi, R., & Peterson, M. A. (2008). Figure–ground segmentation can occur without attention. *Psychological Science, 19,* 660–668.

Kimchi, R., & Razpurker-Apfeld, I. (2004). Perceptual grouping and attention: Not all groupings are equal. *Psychonomic Bulletin & Review, 11,* 687–696.

Knierim, J. J., & Van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology, 67,* 961–980.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology, 71,* 856–867.

Koffka, K. (1935). *Principles of Gestalt psychology.* New York: Harcourt, Brace.

Komatsu, H., & Ideura, Y. (1993). Relationships between color, shape, and pattern selectivities of neurons in the inferior temporal cortex of the monkey. *Journal of Neurophysiology, 70,* 677–694.

Kourtzi, Z., Betts, L. R., Sarkheil, P., & Welchman, A. E. (2005). Distributed neural plasticity for shape learning in the human visual cortex. *PLoS Biology, 3,* 1317–1327.

Kovács, I., & Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure–ground segmentation. *Proceedings of the National Academy of Sciences, 90,* 7495–7497.

Kovács, I., & Julesz, B. (1994). Perceptual sensitivity maps within globally defined visual shapes. *Nature, 370,* 644–646.

Kramer, A. F., & Jacobson, A. (1991). Perceptual organization and focused attention: The role of objects and proximity in visual processing. *Perception & Psychophysics, 50,* 267–284.

Kreiman, G., Fried, I., & Koch, C. (2002). Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proceedings of the National Academy of Sciences, 99,* 8378–8383.

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology, 35,* 71–98.

Lamme, V. A. F. (2003). Why attention and awareness are different. *Trends in Cognitive Sciences, 7,* 12–18.

Lamme, V. A. F., Rodriguez-Rodriguez, V., & Spekreijse, H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cerebral Cortex, 9,* 406–413.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23,* 571–579.

Lamme, V. A. F., & Spekreijse, H. (1998). Neuronal synchrony does not represent texture segregation. *Nature, 396,* 362–366.

Large, M.-E., Aldcroft, A., & Vilis, T. (2005). Perceptual continuity and the emergence of perceptual persistence in the ventral visual pathway. *Journal of Neurophysiology, 93,* 3453–3462.

Lefebvre, C., Jolicoeur, P., & Dell'Acqua, R. (2010). Electrophysiological evidence of enhanced cortical activity in the human brain during visual curve tracing. *Vision Research, 50,* 1321–1327.

Leventhal, A. G., Thompson, K. G., Liu, D., Zhou, Y., & Ault, S. J. (1995). Concomitant sensitivity to orientation, direction, and color of cells in layers 2, 3, and 4 of monkey striate cortex. *Journal of Neuroscience, 15,* 1808–1818.

Li, Z. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proceedings of the National Academy of Sciences, 96,* 10530–10535.

Ling, S., & Carrasco, M. (2006). Sustained and transient covert attention enhance the signal via different contrast response functions. *Vision Research, 46,* 1210–1220.

Liston, D. B., & Stone, L. S. (2008). Effects of prior informatioin and reward on oculomotor and perceptual choices. *Journal of Neuroscience, 28,* 13866–13875.

Livingstone, M. S., & Hubel, D. H. (1987). Phychophysical evidence for separate channels for the perception of form, color, movement and depth. *Journal of Neuroscience, 7,* 3416–3468.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology, 77,* 24–42.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

Mack, A., Tang, B., Tuma, R., Kahn, S., & Rock, I. (1992). Perceptual organization and attention. *Cognitive Psychology, 24*, 475–501.

Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America. A, 7*, 923–932.

Mattingly, J. B., Davis, G., & Driver, J. (1997). Preattentive filling-in of visual surfaces in parietal extinction. *Science, 275*, 671–674.

McAdams, C. J., & Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque area V4. *Journal of Neuroscience, 19*, 431–441.

McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature, 332*, 154–155.

Miller, E. K., Gochin, P. M., & Gross, C. G. (1993). Suppression of visual responses of neurons in inferior temporal cortex of the awake monkey by addition of a second stimulus. *Brain Research, 616*, 25–29.

Mitroff, S. R., & Scholl, B. J. (2005). Forming and updating object representations without awareness: Evidence from motion-induced blindness. *Vision Research, 45*, 961–967.

Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature, 421*, 370–373.

Moore, C. M., & Egeth, H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology. Human Perception and Performance, 23*, 339–352.

Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *Journal of Neuroscience, 14*, 2178–2189.

Mumford, D., Kosslyn, S. M., Hillger, L. A., & Herrnstein, R. J. (1987). Discriminating figure from ground: The role of edge detection and region growing. *Proceedings of the National Academy of Sciences, 84*, 7354–7358.

Neisser, U. (1967). *Cognitive psychology*. Englewood Cliffs, NJ: Prentice-Hall.

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature, 401*, 584–587.

Oram, M. W., & Perrett, D. I. (1992). Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology, 68*, 70–84.

Palanca, B. J. A., & DeAngelis, G. C. (2005). Does neuronal synchrony underlie visual feature grouping? *Neuron, 46*, 333–346.

Palmer, S. E. (1992). Common region: A new principle of perceptual grouping. *Cognitive Psychology, 24*, 436–447.

Palmer, S. E., & Beck, D. M. (2007). The repetition discrimination task: An objective method for studying perceptual grouping. *Perception & Psychophysics, 69*, 68–78.

Palmer, S. E., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review, 1*, 29–55.

Pavlovskaya, M., Sagi, D., Soroker, N., & Ring, H. (1997). Visual extinction and cortical connectivity in human vision. *Cognitive Brain Research, 6*, 159–162.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*, 1136–1169.

Peterson, M. A., Harvey, E. M., & Weidenbacher, H. J. (1991). Shape recognition contributions to figure–ground reversal: Which route counts? *Journal of Experimental Psychology. Human Perception and Performance, 17*, 1075–1089.

Phillips, W. A., & Singer, W. (1997). In search of common foundations for cortical computation. *Behavior and Brain Sciences, 20*, 657–722.

Pooresmaeili, A., Poort, J., Thiele, A., & Roelfsema, P. R. (2010). Separable codes for attention and luminance contrast in the primary visual cortex. *Journal of Neuroscience, 30*, 12701–12711.

Pringle, R., & Egeth, H. E. (1988). Mental curve tracing with elementary stimuli. *Journal of Experimental Psychology. Human Perception and Performance, 14*, 716–728.

Prinzmetal, W. (1981). Principles of feature integration in visual perception. *Perception & Psychophysics, 30*, 330–340.

Prinzmetal, W., Nwachuku, I., Bodanski, L., Blumenfeld, L., & Shimizu, N. (1997). The phenomenology of attention: 2. Brightness and contrast. *Consciousness and Cognition, 6*, 372–412.

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

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience, 19*, 1736–1753.

Richard, A. M., Lee, H., & Vecera, S. P. (2008). Attentional spreading in object-based attention. *Journal of Experimental Psychology. Human Perception and Performance, 34*, 842–853.

Riesenhuber, M., & Poggio, T. (1999a). Are cortical models really bound by the "binding problem"? *Neuron, 24*, 87–93.

Riesenhuber, M., & Poggio, T. (1999b). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019–1025.

Rock, I., & Palmer, S. (1990). The legacy of Gestalt psychology. *Scientific American, 263*(6), 48–61.

Roelfsema, P. R. (2005). Elemental operations in vision. *Trends in Cognitive Sciences, 9*, 226–233.

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Reviews in Neuroscience, 29*, 203–227.

Roelfsema, P. R., Houtkamp, R., & Korjoukov, I. (2010). Further evidence for the spread of attention during contour grouping: A reply to Crundall, Dewhurst, and Underwood (2008). *Attention, Perception, & Psychophysics, 72*, 849–862.

Roelfsema, P. R., Khayat, P. S., & Spekreijse, H. (2003). Subtask sequencing in the primary visual cortex. *Proceedings of the National Academy of Sciences, 100*, 5467–5472.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395*, 376–381.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (2000). The implementation of visual routines. *Vision Research, 40*, 1385–1411.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nature Neuroscience, 7*, 982–991.

Roelfsema, P. R., Lamme, V. A. F., Spekreijse, H., & Bosch, H. (2002). Figure–ground segregation in a recurrent network architecture. *Journal of Cognitive Neuroscience, 14*, 525–537.

Roelfsema, P. R., Scholte, H. S., & Spekreijse, H. (1999). Temporal constraints on the grouping of contour segments into spatially extended objects. *Vision Research, 39*, 1509–1529.

Roelfsema, P. R., & Singer, W. (1998). Detecting connectedness. *Cerebral Cortex, 8*, 385–396.

Roelfsema, P. R., Tolboom, M., & Khayat, P. S. (2007). Different processing phases for features, figures, and selective attention in the primary visual cortex. *Neuron, 56*, 785–792.

Russell, C., & Driver, J. (2005). New indirect measures of "inattentive" visual grouping in a change-detection task. *Perception & Psychophysics, 67*, 606–623.

Salin, P. A., & Bullier, J. (1995). Corticocortical connections in the visual system: Structure and function. *Physiological Reviews, 75*, 107–154.

Sáry, G., Vogels, R., & Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science, 260,* 995–997.

Schall, J. D., & Thompson, K. G. (1999). Neural selection and control of visually guided eye movements. *Annual Reviews in Neuroscience, 22,* 241–259.

Schmidt, K. E., Goebel, R., Löwel, S., & Singer, W. (1997). The perceptual grouping criterion of colinearity is reflected by anisotropies of connections in the primary visual cortex. *European Journal of Neuroscience, 9,* 1083–1089.

Schneider, K. A. (2006). Does attention alter appearance? *Perception & Psychophysics, 68,* 800–814.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80,* 1–46.

Scholte, H. S., Spekreijse, H., & Roelfsema, P. R. (2001). The spatial profile of visual attention in mental curve tracing. *Vision Research, 41,* 2569–2580.

Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: A critical evaluation of the temporal binding hypothesis. *Neuron, 24,* 67–77.

Sharon, E., Galun, M., Sharon, D., Basri, R., & Brandt, A. (2006). Hierarchy and adaptivity in segmenting visual scenes. *Nature, 442,* 810–813.

Sheinberg, D. L., & Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. *Journal of Neuroscience, 21,* 1340–1350.

Sincich, L. C., & Horton, J. C. (2005). The circuitry of V1 and V2: Integration of color, form, and motion. *Annual Reviews in Neuroscience, 28,* 303–326.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Reviews in Neuroscience, 18,* 555–586.

Sireteanu, R., & Rettenbach, R. (1995). Perceptual learning in visual search: Fast, enduring, but non-specific. *Vision Research, 35,* 2037–2043.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature, 400,* 869–873.

Supèr, H., Spekreijse, H., & Lamme, V. A. F. (2001). Two distinct modes of sensory processing observed in monkey primary visual cortex. *Nature Neuroscience, 4,* 304–310.

Sutter, A., Beck, J., & Graham, N. (1989). Contrast and spatial variables in texture segregation: Testing a simple spatial-frequency channels model. *Perception & Psychophysics, 46,* 312–332.

Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science, 262,* 685–688.

Thiele, A., & Stoner, G. R. (2003). Neuronal synchrony does not correlate with motion coherence in cortical area MT. *Nature, 421,* 366–370.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520–522.

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

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Treisman, A. M., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95,* 15–48.

Treisman, A. M., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14,* 107–141.

Treue, S., & Martínez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 399,* 575–579.

Treue, S., & Maunsell, J. H. R. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature, 382,* 539–541.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Nature, 311,* 670–674.

Tsotsos, J. K., Rodríguez-Sánchez, A. J., Rothenstein, A. L., & Simine, E. (2008). The different stages of visual recognition need different attentional binding strategies. *Brain Research, 1225,* 119–132.

Ullman, S. (1984). Visual routines. *Cognition, 18,* 97–159.

van der Velde, F., & de Kamps, M. (2001). From knowing what to knowing where: Modeling object-based attention with feedback disinhibition of activation. *Journal of Cognitive Neuroscience, 13,* 479–491.

Vecera, S. P. (1994). Grouped locations and object-based attention: Comment on Egly, Driver & Rafal (1994). *Journal of Experimental Psychology. General, 123,* 316–320.

Vecera, S. P., & Farah, M. J. (1997). Is visual image segmentation a bottom-up or an interactive process? *Perception & Psychophysics, 59,* 1280–1296.

Vecera, S. P., & O'Reilly, R. C. (1998). Figure–ground organization and object recognition processes: An interactive account. *Journal of Experimental Psychology. Human Perception and Performance, 24,* 441–462.

Vickery, T. J., & Chun, M. M. (2010). Object-based warping: An illusory distortion of space within objects. *Psychological Science, 21,* 1759–1764.

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

von der Malsburg, C., & Schneider, W. (1986). A neural cocktail-party processor. *Biological Cybernetics, 54,* 29–40.

Wannig, A., Stanisor, L., & Roelfsema, P. R. (2011). Automatic spread of attentional response modulation along Gestalt criteria in primary visual cortex. *Nature Neuroscience.*

Watson, S. E., & Kramer, A. F. (1999). Object-based visual selective attention and perceptual organization. *Perception & Psychophysics, 61,* 31–49.

Watt, R. J., & Phillips, W. A. (2000). The function of dynamic grouping in vision. *Trends in Cognitive Sciences, 4,* 447–454.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung, 4,* 301–350.

Wolfe, J. M. (1992). "Effortless" texture segregation and "parallel" visual search are not the same thing. *Vision Research, 32,* 757–763.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238.

Wolfe, J. M. (1999). Inattentional amnesia. In V. Coltheart (Ed.), *Fleeting memories* (pp. 71–94). Cambridge, MA: MIT Press.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research, 37,* 25–43.

Wolfson, S. S., & Landy, M. S. (1998). Examining edge- and region-based texture analysis mechanisms. *Vision Research, 38,* 439–446.

Zeki, S. (1983). Colour coding in the cerebral cortex: The responses of wavelength-selective and colour-coded cells in the monkey visual cortex to changes in wavelength composition. *Neuroscience, 9,* 767–781.

Zucker, S. W., & Davis, S. (1988). Points and endpoints: A size/spacing constraint for dot grouping. *Perception, 17,* 229–247.