The implementation of visual routines

Pieter R. Roelfsema [a,*], Victor A.F. Lamme [a,b], Henk Spekreijse [a]

[a] *Department of Visual System Analysis, Academic Medical Center (UvA), Graduate School of Neurosciences Amsterdam, P.O. Box 12011, 1100 AC Amsterdam, The Netherlands*
[b] *The Netherlands Ophthalmic Research Institute, P.O. Box 12141, 1100 AC Amsterdam, The Netherlands*

Abstract

Many visual tasks can be decomposed into a sequence of simpler subtasks. Ullman suggested that such subtasks are carried out by elemental operations that are implemented by specialized processes in the visual brain [Ullman, S. (1984). Visual routines. *Cognition* (*18*), 97–159]. According to this hypothesis, there are a limited number of elemental operations that, since they can be applied sequentially, may nevertheless give rise to a large number of visual routines. Examples of such elemental operations are visual search, texture segregation and contour grouping. Here we attempt to delineate how such elemental operations are implemented in the visual brain. When an image appears, feedforward processing rapidly leads to an activity pattern that is distributed across many visual areas. Thereafter, elemental operations come into play, and these are implemented by the modulation of firing rates. Firing rate modulations effectuate grouping of neural responses into coherent object representations. Moreover, they permit transfer of information from one operator to the next, which allows flexibility in the sequencing of operations. We discuss how the elemental operations provide a tool to relate cortical physiology to psychophysics, and suggest a reclassification of pre-attentive and attentive processes. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Visual routines; Elemental operations; Firing rate modulations; Object-based attention; Contour integration; Curve tracing; Binding; Texture segregation; Perceptual organization; Visual search

1. Introduction

The human visual system has evolved to subserve multiple purposes. We are able to switch rapidly between visual tasks that are as divergent as driving a car, and reading and interpreting a road map. In the present article we elaborate on a hypothesis by Ullman (1984) that the many visual capabilities are derived from a much smaller set of elemental operations, loosely speaking, the 'instruction set' of the visual brain. According to Ullman's hypothesis, a visual routine can be assembled by the visual system by putting a number of elemental operations in a sequence. Different routines may share some of their elemental operations, and therefore, the visual system can compile a virtually unlimited number of routines. In the present article we outline a new theory on how elemental operations are implemented by the visual brain. We will indicate how the theory can account for a wide range of physiological and psychophysical data.

To illustrate a visual routine we refer to Fig. 1. Suppose that the task is to identify the shape that is connected to the letter B through a curve. Although this task may appear to be somewhat artificial, it is not too unlike other tasks that are broadly used in psychophysics. Indeed, this task can be decomposed in such a way that each elemental operation corresponds to a well-studied psychophysical task. The first elemental operation would be to locate the B, which corresponds to visual search for a letter. The second operation would be to trace the curve connected to the B to its other end. Jolicoeur, Ullman and MacKay (1986, 1991) and Pringle and Egeth (1988) have studied such a task. The third and final elemental operation would be to identify the shape at this position. It should be noted that it is possible to solve this task without making eye movements (the reader may verify this by fixating on the small circle in the middle of the stimulus), although it is easier if eye movements are allowed. We will only

---

* Corresponding author. Fax: +31-20-6916521.
E-mail address: p.roelfsema@ioi.knaw.nl (P.R. Roelfsema)
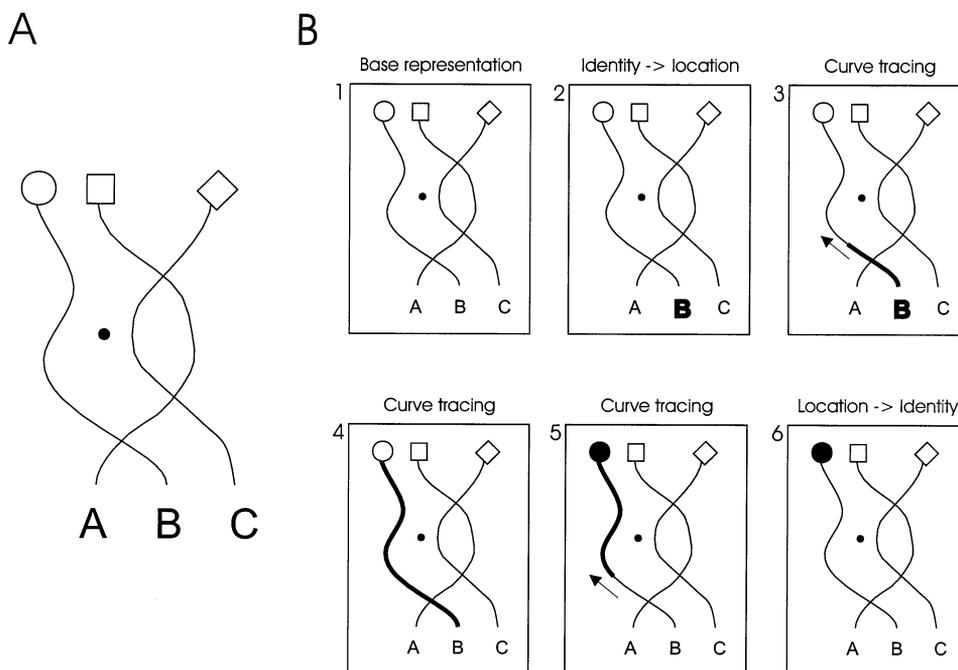
Fig. 1. Example of a visual routine that is composed of three elemental operators. (A) The task is to identify the shape that is connected to the 'B'. (B) First, the base representation of the image is extracted (step 1). The first elemental operation labels the location of the B with an enhanced firing rate (step 2). The second operator labels segments of the curve that starts at this position with an enhanced firing rate, eventually including the shape at the other end (steps 3–5). The third operator identifies this shape (step 6).

discuss the role of eye movements in visual routines briefly, since this subject is covered by a recent article (Ballard, Hayhoe, Pook & Rao, 1997).

Fig. 1B illustrates schematically how the visual cortex may implement such a sequence of elemental operations. Upon presentation of the figure, the visual cortex extracts a base representation, which consists of an activity pattern that is distributed across many cortical areas (not shown in Fig. 1). The base representation includes neurons in retinotopic visual areas such as the primary visual cortex, but also neurons in areas of the inferotemporal cortex, which are involved in object recognition. The defining property of the base representation is that it is computed rapidly (within about 120 ms, see Section 2), in a single feedforward sweep of activity through the hierarchy of visual areas. As soon as the base representation has been computed, elemental operations can be applied.

In our example, the first elemental operation can be called a type to place conversion, since the location needs to be found of an object whose type (the letter B) has been specified. It has been shown that neurons in several areas of the monkey brain that are selective for the identity of the target object exhibit an enhanced firing rate during visual search (Chelazzi, Miller, Duncan & Desimone, 1993; Rainer, Asaad & Miller, 1998). Other findings indicate that neurons selective for the location of the target object also enhance their firing rate during search (Schall & Hanes, 1993; Gottlieb,

Kusunoki & Goldberg, 1998; Schall & Thompson, 1999). Thus, both neurons that are selective for the location of the target object as well as neurons selective for its identity are eventually labeled by an enhanced firing rate, and this is the end result of the first elemental operation (Fig. 1B). The second elemental operator has to trace the curve to its other end. It can take the locus of enhanced activity that was brought about by the first operator as a starting point. From there, the operator has to label neurons that respond to the successive segments of the same curve. We have recently recorded from the visual cortex of monkeys performing a curve tracing task, and found that responses of neurons to the various segments of a curve that should be traced are labeled by an enhanced firing rate (Roelfsema, Lamme & Spekreijse, 1998). The third basic operator may be called a place to type conversion, and this conversion should be carried out at the end of the traced curve. Again, this operator may take advantage of the rate enhancement produced by the previous elemental operation in order to avoid processing of shapes at erroneous locations. In a number of studies it has been shown that if monkeys have to identify a single object among others, neurons that respond to the selected location have a firing rate that is enhanced relative to that of neurons responding to non-selected locations (Moran & Desimone, 1985; Motter, 1993, 1994a; Treue & Maunsell, 1996; Luck, Chelazzi, Hillyard & Desimone, 1997; Vidyasagar, 1998; McAdams

& Maunsell, 1999). The weaker neural responses to non-selected locations prevent their interference during the identification of the target shape (see Section 3.4.1). It should be noted that the elemental operations of our example are all associated with a modulation of firing rates in various areas of the visual cortex.

These rate modulations have been attributed to visual attention. Indeed, each of the elemental operations of our example corresponds to an established psychophysical task, and there is evidence for the involvement of visual attention in all of these tasks. First, there is an enormous amount of psychophysical evidence for the involvement of attention in visual search (reviewed by Posner & Presti, 1987; Treisman, 1988; Wolfe, 1992; Desimone & Duncan, 1995). Second, curve tracing is associated with attention to the curve that is traced (as will be discussed below). Third, attention is necessary for the identification of a visual shape at a prespecified location in a crowded display (Eriksen & Yeh, 1985; van der Heijden, Wolters, Groep & Hagenaar, 1987). Thus, all the elemental operations that comprise this visual routine appear to correspond to shifts of attention.

Many concurrent theories of visual perception distinguish between pre-attentive and attentive stages of visual information processing (Treisman & Gelade, 1980; Posner & Presti, 1987; Allport, 1989; Cave & Wolfe, 1990). According to these theories, pre-attentive processing is automatically and rapidly applied to the entire visual field, and it is therefore sometimes referred to as *early vision*. Perceptual grouping phenomena, which allow for effortless segregation of objects from their background are traditionally ascribed to the pre-attentive stage (Treisman & Gelade, 1980). When pre-attentive vision does not succeed in the segregation of objects from the background, the attentive system comes into play, but at the cost of prolonged processing times. It is tempting to identify the base representation with the pre-attentive stage, and the elemental operations with the attentive stage. Indeed, many groupings are formed in the base representation, and these will be referred to as *base groupings*. Base groupings are available as soon as the base representation has been computed, because they reflect the tuning of individual neurons. The scope of these base groupings is limited, and does not suffice in many visual tasks. In these tasks further groupings need to be formed by elemental operations. This second category of grouping operations will be called *incremental groupings*. However, some of the incremental groupings of our theory are traditionally attributed to the pre-attentive stage. This implies that the pre-attentive stage is not identical to the base representation.

In order to avoid confusion, and in accordance with Ullman (1984), we will rather use the term *incremental representation* for the neuronal activity pattern that is modified during the execution of visual routines. Thus, the incremental representation consists of rate modulations, which reflect intermediate results of operators that have been applied. Subsequent operators are sensitive to the actual content of the incremental representation. Therefore, rate modulations form a mechanism by which information is transferred from one operator to the next. It is the aim of the present article to review evidence in favor of some of the elemental operations and to suggest how they are implemented by the visual brain.

2. The base representation

Before visual routines can be applied, an early representation of the image has to be extracted by the visual cortex. This representation has been called base representation (Ullman, 1984) and is related to what Marr (1982) referred to as raw primal sketch[1]. However, our base representation differs from that of earlier theories, by the inclusion of neurons in higher areas of the visual cortex, such as the inferior temporal and parietal cortex. In our formulation, the difference between the base representation and the incremental representation is that the former is computed rapidly. Thus, the base representation is only early in time, but it is not restricted to the lower areas of the visual cortex.

The many areas of the visual cortex can be arranged in hierarchical scheme, on the basis of a difference between the layout of feedforward connections and feedback connections (Felleman & Van Essen, 1991). Feedforward connections run from lower to higher areas, and provide the visual drive of cortical neurons. The pattern of feedforward connections determines the location and extent of what is called the 'classical receptive field', i.e. the region from which a neuron can be activated by visual input (Allman, Miezin & McGuinness, 1985). The size of the classical receptive field increases in higher areas, because neurons receive afferents from neurons in lower areas with receptive fields that are displaced relative to each other. Thus, the classical receptive field of a cortical neuron corresponds to the aggregation of the receptive fields of neurons at lower levels projecting to it (Fig. 2A) (Salin & Bullier, 1995). An independent measure of the hierarchical level of a visual area can be derived from the latency of the first spikes that are evoked by a novel stimulus (Raiguel, Lagae, Gulyás & Orban, 1989; Dinse & Krüger, 1994; Nowak, Munk, Girard & Bullier, 1995). Visual latencies are longer for areas that are separated

---

[1] Marr (1982) distinguished between a raw primal sketch and a primal sketch. The latter also included perceptual groups that form on the basis of Gestalt criteria. These do not belong to what we call base representation.

from the input by more synapses. The latency of the first responses to a stimulus at a particular level of the processing hierarchy is typically about 10 ms longer than the latency at the preceding cortical level. The typical interval between two spikes of a single cortical neuron is longer than 10 ms. Therefore, it has been argued that neurons can only transmit a single spike to the next hierarchical level before cells at that level start to fire (Oram & Perrett, 1992; Tovée, 1994). This implies that the first spikes at a particular level L can only be influenced by the first spikes at level L-1, which, in turn, are evoked by the first spikes at level L-2[2]. Thus, the ensemble of neurons that participate in the first sweep of activity through the hierarchy of visual areas is primarily determined by the pattern of feedforward connections. This is the activity pattern that will be referred to as base representation.

The first spikes provide an indication of what is achieved by the feedforward connections, in terms of neural computation. The degree of tuning of the first
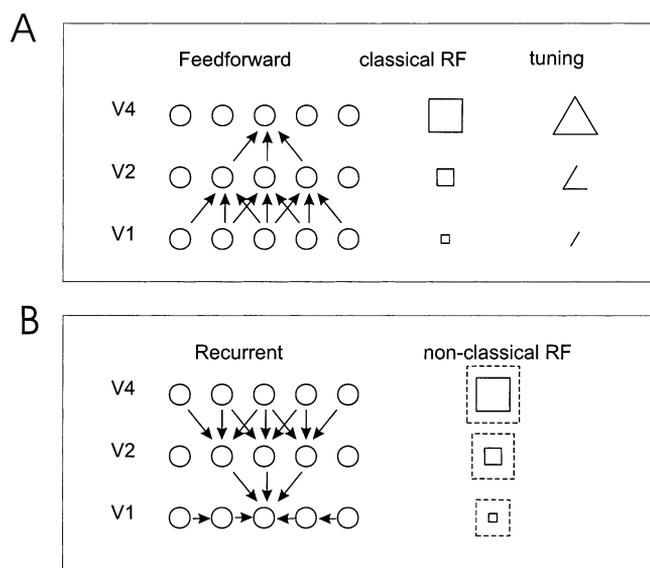

Fig. 2. Connections that contribute to the base and incremental representation. (A) Feedforward connections define the base representation. They determine a neuron's classical receptive field and its tuning. When the visual cortical hierarchy is ascended, receptive fields become larger and tuning becomes more complicated. (B) The incremental representation consists of response modulations that may originate outside the classical receptive field of neurons. These modulatory effects are not present during the cell's initial response to a novel stimulus, because they reach it through a longer route that involves lateral and feedback connections.

[2] A valid objection is that there are also feedforward connections that skip one or two levels. These would allow the second or third spike at level L to influence the first spike at level L + 2 or L + 3. The main point remains that the first spikes at any level are mainly determined by the layout of the feedforward connections, since these spikes cannot be influenced by feedback, and since there is also little time for within level connections to exert their influence.

spikes is remarkable in many cases that have been studied. For example, the first spikes in the inferior temporal cortex exhibit tuning to complicated stimuli such as faces (Oram & Perrett, 1992; Tovée, Rolls, Treves & Bellis, 1993; Kobatake & Tanaka, 1994). The activation of neurons that are selective for a face can be conceived of as a grouping operation, by which contours forming the mouth, eyes, and other visible attributes are grouped together by the visual system. These *base groupings* are available as soon as the feedforward sweep of activity reaches the inferior temporal cortex. Similarly, the first spikes evoked in area MST, an area of the parietal cortex contributing to motion perception, are tuned to complex patterns of optical flow (e.g. Graziano, Andersen & Snowden, 1994), thus grouping together the component motions. The base representation comprises the initial activity in all these visual areas. Nevertheless, the scope of the base groupings has to be limited. It is unlikely that there are cells in higher visual areas that are tuned to, for example, contour configurations of arbitrary complexity. The number of possible configurations readily outnumbers the cells of the entire visual cortex (von der Malsburg, 1981).

The first bottom-up sweep of activity also activates inhibitory neurons (e.g. Shao & Burkhalter, 1996). These feedforward inhibitory effects presumably provide an essential contribution to the tuning properties of neurons. In addition, they allow mutual inhibition among the representations of multiple objects. The magnitude of the response to a stimulus inside the classical receptive field is reduced when an additional, suboptimal stimulus is introduced into or close to the receptive field (Allman et al., 1985; Nelson & Frost, 1985; Schein & Desimone, 1990; Knierim & Van Essen, 1992; Miller, Gochin & Gross, 1993; Sengpiel, Baddeley, Freeman, Harrad & Blakemore, 1998; Reynolds, Chelazzi & Desimone, 1999). The region over which this response reduction is evoked scales with the size of the receptive field (Kastner, De Weerd, Desimone & Ungerleider, 1998). Since this response reduction is present from the onset of the visual responses, we include these influences in the base representation, although they are evoked from a region that is larger than the classical receptive field. This implies that the base representation of a crowded display does not equal the superposition of all activity patterns that would be evoked by the constituent objects if they were presented in isolation (this process which we call feedforward counterbalancing is discussed in more detail in Section 3.4.1).

3. The elemental operations

Neurons in any visual area may, in principle, receive information from the entire visual field through horizontal and feedback connections, even if their classical

receptive field is very small. These horizontal and feedback connections have a modulatory effect on a cell's response. Moreover, the route through which this information reaches a neuron is longer than that through the feedforward connections that define the classical receptive field (Fig. 2B). Therefore, these modulatory effects occur after an additional delay. Such delayed influences from beyond the classical receptive field have been uncovered in a number of studies (Lamme, 1995; Gottlieb et al., 1998; Roelfsema et al., 1998; Vidyasagar, 1998; Schall & Thompson, 1999), and are attributed to the elemental operations.

Elemental operations are required if a visual task is not solved during the first feedforward sweep of activity. In this article we will distinguish two broad categories of elemental operations. The first category of operations provides the incremental groupings, partial solutions to what is usually called the binding problem. This is the problem faced by the visual system in recovering all features that belong to a single visual object (von der Malsburg, 1981; Singer, 1993; Treisman, 1996). Visual cortical neurons are typically broadly tuned to one or a few feature dimensions, such as color, motion or shape. In the presence of multiple objects, the binding problem is equivalent to identifying the population of these broadly tuned neurons that are activated by *one* of the objects. Such a group of neurons is usually called an assembly (von der Malsburg, 1981). Thus, the search for the letter B in the example of the introduction (Fig. 1) corresponds to binding a location to the identity of the B. Similarly, curve tracing can be characterized as binding spatially separate contour segments into a coherent representation of the curve that is traced. In this section, curve tracing will serve as a prototype for an elemental operation of the first category. The neurophysiological mechanisms of curve tracing will first be discussed in some detail. Then we will generalize these mechanisms to other elemental operations of the first category. Thereafter, we will discuss the elemental operations of the second category. This is a more heterogenous class of 'maintenance' operations that preserve the incremental representation during changes in the visual image and during eye movements.

3.1. Contour grouping

Contour grouping will serve as a prototype operation of the first category. It is straightforward to situate the binding problem in an environment composed of contours (such as a jungle or a electronic print). Contour grouping has to occur in the incremental representation, because the base representation does not contain neurons tuned to arbitrary contour configurations. This implies that complicated curves need to be represented by assemblies of neurons. Such assemblies include neu-

rons at the lower levels of the visual cortical hierarchy, which encode the properties of each of the shorter contour segments that form a curve. These cells convey information about the precise location and orientation of the contour segment inside their small receptive field. In addition, the assemblies include neurons in higher visual cortical areas representing more abstract curve features, such as corners (Kobatake & Tanaka, 1994; Pasupathy & Connor, 1999). Cells at these higher levels do not convey information about the precise location of these features. But taken together, the various levels of the cortical hierarchy may form a relatively complete representation of a curve.

Unfortunately, such a distributed representation suffers from what von der Malsburg (1981) called the 'superposition catastrophe' if there are multiple curves in the image. Neurons that respond to features of one of the curves are, in principle, indistinguishable from neurons that respond to another curve. Thus, the collection of responses signals a set of curve features that are present in the image, but information about which feature belongs to which curve is lost. These additional groupings are formed in the incremental representation, for a single curve at a time. Contour grouping operations should exploit the Gestalt criteria of colinearity and connectedness, because closely spaced segments of the same curve are typically colinear and connected to each other (Kellman & Shipley, 1991; Field, Hayes & Hess, 1993; Palmer & Rock, 1994; Polat & Sagi, 1994). In this section, we will propose how the visual brain implements this elemental grouping operation.

3.1.1. The psychophysics of curve tracing

Curve tracing tasks probe the capability of subjects to indicate whether two contour segments belong to a single curve. Jolicoeur et al. (1986, 1991) and Pringle and Egeth (1988) investigated curve tracing using stimuli that consisted of two curves and two markers. One of the curves had to be traced in order to determine whether the markers were on the same or on different curves. An important result was that the reaction time depended on the distance between the markers. If the markers were on the same curve, the reaction time increased almost linearly with the distance between them. Jolicoeur et al. (1986) varied the distance between markers as measured along the curve, while holding the euclidean distance constant. Remarkably, the increase in reaction time still occurred, as if the curve tracing operator had to follow the entire curve. Moreover, tracing speed depended on the properties of the curve that was traced. Tracing speed is reduced if the curves have a high curvature or if they are close to other image elements (Jolicoeur et al., 1991). Jolicoeur et al. (1991) and McCormick and Jolicoeur (1991) accounted for these results by identifying the curve tracing operator with focal attention. According to their model, an

attentional zoom lens starts at one of the markers, and shifts along the curve until the other marker is reached. They suggested that the size of the zoom lens is varied during the tracing operation in order to ensure that only a single curve is within its aperture. When the traced curve is well separated from other image elements the zoom lens is wide, and few shifts are sufficient to trace a long distance. In contrast, if curves are close together or if the traced curve has a high curvature, the zoom lens is narrow, and more shifts are necessary to travel a fixed distance along the curve. This model gave a good quantitative fit to the pattern of reaction times observed during curve tracing (McCormick & Jolicoeur, 1991). We note, however, that such a model does not predict that all segments of a curve are illuminated together by the attentional spotlight, because it includes only a few segments of the curve at a time. In the present article we deviate from this hypothesis by suggesting that curve tracing is not related to shifts of *focal* attention, but rather to the spread of *object-based* attention among the various segments of the relevant curve (see also Roelfsema & Singer, 1998). We will suggest that an attentional label spreads along the entire curve in order to group its segments into a coherent representation.

3.1.2. Implementation of the curve tracing operator

We have argued previously that visual cortical neurons that respond to a single curve distribute a label among each other during curve tracing (Roelfsema & Singer, 1998). This label should spread selectively to neurons that respond to segments of the same curve. The corticocortical connections in the primary visual cortex reflect Gestalt criteria because they predominantly interconnect neurons tuned to nearby line elements that are approximately colinear (Bosking, Zhang, Schofield & Fitzpatrick, 1997; Schmidt, Goebel, Löwel & Singer, 1997). We therefore suggested that these connections contribute to the selectivity of label spreading. We have recently identified the label as an enhanced firing rate (see below). Our hypothesis is illustrated in Fig. 3, which shows a single layer of orientation selective neurons (rectangles with rounded corners). There are connections between neurons with colinear receptive fields (smaller rectangles), and also between neurons with overlapping receptive fields and a non-orthogonal orientation preference (these connections are not shown). The patterns at the top of Fig. 3 activate a subset of neurons that have a contour segment with the appropriate orientation in their classical receptive field. These cells, whose receptive fields are shown in gray, belong to the base representation. We propose that the curve tracing process is carried out by spreading an enhanced firing rate through the network of corticocortical connections. The selectivity of this process can be warranted if the label only spreads among neurons that belong to the base representation. Thus, only connections between neurons of the base representation propagate the label (black connections), whereas the other connections are not activated (white connections). We refer to the ensemble of enabled connections as the *interaction skeleton*. The curve tracing operation can be initiated by inserting the rate enhancement at one of cells that respond to the curve that should be traced. It is easy to see in Fig. 3 that this label will eventually spread to the cells that respond to the various segments of this curve. A slightly altered image, as shown in Fig. 3B, is associated with a slightly different base representation, and therefore, with a different interaction skeleton. For this image, the rate enhancement will also spread to the appropriate neurons. Note that a local change, confined to the center of the image, results in a global change in the interaction skeleton, and a similar change in the label spreading operation (Roelfsema & Singer, 1998).

How can this proposal account for the dependence of tracing speed on the distance between curves (Jolicoeur et al., 1991)? We assume that the label is simultaneously spread at multiple levels of the visual cortical hierarchy (as was previously proposed by Edelman, 1987). This hypothesis is illustrated in Fig. 4, which shows receptive fields of V1 and V2 neurons that are incorporated into the base representation of the stimulus of Fig. 3A. In Section 2 we indicated that neurons that have multiple image components in their receptive field receive inhibitory feedforward input (feedforward counterbalancing, see Section 3.4.1). This causes a hole in the base representation of area V2, at the location where the two curves come closest. As in Fig. 3, the intra-areal connections are between neurons that respond to approximately colinear contour segments. The length of these connections is relatively constant across visual areas (Salin & Bullier, 1995). However, higher areas have a representation of the visual field that is more compressed. This implies that an axon with a constant length traverses more of the visual field in higher visual areas than in lower visual areas. Thus, in higher areas larger distances of the visual field can be crossed in just a few synapses, and the labeling operation can progress faster. Initially, in the example of Fig. 4, labeling is therefore fastest in area V2. We assume that the rate enhancement also spreads between neurons with overlapping receptive fields at different hierarchical levels, through feedforward and feedback connections, which is in accordance with anatomical data (Salin & Bullier, 1995). Feedback connections relay the enhanced labeling speed of area V2 to area V1 (double arrow at $t = 2$). As soon as the hole in the V2 base representation is reached, however, label spreading is confined to area V1, and this reduces its speed. At $t = 5$, V1 neurons at
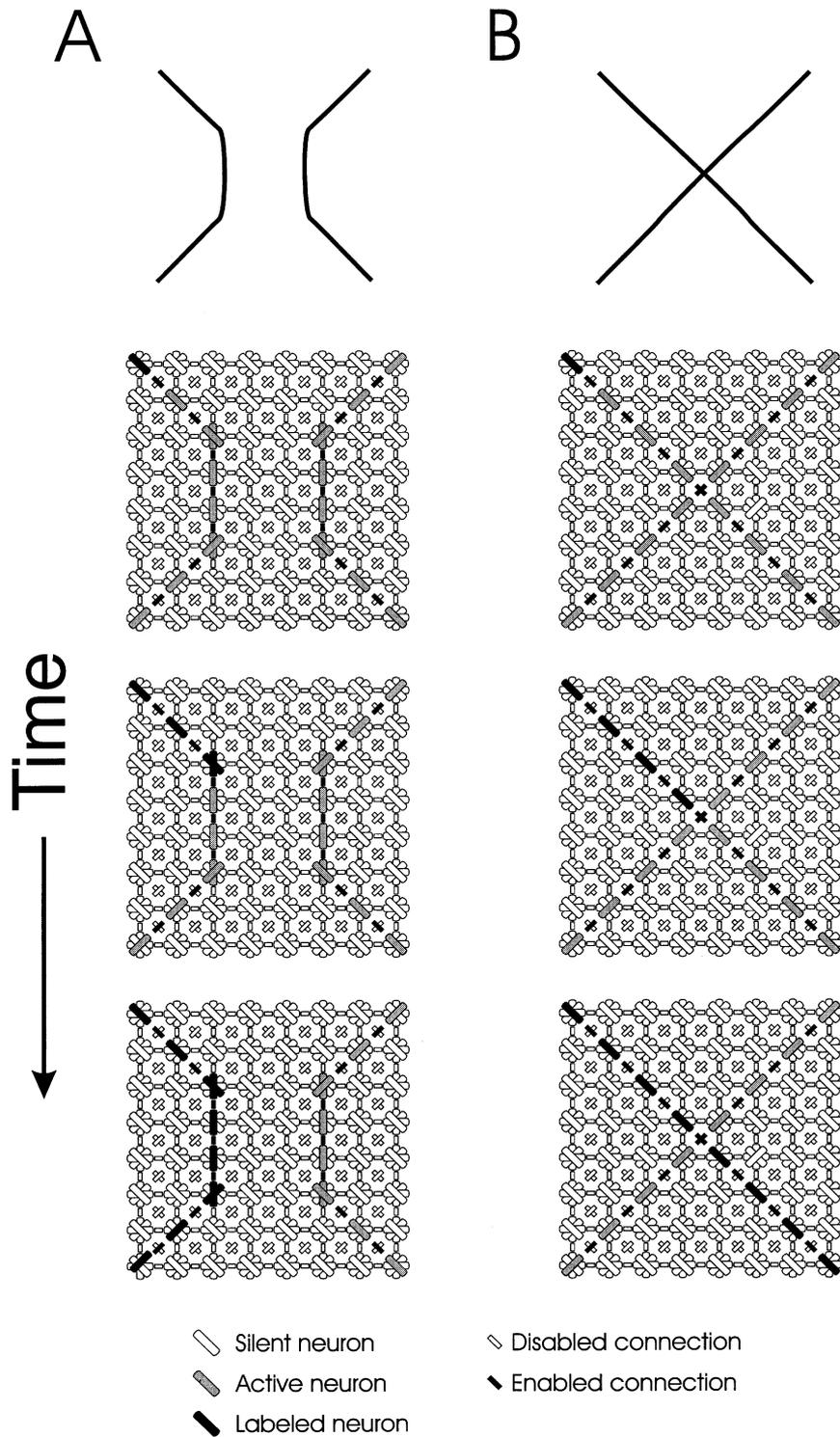
Fig. 3. Implementation of the curve tracing operator. Neurons selective for various orientations are shown as rectangles with rounded corners. Neurons that are activated by the images (top) form the base representation and are shown in gray and black. Horizontal connections (small rectangles) link neurons with nearby receptive fields that are tuned to colinear orientations. Neurons tuned to different orientations but with overlapping receptive fields are also interconnected (not shown). A label (black) is spread through these connections, but only among activated cells. This implies that connections between activated cells are enabled (black). If a label is injected at one of the activated cells it will spread to all neurons that respond to segments of the same curve.

the position of the hole in V2 have received the rate enhancement (double arrow), and the hole is filled in subsequently. The process by which such holes can be filled due to labeling at a lower level has been called 'label evaluation' (Roelfsema, 1998). Label evaluation will be described below in some more detail (Section 3.4.1). After the hole has been filled, the spread of the rate enhancement can be taken up again by V2, and its speed increases accordingly.

The scheme of Fig. 4 is simplified, because higher cortical areas have been omitted. Additional hierarchical levels speed up curve tracing even further, in particular if the curve that should be traced is well isolated in the visual field. These proposals are in accordance with an analysis by Mahoney and Ullman (1988), who referred to these larger curve segments that are extracted during the computation of the base representation as 'chunks'. A recent study by Vecera & Farah (1997) provides some indications for the representation of such chunks. The speed of curve tracing was compared between familiar and unfamiliar contour shapes. For example, the time to decide whether two markers were on a single contour was investigated for contours belonging to normal and inverted letters. Tracing speed
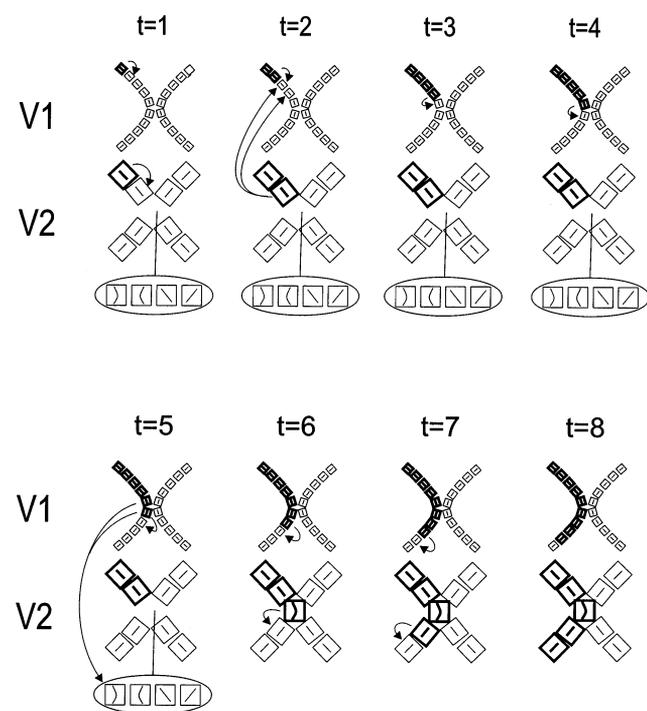

Fig. 4. Simultaneous label spreading in areas V1 and V2. Receptive fields are shown of neurons that belong to the base representation. The proximity of the curves causes a hole in the base representation of area V2. The speed of labeling is highest in area V2, but at the hole labeling can only occur in area V1. When labeling in area V1 has crossed this location, the inhibition is relieved for one of the V2 neurons and the hole is filled in (at $t = 6$). Ellipse, neurons in area V2 that receive feedforward excitatory and inhibitory input.

was highest for contours forming a normal letter. This finding has important implications. First, it shows that areas related to object recognition may provide chunks, which speed up the curve tracing process. When neurons representing such chunks exhibit the rate enhancement, responses to all contour segments that form the chunk may also rapidly receive the label through feedback connections (as in Fig. 4). Second, these results suggest that chunks are formed by visual experience.

3.1.3. Physiological evidence for a labeling operation

In order to investigate the validity of our theory, we recently recorded from the primary visual cortex of monkeys that were trained to perform a curve tracing task (Fig. 5) (Roelfsema et al., 1998). The monkeys had to maintain visual fixation at a point on a computer monitor, and to trace the curve connected to this point. The animals had to report, which of two circular targets was at the other end of this curve by making an eye movement to it, and we were interested in the neural activity in the episode before the monkey moved his eyes. We used pairs of complementary stimuli in which one or the other of two curves was connected to the fixation point. The difference between stimuli was typically far from the neurons' classical receptive field. Therefore, the content of the receptive field was identical for the two stimuli (Fig. 5B). However, for one of the stimuli the receptive field was on the curve that had to be traced, and for the other it was on the distractor curve. On average, the firing rate of the neurons was 30% stronger if their receptive field was on the traced curve. This enhancement occurred for responses to the various segments of this curve, and even if it crossed another curve (as in Fig. 3B), suggesting that the entire cortical representation of the traced curve 'lights up' during curve tracing.

These results indicate that curve tracing is associated with the spread of a label through the network of corticocortical connections. The identity of this label is an enhanced firing rate. Neurons started to respond about 40 ms after stimulus presentation, which is a typical latency for neurons in the primary visual cortex. These initial responses reflect the feedforward input from the lateral geniculate, and form the base representation. The rate enhancement due to curve tracing occurs later, at a latency longer than 150 ms (Fig. 5C). We recently confirmed that segments close to the fixation point are labeled earlier than segments that are more remote, as predicted by the theory (Roelfsema & Spekreijse, 1999). The V1 neurons could only be informed about the difference between complementary stimuli through polysynaptic pathways, since this difference occurred at a location that was remote from their classical receptive field (Fig. 2B). Clearly, the rate enhancement belongs to the incremental representation.
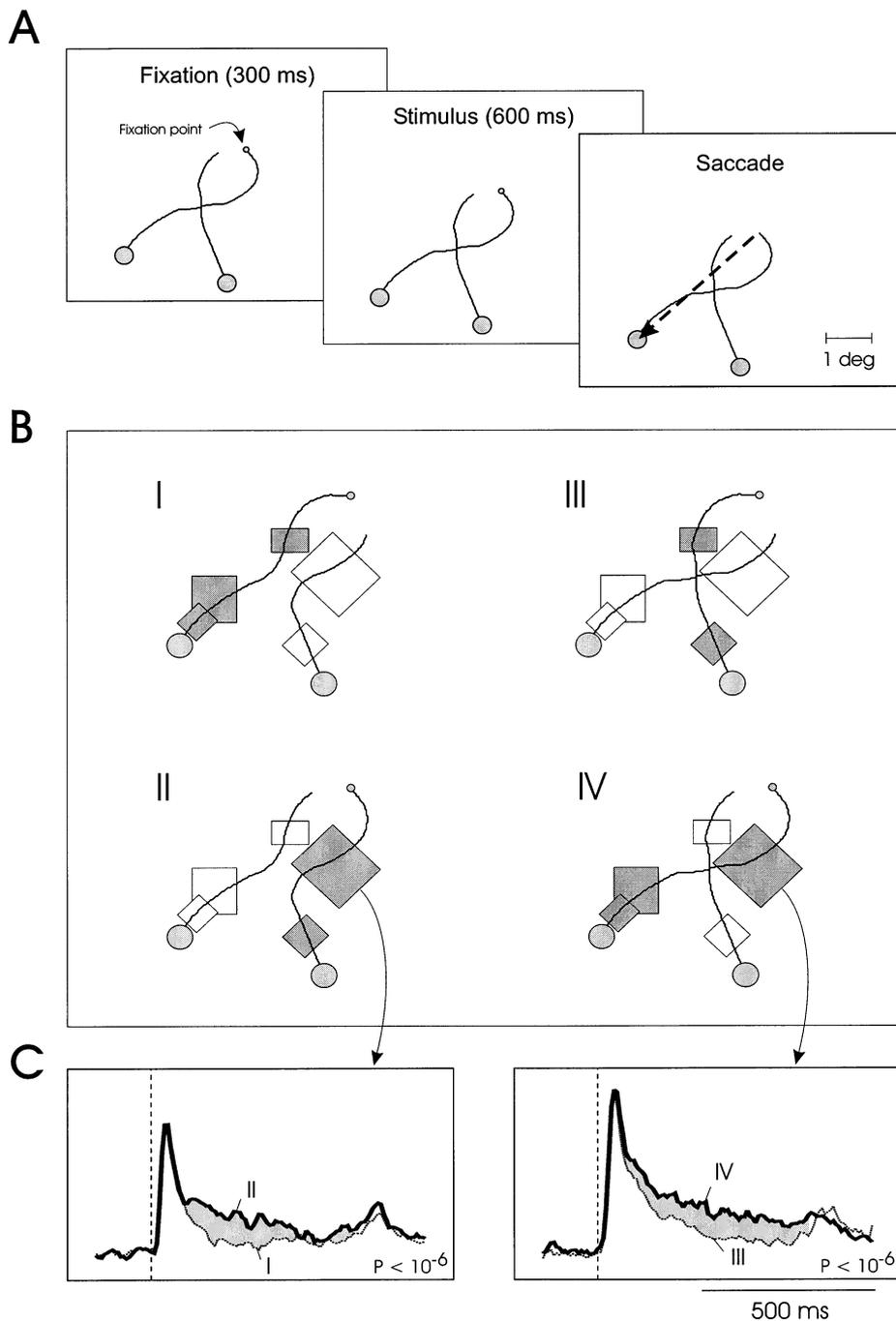
Fig. 5. Response modulations in the primary visual cortex during curve tracing. (A) Behavioral task. Monkeys had to fixate a point in the middle of a display. While the monkey maintained fixation, a stimulus was shown for 600 ms that consisted of two curves and two circles. One of the curves (target curve) connected the fixation point to one of the circles (target circle). The other curve and circle were distractors. Hereafter, the fixation point was removed and the monkey had to make an eye movement to the target circle. Neural responses were only evaluated in the interval during which the monkey maintained fixation. (B) Receptive fields of five multineuron recordings in area V1. Receptive fields of neurons that exhibited a response enhancement are shown in gray. Note that the response enhancement labels the various segments of the target curve. (C) Responses of one of the cell groups. Thick line, response to the target curve; thin line, response to the distractor curve; gray region, response modulation. The vertical stippled line indicates stimulus onset (modified from Roelfsema et al., 1998).

3.1.4. Involvement of visual attention in curve tracing

Other physiological studies have revealed comparable rate modulations in various visual cortical areas, during tasks in which attention had to be directed to one of a number of visual objects (Bushnell, Goldberg & Robinson, 1981; Moran & Desimone, 1985; Motter, 1993; Treue & Maunsell, 1996; Luck et al., 1997; Vidyasagar, 1998). These studies agree that neural re-

sponses to attended items are enhanced relative to responses to unattended items. If the rate enhancement during curve tracing is also a correlate of attention, a traced curve should be perceptually enhanced relative to other curves. We have recently investigated the validity of this prediction using a dual task design (Scholte, Lamme, Spekreijse & Roelfsema, 1999). The primary task of the subjects was to trace one of two curves. Some segments of the curves were colored, and it was the secondary task to report one of the colors. Colors of the traced curve were reported much more reliably than colors of the distractor curve. This provides direct evidence for the involvement of attention in curve tracing.

The physiological data indicate that the end result of the tracing operation is a state in which neural responses to all segments of the target curve are simultaneously enhanced, even this curve overlaps with another curve. Thus, curve tracing is associated with *object-based* attention. Temporal delays have to be attributed to the gradual spread of attention along the curve that is traced. This implies a 'growth cone' model of attentional selection. At the location of the interaction skeleton where the rate enhancement spreads, new contour segments are added to the representation of the attended curve.

3.1.5. Curve tracing and perceptual grouping

We suggested that during curve tracing contour segments are grouped into extended curves. This grouping operation is apparently associated with visual attention to the curve that is being traced. These findings fit remarkably well in the feature integration theory of Treisman and co-workers (e.g. Treisman & Gelade, 1980). In this theory attention functions as a glue that integrates all features that belong to a single perceptual object into a coherent representation of that object. In the feature integration theory the emphasis is on focal attention that needs to be directed to the spatial location of the object. In case of curve tracing, such an operation fails if the traced curve cannot be included in the focus of attention without avoiding the intrusion of other image components. However, it is a straightforward extension of the feature integration theory to assume that object-based attention fulfills an equivalent role in situations in which the features of an object cannot be selected on the basis of the spatial location alone, for example because the object is spatially overlapping with another object (e.g. Duncan, 1984). Such an interpretation is supported by our results.

3.1.6. Is contour grouping attentive or pre-attentive?

There is a profound implication of these findings for theories of visual perception that are based on the distinction between pre-attentive and attentive processing stages, which was originally introduced by Neisser (1967) (e.g. Treisman & Gelade, 1980; Kahneman & Henik, 1981; Duncan, 1984; Posner & Presti, 1987; Duncan & Humphreys, 1989; Wolfe, 1994). These theories suggest that Gestalt criteria are applied pre-attentively and with unlimited capacity, in order to define perceptual groups. Curve tracing should occur pre-attentively, according to these theories, because it is equivalent to parsing the image on the basis of the Gestalt criteria of colinearity and connectedness. Our analysis indicates that neurons responding to the various segments of a single curve are indeed linked by the interaction skeleton as soon as the base representation is computed. However, this implicit linkage is without effect up to the moment that the label spreading operation starts. Only after the spread of the rate enhancement a collection of contour segments is grouped *explicitly* into a coherent representation of a single curve. Explicit grouping therefore takes time, and is associated with attention to the traced curve. This implies that the Gestalt criteria of colinearity and connectedness cannot be applied with unlimited capacity during curve tracing.

A number of recent studies showed that a row of colinear oriented elements readily stands out from a background of elements with a random orientation (e.g. Field et al., 1993; Kovács & Julesz, 1993). Thus, these studies support effortless contour grouping. How can these results be reconciled with curve tracing data? We believe that the contradiction between these 'pre-attentive' contour grouping tasks and curve tracing is related to a difference between distractors (see also Roelfsema, Scholte & Spekreijse, 1999). In the studies reporting rapid and effortless contour grouping, distractors had a random orientation. There are feedforward algorithms for the detection of contours defined by colinear line elements among randomly oriented background elements (e.g. Gigus & Malik, 1991). Thus, this form of grouping may actually be supported by the base representation. In contrast, in curve tracing tasks there are a number of curves with equally colinear segments. These feedforward algorithms do not provide a solution in curve tracing tasks, because the degree of colinearity need not differ between the curve that is traced and other curves. In fact, the traced curve may even be more contorted than some of the distractors.

3.2. Architecture of the interaction skeleton in other feature domains

Curve tracing served as an example of how the binding problem can be solved by applying an elemental operation. It is our aim to indicate how these mechanisms can be generalized to solve binding prob-
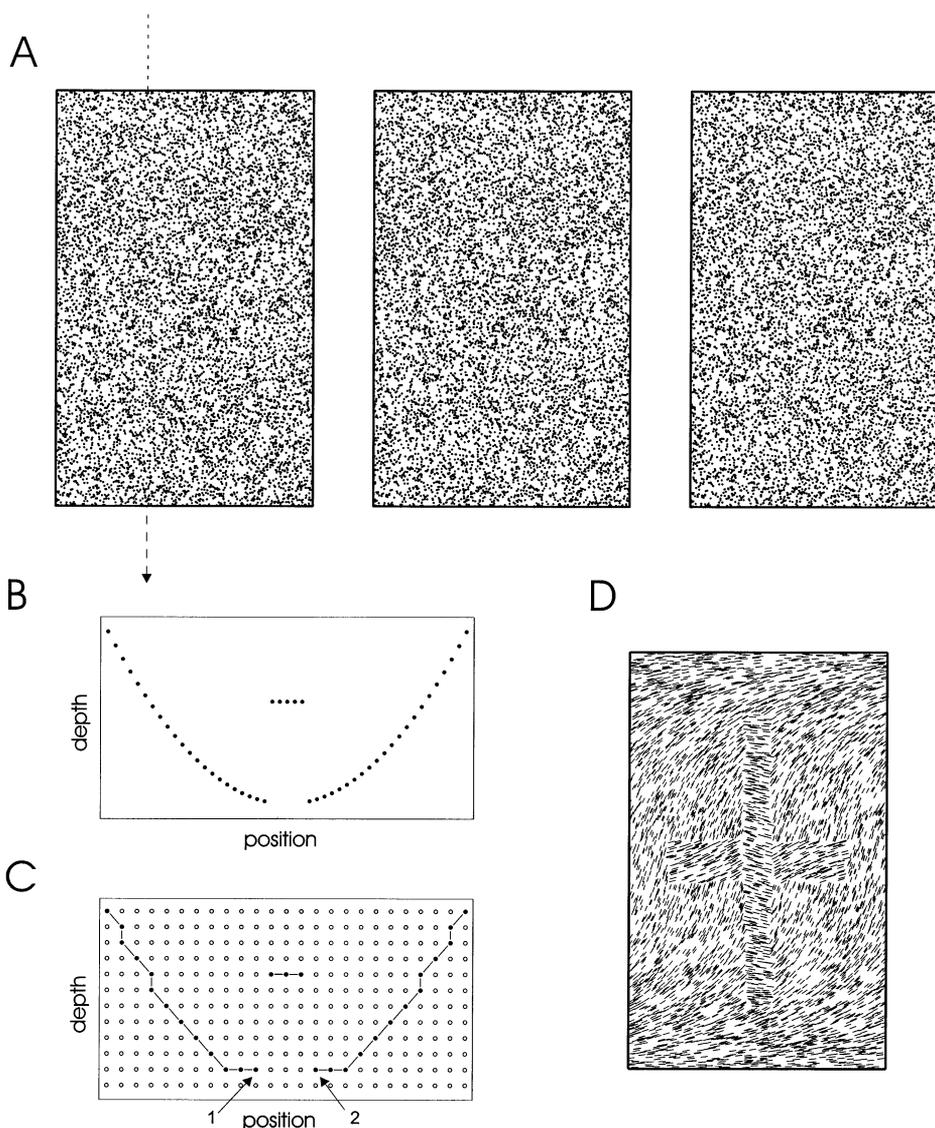
Fig. 6. Label spreading in two dimensions. (A) The left two images should be fused with uncrossed eyes, and the right two images by crossing the eyes. (B) Depth profile along a line in the image. (C) The interaction skeleton consists of connections between activated neurons (black). (D) Similar image, but here coherent image regions are defined by orientation rather than depth.

lems in other feature domains[3]. Let us therefore isolate the essential ingredients of our scheme. First, the base representation is computed. It contains base groupings that are hardwired into the system. Further groupings are only implicitly available, because neurons that are likely to respond to features that belong to a single

object are linked by corticocortical connections. The base representation subdivides these corticocortical connections into a disabled category and an enabled category. Enabled connections make up the interaction skeleton. By the time that the base representation has been computed, neurons that respond to features of the same object are linked by the interaction skeleton. This also holds for neurons that respond to widely separated image elements, although these are only indirectly connected through a chain of cells responsive to interspersed image elements. A rate enhancement has to spread through the interaction skeleton in order to make these additional groupings explicit.

It is relatively straightforward to incorporate additional grouping criteria into the architecture of corticocortical connections. One of the Gestalt rules states, for

---

[3] Many of our conjectures are closely related to previous theories on the role of synchronization in feature binding (e.g. Sporns, Gally, Reeke, Jr & Edelman, 1989; Hummel & Biederman, 1992; Schillen & König, 1994; Roelfsema et al., 1996). In these theories it was suggested that synchronization is the label that is distributed among neurons that respond to attributes of a single object. We deviate from these earlier theories by suggesting that the label is a rate enhancement. However, the architecture of the connections responsible for spreading the synchrony and the rate label can be virtually identical (see also Roelfsema & Singer, 1998).

example, that image elements with a similar shape, color, depth or motion are likely to be grouped together. This rule can be implemented by interconnecting neurons that are tuned to similar attributes with label spreading connections. This hypothesis is supported by anatomical data, since intracortical connections have indeed been found to selectively interconnect neurons with a similar orientation or color sensitivity (Livingstone & Hubel, 1988; Gilbert & Wiesel, 1989; Malach, Tootell & Malonek, 1994).

3.2.1. Grouping by disparity

Fig. 6A,B illustrates how this would account for grouping of depth planes. If the stereogram is fused, a horizontal and a vertical bar can be seen, both of which are slightly bent. The bars are superimposed on a curved plane. Note that in this situation a label indicative of assembly membership has to be spread in a region instead of along a curve. In his original paper on visual routines, Ullman (1984) referred to the spread of a label in two dimensions as *coloring*. In this section we will suggest how coloring can be implemented. Fig. 6B shows the range of disparities occurring along a line trough the image. The disparity of the horizontal bar is well within the range of disparities of the background plane. Nevertheless, the bar segregates from the plane because the change of disparity is abrupt at the borders of the bar, whereas disparity changes gradually at other locations. Fig. 6C illustrates an array of neurons that respond to the depth profile of Fig. 6B. Neurons positioned above each other are tuned to different disparities, and neurons that are laterally displaced have a different position of their receptive field. It is assumed that corticocortical connections link neurons with closely spaced receptive fields and a similar disparity tuning. Only connections among active cells are shown. These connections form the interaction skeleton. Note that even neurons with highly different disparity tuning are linked indirectly by the interaction skeleton, through neurons with an intermediate disparity tuning. The simple architecture of corticocortical connections of Fig. 6B does not provide a direct link between the neurons labeled 1 and 2 which have their receptive field in the background plane. Nevertheless, these neurons are connected through a chain of neurons which have their receptive fields in the background plane, but outside the image slice illustrated in Fig. 6B,C. A slightly more sophisticated wiring scheme would also provide a direct link between the respective cells. The selectivity of these connections could depend on grouping criteria such as co-planarity. This wiring scheme would also interconnect neurons that respond to the two segments of the horizontal bar, and as a result three perceptual groups for the image of Fig. 6A would emerge, one group for each bar, and one group for the background plane.

In Fig. 6, the disparity changes abruptly at the borders between image regions. These borders can be detected by a local process that is sensitive to large disparity gradients. Such a process can be carried out in two ways. First, it can be carried out by a feedforward operator, in the base representation. Second, it can be carried out by lateral inhibition. This is a rapid process that can apply in parallel to the entire image, but that nevertheless belongs to the incremental respresentation. In order to recover a curved surface, however, many local similarities of disparity need to be integrated. This can *only* be carried out in the incremental representation, for example by spreading a rate enhancement among neurons that are connected to each other through the interaction skeleton. Thus, only after the rate enhancement is spread among all neurons that respond to a single surface, that surface is grouped together as a coherent region, and segregates from the rest of the image. Curved surfaces may also contain regular, planar regions that can be detected in the base representation. Such 'surface chunks' may speed up the labeling operation, in analogy with Fig. 4.

A strong prediction of our theory is that attention is the psychological manifestation of the spread of the rate enhancement. Psychophysical evidence in support of these proposals has been reported by He & Nakayama (1995), who showed that visual attention indeed tends to spread along surfaces that are tilted in depth.

3.2.2. Orientation contrast

Fig. 6D generalizes these ideas into another feature domain. The image is similar, but now the horizontal and vertical bars are defined by orientation contrast. Images such as these have been studied by Nothdurft (1992). Nothdurft found that image regions with a local orientation contrast can be rapidly detected in spite of the variation of orientations within such a region or within the background. In analogy with what was discussed above, we would suggest that borders at which the orientation changes abruptly are detected rapidly. However, in order to identify the image elements that belong to a coherent image region, border detection does not suffice. In Fig. 6D the vertical bar also segregates from the horizontal bar. We suggest that grouping of all image elements that belong to the vertical bar costs additional time, because it depends on the spreading of a rate enhancement among the responsive neurons. We recently obtained direct physiological evidence for these proposals (Lamme, Rodriguez-Rodriguez & Spekreijse, 1999). Responses of neurons in the primary visual cortex of awake monkeys to textures such as that shown in Fig. 7A were investigated. Stimuli either contained a square region that segregated from the background on the basis of orientation contrast, or consisted of a texture with a homogeneous
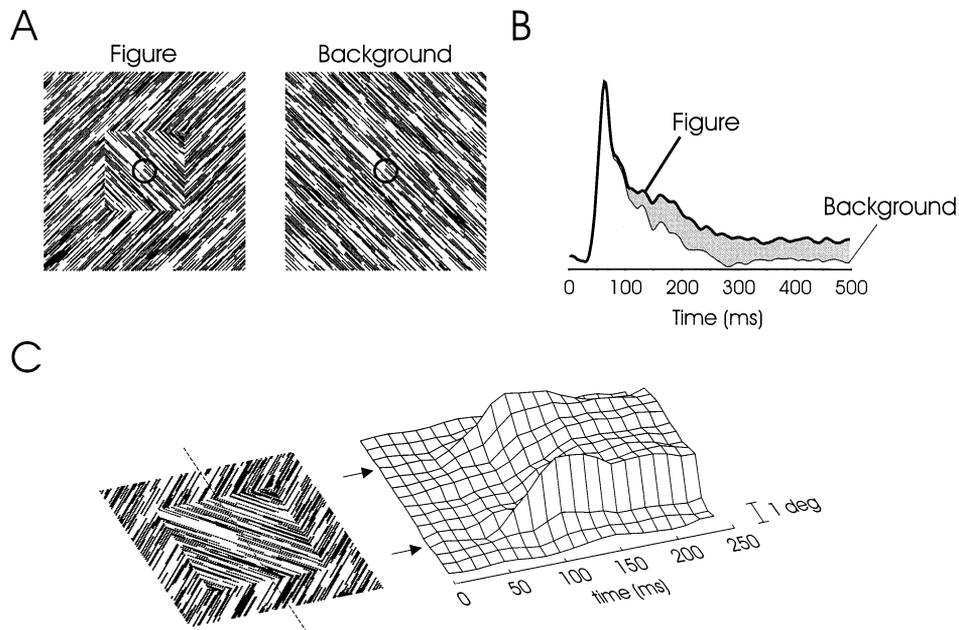
Fig. 7. Response enhancement in the primary visual cortex during texture segregation. (A) Stimuli shown to an awake monkey while recordings were obtained in area V1. The receptive field (circle) was centered on a square figure or on a homogeneous background. Note that the contents of the receptive field was identical for the two stimuli. (B) Population responses to the figure (thick line) and to the background (thin line). The gray area depicts the modulation. (C) The modulation was isolated by subtracting responses to the background from those to the figure for several positions of the receptive field relative to the texture square. Response enhancement to the edges occurred earlier than enhancement to the interior of the square (modified from Lamme et al., 1999).

orientation. Care was taken to ensure that the classical receptive field of the neurons (circle in Fig. 7A) contained identical texture elements in the two stimulus conditions. Fig. 7B compares the average responses of a large population V1 cells between the stimulus containing a square and the background stimulus. Responses to the texture square were enhanced. If the receptive fields were centered on the texture square, as in Fig. 7B, the response enhancement occurred after a latency of approximately 100 ms. Fig. 7C illustrates the dependence of the response enhancement on the position of the receptive field relative to the square. The response enhancement was computed for various positions of the square by subtracting responses to the background stimulus from responses to the stimulus that contained the square (gray region in Fig. 7B). If the receptive field fell on the edge of the square, response enhancement occurred early, starting after a latency of about 60 ms. However, if the receptive field fell on the interior of the square, response enhancement occurred after an additional delay (Lee, Mumford, Romero & Lamme, 1998; see also Lamme et al., 1999). These data suggest that the texture borders are detected by a rapid process in the incremental representation, that can be based on lateral inhibition. Additional time is required for the spread of the rate label across texture regions in order to group such coherent regions together. Moreover, these results suggest that borders between texture regions may function as seeds for the process of label spreading.

3.2.3. Other feature domains

There is physiological as well as psychophysical evidence for the generalization of these conjectures to other feature domains. In a study in awake monkeys, Lamme (1995) observed a rate enhancement in area V1 for figures that differed from the background in their direction of motion. Similarly, Zipser, Lamme and Schiller (1996) obtained rate enhancements in area V1 for figures defined by differences in luminance, color or disparity. In our theory, these rate modulations should be accompanied by attention for the respective image regions. A large number of psychophysical results support the hypothesis that attention tends to spread within image regions that cohere in these feature domains. Driver and Baylis (1989), for example, showed that attention spreads among image elements with a similar movement direction. There is also substantial evidence for the spread of attention among image elements with a similar color (Prinzmetal, 1981; Harms & Bundesen, 1983; Kramer & Jacobson, 1991; Prinzmetal & Keysar, 1989). Indeed, observations of this type inspired the Gestalt psychologists to formulate their laws of perceptual grouping (reviewed by Rock & Palmer, 1990).

3.3. Linking by cells that are tuned in multiple features domains

We have suggested previously that neurons tuned to multiple feature domains provide links in the interaction
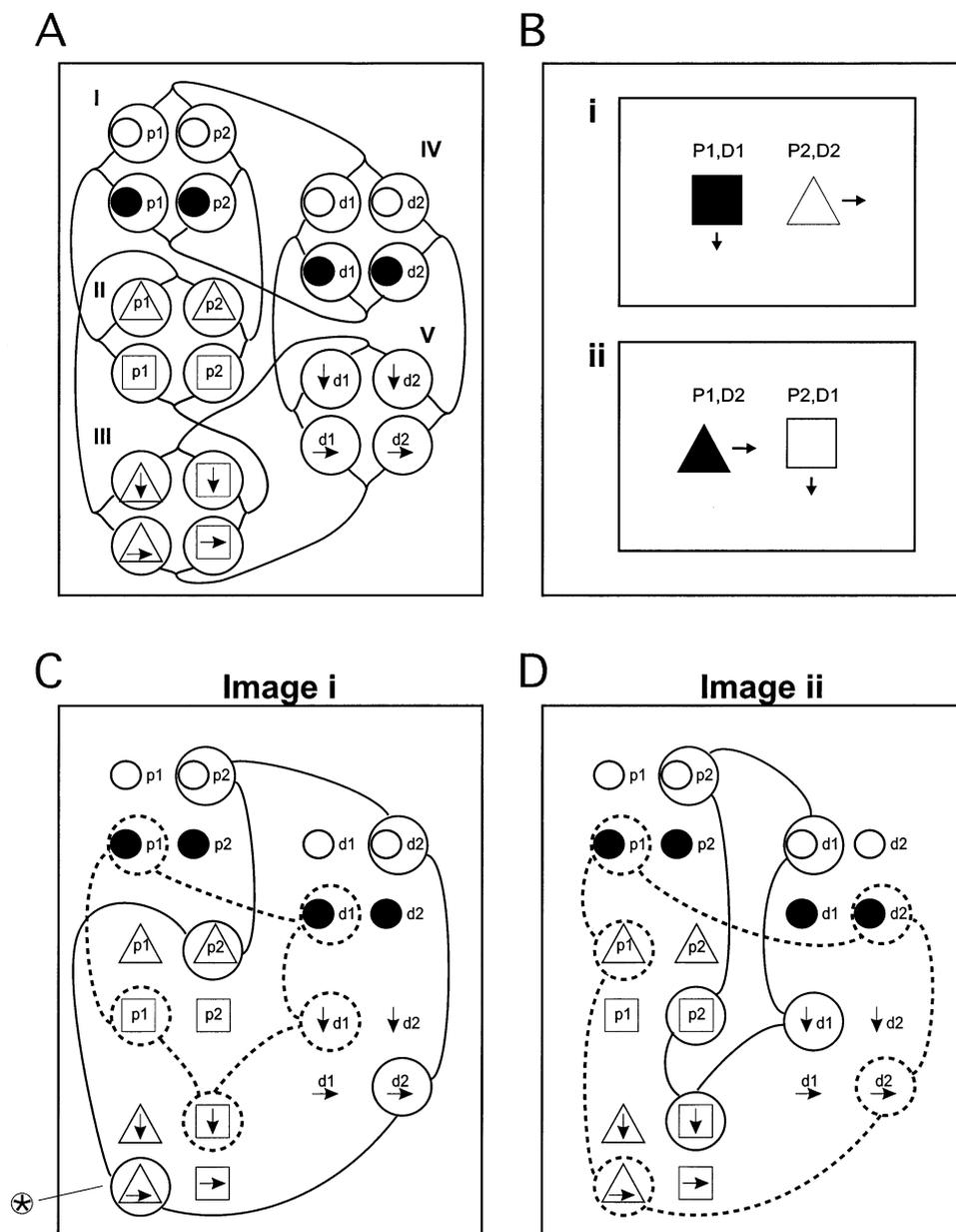
Fig. 8. Label spreading among neurons tuned to different feature domains. (A) Five clusters of neurons that are tuned to conjunctions in two feature domains. Clusters are tuned to location × color (I), location × shape (II), shape × direction of motion (III), color × depth (IV), and depth × direction of motion (V). Connections exists between neurons with a similar tuning on a shared feature domain. P1 and P2, two position in the visual field. D1 and D2, two disparities. (B) Two visual images composed of the same features, but in different combinations. (C,D) Base representation and interaction skeleton evoked by stimulus i (C) and ii (D). The interaction skeleton consists of the subset of connections that interconnect activated cells (circles). For clarity, disconnected compartments of the interaction skeleton are indicated by stippled and continuous connections. The visual system would have to spread an enhanced firing rate within one of these compartments to make additional feature conjunctions explicit.

skeleton that are essential for the formation of assemblies of neurons tuned to the diversity of features that may belong to a single object (Roelfsema, Engel, König & Singer, 1996; Roelfsema, 1998). A general finding in cortical physiology is that the firing rate of neurons can be influenced by variations along multiple feature dimensions (Zohary, 1992). For example, the neurons in the primary visual cortex are selective for both the

orientation and the location of a contour, and a large proportion of these cells is in addition selective for the disparity, color or movement direction of stimuli (Leventhal, Thompson, Liu, Zhou & Ault, 1995). This also holds in other visual areas. Two neurons may therefore both be tuned to particular, shared feature domains and at the same time exhibit tuning to distinct domains. Connections between neurons that exhibit a

similar selectivity in shared feature domains provide the necessary label spreading connections. Fig. 8A illustrates this for a network that consists of five groups of neurons that are selective for color, motion, shape, disparity and spatial location. Each neuron is selective for only two of these features. Neurons in cluster I, for example, are tuned to color and spatial location (i.e. they have a small classical receptive field). Neurons in cluster II are tuned to spatial location and shape. These two clusters are both tuned to spatial location, and neurons with a similar receptive field location are linked. We call the feature domain that is shared between cell clusters a 'linking dimension' (Roelfsema et al., 1996). Thus, spatial position is the feature that links clusters I and II. Similarly, color is the linking dimension between clusters I and IV, and this connection scheme is reiterated between the other clusters. The base representation evoked by image i (of Fig. 8B) is shown in Fig. 8C. Activated cells are indicated by circles. Only connections between activated neurons may spread an enhanced firing rate, and they form the interaction skeleton. Note that the interaction skeleton links all neurons that are activated by the same object. When the enhanced firing rate is spread within a connected compartment of the interaction skeleton, all neurons that respond to features of the same object are labeled. Thereby additional conjunctions like, for example, between color and motion are computed. These conjunctions are not represented by dedicated cells. Importantly, additional conjunctions can only be computed for a single object at a time. Illusory conjunctions result if the enhanced firing rate is simultaneously spread in two compartments of the interaction skeleton.

Image ii is a slightly different image, in which the same features are present, but in different combinations. Its base representation deviates from that of image i, because of differences in explicitly encoded feature conjunctions. The activity pattern enables a different subset of connections, such that appropriate neurons are linked by the interaction skeleton. Local changes in the pattern of activated cells result in global changes in the interaction skeleton, and similar changes in the tag spreading operations (as in Fig. 3). The neurons in Fig. 8 are tuned to only two feature dimensions. Conjunctions between more than two features have to be computed in the incremental representation. The number of neurons is therefore far less than the number that would be required if neurons would have to be as sharply tuned on all feature dimensions at the same time (e.g. von der Malsburg, 1981).

3.4. Feedforward and feedback

So far, we emphasized the similarities between solutions for a variety of binding problems including curve tracing (Fig. 3), surface segmentation (Fig. 6C) and binding between different feature domains (Fig. 8). There are, obviously, also differences between these elemental operations. One category of differences is related to the organization of the visual system. The connections of Figs. 3 and 6C, on the one hand, are between neurons of the same visual area. We will not repeat what was said about lateral connections in relation to these figures. The connections in Fig. 8, on the other hand, run between different areas. A number of these connections would be asymmetric, because they interconnect areas at different hierarchical levels. Cell clusters that are selective for spatial location (clusters I and II), for example, are situated at lower hierarchical levels, where receptive fields are small. Other cell clusters are not selective for spatial position (III, IV and V), and would be located at a higher level. In this section we will outline some implications of these differences.

3.4.1. Label evaluation and the feedforward spread of the rate enhancement

It is essential that the firing rate label spreads only among neurons that respond to features of the same visual object. To achieve this selectivity the visual system has to exploit conjunctions that are available in the base representation (as in Fig. 8). A feature conjunction that is widely available in early visual areas is that between spatial location and other features, due to the small receptive fields. Spatial location is therefore an important linking dimension in early visual areas. This is supported by anatomical data, since corticocortical connections preferentially interconnect neurons with overlapping receptive fields.

In order to compute other feature conjunctions that are invariant for spatial location *in the base representation* (i.e. before label spreading) spatial location can also be exploited. This is illustrated in Fig. 9, which compares strategies for the computation of conjunctions between color and shape. Neurons at level I have spatially restricted receptive fields. In the first strategy, spatial pooling occurs at level II (Fig. 9A). This is followed by a stage at which feature conjunctions are computed by thresholding (neurons at level III are only activated if both afferents provide input). Unfortunately, this leads to illusory conjunctions for images that contain multiple objects, because information about the features that go together is lost at the pooling stage (Fig. 9B). The alternative is to compute conjunctions at each location separately, and postpone the stage of spatial pooling (Fig. 9C). This prevents illusory conjunctions, but at the cost of having to reiterate conjunction computations for each location of the image. This cost becomes prohibitive for conjunctions of arbitrary complexity (e.g. von der
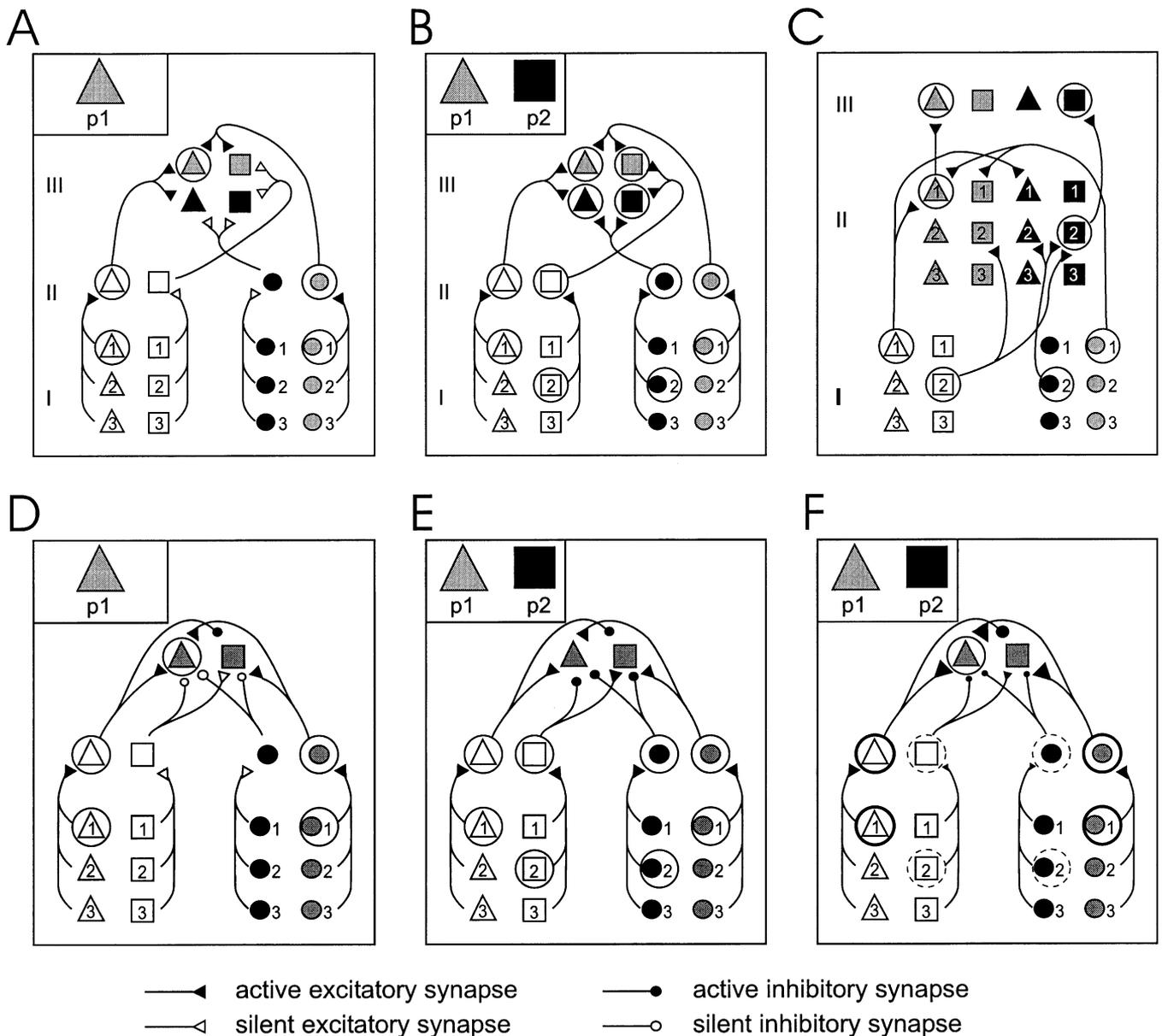
Fig. 9. Strategies for the computation of feature conjunctions. The aim is to compute conjunctions between shape and color, invariant for location. (A) Neurons at level I have non-overlapping receptive fields (at locations 1–3). In case of a single object, it is possible to pool information across locations at stage II, and to compute conjunctions at stage III by requiring both inputs to be active. Circles, active neurons. Black triangles, active excitatory synapses. White triangles, silent excitatory synapses. (B) In the presence of multiple objects this strategy results in the activation of neurons tuned to illusory conjunctions, i.e. feature conjunctions not present in the image. (C) Illusory conjunctions are avoided if conjunctions are computed before spatial pooling. (D) Same as A, but feedforward inhibitory connections are added between neurons of levels II and III that are tuned to incompatible features. Small circles, inhibitory synapses. (E) In the presence of two objects, the inhibitory input to level III cells counteracts the excitatory input, and they remain silent. This is called feedforward counterbalancing. (F) An enhanced firing rate spreads at level I among neurons tuned to the color gray and to the shape of a triangle, exploiting spatial location as a linking dimension. Excitatory input now outweighs the inhibitory input for neurons encoding the conjunction between gray and triangle, but not for neurons that are tuned for other conjunctions.

Malsburg, 1981). Nevertheless, it is a valuable strategy for the computation of simple feature conjunctions in the base representation.

A second alternative strategy, which we call *feedforward counterbalancing*, is illustrated in Fig. 9D–F. Inhibitory connections between incompatible features are added in the projection from level II to level III[4]. If a

_____

[4] A similar connectivity scheme based on a balance between excitatory and inhibitory projections has recently been proposed for the feedforward projection from the LGN to the primary visual cortex to explain contrast invariant orientation selectivity (Troyer, Krukowski, Priebe & Miller, 1998).

single object is presented, neurons at level III are activated during the feedforward sweep of activity (Fig. 9D). However, inhibitory and excitatory influences are balanced in the presence of multiple objects (Fig. 9E). Thus, the base representation of multiple objects does not equal the superposition of activity patterns that are evoked by the same objects if presented in isolation. Conjunctions now need to be computed in the incremental representation (incremental groupings), and this can only occur for one of the objects at a time. To achieve this, the enhanced firing rate needs to be injected at one of the features that is extracted in the base representation. Fig. 9F shows the situation that occurs when an enhanced firing rate is injected at spatial location 1 of level I, and spreads by utilizing spatial location as a linking dimension (label spreading connections have been omitted for clarity). Thereby, the conjunction between a shape and a color is made explicit for one of the objects. The level III neuron that is tuned to this conjunction will be activated subsequently, because the balance between excitation and inhibition is shifted in its favor. We refer to the activation of neurons at higher hierarchical levels due to label spreading at lower levels as 'label evaluation' (Roelfsema, 1998).

The transition from spatially tuned responses to translationally invariant responses is achieved in a single step in the schematic illustration of Fig. 9. A more realistic model would consist of a larger number of hierarchical levels (e.g. Fukushima, 1980). At each hierarchical level receptive fields are larger and tuning is more intricate than at the preceding level. In such a model, the proposed mechanism can be utilized, provided that inhibitory interactions are invoked from a region that is as large as the excitatory convergence zone. A prediction of such a model is that the complexity of base groupings, feature conjunctions that are computed in the base representation, depends gradually on the spacing between image elements.

When there is a single object in the visual field, feedforward counterbalancing permits the computation of highly elaborate feature conjunctions in the base representation. Recent findings indicate that the first spikes of neurons in the inferotemporal cortex, that occur about 80 ms after the appearance of a novel stimulus, are indeed tuned to highly complex feature conjunctions (Oram & Perrett, 1992; Rolls & Tovée, 1994). The utility of purely feedforward architectures for the recognition of isolated objects has also been demonstrated in neural network models. LeCun, Jackel, Bottou, Brunot, Cortes, Denker et al. (1995), for example, studied feedforward neural networks that can recognize handwritten digits. The model predicts that the selectivity of the first spikes breaks down if there are multiple image components within a single receptive field. In this situation, the only base groupings remain

that are reiterated across locations (as in Fig. 9C), and more elaborate conjunctions depend on elemental operations.

There is physiological evidence in support of feedforward counterbalancing. Many studies have shown that the magnitude of a neuron's response is reduced when multiple stimuli fall in its receptive field (Moran & Desimone, 1985; Sato, 1988; Miller et al., 1993; Treue & Maunsell, 1996). Larger receptive fields explain why inhibitory interactions among the neural responses to multiple stimuli are most pronounced in higher visual areas (Moran & Desimone, 1985; Desimone & Duncan, 1995; Kastner et al., 1998). Cueing the animal to the location of one of the stimuli relieves the inhibitory effects. Now the neurons respond as if the cued stimulus was presented in isolation (Moran & Desimone, 1985; Treue & Maunsell, 1996; Reynolds et al., 1999). In these situations, the labeling process should be carried out at lower hierarchical levels, where receptive fields are small enough to contain only a single object. Neurons in lower visual areas indeed exhibit an enhanced firing rate if their receptive field coincides with the cued location (Motter, 1993, 1994a; Treue & Maunsell, 1996; Luck et al., 1997; Vidyasagar, 1998; McAdams & Maunsell, 1999).

It should be stressed, however, that spatial location is not the only linking dimension available at the lower hierarchical levels. During curve tracing, for example, the rate label is spread on the basis of connectedness and colinearity. In Fig. 4 the hole in the base representation in V2 is caused by feedforward counterbalancing (as in Fig. 9E), but label evaluation eventually fills in the hole in the incremental representation.

3.4.2. Guided search and the top-down spread of rate enhancement

Many instances of visual search correspond to a type to place conversion. A feature (or feature conjunction) that is encoded in a translationally invariant manner has to be converted into the location at which this feature can be found. This can be achieved by the spread of an enhanced firing rate from areas with translationally invariant tuning to areas that are tuned to spatial location. In order to locate the triangle in Fig. 8C, for example, it would suffice to inject the enhanced firing rate at neurons tuned to the shape of a triangle (star in Fig. 8C). After the spread of the rate label, neurons in clusters I and II that are selective for its spatial location will be labeled by the enhanced firing rate.

Direct psychophysical evidence for the occurrence of a type to place conversion during visual search has been obtained by Kim and Cave (1995), using a dual task design. The primary task was to report the presence or absence of a target item in a search display. The location of the target item was irrelevant for a correct
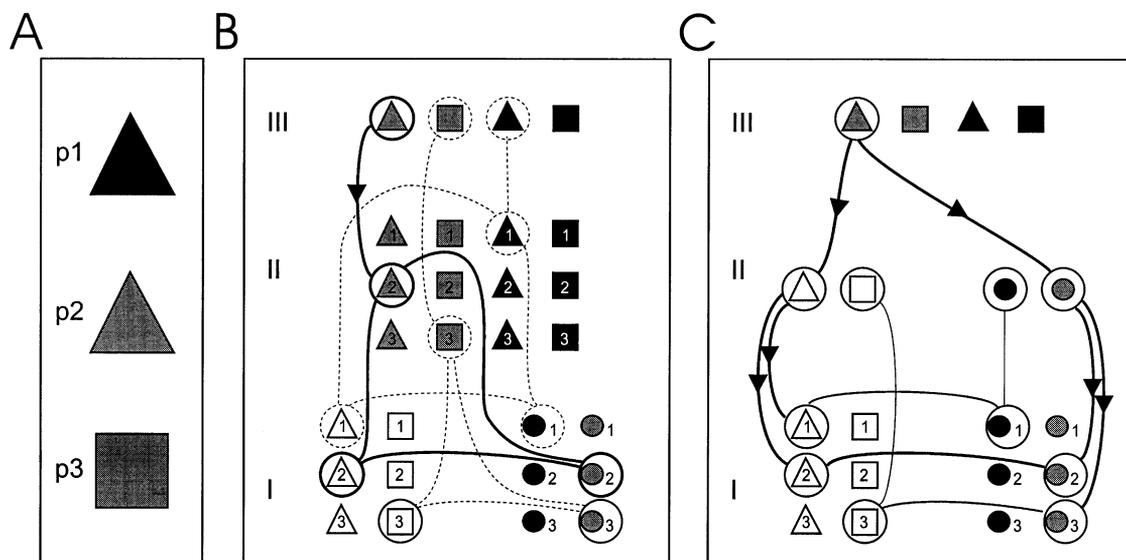
Fig. 10. Type to place conversion. (A) Search display in which the location of the gray triangle (p2) has to be recovered. (B) Same network as in Fig. 10C. In this network the feature conjunctions are computed in the base representation. The interaction skeleton consists of three disconnected networks of enabled connections. The rate enhancement is injected at level III, and can only spread to neurons with receptive fields at the appropriate location (thick lines). (C) Same network as in Fig. 10D–F. In this network the conjunction is not computed by the base representation. The interaction skeleton contains links between neurons that respond to different objects. Therefore, the rate enhancement also spreads to neurons with receptive fields at erroneous locations, if it is injected at level III.

response. In a proportion of trials, however, the search display was followed by a detection probe at a location that had been occupied by one of the items of the search display. The identification of probes at a location that had been occupied by a target item was faster and more accurate. This indicates that the location of a target item is identified during visual search, even if it is irrelevant for a correct response. Moreover, a trace of the type to place conversion persists after the disappearance of the search display, and enhances the identification of a probe at the labeled location. This trace provides evidence for the sensitivity of subsequent operators for the contents of the incremental representation.

The efficiency by which the rate label spreads to the appropriate lower level cells depends on whether the feature or feature conjunction that should be found is computed in the base representation. Features or conjunctions that are computed in the base representation can be found efficiently[5]. Fig. 10B illustrates this for the search for a conjunction between shape and color. It reproduces the network of Fig. 9C that reiterated the computation of conjunctions for a large number of positions in the visual field. If an enhanced firing rate is injected at the neuron encoding the required conjunction, it spreads selectively to lower level neurons that are tuned to the location at which this feature conjunc-

tion can be found. Search for a single feature would proceed similarly, as long as it is detected in parallel across the visual image. In the alternative architecture, in which the conjunction was detected in the incremental representation, search is less efficient, because the interaction skeleton does not provide a one-to-one link between the feature conjunction and the location at which this conjunction can be found (Fig. 10C). Consequentially, the enhanced firing rate also spreads to locations that do not contain the required conjunction, but only one of the features. Note, however, that the location that contains both features receives more top-down activation than the locations at which there is only a single feature. Here our proposal agrees with the 'guided search theory' of Wolfe and co-workers (Cave & Wolfe, 1990; Wolfe, 1994).

We want to notice, once more, that the transition from a retinotopic to a translationally invariant representation is gradual. The degree to which feature conjunctions are computed in the base representation depends gradually on the complexity of the image elements and their spacing. This dependence yields the testable prediction that search efficiency depends on the distance between a target item and its nearest distractors, even if the number of display items is constant.

Physiological evidence supports the involvement of rate modulations in the type to place conversion. As soon as the identy of the object that is searched has been specified, neurons in the inferotemporal and frontal cortex with a corresponding selectivity start to fire (Chelazzi et al., 1993; Rainer et al., 1998). After the

---

[5] In Figs. 10 and 11 we address the computation of conjunctions between color and shape. This choice is only for convenience. We do not wish to speculate on the extent that this particular conjunction is indeed computed in parallel across the visual scene. These features can be substituted by others without changing the arguments.

appearance of the search display, responses of inferotemporal neurons that are tuned to the identity of the target object are also enhanced. This response enhancement is delayed relative to the initial visual response (Chelazzi et al., 1993). It should propagate to earlier visual areas. Direct evidence in favor for a feature selective top-down enhancement of responses in retinotopic areas has been obtained by Motter (1994a). In this study a cue indicated the color of items that were relevant for the monkey. In area V4, responses to all items of the relevant color were enhanced after a delay of 150–200 ms. Because these neurons have restricted receptive fields, the response enhancement labels the location of target items.

After the type to place conversion, the spatial information can be used to foveate the target object. To this aim, the rate enhancement has to be spread to areas involved in the generation of eye movements. Many neurons in these areas have spatially restricted receptive fields, and therefore, spatial location can be used as the linking dimension. The generation of an eye movement to a target object that is embedded in a search display has been investigated in a number of studies (Schall & Hanes, 1993; Thompson, Bichot & Schall, 1997; Gottlieb et al, 1998; Schall & Thompson, 1999). After the onset of the search display, the initial responses evoked by the target and distractors are equally strong in these areas. Additional time is required before neurons responding to the target object exhibit a response enhancement.

3.5. Maintenance operations

In the above we disregarded a number of issues that are potentially harmful for a visual routine. An alteration of the visual image or an eye movement will change the base representation, and may erase the pattern of rate modulations. Thereby, results of previously applied operators would be lost. The second category of elemental operations provides mechanisms that protect the incremental representation against these events. On the other hand, rate modulations should not persist indefinitely, but need to be removed at some point in time. These operations are also included in the second category.

3.5.1. Removal of the rate label
The elemental operations of the first category can be implemented by the spread of an enhanced firing rate through the interaction skeleton. The persistence of the rate modulation after the completion of an elemental operation is beneficial in some tasks, because it may provide the point of departure for the subsequent operation. However, the rate enhancements may have to be removed in other tasks. In our example routine of Fig. 1, for example, the rate label should be removed from

the representation of the letter B by the time of the place to type conversion (at step 6 in Fig. 1B). To our knowledge, little is known about the physiological mechanisms that are responsible for the removal of the enhanced firing rate. A number of non-exclusive possibilities come to mind. First, rate enhancements may decay passively, with time constants that could even differ between visual areas. Second, rate modulations may be suppressed as a consequence of the application of subsequent operators. Third, the visual system may even possess specific mechanisms to actively suppress existing rate enhancements.

We are aware of only a single study by Motter (1994b) in which the physiological correlates of the removal of attention were studied. In his study, monkeys were cued with respect to the color of relevant image elements. Responses in area V4 to relevant image elements were enhanced (as was discussed above). In a subset of trials, however, the color cue was changed, rendering a different subset of image elements relevant. The first effect of the change in the color cue was an enhancement of responses to items that gained relevance, after 150–200 ms. After an additional delay, response enhancement disappeared for the items that lost relevance, 300 ms after the change in the color cue. The disappearance of the response enhancement might be an indirect consequence of the response enhancement for items gaining relevance, altering the balance between excitatory and inhibitory input. However, the data do not permit a distinction between passive decay and active suppression.

Ullman (1984) reserved an elemental operation, which he called marking, for keeping track of locations that had been labeled previously. The fate of image elements from which attention is removed depends on the visual task. If an *irrelevant* cue, like a luminance increment, summons attention to one of a number of image elements, processing of this element is impaired at a later point in time (Tipper, Weaver, Jerreat & Burak, 1994). Removing attention from an image element appears to reduce the ability of this item to attract attention. If such an inhibitory process extends into the next trial it is called negative priming (reviewed by Fox, 1995). However, negative priming is not the only possible outcome when an attended item reappears on a subsequent trial.

3.5.2. Maintenance of the incremental representation across changes in the visual image
Facilitation is often obtained if an attended image feature needs to be attended again on a subsequent trial. Thus, features can also be *positively* primed if there is no need to remove attention (e.g. Maljkovic & Nakayama, 1994, 1996; Stadler & Hogan, 1996). Positive priming indicates that attributes of the incremental representation survive during the intertrial interval. In

most tasks, the visual image changes drastically in such an episode, and corresponding alterations occur in the base representation. We attribute the information that persists in spite of these changes in the base representation to working memory. Working memory is also required if the base presentation changes *during* a trial due to a change in the visual input.

Spatial cueing is a widely used paradigm that probes the maintenance of attentional traces across changes in the visual image. Spatial cues for directing attention are typically separated in time from subsequent stimuli. The spatial signal has to persist during this interval, implying a form of working memory. Correlates of the spatial working memory have been found in the prefrontal cortex (Funahashi, Chafee & Goldman-Rakic, 1993; Wilson, O'Scalaidhe & Goldman-Rakic, 1993; Rainer et al., 1998), the parietal cortex (Gnadt & Mays, 1995; Colby, Duhamel & Goldberg, 1996), and even in early visual areas (Luck et al., 1997). A thorough discussion of the intimate relationship between working memory and visual attention is beyond the scope of the present article. Nevertheless, spatial working memory is an elemental operation that permits additional flexibility in the sequencing of other operations. Neurons that maintain spatial information can inject the enhanced firing rate at the appropriate position in lower visual areas as soon as a new stimulus appears and its base representation has been computed. Thereby, working memory can, for example, provide the spatial input of a place to type conversion.

Conversely, working memory may also provide the input for a type to place conversion, if the target *identity* was specified at an earlier point in time. Neuronal activity that maintains the identity of a target object across temporal delays has been found in the prefrontal (Rainer et al., 1998) and inferotemporal cortex (Chelazzi et al., 1993; Fuster, 1997) of the monkey. After presentation of a search display, these neurons can initiate the type to place conversion.

3.5.3. Eye movements

In free viewing conditions, an eye movement will often result from the attentive selection of a location. This disrupts the base and incremental representation in early visual areas, due to the shift of receptive fields. Fortunately, there are neurons in higher visual areas with a receptive field that is anchored in head coordinates rather than in retinal coordinates (e.g. Duhamel, Bremmer, BenHamed & Graf, 1997; Colby, 1998). Rate enhancements exhibited by these neurons are still valid after an eye movement. Similarly, rate enhancements of neurons in areas of the inferotemporal cortex, which have very large receptive fields, can also be preserved during an eye movement. When these rate enhancements are fed back to retinotopic areas after the computation of the new base representation, attended objects can be rescued.

Ballard et al. (1997) and Pylyshyn (1988) suggested that eye and hand movements can also be beneficial for visual routines, because they introduce additional flexibility in the orchestration of elemental operations. The basic idea is that the eye and hand position provide 'deictic' pointers to locations of interest. The fovea or a finger can be moved to a relevant location in order to mark it for future reference. An operator that occurs at a later point in the routine can retrieve information from the fovea, or about an image element that is close to, for example, the index finger. Thereby, objects can be marked without burdening working memory.

4. Visual routines: sequences of elemental operations

In the above, we have described a subset of the available elemental operations. It is probable that future work will isolate additional operators. For example, one class of operators that was only briefly mentioned is related to border detection (Sections 3.2.1 and 3.2.2). Nevertheless, we hypothesize that the number of elemental operations is finite. In contrast, it should be impossible to provide an exhaustive survey of visual routines. Their combinatorial richness derives from the possibility to put an arbitrary number of elemental operations into a sequence (Ullman, 1984). So far, we have illustrated a single visual routine, in Fig. 1. In this section we give further illustrations of how visual routines could be assembled for other tasks.

4.1. Sample routines composed of a few elemental operations

In a recent study, Vidyasagar (1998) investigated the activity of neurons in the primary visual cortex in a task in which monkeys had to report the orientation of one of a number of gratings. There was a green signal dot above the relevant grating, which was absent for the other gratings. A simple visual routine for this task could be composed of (1) a type to place conversion for the signal dot, and (2) a place to type conversion for the grating next to the signal dot. Results were consistent with this interpretation. Responses of neurons in the primary visual cortex to the relevant grating were enhanced. This response enhancement did not occur for the initial visual responses, but at a latency of at least 100 ms after the appearance of the stimulus array.

The next example is from a study by Rainer et al. (1998). They recorded from the prefrontal cortex of monkeys performing a delayed-matching-to-sample task. Two images were shown that were separated by an interstimulus interval. Both images contained three shapes. Within a block of trials, only one of the shapes was relevant. However, the relevant shape changed between blocks. In each trial, the monkey had to remem-
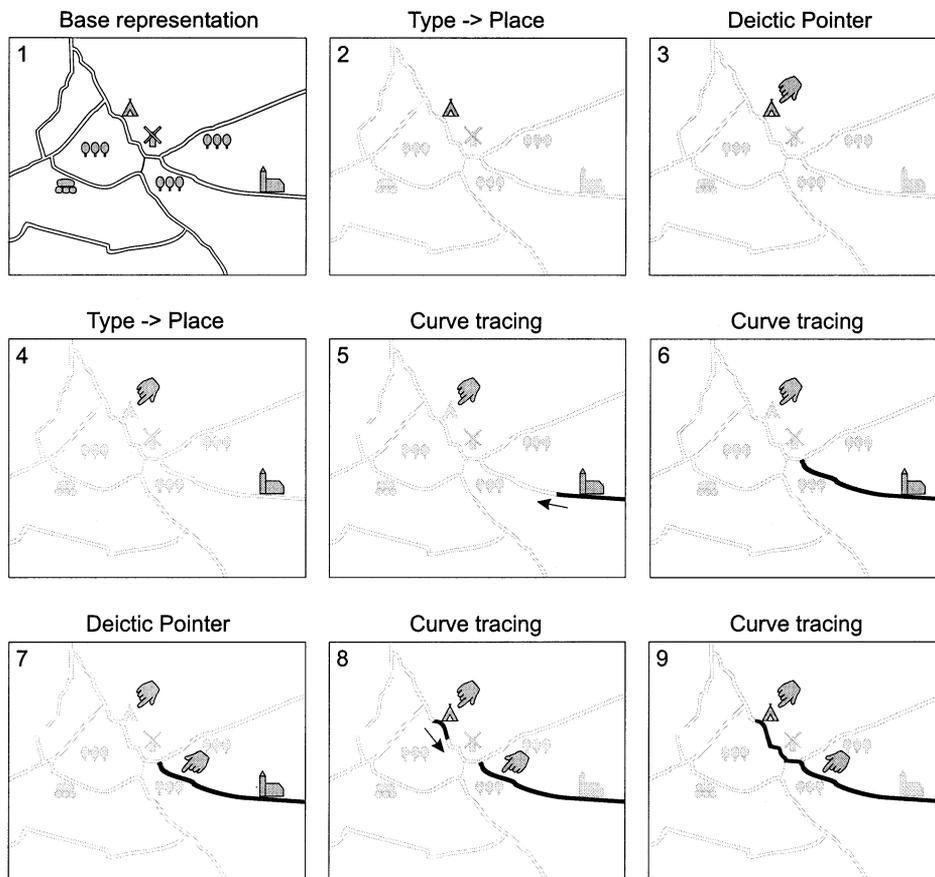
Fig. 11. Visual routine for the interpretation of a road map. Step 1, Computation of the base representation; step 2, search for the camping site, whose location is labeled; step 3, the index finger is moved to the labeled location; step 4, search for the church, steps 5 and 6, curve tracing until a T-junction is encountered; step 7, the other index finger is moved to the T-junction; steps 8 and 9, curve tracing, starting at the location of the camping site.

ber the location of the relevant shape in the first image. The second display was shown after a delay, and the animal had to report whether the relevant shape reappeared at the same location. A visual routine for this task might consist of the following elemental operations. First, a type to place conversion for the relevant shape in the first image in order to identify the relevant location. The activity of the prefrontal neurons started to reflect the relevant location after a latency of 150 ms. A second operator would be required that marks the relevant location during the ensuing interstimulus interval (this operator is related to working memory, see Section 3.5.2). Many neurons in the prefrontal cortex that were selective for the marked location indeed maintained their discharge during the interstimulus interval (Rainer et al., 1998). The third operator, a place to type conversion at the marked location, could be applied after the appearance of the second image. The monkey's report should depend on the match between this shape and the relevant shape.

A second remarkable result of the study by Rainer et al. (1998) was derived from a comparison between blocks of trials in which different shapes were relevant.

A proportion of prefrontal neurons was tuned to one of the shapes. These neurons maintained an enhanced firing rate that even persisted between trials, and for an entire block of trials. These cells appear to reflect working memory for the target shape.

4.2. Interpretation of a road map

In order to gain insight in an elemental operator, it is a good strategy to isolate it from other operators. The routines that were discussed so far may seem to be somewhat artificial, because they were relatively simple and could be decomposed in one or a few operators. It is more difficult to study the contributions of each of the operators in more complicated routines. Moreover, more elaborate tasks may permit inter-individual differences in the order and even identity of operators that are applied. In order to illustrate the versatility of the approach, however, we will show how elemental operations might be applied to solve a more elaborate and natural task: the interpretation of a road map.

Suppose that the task is to drive to a camping site and that our subject has just passed a church. The

subject's road map might look like Fig. 11. In Fig. 11 the first operator that is applied is a search for the camping site. This location is labeled by an enhanced firing rate (step 2). In our example, the subject moves his index finger to the labeled location for future reference (step 3) (the finger is a *deictic pointer* in the terminology of Ballard et al., 1997). The location of the finger, which is easily retrieved, permits the removal of this label in order to reduce interference while a subsequent operator is applied, the search for a church (step 4). After the church has been located, a curve tracing operator is launched, labeling the road in the direction of the index finger (steps 5 and 6). Again, the tracing operation can take the enhanced firing rate from its predecessor as a starting point. When the tracing operator arrives at the T-junction, a number of possibilities emerge. First, the tracing operator may proceed in the correct direction, i.e. the direction of the index finger. Second, it may proceed in the erroneous direction. It may even simultaneously proceed in both directions. We have no firm hypothesis about what happens in these situations, although they are experimentally testable[6]. Our subject takes no chances, and puts his other index finger at the T-junction (step 7). Then he starts to trace from the location of the first index finger (step 8). As soon as the T-junction is reached, the entire route has been labeled (step 9).

It is clear that the solution in Fig. 11 is far from unique. In addition, a number of crucial issues are neglected. First, a typical subject would make a number of eye movements while studying the map. This, however, need not be detrimental for the labeling operation, since the pattern of rate modulations may be remapped after an eye movement (Duhamel, Colby & Goldberg, 1992; see Section 3.5.3). Second, there are unresolved issues related to the scope and applicability of the operators themselves. Is it necessary to remove the rate label from previous positions, while searching for a shape like a church? How clever is the curve tracing operator, can it simultaneously proceed in multiple directions? Third, it is unlikely that such a complicated routine is completely precompiled. At many instances multiple operators apply, and it is an open question how the appropriate operator is selected. In the present article the focus is on the implementation of the operators themselves. We will, however, now indicate a number of important issues that are left open by our proposals.

---

[6] We have recently studied the spread of attention at intersections between a curve that is traced and an irrelevant curve, using the methods described in Section 3.1.4 (Scholte et al., 1999). Some attention appears to enter the irrelevant curve if its crosses the traced curve.

5. Open questions

A first issue that we neglected is what Ullman (1984) called the 'assembly problem'. What glues a sequence of operations together into a single routine? In simple and highly repetitive tasks, routines may consist of fixed sequences of elemental operators. In contrast, if tasks are more complicated (as in Fig. 11) the order of the operations may have to vary from trial to trial. A visual routine may even have to be aborted if unexpected image elements appear that capture attention (Egeth & Yantis, 1997). How these events should be handled is left open (for approaches to these issues, see Newell, 1990; Whitehead & Ballard, 1991; Anderson & Lebiere, 1998). Nevertheless, even in complicated tasks some subsequences of operators may occur more often than others, and embedding operators in subsequences may be useful. The physiological mechanisms that are responsible for the orchestration of operators are unknown. Some insights may be gained from a related issue, the organization of limb movements into longer sequences. Tanji and Shima (1994) found neurons in the supplementary motor area of awake monkeys that were activated while specific sequences of arm movement were planned. At present it is open whether comparable neurons exist that arrange elemental operators into larger sequences.

A source of possible controversy is the dichotomy between the base and incremental representation. We identified the base representation with the first feedforward sweep of activity through the hierarchy of visual areas. Although it proceeds rapidly, recurrent interactions may be initiated in early visual areas, while the base representation is still under construction in higher areas. This implies that the base and incremental representation are not completely separable in time.

A related open question is the degree of invariance of the base representation. Is the base representation of a novel stimulus independent of previously applied operators? Physiological evidence suggests that it is not. Thompson et al. (1997), for example, recorded from the frontal eye fields of monkeys. The color of the fixation point instructed the animals to either make an eye movement to a stimulus that appeared in a neuron's receptive field or to maintain fixation. The instruction to make an eye movement enhanced the initial responses of a large proportion of neurons at latencies shorter than 80 ms. These early responses are presumably evoked by the feedforward sweep of activity. The instruction to make an eye movement had similar effects on the initial responses of parietal neurons (Bushnell et al., 1981; Colby et al., 1996). Moreover, spatial cueing also influences initial visual responses in early visual areas. Early visual responses to a cued location were enhanced in area V4 (Luck et al., 1997). This implies that traces of previously applied operators

may modify the base representation of a subsequent stimulus. These traces presumably reflect the persistence of rate modulations after the disappearance of an image, which permits the maintenance of relevant spatial locations and identities in working memory (Chelazzi et al., 1993; Fuster, 1997; Rainer et al. 1998). These traces apparently start to exert their effect during the construction of the base representation of a novel image, thus supporting the transmission of information between operators that are applied on subsequent images. We suggested that these effects are intimately related to priming, which characterizes the persistence of traces across trials (Section 3.5).

6. Relation to previous theories

In order to compare our model with concurrent theories on visual processing, we first summarize its key components. Thereafter, we indicate where our proposals deviate from other theories.

After the appearance of an image, there is a feedforward sweep of activity through the hierarchy of visual areas. This sweep of activity constitutes the base representation, which includes neurons in early as well as higher visual areas. If there is a single object, many complicated conjunctions can be computed in the base representation. We refer to these conjunctions as base groupings. In the presence of multiple objects, however, feedforward counterbalancing truncates the feedforward sweep. This prevents the activation of neurons that are tuned to unsupported conjunctions. Additional conjunctions depend of the application of elemental operators, and are called incremental groupings. These operators spread a rate enhancement through the network of corticocortical connections. Only connections between active neurons spread the rate enhancement, and they form the interaction skeleton. Incremental groupings may include conjunctions between spatially separate contour segments of a single object, image elements belonging to a single surface, and conjunctions between different feature domains. High-level visual functions require visual routines, which consist of sequences of elemental operations. Elemental operators make their results available to subsequent operators through the pattern of rate modulations, or by moving the eye or hand to a relevant spatial location.

6.1. Theories about visual routines

Our proposals deviate in a number of respects from the original theory on visual routines by Ullman (1984). The most important difference is between levels of description. Ullman's formulation was largely at a computational level, whereas our proposals are concerned with the implementation of visual routines at a neu-

ronal level. A second important difference is related to what is included in the base representation. Ullman (1984) included only primitive features in the base representation. In our formulation, the base representation consists of the first feedforward sweep of activity through the hierarchy of visual areas. Thus, the base representation also includes the activity of neurons at higher hierarchical levels that are selective for complicated shapes or patterns of motion. This decision is justified by the following arguments. First, the feedforward sweep of activity proceeds very rapidly. Some neurons in the inferotemporal cortex respond after a latency of 80 ms (Oram & Perrett, 1992), a latency at which a proportion of V1 neurons have not even fired (Raiguel et al., 1989; Nowak et al., 1995). Second, the rules that determine the interplay between activity and functional connectivity are similar in lower and higher visual areas. Although the feature domains that determine the selectivity of corticocortical connections differ between areas, it is the base representation that enables a subset of these connections, thereby defining the interaction skeleton. As a result, the proposed mechanisms account for grouping operations that are as diverse as curve tracing, labeling of image regions defined by texture, depth or motion, and even the grouping of responses of neurons that are tuned to different feature domains. Operators such as curve tracing and region labeling, which were classified by Ullman (1984) as distinct are implemented by similar mechanisms.

Ballard et al. (1997) have argued that the actual position of eyes and limbs allow additional flexibility in the sequencing of elemental operations. As soon as such a deictic pointer is brought to a relevant image location, this location need not be held in working memory, but may nevertheless be easily retrieved if it is required at a later point in time. It may seem tempting to suggest that eye movements also reduce the need for other elemental operations such as curve tracing or the place to type conversion. For example, a curve can be traced by making a sequence of small eye movements or by moving a finger along it. However, in these situations moving the eyes or the finger to a location on the traced curve actually depends on the correct segmentation of the visual image. In other words, mental curve tracing should precede the movement. Nevertheless, eye or hand movements can assist during curve tracing. First, locations at which curves come into close proximity can be foveated in order to improve visual resolution. Second, it can be a good strategy to move the eyes or a finger to a location midway on a curve if it is long and contorted. Thereafter, the tracing operation may resume at this position. Third, it can be advantageous to mark a location where a curve splits, and to start tracing from another location (as in Fig. 11).

Another elemental operation that may benefit from eye movements is the place to type conversion. Foveating the object of interest is a natural strategy to exclude interference from other visual objects. However, even if the object is at the fovea, it may have to be segregated from image components that are spatially overlapping or occluding. In these situations eye movements fail to isolate the object, and its recognition depends on the exploitation of other grouping criteria, such as colinearity and connectedness.

6.2. Models of visual attention

It is remarkable that the elemental operations that have been investigated physiologically are all associated with a relative enhancement of the firing rate of visual cortical neurons. At a psychological level of description, these operators involve visual attention. This provides strong support for the conjecture that rate enhancements in visual cortical areas constitute the physiological correlate of visual attention (Maunsell, 1995; Desimone & Duncan, 1995). At first sight, the base representation appears to map onto pre-attentive vision and the incremental representation onto attentive vision. At close scrutiny, however, this mapping is far from one-to-one.

Many grouping operations that are traditionally attributed to pre-attentive vision occur in the incremental representation. This can be illustrated for curve tracing, which corresponds to contour grouping on the basis of colinearity and connectedness. As soon as the base representation has been computed the groundwork has been done, since neurons that respond to the various segments of the same curve are linked by the interaction skeleton (Fig. 3). However, this linking only becomes explicit, i.e. manifest for the rest of the visual brain, after the spread of the rate label. This labeling operation takes time and belongs to the incremental representation. Similar arguments hold for the grouping of image elements into two-dimensional regions. Coherent image regions can be defined by a gradual change in a feature such as depth or orientation (Fig. 6). Abrupt discontinuities in feature gradients can be detected rapidly. But the spread of the rate label in order to 'color' an image region as coherent takes time. These aspects of our theory distinguish it from concurrent theories of visual attention, because they assume that perceptual grouping applies with unlimited capacity (e.g. Treisman & Gelade, 1980; Duncan & Humphreys, 1989). On the other hand, advanced feature conjunctions that traditional theories would classify as unattainable by pre-attentive vision, may nevertheless by computed by the base representation, in particular if the respective image component is well isolated in the visual field (Oram & Perrett, 1992; Rolls & Tovée, 1994). We tried to avoid confusion by using the terms base and incremental representation instead of pre-at-

tentive and attentive vision. In order to bring the dichotomies into correspondence the following two modifications are required. First, activity in higher visual areas needs to be included into the pre-attentive stage. Second, many Gestalt rules of perceptual grouping have to be transferred from the pre-attentive to the attentive stage.

Our theory agrees with the feature integration theory of Treisman en co-workers (Treisman & Gelade, 1980; Treisman, 1988) that attention is required to bind the various features of an object into a coherent representation. In feature integration theory, a master map of locations is employed to direct attention. A spotlight groups together the various features that are present at a single position in the visual field (for a modeling approach see Mozer, 1991). In our theory, spatial location is also an important linking dimension, but other grouping criteria are equally essential. It therefore provides a natural blend between object-based theories (Duncan, 1984) and space-based theories of visual attention (Eriksen & St. James, 1986; Treisman, 1988; LaBerge & Brown, 1989; Kramer & Jacobson, 1991).

In many other psychological theories, attention features as a 'magical substance' with limited supply. Its availability determines the quality and speed by which cognitive operations are executed. Theories on the implementation of cognitive operations attempt to replace this account by more specific mechanisms. Our conjecture is that attention is the psychological correlate of a collection of elemental operations, which are all associated with the spread of a rate enhancement among visual cortical neurons.

References

Allman, J., Miezin, F., & McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local–global comparisons in visual neurons. *Annual Reviews in Neuroscience*, 8, 407–430.

Allport, A. (1989). Visual attention. In M. I. Posner, *Foundations of cognitive science* (pp. 631–682). Cambridge MA: MIT Press.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20, 723–767.

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*, 15, 2112–2127.

Bushnell, M. C., Goldberg, M. E., & Robinson, D. L. (1981). Behavioral enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. *Journal of Neurophysiology*, 46, 755–772.

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, 22, 225–271.

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

Colby, C. L., Duhamel, J.-R., & Goldberg, M. E. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *Journal of Neurophysiology*, 76, 2841–2852.

Colby, C. L. (1998). Action-oriented spatial reference frames in cortex. *Neuron*, 20, 15–24.

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

Dinse, H. R., & Krüger, K. (1994). The timing of processing along the visual pathway in the cat. *Neuroreport*, 5, 893–897.

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.

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye-movements. *Science*, 255, 90–92.

Duhamel, J. R., Bremmer, F., BenHamed, S., & Graf, W. (1997). Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature*, 389, 845–848.

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 Reviews*, 96, 433–458.

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

Egeth, H. E., & Yantis, S. (1997). Visual attention: control, representation, and time course. *Annual Reviews of Psychology*, 48, 269–297.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Perception & Psychophysics*, 40, 225–240.

Eriksen, C. W., & Yeh, Y.-Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology*: *Human Perception and Performance*, 11, 583–597.

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.

Fox, E. (1995). Negative priming from ignored distractors in visual selection: a review. *Psychonomic Bulletin & Review*, 2, 145–173.

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

Funahashi, S., Chafee, M. V., & Goldman-Rakic, P. S. (1993). Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. *Nature*, 365, 753–756.

Fuster, J. M. (1997). Network memory. *Trends in Neurosciences*, 20, 451–459.

Gigus, Z., & Malik, J. (1991). Detecting curvilinear structure in images. U.C. Berkeley CSD Technical Report 91/619.

Gilbert, C. D., & Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Journal of Neuroscience*, 9, 2432–2442.

Gnadt, J. W., & Mays, L. E. (1995). Neurons in monkey parietal area LIP are tuned for eye-movement parameters in three-dimensional space. *Journal of Neurophysiology*, 73, 280–297.

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, 391, 481–484.

Graziano, M. S. A., Andersen, R. A., & Snowden, R. J. (1994). Tuning of MST neurons to spiral motions. *Journal of Neuroscience*, 14, 54–67.

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 of the USA*, 92, 11155–11159.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, 99, 480–517.

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.

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

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science*, 282, 108–111.

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

Kim, M.-S., & Cave, K. R. (1995). Spatial attention in visual search for features and feature conjunctions. *Psychologicl Science*, 6, 376–380.

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.

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 of the USA*, 90, 7495–7497.

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.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Reviews*, 96, 101–124.

Lamme, V. A. F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, 15, 1605–1615.

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.

LeCun, Y., Jackel, L., Bottou, L., Brunot, A., Cortes, C., Denker, J., Drucker, H., Guyon, I., Müller, U., Säckinger, E., Simard, P., Vapnik, V., et al. (1995). Comparison of learning algorithms for handwritten digit recognition. In F. Fogelman, & P. Gallinari, *International conference on artificial neural networks* (pp. 53–60). Paris: no. EC2 & Cie.

Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. F. (1998). The role of the primary visual cortex in higher level vision. *Vision Research*, 38, 2429–2454.

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.

Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science*, *240*, 740–749.

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.

Mahoney, J. V., & Ullman, S. (1988). Image chunking defining spatial building blocks for scene analysis. In Z. Pylyshyn, *Computational processes in human vision: an interdisciplinary perspective* (pp. 169–209). Norwood NJ: Ablex.

Malach, R., Tootell, R. B. H., & Malonek, D. (1994). Relationship between orientation domains, cytochrome oxidase stripes, and intrinsic horizontal connections in squirrel monkey area V2. *Cerebral Cortex*, *4*, 151–165.

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*, 657–672.

Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. the role of position. *Perception & Psychophysics*, *58*, 977–991.

Marr, D. (1982). *Vision*. New York: W.H. Freeman and Company.

Maunsell, J. H. R. (1995). The brain's visual world: representation of visual targets in cerebral cortex. *Science*, *270*, 764–769.

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

McCormick, P. A., & Jolicoeur, P. (1991). Predicting the shape of distance functions in curve tracing: Evidence for a zoom lens operator. *Memory & Cognition*, *19*, 469–486.

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.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*, 782–784.

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology*, *70*, 909–919.

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

Motter, B. C. (1994b). Neural correlates of feature selective memory and pop-out in extrastriate area V4. *Journal of Neuroscience*, *14*, 2190–2199.

Mozer, M. C. (1991). *The perception of multiple objects*. Cambridge, MA: MIT Press.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton Century Crofts.

Nelson, J. I., & Frost, B. J. (1985). Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex. *Experimental Brain Research*, *61*, 54–61.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Nothdurft, H. C. (1992). Feature analysis and the role of similarity in preattentive vision. *Perception & Psychophysics*, *52*, 355–375.

Nowak, L. G., Munk, M. H. J., Girard, P., & Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Visual Neuroscience*, *12*, 371–384.

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.

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

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology*, *82*, 2490–2502.

Polat, U., & Sagi, D. (1994). The architecture of perceptual spatial interactions. *Vision Research*, *34*, 73–78.

Posner, M. I., & Presti, D. E. (1987). Selective attention and cognitive control. *Trends in Neurosciences*, *10*, 13–17.

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., & Keysar, B. (1989). Functional theory of illusory conjunctions and neon colors. *Journal of Experimental Psychology: General*, *118*, 165–190.

Pylyshyn, Z. (1988). Here and there in the visual field. In Z. Pylyshyn, *Computational processes in human vision: an interdisciplinary perspective* (pp. 210–236). Norwood NJ: Ablex.

Raiguel, S. E., Lagae, L., Gulyás, B., & Orban, G. A. (1989). Response latencies of visual cells in macaque areas V1, V2 and V5. *Brain Research*, *493*, 155–159.

Rainer, G., Asaad, W. F., & Miller, E. K. (1998). Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature*, *393*, 577–579.

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

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

Roelfsema, P. R., Engel, A. K., König, P., & Singer, W. (1996). The role of neuronal synchronisation in response selection: a biologically plausible theory of structured representations in the visual cortex. *Journal of Cognitive Neuroscience*, *8*, 603–625.

Roelfsema, P. R. (1998). Solutions for the binding problem. *Zeitschrift für Naturforschung*, *53c*, 691–715.

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

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

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

Roelfsema, P. R., & Spekreijse, H. (1999). Correlates of a gradual spread of attention over a traced curve in macaque area V1. *Society for Neuroscience Abstracts*, *25*, 4.

Rolls, E. T., & Tovée, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London Series B*, *257*, 9–15.

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

Sato, T. (1988). Effects of attention and stimulus interaction on visual responses of the inferior temporal neurons in macaque. *Journal of Neurophysiology*, *60*, 344–364.

Schall, J. D., & Hanes, D. P. (1993). Neural basis of saccade target selection in frontal eye field during visual search. *Nature*, *366*, 467–469.

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

Schein, S. J., & Desimone, R. (1990). Spectral properties of V4 neurons in the macaque. *Journal of Neuroscience*, *10*, 3369–3389.

Schillen, T. B., & König, P. (1994). Binding by temporal structure in multiple feature domains of an oscillatory neuronal network. *Biological Cybernetics*, *70*, 397–405.

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.

Scholte, H. S., Lamme, V. A. F., Spekreijse, H., & Roelfsema, P. R. (1999). Atention in mental curve tracing task is not zoom-lens or spotlight based. *Investigative Ophthalmology & Visual Science*, *40*, S51.

Sengpiel, F., Baddeley, R. J., Freeman, T. C. B., Harrad, R., & Blakemore, C. (1998). Different mechanisms underlie three inhibitory phenomena in cat area 17. *Vision Research*, *38*, 2067–2080.

Shao, Z., & Burkhalter, A. (1996). Different balance of excitation and inhibition in forward and feedback circuits of rat visual cortex. *Journal of Neuroscience*, *15*, 7353–7365.

Singer, W. (1993). Synchronization of cortical activity and its putative role in information processing and learning. *Annual Reviews in Physiology*, *55*, 349–374.

Sporns, O., Gally, J. A., Reeke, G. N. Jr, & Edelman, G. M. (1989). Reentrant signaling among simulated neuronal groups leads to coherency in their oscillatory activity. *Proceedings of the National Academy of Sciences of the USA*, *86*, 7265–7269.

Stadler, M. A., & Hogan, M. E. (1996). Varieties of positive and negative priming. *Psychonomic Bulletin & Review*, *3*, 87–90.

Tanji, J., & Shima, K. (1994). Role for supplementary motor area cells in planning several movements ahead. *Nature*, *371*, 413–416.

Thompson, K. G., Bichot, N. P., & Schall, J. D. (1997). Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *Journal of Neurophysiology*, *77*, 1046–1050.

Tipper, S. P., Weaver, B., Jerreat, L. M., & Burak, A. L. (1994). Object-based and environment-based inhibition of return of visual attention. *Journal of Experimental Psychology*: *Human Perception and Performance*, *20*, 478–499.

Tovée, M. J., Rolls, E. T., Treves, A., & Bellis, R. P. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. *Journal of Neurophysiology*, *70*, 640–654.

Tovée, M. J. (1994). How fast is the speed of thought? *Current Biology*, *4*, 1125–1127.

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

Treisman, A. M. (1988). Features and objects: the fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, *40A*, 201–237.

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

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

Troyer, T. W., Krukowski, A. E., Priebe, N. J., & Miller, K. D. (1998). Contrast-invariant orientation tuning in cat visual cortex: thalamocortical input tuning and correlation-based intracortical connectivity. *Journal of Neuroscience*, *18*, 5908–5927.

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

van der Heijden, A. H. C., Wolters, G., Groep, J. C., & Hagenaar, R. (1987). Single-letter recognition accuracy benefits from advance cuing of location. *Perception & Psychophysics*, *42*, 503–509.

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

Vidyasagar, T. R. (1998). Gating of neuronal responses in macaque primary visual cortex by an attentional spotlight. *Neuroreport*, *9*, 1947–1952.

von der Malsburg, C. (1981). *The correlation theory of brain function*. *Internal report 81-2*. Göttingen: Max-Planck-Institute for Biophysical Chemistry.

Whitehead, S. D., & Ballard, D. H. (1991). Learning to perceive and act by trial and error. *Machine Learning*, *7*, 45–83.

Wilson, F. A. W., O'Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, *260*, 1955–1958.

Wolfe, J. M. (1992). The parallel guidance of visual attention. *Current Directions in Psychological Science*, *1*, 124–128.

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

Zipser, K., Lamme, V. A. F., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *Journal of Neuroscience*, *16*, 7376–7389.

Zohary, E. (1992). Population coding of visual stimuli by cortical neurons tuned to more than one dimension. *Biological Cybernetics*, *66*, 265–272.