ALGORITHMS FOR THE DETECTION OF CONNECTEDNESS AND THEIR NEURAL IMPLEMENTATION

ALGORITHMS FOR THE DETECTION OF CONNECTEDNESS AND THEIR NEURAL IMPLEMENTATION

PIETER R. ROELFSEMA, SANDER BOHTE, HENK SPEKREIJSE

Graduate School Neurosciences Amsterdam, Laboratory of Medical Physics, and Netherlands Ophthalmic Research Institute (KNAW), dept. Visual System Analysis, Academic Medical Center (UvA), P.O. Box 12141, 1100AC Amsterdam, The Netherlands

Connectedness is one of the most important grouping criteria that allow the visual system to segregate objects from each other and from the background. We review algorithms for the detection of connectedness from a physiological point of view. Connectedness detection by a feedforward network is physiologically implausible. Instead, we present evidence that visual cortical neurons label connected image regions serially, by exhibiting an enhanced firing rate. This suggests an intimate relationship between connectedness detection and visual attention.

1   Introduction

Typical natural visual scenes contain many objects, which need to be segregated from each other and from the background. Components of a single object are usually related to each other by numerous perceptual grouping criteria, including colinearity and connectedness. The visual system is well adapted to exploit these grouping criteria in order to recover image regions occupied by one of the objects. Indeed, the visual system readily groups image regions together that are connected to one another[1], as is illustrated in Fig. 1. Image elements that are connected to one another are easily segregated from other image regions. This is important, since other functions of the visual system, like object recognition, depend on proper segmentation of the visual image. In this chapter we will consider the physiological mechanisms that allow the visual system to detect connections among image regions.

Minsky and Papert[2] were among the first to recognize that connectedness detection is not a trivial task for a neural network, like the visual cortex. They studied the detection of connectedness by perceptrons, simple neural networks that in many ways resemble the multilayered feedforward neural networks that have become rather popular nowadays[3,4]. Minsky and Papert were able to show that the computation of connectedness using perceptrons is problematic in that the number of
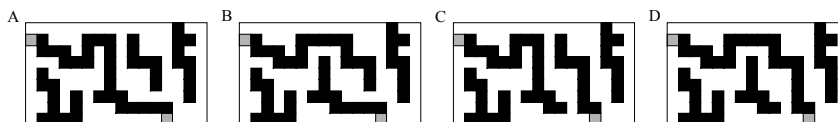
Figure 1: Connectedness as a grouping criterion. A-D, Subtle and local changes in the visual image may exert a pronounced effect on the process of image segmentation. In A the two grey tiles are connected and perceived as belonging to a single image region. Local alterations in the image, as in B and C, give rise to a different segmentation of the image: now the grey tiles belong to different image components. D, When both alterations are combined the grey tiles are connected once more, illustrating that for the computation of connectedness many local decisions need to be integrated.

units required for the computation scales very unfavourably with the size of the retina on which images are displayed. Using perceptrons, the number of units that would be required for the computation of connectedness on a realistic retina is likely to exceed the number of neurons of the entire visual cortex. In the present chapter we will argue that the difficulties that perceptrons have with computing connectedness are shared, to some extent, by other multilayered *feedforward* network architectures.

In contrast, serial algorithms, which can be implemented using *recurrent* networks, offer a rather cheap solution for the computation of connectedness in terms of the number of processing units required, even for retinas of realistic size as was shown, for example, by Minsky and Papert[@2]. Ullman[@5] also suggested that the visual system should be equipped with a special serial contour-tracing operator in order to detect connectedness relationships. We will review evidence in favour of a particular implementation of this contour-tracing operator. Our hypothesis is that a special label is spread among all neurons that respond to a connected image component. The label distinguishes the neurons from cells that respond to other, disconnected image components.

Although our computational considerations strongly suggest that recurrence is crucial for connectedness detection, we also provide psychophysical and physiological evidence in support of our hypothesis. Processing in a feedforward network is completed as soon as information arrives at the top layer of the network. In contrast, processing in a recurrent network is not completed within a predetermined amount of time. We review evidence that the detection of connections between image regions by human observers is associated with prolonged reaction times. Moreover, we demonstrate that neurons at the earliest levels of the visual processing hierarchy are influenced from outside their classical receptive field during the detection of connectedness. Therefore, there is also direct evidence for the recurrence of visual cortical processing.

2

2 Representation of connectedness

Before considering how connectedness should be computed, it should be discussed how connectedness is represented by visual cortical neurons. One of the strategies is to represent connectedness with dedicated neurons. The visual cortex may contain 'connectedness detectors'; neurons whose activity depends on connections between image regions. However, there are a number of disadvantages associated with representing connectedness using dedicated cells. First, realistic images are usually composed of many connected and disconnected regions. It is therefore critical how these putative connectedness detectors indicate which of these regions are connected and which are not. An obvious and straightforward solution to this ambiguity is to reserve a neuron for each combination of image locations that may be connected. However, since the number of combinations of image locations scales as the square of the sampling resolution, the visual cortex would have to reserve a huge number of cells for the representation of connectedness. Even more unrealistic cell numbers are required when connectedness relationships among more than two image locations need to be represented by dedicated cells. This indicates that connectedness represents a severe case of the 'binding problem' in the spatial domain.

Second, properties of connectedness detectors constrain their location in the visual hierarchy. If a cell's response depends on the presence or absence of a connection between two image regions then, by definition, its receptive field includes these image regions and the path between them. This implies that the receptive field of a connectedness detector should comprise the entire visual field, since a connecting path may run anywhere. At face value, this indicates that cells sensitive to connectedness among image regions must reside in the higher areas of the visual cortex, where receptive fields are largest. This is at odds with the fact that grouping processes are an early, 'preattentive' visual function on which much of the subsequent visual processing depends, at least when one adopts a primarily feedforward view of visual cortical processing (e.g. ref. @6).

An alternative representation of connectedness is to label neurons that respond to a connected image component. Two labels that are plausible from a physiological point of view are an enhanced firing rate (Fig. 2A), and synchrony of neural discharges (Fig. 2B). The synchronization label was proposed by Abeles[@7] and von der Malsburg[@8,@9] and, in a preliminary form, also by Milner[@10]. These authors pointed out that the precise timing of action potentials could be utilized in order to label a group of neurons that respond to a single object. The hypothesis was put forward that neurons that respond to the various features of a single image
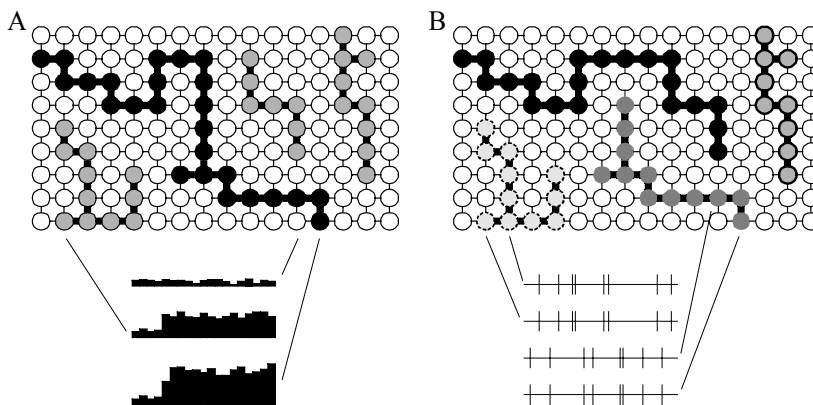
Figure 2: Representation of connectedness using labels. A, Neurons that respond to pixels belonging to one of the connected image components have a higher firing rate (black) than neurons responding to other components (grey). Unresponsive cells are shown in white. B, Neurons that respond to the same image component fire their action potentials (vertical ticks) at the same time, whereas cells that respond to different image components do not fire in synchrony. Note that neuronal synchrony can provide a different label for each image component (as indicated by linestyle and shading).

component should synchronize their discharges on a fine temporal scale, whereas neurons that respond to different image components should not fire in synchrony. In the case of connectedness detection this proposal suggests that neurons responsive to connected image components fire in synchrony, whereas neurons responding to disconnected image components do not. Synchronization provides multiple labels, and therefore, each connected image component can, in principle, receive a different label (Fig. 2B).

An enhanced firing rate is the second candidate tag. According to this proposal, neurons that respond to a connected subset of the visual image have an enhanced firing rate relative to neurons that respond to the rest of the image (Fig. 2A). There is physiological evidence for the involvement of variations in firing rate in perceptual grouping. Studies in awake monkeys have revealed that neurons in the primary visual cortex exhibit stronger responses to image elements that belong to a figure than to image elements that belong to the background[11,12]. A likely psychological correlate of this label is visual attention for the respective image region. Responses of visual cortical neurons to an attended object are enhanced relative to responses to objects that are not attended[13-16]. Enhanced firing rates provide only a single label. When neurons that respond to disconnected image regions would also exhibit an enhanced firing rate, the respective image components are spuriously labelled as

connected.

3. Algorithms for the detection of connectedness

Further constraints on connectedness detection can be derived from computational considerations. Feedforward networks appear to be particularly inefficient for the computation of connectedness. The most primitive representative of such a feedforward network is the perceptron, which has been shown to be maladapted for computing connectedness by Minsky and Papert[@2]. Perceptrons have been succeeded by parallel distributed processing (PDP) networks for which powerful training algorithms exist, like the ubiquitously used backpropagation algorithm[@4]. We therefore investigated whether these feedforward networks are better equipped to compute connectedness.

3.1 Feedforward networks with a single hidden layer

As a first approach, a feedforward network with three layers was trained to detect connectedness between two opposite corners of a square retina (Fig. 3). The first layer of the network received input from the retina. The second layer was composed of units that were connected to all neurons in the first layer. These units will be called hidden units. The third layer consisted of a single decision unit. The task was to detect a path between opposite corners that was composed of adjacent black pixels. The decision unit should only be activated when the opposite corners were connected to each other. During the training procedure patterns like the images of Fig. 3A were given as input to the network, and the synaptic weights were adjusted according to a version of the back-propagation rule[@4]. Typically, a subset of all possible patterns was used as training set, and subsequently the network was tested with a different set of images, which included some images that were not in the training set. This allowed us to study whether rules learned by the network would also apply to novel images. We were interested in the number of hidden units that was required to solve the task, and its dependence on the size of the retina.

The minimal number of hidden units for a retina of 4x4 pixels, for example, was found to be 20. The back-propagation rule did not find a correct solution when a smaller number of hidden units were used. In order to understand this strict dependence of a solution on the number of hidden units, the pattern of connections between the hidden units and the input layer was investigated for a network that had converged to a correct solution. Interestingly, units in the second layer were tuned to
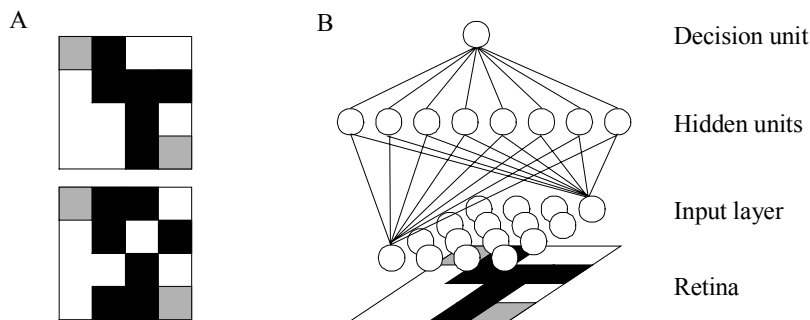
Figure 3: Three-layer feedforward network for connectedness detection. A, Two of the images on which the network was trained. In the upper image the two grey corner points were connected, but in the lower image they were not. B, The network consisted of three layers. Neurons of the input layer were connected to all units of the hidden layer. During training the strength of all connections was changed according to the back-propagation rule.

an entire path connecting the two corner positions (Fig. 4A). Hidden units were selective for what we call a *principal path*. Principal paths are interrupted as soon as the color of an arbitrary black pixel is changed to white. For the 4x4 retina the number of principal paths is 20, as is shown in Fig. 4A. Thus, the strategy learned by the network depends on the detection of principal paths. Indeed, all images in which the two corner points are connected contain at least a single principal path. Therefore, the two corner points are not connected, if all hidden units remain silent. The number of hidden units compares favorably to the 2^{14} possible input images in which the two corner points are not varied. The back-propagation rule captured a significant amount of the structure inherent in the task. This rule based on the detection of principal paths even worked for novel images, which did not appear in the training set. Unfortunately, the number of hidden units required for this solution grows exponentially with increasing image size, since principal paths can be winding (Fig. 4B). This presumably explains why, in our hands, the backpropagation failed to find solutions for retinas of 5x5 pixels or larger.

The results obtained with our feedforward network with a single hidden layer illustrate that connectedness is indeed a non-trivial property of the visual image. In the case of a 4x4 retina the network was only trained to detect the presence of a connection between two corner points. The network would require additional units to detect connections between arbitrary locations in the image. However, the largest problem of the solution by the three-layer network is the exponential growth of the required number of neurons with increasing image size. Fortunately, this exponential
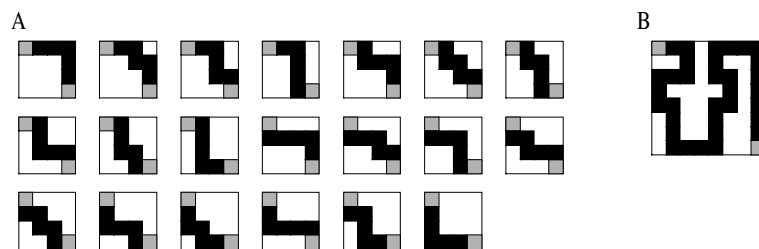
Figure 4. A, Receptive fields of the 20 units of the hidden layer. All units were selective for a principal path: an entire path connecting the two grey pixels. B, Example of a principal path for an 8x8 retina.

growth is not a property of feedforward networks in general. In the remainder of this section we will present feedforward networks with more layers between input and output. It will be seen that these network architectures require less units. However, all further algorithms that are presented (and that we are aware off) have a more natural implementation in a recurrent network than in a feedforward network. Therefore, we will first discuss the recurrent versions of these algorithms, and subsequently, how they could be implemented in a feedforward network.

3.2 Networks composed of connectedness detectors

Connectedness can be computed using connectedness detectors, units that signal the presence of a connection between two locations in an image. Representational considerations dictate that a unit is reserved for each pair of image locations, as was discussed above. The design of such a network is illustrated in Fig. 5. Neurons in the first layer of the network (layer 0) are activated in the first iteration of the algorithm. They signal adjacent black pixels, and are reproduced for each pair of adjacent image locations. Neurons in the next layer (layer 1) are selective for a connection between image locations that are separated by two pixels. A unit (P_i-P_j) of layer 1, which is selective for a connection between image locations P_i and P_j, can be activated from layer 0. This occurs when one of the neighboring positions P_k of both P_i and P_j contains a black pixel. In this case (P_i-P_k) and (P_j-P_k) are in layer 0 and are activated in the first iteration. Therefore, (P_i-P_j) is activated in iteration 2. This wiring scheme is repeated in higher layers. The algorithm proceeds by activating all cells (P_i-P_j) for which there is a location P_k, such that (P_i-P_k) and (P_j-P_k) are both active. When P_i to P_j are connected through the shortest possible path, there is always a location P_k, such that (P_i-P_k) and (P_j-P_k) are the next lower layer. In this

7

A

Layer n $(2^{n-1}+1 \leq d \leq 2^n)$

Layer 2 $(3 \leq d \leq 4)$

Layer 1 $(d = 2)$
Layer 0 $(d = 1)$

Retina

B

$P_i\text{-}P_j$

$P_i\text{-}P_k \wedge P_j\text{-}P_k$ $P_i\text{-}P_l \wedge P_j\text{-}P_l$

$(P_i\text{-}P_k)$ $(P_j\text{-}P_k)$ $(P_i\text{-}P_l)$ $(P_j\text{-}P_l)$

X

Π_ι Π_φ

1   2   3   4   5
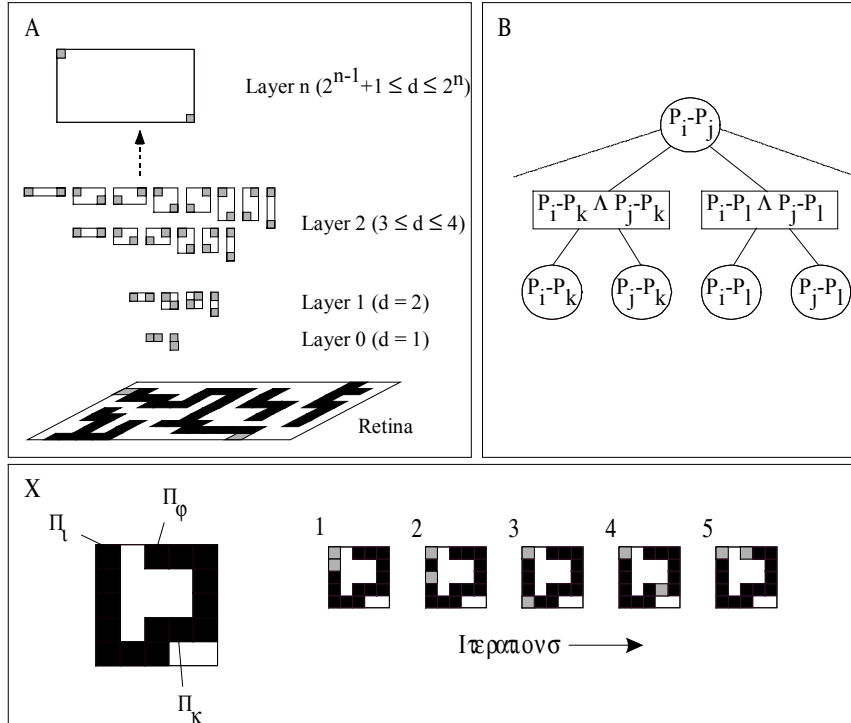
Π_κ

Ιτερατιονσ $\longrightarrow$

Figure 5: Detection of connectedness using dedicated neurons. A, Neurons are reserved for each combination of image positions. Neurons detecting connections between adjacent image positions are situated in layer 0. In subsequent layers connections between positions that are separated by a larger distance (d) in the image are represented. B, In each iteration of the algorithm those units $(P_i\text{-}P_j)$ are activated for which there is an image location P_k, such that neurons $(P_i\text{-}P_k)$ and $(P_j\text{-}P_k)$ are both active. This can be achieved by reserving a linear threshold unit for each triplet (P_i, P_j, P_k) of image locations. Note that neurons $(P_i\text{-}P_k)$ and $(P_j\text{-}P_k)$ can, in principle, be located in a higher layer than $(P_i\text{-}P_j)$. (**C**) Image, in which locations P_i and P_j are connected via a detour through P_k. Right, subset of units that are activated in successive iterations. Note that the rightmost unit, which is activated in the 5th iteration, is located in layer 1 of the network (see A).

situation $(P_i\text{-}P_j)$ will be activated one iteration later than $(P_i\text{-}P_k)$ and $(P_j\text{-}P_k)$.

An important observation can be made for situations in which image locations are connected via a detour. Suppose that P_i and P_j are separated by 2 pixels, but connected through P_k, which is separated by a larger distance in the image (Fig. 5C). A connection between P_i and P_k on the one hand and a connection between P_j and P_k on the other hand, implies a connection between P_i and P_j. Thus, neurons $(P_i\text{-}P_k)$

and $(P_j\text{-}P_k)$ should activate $(P_i\text{-}P_j)$ (Fig. 5B). However, units $(P_i\text{-}P_k)$ and $(P_j\text{-}P_k)$ are in a higher layer of the network than $(P_i\text{-}P_j)$, which is in layer 1. Thus, activation of $(P_i\text{-}P_j)$ depends on feedback from higher layers. Indeed, the order in which connectedness detectors $(P_a\text{-}P_b)$ are activated is determined by the distance between P_a and P_b as measured along the current path, rather than by the actual distance between P_a and P_b. It can be seen that if P_a and P_b are connected by a path of length L, then $(P_a\text{-}P_b)$ is activated in iteration t, where $t = {}^2\log L + 1$. This has two important implications. First, feedback from higher to lower layers is essential for images with pixels that are connected via a detour. Second, the order in which connectedness detectors are activated depends on the layout of paths in the image. These considerations imply that there is an inherent seriality in the detection of connectedness (see also ref. @5).

The algorithm requires $N^2 \cdot M^2$ units for an N·M retina. There are 10^6 axons in the optic nerve, and this provides a rough estimate of N·M in human vision. In order to implement this algorithm, the human visual cortex would have to reserve about 10^{12} neurons, which is implausible from a physiological point of view. Unfortunately, the activation rule by which neurons $(P_i\text{-}P_k)$ and $(P_j\text{-}P_k)$ activate $(P_i\text{-}P_j)$ is non-linear, and therefore, $N^2 \cdot M^2$ linear threshold units cannot carry out the algorithm[a]. It can be implemented using $O(N^3 \cdot M^3)$ linear threshold units (Fig. 5B). Therefore, the 10^{12} neurons required for implementing this algorithm in the human visual system provides a lower bound, since it seems unlikely that the non-linear version of the dendritic integration rule can be carried out by cortical neurons.

Any serial algorithm that is completed within a limited number of iterations can also be carried out by a feedforward network. Therefore, it is possible, albeit somewhat artificial, to implement this algorithm in a feedforward network. For the implementation, a layer is reserved for each iteration of the algorithm. The maximal number of iterations is $O(\log(N\cdot M))$, since the maximal length of a path in the image is $O(N\cdot M)$. Units selective for a particular connection $(P_a\text{-}P_b)$ have to reappear in almost every layer, because the order in which connectedness detectors are activated varies between images. A feedforward network composed of linear threshold units would require $O(N^3 \cdot M^3 \log (N\cdot M))$ units in $O(\log (N\cdot M))$ layers. Nonetheless, the number of units scales much more favorably with the size of the retina than in the three-layer feedforward network that was discussed above.

---

[a] Linear threshold units sum their inputs after scaling by the synaptic weights and are activated when the sum crosses a threshold (e.g. ref @2).

3.3 Connectedness detection using labels

Connectedness can also be represented by labelling neurons that respond to a connected image region. The most powerful algorithms that are available for the computation of connectedness on parallel computers use such labels (e.g. ref. @17). These algorithms are rather economic in terms of the number of units that are required. Here, we discuss a simple version of a label-spreading algorithm, which is readily implemented in a neural network with a single layer. Cells in the network of Fig. 6 are activated from a particular image location, and importantly, units that receive input from adjacent image positions are reciprocally connected. The algorithm presupposes that a tag is spread through these 'horizontal' connections, but only among units that are activated by the image. This implies that a connection is only effective if both the pre- and postsynaptic cell is active. It is convenient to refer to the subset of connections that are switched on in this way as the *interaction skeleton* (thick lines in Fig. 6C,D). At step one of the algorithm, one of the responsive neurons receives the tag. It is easy to see that in subsequent iterations the tag spreads to precisely those neurons that respond to connected image regions. The algorithm relies on the use of two distinct signals for each position of the image. The first signal which depends on the colour of the pixel at a particular image location is the activation of network units. The second signal is the label that spreads through the interaction skeleton, the nature of which will be discussed in a later section of this chapter. Note that the spread of the tag is not determined by local features of the image but, in accordance with the nature of connectedness, by the global pattern of activated cells, which determines the composition of the interaction skeleton. Small differences in the image may lead to drastic changes in connectedness relations among image regions (as was demonstrated in Fig. 1). However, these differences are reflected in equivalent changes in the composition of the interaction skeleton (Fig. 6).

The number of units necessary to carry out the algorithm equals N·M for a retina of size N·M, and does not require more neurons than is physiologically plausible. However, the number of iterations required by the algorithm is proportional to the length of the path where tag-spreading is initiated. Thus, it requires more iterations than the algorithm based on connectedness detectors that was discussed above. It can be speeded up, however, by the addition of additional layers, as will be discussed in the next section.

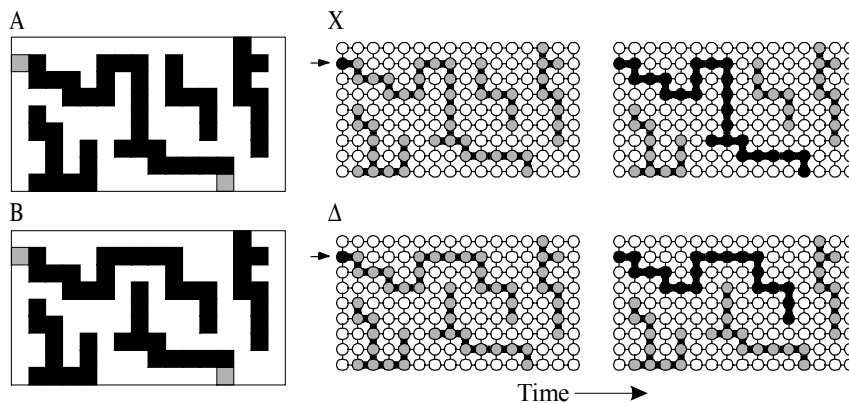The tag-spreading algorithm can also be improved by the introduction of

Figure 6: A tag-spreading algorithm for connectedness detection. A, Image in which the two grey tiles are connected. B, Slightly modified image in which the two grey tiles are not connected. C, Schematic illustration of the interaction skeleton. Neurons (circles) with adjacent receptive fields are reciprocally coupled (lines). Neurons that are activated by the image have been indicated in grey. The interaction skeleton consists of those tag-spreading connections that link active neurons (thick lines). In order to detect connectedness, a label (black) is spread among activated neurons, starting at the arrow. The label will only propagate to neurons that respond to image locations that are connected to the tile that is indicated by the arrow. Note that the tag spreading process could be initiated anywhere in the interaction skeleton. D, A local change (between C and D) in the image results in a global change in the composition of the interaction skeleton, and in an equivalent change in the tag spreading process.

multiple labels. This allows the network to represent multiple connected image components at the same time. Shiloach and Vishkin[@17] showed that the number of iterations can be reduced to O(log (N·M)) by using multiple labels on a parallel computer. However, in order to obtain this computational speed, each unit needs to be connected to every other unit, and the number of available labels has to be proportional to the number of pixels. We refrain from going into the details of these extremely cost-effective algorithms, because they lack physiological plausibility.

Obviously, tag-spreading can also be carried out by a feedforward network. For each iteration of the recurrent algorithm a layer should be reserved in the feedforward network. Therefore, the number of layers in the feedforward network should be proportional to N·M, since the maximal number of iterations of the recurrent algorithm is proportional to the longest possible path, which is O(N·M). Thus, the feedforward version of the algorithm requires O(N·M) layers with N·M units per layer. The total number of units is $O(N^2 \cdot M^2)$, which is physiologically implausible. Nonetheless, this feedforward algorithm can be carried out by
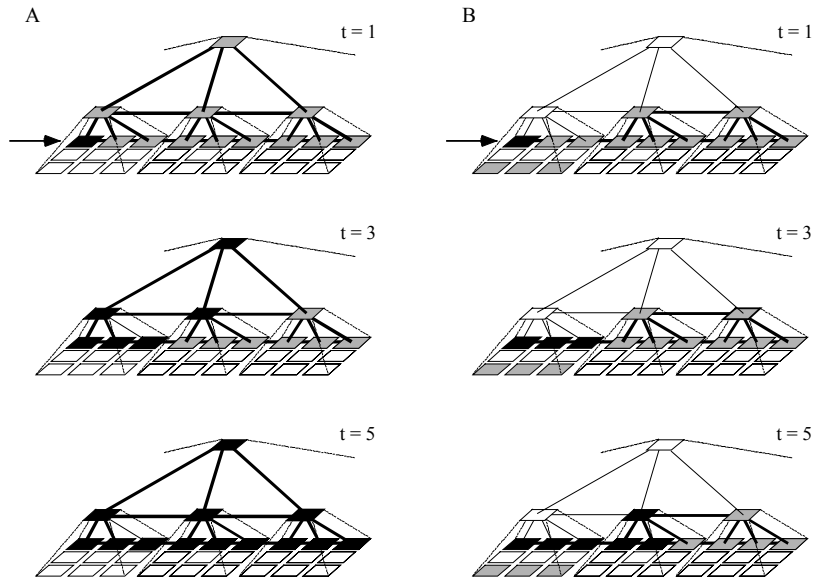
11

Figure 7: Pyramid algorithm for the detection of connectedness. A, Three layers of a pyramid network are shown. Neurons in higher layers are reciprocally connected to 9 neurons of the next lower layer. These neurons are only active if it can be excluded (on the basis of simple heuristics) that there are multiple disconnected image components in their receptive field. The interaction skeleton is shown as thick lines. Only a subset of the connections to the neurons in the upper layer is depicted. Note that the higher layers speed up tag-spreading. B, The presence of disconnected image components deactivates part of the pyramid. This retards tag-spreading considerably.

$O(N^2 \cdot M^2)$ linear threshold units, which is less than the $O(N^3 \cdot M^3 \log (N \cdot M))$ required by the feedforward network composed of connectedness detectors, that was discussed above.

3.4 Pyramid algorithms

In regions where the path along which the label should spread is relatively straight, and different image components remote, additional layers in the network can speed up tag-spreading. Edelman[@18] described an algorithm that takes advantage of this possibility, and can be implemented on a pyramid computer. Fig. 7 illustrates such a pyramid algorithm, schematically. The lowest layer of the pyramid corresponds to

the tag-spreading network described in the previous section. However, additional layers are included, which receive input from a number of neurons (9 neurons in Fig. 7) of the next lower layer. Units in higher layers have larger receptive fields, but are only active when it can be verified, on the basis of simple heuristics, that there is a single connected image component in their receptive field. An example of such an heuristic is the requirement that all pixels within the receptive field are on the same horizontal or vertical row (Fig 7A)[a]. More complicated heuristics, which also can be computed by the pyramid in a feedforward way, have been described previously[@18,@19]. Valid heuristics establish that there is a single connected component within the receptive field of neuron. However, when a neuron remains silent, disconnected image components within its receptive field may nonetheless be connected, for example through a detour outside the receptive field. Activity in higher layers can speed up the tag-spreading process considerably, as is illustrated in Fig. 7A. Since receptive fields are larger in higher layers, adjacent receptive fields are separated by a larger distance in the visual field. Thus, larger distances in the visual field are bridged within a single iteration. For straight paths, the time that is required by a pyramid algorithm can be reduced to O(log(L)), where L is the length of the path that should be labelled[@18]. However, when the path is not straight, or other image elements are nearby, tag-spreading requires high resolution information, and only occurs in the lower layer(s) of the pyramid. This reduces the speed to a few pixels per iteration (Fig. 7B). Thus, the time required by a pyramid algorithm depends on the visual input, and varies between O(log(L)) and O(L).

3.5 Comparison of the various algorithms

Obviously, we cannot discuss the plethora of algorithms that can be designed for the detection of connectedness. It is possible, if not likely, that algorithms which require fewer units or which are faster will be found in the future. Nonetheless, we believe that our considerations strongly indicate that recurrent networks have advantages over feedforward networks for the computation of connectedness. Networks that compute connectedness using 'connectedness detectors' as well as networks that use tag-spreading benefit from recurrence. This is presumably related to an inherent seriality in the computation of connectedness (see ref. @5).

---

[a] For human vision, equivalent requirements are that (1) all image elements within a receptive field are approximately colinear, and (2) there are no line terminations.

It is our aim in the remainder of this chapter to compare the algorithms that were introduced to physiological and psychophysical data. Let us therefore recapitulate the distinctive features of the various algorithms. First, algorithms differ in the number of image regions in which connectedness computations take place simultaneously. The number of image regions that can be labelled by tag-spreading algorithms is equal to the number of available labels. In contrast, algorithms that use connectedness detectors exhibit no limitation in this respect, since computations can be performed in parallel on all image regions. Second, there are differences in the number of iterations required by the various network types, and their dependence on the visual input. This should allow a comparison between the number of iterations that algorithms require, and human reaction times in tasks that require connectedness detection. Third, algorithms implemented in recurrent networks predict that the effects of feedback, and within-layer interactions can be measured at the earliest levels of the visual cortical processing hierarchy. In contrast, neurons in the early layers of the feedforward implementations that were discussed above are tuned to local features of the image. Fourth, there are differences in the number of units that are required to perform the computations. In this respect tag-spreading algorithms are clearly superior over algorithms based on connectedness detectors. Tag-spreading allows the network to segregate an entire connected image component from other components and the background. This is beneficial for subsequent visual processes like, for example, object-recognition. In the network composed of connectedness detectors, on the other hand, only pair-wise connections are explicitly represented. The required neural hardware would grow beyond bounds if dedicated units should represent triplets or larger numbers of connected pixels. In other words, connectedness representation is the spatial version of the binding problem.

4 Physiology of connectedness detection

There is, to our knowledge, no physiological evidence bearing on the computation of connectedness using connectedness detectors. Therefore, it is imperative to restrict our discussion of physiological data to the tag-spreading algorithms.

4.1 Evidence for synchrony as a label for connectedness

It has been suggested that neuronal synchrony might label responses to related

features and demarcate these responses from responses to unrelated features[@7-@10,@20]. The suggestion that synchrony might label connected image regions is in accordance with this hypothesis. Thus, neurons responding to a connected image region should fire their action potentials at approximately the same time, whereas neurons responsive to unconnected image regions should not fire in synchrony (Fig. 2B). Neuronal synchrony allows simultaneous labelling of multiple image regions, since synchronization can occur selectively between the responses to connected image regions, and not between responses to non-connected regions. There is physiological evidence in support of this hypothesis. Neurons in the primary visual cortex fire in synchrony when they respond to segments of the same visual contour. However, when the same neurons respond to separate contour segments the strength of synchrony is reduced[@21-@24].

A prediction from this hypothesis is that whenever synchronization among visual cortical neurons is disturbed, the detection of connectedness should be impaired. This begs the question whether there are *selective* disturbances of neuronal synchrony, sparing further aspects of visual cortical processing. Strabismic amblyopia appears to be such a disorder. It is a condition that develops in a proportion of humans and animals with a squint. In order to avoid double vision, these individuals develop a strategy in which they continuously suppress vision with one of their eyes, and fixate objects with the other eye. The non-fixating eye develops a form of amblyopia that is characterised by a reduced acuity, and distortions of visual perception[@25,@26]. In contrast to other types of amblyopia[@27], strabismic amblyopia is not associated with a loss of neurons that are activated through the affected eye. Paradoxically, fine gratings that cannot be perceived with the amblyopic eye still evoke responses from visual cortical neurons that are as vigorous as responses evoked through the normal eye[@28-@31]. In a recent study, patterns of synchrony were explored among neurons of the primary visual cortex of cats with strabismic amblyopia[@32]. The strength of synchrony among responses evoked through the amblyopic eye was lower than the strength of synchrony among responses evoked through the normal eye. Furthermore, the loss of synchrony among responses of neurons activated through the amblyopic eye was most pronounced for fine gratings, which exceeded the threshold of the amblyopic eye. These results indicate that strabismic amblyopia is indeed associated with a selective loss of neuronal synchrony among neurons at the lowest level of the visual cortical hierarchy[@32].

This disruption of tag-spreading should be associated with a failure of connectedness detection in patients with strabismic amblyopia. Hess et al.[@33] explored visual distortions that patients with strabismic amblyopia perceive when viewing with their affected eye. They asked amblyopic observers to look with their
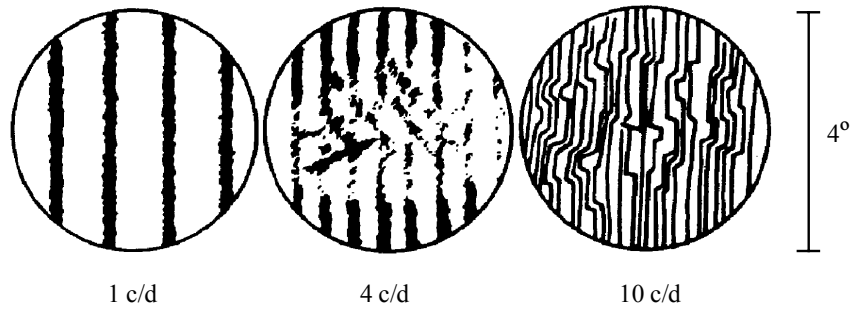
1 c/d   4 c/d   10 c/d

Figure 8: Spatial distortions of amblyopic vision. Gratings of various spatial frequencies were shown to the amblyopic eye of a patient with strabismic amblyopia. After looking at these patterns, the patient was asked to open the normal eye and to draw his impression of the visual stimulus. Shown in this figure are the drawings of gratings with different spatial frequencies. The scale bar denotes the size of the presented stimuli in degrees of visual angle. Note that with increasing spatial frequency the distortions become more severe. The patient reports on seeing connections between contour segments that are actually disconnected. Conversely, to the amblyopic eye elongated contours appear interrupted (reproduced with permission from ref. @33).

amblyopic eye at gratings of various spatial frequencies and to make drawings of the distorted percepts. Examples of such drawings are shown in Fig. 8. Apparently, amblyopic patients fail to perceive connections among contours that are connected, and vice versa, perceive connections among contours that are actually disconnected. These visual distortions are most pronounced for fine gratings. Taken together, these results indicate that a loss of synchrony is associated with a pronounced deterioration of connectedness detection.

4.2 Evidence for rate enhancement as a label for connectedness

The second, physiologically plausible label that the visual cortex could utilize is an enhanced firing rate. Physiological studies have shown that the visual cortex can indeed label image regions using an enhanced firing rate[11],[12]. Thus, according to this proposal, neurons that respond to a particular connected image region should exhibit a response enhancement relative to neurons that respond to other image regions (Fig. 2A).

In order to strengthen the link between connectedness detection and tag-spreading, we have recently started to record the activity of neurons in the primary visual cortex of awake monkeys who were trained in a task that requires the detection of connections among image regions[34]. The animals had to find a visual

target, which was connected to a fixation point through a curve (Fig. 9A). Activity in the primary visual cortex was recorded while the monkeys performed this task. The firing rate of neurons in the primary visual cortex was influenced by the connectedness relationships. In particular, neurons exhibited an enhanced response when the curve that connected the fixation point to the target crossed their receptive field. Thus, the curve that was relevant for the monkeys, since it indicated the correct eye movement, was labelled using an enhanced firing rate. Response enhancement occurred after a latency of about 200 ms, which is long after the onset of visual responses in the primary visual cortex. The additional delay is presumably related to the temporal constraints on the tag-spreading process. The primary visual cortex is at the lowest level of visual cortical processing hierarchy. Therefore, these results provide direct evidence for the recurrence of tag-spreading.

4.3 Temporal constraints on tag-spreading

The results in cats and humans with amblyopic vision suggested neuronal synchrony as a label for connected image regions, whereas the results of the trained monkeys indicated that rate enhancement is the label. We would like to propose that temporal constraints on the tag-spreading process need to be taken into account in order to understand this discrepancy.

Let us first consider how the synchrony label spreads through the interaction skeleton. Closely located neurons in the visual cortex exhibit a tendency to fire their action potentials at approximately the same time[@23]. This suggests that local cortical circuits are laid out to allow for cooperativity of firing among these local cell groups (see also ref. @35). Thus, neurons that have reached the firing threshold are likely to activate others that are also close to the threshold rapidly, thus starting a burst of activity[@36]. It seems likely that such bursts, or neuronal avalanches as they may be called, travel some distance along the interaction skeleton (Fig. 6, 7). Then they die out, for example because inhibition builds up or because they collide with avalanches that started at other positions. These bursts of activity would provide a mechanism for synchronizing neurons that are connected through the interaction skeleton. The course that an avalanche can take is constrained by the layout of the interaction skeleton. Neurons that are not connected by the interaction skeleton cannot participate in the same avalanche, and will not synchronize their discharges.

In order to synchronize neurons that are separated by a large number of synapses, an avalanche should have enough time to develop. Many studies have reported burst durations of approximately 8-10 ms[@20-@24,@37,@38]. In such a short time an avalanche cannot, with all likelihood, spread further than 6-8 synapses. However, a prolongation of the average avalanche duration may help to synchronize neurons

17

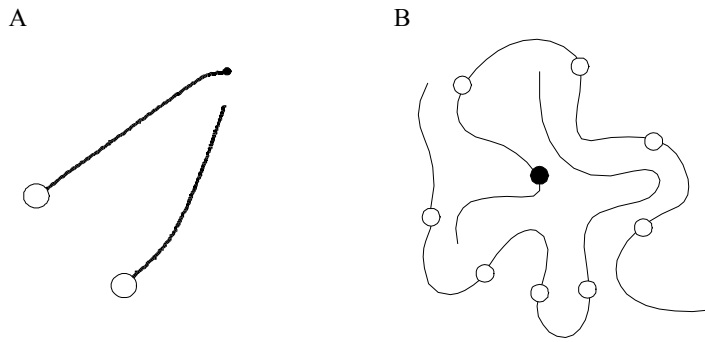A   B

Figure 9: Tasks that probe connectedness detection. A, Task used in ref. @34. Monkeys were trained to fixate a small point (black circle) and to make an eye movement to a target (white circle) that was connected to the fixation point. B, Curve tracing task employed by Jolicoeur and coworkers (ref. @19,@42,@43). The image consists of a pair of curves, one of which is connected to the fixation point (black circle). A target (white circle) could appear on the same curve as the fixation point at various distances. The circular target could also appear on the other curve. During the actual experiment the target appeared only at a single of these eight possible locations. Moreover, the curves differed between successive trials. Subjects were required to indicate whether the target was on the same or on a different curve as the fixation point.

that are separated by many synapses. Nelson *et al.*[@39] found that visual cortical neurons synchronize their discharges on a variety of timescales. In their study the synchronization patterns were subdivided in three categories: T-, C- and H-type coupling. T-type coupling (T for tower) was most precise, and associated with an average burst duration of 3 ms. The average precision of synchronization for C-type coupling (C for castle) was 30 ms, and for H-type coupling (H for hill) 400 ms. The spread of synchrony was found to be larger for H- and C- type coupling than for T-type coupling. Thus, longer avalanches are associated with a more extensive spread of synchrony through the interaction skeleton. The range of possible avalanche durations can be extended even further, since rate changes that covary among neurons on a timescale of even seconds have been found[@40,@41].

There is, however, also a disadvantage associated with long avalanche durations. A new burst of activity should only be initiated after the last avalanche died out. Therefore, long avalanche durations need to be accompanied by a low probability of their initiation. If by chance avalanches simultaneously occur in unconnected regions of the interaction skeleton, the respective image regions are spuriously labelled as if they are connected. Thus, when the avalanche duration is too long, disconnected image components cannot be labelled simultaneously on a perceptually relevant time scale. There is a trade-off between the duration of avalanches, the frequency with

which they can be initiated, and the probability of spurious synchronization. One of the virtues of the synchrony label, the possibility for simultaneous representation of multiple objects on a perceptual timescale, is sacrificed with very long avalanche durations. However, this cannot be avoided if image components are elongated, and many synapses need to be bridged by the tag-spreading process. More prolonged changes in firing rates in a connected compartment of the interaction skeleton are a likely correlate of shifts of visual attention[@13-@16]. It is plausible that image components encoded by neurons that did not participate in avalanches for a perceptually relevant time-span are temporarily excluded from conscious visual perception. At these timescales the synchrony label is similar, if not identical, to the label of enhanced firing rates. The only distinction between the two labels is the timescale on which neurons in unconnected regions of the interaction skeleton may alternate, the timescale for the synchrony label being considerably shorter than the timescale of attentive shifts. Thus, the two labelling mechanisms have a highly similar computational structure, and can rely on the same architecture of tag-spreading connections.

Taken together, the physiological data strongly suggest that the visual cortex evaluates connectedness relationships using tag-spreading. In order to strengthen this conclusion, and to discriminate between the various tag-spreading algorithms we will now review human performance in tasks that require connectedness detection.

5 The psychophysics of connectedness detection

Jolicoeur and coworkers[@19,@42,@43] investigated human performance in curve tracing tasks. In these tasks, observers have to judge whether a circular targets is on the same, or on a different curve as a fixation point (Fig. 9B). Thus, tracing tasks probe the capabilities of humans in detecting connections between image regions. Reaction times exhibit an almost linear dependence on the distance between target and fixation point, when both are on the same curve[@19,@42]. Large distances between the fixation point and the target, as measured along the curve, can increase the reaction time by as much as 300 ms. These results provide evidence against algorithms which are completed within a fixed number of iterations, and also against algorithms which require $O(\log(L))$ iterations (where L is the length of the curve to be traced).

The performance of the pyramid algorithm, on the other hand, is largely consistent with the pattern of human reaction times. A particularly important result in this respect is the finding that the speed of tracing depends on the curvature of a curve, and on its distance from other image components. Indeed, when a curve is contorted or other curves are nearby, tracing speed is reduced[@19]. Moreover, reaction

time data suggest that curve tracing can only be performed for a single curve at a time[43]. This is evidence against algorithms based on connectedness detectors, because these algorithms can simultaneously perform their calculations in different image regions. It also argues against tag-spreading algorithms that utilize multiple tags, since this also allows simultaneous computations for disconnected image regions.

In summary, results on curve tracing largely corroborate the physiological data, which suggested that connectedness is detected by tag-spreading. The algorithm based on connectedness detectors is inconsistent with a linear dependence of reaction times on the length of the curve to be traced, and erroneously predicts that tracing can take place in multiple image regions at the same time. With respect to the tag-spreading algorithms, the reaction time data are in good agreement with the predictions of the pyramid algorithm with a single label.

6 Concluding remarks

We have indicated how connectedness can be represented efficiently and economically by labelling responses of neurons to connected image regions. Tag-spreading algorithms evaluate connectedness relationships by spreading these labels through the interaction skeleton. The architecture of corticocortical connections is in accordance with the requirements of the tag-spreading algorithms. Indeed, synchronization between spatially separate neurons depends on the integrity of corticocortical connections (reviewed in ref. @44). The physiological and psychophysical data reviewed here are also in agreement with the tag-spreading algorithms.

When an elongated curve is traced, neurons that respond to it exhibit an enhanced firing rate[34]. Rate modulations are well established correlates of shifts of visual attention[13-16]. Thus, the propagation of the enhanced firing rate label along the interaction skeleton might be associated with a spread of visual attention over the respective connected image component. There is psychophysical and neuropsychological evidence in support of this hypothesis.

Patients with lesions of the parietal cortex sometimes fail to perceive visual objects, when other objects are presented at the same time. This symptom is called extinction, and is usually interpreted as a failure to shift visual attention from one object to another[45]. Extinction can be prevented, however, when the respective image components are connected to each other[46], which suggests that the connection aids in propagating visual attention from one image component to another. Experiments in healthy individuals provide evidence that is in accordance with this interpretation. Kramer and Jacobson[47] asked subjects to discriminate

between features of contour segments. Different linestyles of a target contour were associated with right and left button presses. A target contour could be flanked by distractors, which possessed response-compatible or response-incompatible features. On trials with response-incompatible distractors, for example, a target contour with a linestyle that required a left button response would be flanked by distractors which required a right button response. These response-incompatible distractors increased reaction times substantially, and in particular when they were connected to the target contour. Thus, it is difficult to disregard irrelevant contour segments that are connected to an attended contour. Taken together, these results strengthen the proposal that visual attention has a tendency to spread from attended image regions to image elements that are connected to these regions, as a direct consequence of the tag-spreading process.

One of most important functions of visual attention is to select image features that belong to a single object[48]. Connectedness is one of the Gestalt criteria which allow the visual system to decide which image elements belong to a single object. Therefore, spreading an enhanced firing rate through the interaction skeleton provides a mechanism for the attentive selection of entire objects, in other words, object-based attention (see also ref. [49]).

Acknowledgements

References

[1]. I. Rock and S. Palmer, *Sci. Am.* **263** (6), 48 (1990).

[2]. M.L. Minsky & S.A. Papert, *Perceptrons* (MIT Press, Cambridge, 1969).

[3]. F. Crick, *Nature* **337**, 129 (1989).

[4]. D.E. Rumelhart and J.L. McClelland, *Parallel distributed processing* (MIT Press, Cambridge, 1986).

[5]. S. Ullman, *Cognition* **18**, 97 (1984).

[6]. E.T. Rolls and M.J. Tovée, *Proc. R. Soc. Lond. B* **257**, 9 (1994).

[7]. M. Abeles, *Local cortical circuits* (Springer-Verlag, Berlin, 1982).

[8]. C. von der Malsburg, *The correlation theory of brain function* (Internal report 81-2,

Max-Planck-Institute for Biophysical Chemistry, Göttingen, 1981).

@9. C. von der Malsburg and W. Schneider, *Biol. Cybern.* **54**, 29 (1986).

@10. P.M. Milner, *Psychol. Rev.* **81**, 521 (1974).

@11. V.A.F. Lamme. *J.Neurosci.* **15**, 1605 (1995).

@12. K. Zipser, V.A.F. Lamme and P.H. Schiller. *J.Neurosci.* **16**, 7376 (1996).

@13. J.H.R. Maunsell, *Science* **270**, 764 (1995).

@14. W.T. Newsome, *Curr. Biol.* **6**, 357 (1996).

@15. M.I. Posner and S. Dehaene, *Trends Neurosci.* **17**, 75 (1994).

@16. S.P. Wise and R. Desimone, *Science* **242**, 736 (1988).

@17. Y. Shiloach and U. Vishkin, *J. Algorit.* **3**, 57 (1982).

@18. S. Edelman, *Comp. Vision Graph. Image Process.* **40**, 169 (1987).

@19. P. Jolicoeur, S. Ullman and M. Mackay, *J. Exp. Psychol.: Hum. Percept. Perform.* **17**, 997 (1991).

@20. W. Singer and C.M. Gray, *Annu. Rev. Neurosci.* **18**, 555 (1995).

@21. M. Brosch, R. Bauer and R. Eckhorn, *Cereb. Ctx.* **7**, 70 (1997).

@22. W.A. Freiwald, A.K. Kreiter and W. Singer, *Neurorep.* **6**, 2348 (1995).

@23. C.M. Gray, P. König, A.K. Engel and W. Singer, *Nature* **338**, 334 (1989).

@24. M.S. Livingstone, *J. Neurophysiol.* **75**, 2467 (1996).

@25. R. Sireteanu, W.D. Lagrèze, D.H. Constantinescu, *Vision Res.* **33**, 677 (1993).

@26. G.K. von Noorden, *Binocular vision and ocular motility. Theory and management of strabismus.* (Blackwell, Oxford, 1990).

@27. D.H. Hubel and T.N. Wiesel, *J. Physiol.* **206**, 419 (1970).

@28. C. Blakemore and F. Vital-Durand, *Ophtal. Physiol. Opt.* **12**, 83 (1992).

@29. Y.M. Chino, M.S. Shansky, W.L. Jankowski and F.A. Banser, *J. Neurophysiol.* **50**, 265 (1983).

@30. D.P. Crewther and S.G. Crewther, *Exp. Brain Res.* **79**, 615 (1990).

@31. R.D. Freeman and T. Tsumoto, *J. Neurophysiol.* **49**, 238 (1983).

@32. P.R. Roelfsema, P. König, A.K. Engel, R. Sireteanu and W. Singer, *Eur. J. Neurosci.* **6**, 1645 (1994).

@33. R.F. Hess, F.W. Campbell and T. Greenhalgh, *Pflügers Arch.* **377**, 201 (1978).

@34. P.R. Roelfsema, V.A.F. Lamme and H. Spekreijse, *Soc. Neurosci. Abstr.* **23**, 1544 (1997).

@35. D.C. Somers, S.B. Nelson and M. Sur, *J. Neurosci.* **15**, 5448 (1995).

@36. P.C. Bush and R.J. Douglas, *Neural Comp.* **3**, 19 (1991).

@37. R. Eckhorn *et al, Biol. Cybern.* **60**, 121.

@38. D.Y. Ts'o, C.D. Gilbert and T.N. Wiesel, *J. Neurosci.* **6**, 1160 (1986).

@39. J.I. Nelson, P.A. Salin, M.H.J. Munk, M. Arzi and J. Bullier, *Vis. Neurosci.* **9**, 21

(1992).

@40. F. Amzica and M. Steriade, *J. Neurophysiol.* **73**, 20 (1995).

@41. M. Bach and J. Krüger, *Exp. Brain Res.* **61**, 451 (1986).

@42. P. Jolicoeur, S. Ullman and M. MacKay, *Mem.Cognit.* **14**, 129 (1986).

@43. P. Jolicoeur In *Computational Processes in Human Vision: An Interdisciplinary Perspective*, ed. Z. Pylyshyn, p. 133-168 (Ablex, Norwood NJ, 1988).

@44. P.R. Roelfsema, A.K. Engel, P. König and W. Singer, *J. Cognit. Neurosci.* **8**, 603 (1996).

@45. M.J. Farah, *Visual Agnosia* (MIT Press, Cambridge, 1990).

@46. G.W. Humphreys and M.J. Riddoch, In *Attention and Performance XIV*, pp. 143-162 (MIT Press, Cambridge, 1993).

@47. A.F. Kramer and A. Jacobson, *Percept. Psychophys.* **50**, 267 (1991).

@48. A.M. Treisman and G. Gelade, *Cogn. Psychol.* **12**, 97 (1980).

@49. J. Duncan, *J. Exp. Psychol.: Gen.* **113**, 501 (1984).