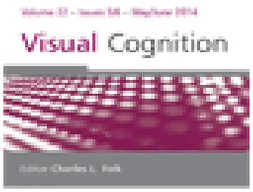

Visual Cognition

ISSN: 1350-6285 (Print) 1464-0716 (Online) Journal homepage: https://www.tandfonline.com/loi/pvis20

VAM: A neuro-cognitive model for visual attention control of segmentation, object recognition, and space-based motor action

Werner X. Schneider

To cite this article: Werner X. Schneider (1995) VAM: A neuro-cognitive model for visual attention control of segmentation, object recognition, and space-based motor action, Visual Cognition, 2:2-3, 331-376, DOI: 10.1080/13506289508401737

To link to this article: https://doi.org/10.1080/13506289508401737

VISUAL COGNITION, 1995, 2 (2/3), 331–375

VAM: A Neuro-cognitive Model for Visual Attention Control of Segmentation, Object Recognition, and Space-based Motor Action

Werner X. Schneider

Ludwig-Maximilians-University, Munich, Germany

This paper introduces a new neuro-cognitive Visual Attention Model, called VAM. It is a model of visual attention control of segmentation, object recognition, and space-based motor action. VAM is concerned with two main functions of visual attention—that is "selection-for-object-recognition" and "selection-for-space-based-motor-action". The attentional control processes that perform these two functions restructure the results of stimulus-driven and local perceptual grouping and segregation processes, the "visual chunks", in such a way that one visual chunk is globally segmented and implemented as an "object token". This attentional segmentation solves the "inter- and intra-object-binding problem". It can be controlled by higher-level visual modules of the what-pathway (e.g. V4/IT) and/or the where-pathway (e.g. PPC) that contain relatively invariant "type-level" information (e.g. an alphabet of shape primitives, colors with constancy, locations for space-based motor actions). What-based attentional control is successful if there is only one object in the visual scene whose type-level features match the intended target object description. If this is not the case, where-based attention is required that can serially scan one object location after another.

VAM's basic architecture and processing dynamics explain a large data base from experimental psychology, namely "similarity effects" (Duncan & Humphreys), "local feature contrasts" (Nothdurft), and "conjunction search"

Requests for reprints should be sent to Werner X. Schneider, Lugwig-Maximilians-University, Department of Experimental Psychology, Leopoldstrasse 13, 80802 Munich, Germany. E-mail: wxs@mip.paed.uni-muenchen.de.

I wish to thank Alan Allport, Heiner Deubel, Rainer Goebel, Lex van der Heijden, and an anonymous reviewer for extremely helpful and constructive comments; Nike Hucke, Alexandra Tins, and especially Heidi John for their indispensable advice in improving the language; Claus Bundesen and Hitomi Shibuya for a perfectly organized and very stimulating conference; and, finally, Wolfgang Prinz for his constant support.

(Treisman). Furthermore, "spatial precueing" (Posner), a "categorial effect in lateral masking" (Styles & Allport), and the "coupling between saccades and object recognition" (Deubel & Schneider) are explicated with the same attentional mechanisms. Moreover, a common neurophysiological interpretation of stimulus-driven and attentional segmentation is given that relies on synchronized neuronal activation (Milner, Malsburg, Singer, Eckhorn), and the results of single-cell studies on visual attention (Moran & Desimone, Motter, Chelazzi et al.) are discussed in relation to the VAM mechanisms. VAM's explanatory range and predictive capabilities are demonstrated by providing a new perspective on "Eriksen-interference" and on the "coupling between space-based motor action (e.g. saccades) and object recognition". A number of new predictions are made for both experimental situations. Finally, VAM's relationship to other theories of visual attention (Treisman, LaBerge et al., Kosslyn et al., Olshausen et al., Goebel, Wolfe, Van der Heijden, Bundesen, Duncan et al.) is analysed and evaluated. The paper closes with a discussion of challenging issues and open questions.

The Current Status of Visual Attention Research

"Visual attention" is currently a highly active research area in experimental psychology, the neurosciences, and in computational modelling. In experimental psychology, the field consists of data and theories on visual search (for summaries, see Duncan & Humphreys, 1989; Treisman, 1988; Wolfe, 1992), spatial precueing (see Eriksen, 1990; Posner, 1980; Van der Heijden, 1992), selective report from multi-item displays of brief duration (see Bundesen, 1990; Van der Heijden, 1992), and diverse interference paradigms (see Eriksen & Eriksen, 1974; Tipper, 1985; Van der Heijden, 1992). Neuropsychological research on visual attention is mainly concerned with the effects of parietal and frontal lesions in humans, such as the "neglect" phenomena (see Allport, 1993; Posner & Peterson, 1990). Within the neurosciences, the neurophysiological approach focuses on task-dependent effects of visual attention manipulations at the single-cell level (see Allport, 1993; Posner & Petersen, 1990), at the level of event-related potentials (e.g. Hillyard, Munte, & Neville, 1985), and at the level of metabolic activation of brain areas (e.g. PET: see LaBerge & Buchsbaum, 1990). Finally, the area of computational modelling, especially the neural network approach, gives us an idea of how visual attention processes might be implemented in the neural hardware of a primate brain (see Goebel, 1993; Niebur, Koch, & Rosin, 1993; Olshausen, Anderson, & Van Essen, 1993; Phaf, Van der Heijden & Hudson, 1990).

This abundance of knowledge might induce the expectation that there is strong convergence towards one common theory or model of visual attention. But this is not the case. Instead, the impression of scattering and divergence prevails, at least at the level of individual models. Several factors are responsible. (a) Different domains of visual attention data are covered by different

theories, and even within a domain there are incompatibilities (see Schneider, 1993). For instance, visual search is, on the one hand, the subject of competing theories of Treisman (1988), Wolfe & Cave (1990), and Duncan & Humphrey (1989); on the other hand, spatial precueing effects are the topic of the theories of LaBerge and Brown (1989) and Van der Heijden (1992). The overlap between the two classes of theories is not very large. (b) Current theories of visual attention are formulated at different levels of abstraction. For instance, Bundesen (1990) has offered an elegant mathematical model with a large scope, and Olshausen et al.'s (1993) model specifies in a neurally based way the basic computational operations of visual attention in shape-based recognition. Therefore not all theories or models can be directly compared and tested on an empirical basis. (c) Allport's (1993) excellent recent review on the neuro-cognitive data basis of visual attention has revealed a complicated picture that rules out any simple or parsimonious model.

VAM: An Overview

What would an adequate move in such a situation be? Is it in vain to hope for a unifying set of visual attention mechanisms? Are there separate and independent visual attention functions and mechanisms for different data domains? In this paper I argue the opposite and opt for a unified neuro-cognitive Visual Attention Model (VAM). It is a model about visual attention control of segmentation, object recognition, and space-based motor action. VAM assumes two main functions of visual attention—namely, "selection-for-object-recognition" and "selection-for-space-based-motor-action".[1] The attentional control processes that perform these two functions restructure the results of stimulus driven and local perceptual grouping and segregation processes, the "visual chunks", in such a way that one visual chunk is globally segmented and implemented as an "object token". This attentional segmentation solves the "inter- and intra-object-binding problem". These problems arise because visual chunks are locally segmented at the first low-level cortical stage (V1), but they are initially not globally segmented at higher levels of the visual system—that is, within the "type-level" modules of the "where"- (e.g. PPC) and "what"-pathway modules (e.g. V4/IT). However, object recognition and space-based motor actions require global segmentation of visual information. This means that information about one object has to be distinguished from information about other objects at these higher levels. This state of distinctness corresponds to an "object token" and solves the intra- and inter-object-binding problems. It is argued that the main function of visual attention is to implement such a token that contains globally segmented information about one object. If attention is endogenously controlled, for example by instructing a subject to saccade to a red object, then an attentional signal is applied to the higher type-level representations (e.g. V4/IT) that

[1]These terms follow Allport (1987), who introduced the term "selection-for-action".

correspond to the selection attribute (colour red). This signal propagates back to the retinotopic low-level structure V1, or, more precisely, to the visual chunk that shares the higher-level type feature (colour red) of the selection attribute. V1 acts then as a location-based distributor of the attentional signal and sends it out to all higher type-level modules of the what- and where-pathway. Thus, information at all levels of the visual system that is "tagged" with this attentional signal is segmented globally (distinguished) and implements an object token. If this token contains information about one object, then the recognition process can work successfully and the space-based motor system receives the spatial parameters for action (e.g. the endpoint for a saccade to a red object). But if the token contains information from more than one object (binding problems!) then "where-based" attentional control is required. It selects information of one region (e.g. corresponding to one red object) in V1, and global segmentation is achieved for this information and the corresponding "implicit" chunk. The where-based attentional control can serially scan regions in V1 until the target object is found and recognized.

THE FUNCTIONS OF VISUAL ATTENTION: "SELECTION-FOR-OBJECT-RECOGNITION" AND "SELECTION-FOR-SPACE-BASED-MOTOR-ACTION"

What are the functions of visual attention? This question (see Allport, 1987; Neumann, 1987) is often neglected in theories about visual attention but should be answered first before attempts are made to specify the mechanisms (see, for example, Marr, 1982, for forceful arguments). An analysis of the literature reveals at least two kinds of functions: "selection-for-action" (Allport, 1987; Neumann, 1987) and "selection-for-object-recognition"[2] (e.g. Goebel, 1993; LaBerge & Brown, 1989; Olshausen et al., 1993). To my knowledge, none of the existing theories has attempted to take both functions as a point of departure. The Visual Attention Model, VAM, however, attempts to do so.

The first function—*selection-for-action*—is further specified by Neumann (1987) as the problem of "parameter specification" in motor action control. For instance, grasping a certain object (such as a glass of beer) presupposes that its spatial coordinates are used to compute the movement to the target object. If the visual field contains several suitable objects (such as several glasses of beer), then only the spatial coordinates of the intended object (one's own glass) should be "temporarily coupled" (Allport, 1987) to the motor control structure. More generally, space-based motor actions like grasping or making a saccadic eye movement require a mechanism that supplies the motor system only with the spatial parameters of the intended target object (for example, the end position of

[2]Selection-for-feature-integration" is a third suggestion function (e.g. Treisman & Gelade, 1980; Treisman, 1988); see the final Section for comments on the feature-integration theory.

the grasping trajectory). Therefore "selection-for-space-based-motor-action" is considered a central function of visual attention.[3]

The second function can be called, in a similar vein, *selection-for-object-recognition*. This function is based on the assumption that basic-level object recognition is a computationally costly operation for the primate brain that cannot be applied to all objects of the visual field in parallel (e.g. Goebel, 1993; Hummel & Biederman, 1992; LaBerge & Brown, 1989; Neisser, 1967; Olshausen et al., 1993). Therefore, only one or a few objects at a time are recognized. The required selection process is ascribed to visual attention. This function is controversial; some authors (e.g. Van der Heijden, 1992) have denied any role of visual attention for object recognition because they do not assume a "capacity limitation" (unlike, for example, Broadbent, 1958) for this type of process.This means that the computing system should be able to recognize all objects of the visual field in parallel. Is this a realistic claim for the primate brain? Is "selection-for-object-recognition" merely a pseudo-problem?

Analysing the problems a visual object recognition system has to solve (e.g. Biederman, 1987; Humphreys & Bruce, 1989; Marr, 1982; Ullman, 1989) should give a first answer. Let us consider shape-based basic-level recognition of objects. The retinal projection a single object produces can vary in location, size, and orientation depending on the perspective and the distance between the viewer and the object. This produces very different input patterns at the retinal level. The process comparing the stored representation of an object in memory with the "to-be-matched" input representation has to cope with these variations. The suggested solutions to these invariance problems—either "alignment operations" (e.g. Ullman, 1989) or "recognition-by-components operations" (e.g. Biederman, 1987)—have one common implication. They cannot be applied to all objects of a natural scene (within the acuity limits) in parallel. Instead, these invariance-generating operations can be carried out for only one object (or a few objects) at a time. Furthermore, those models of object recognition mainly concerned with the computational operations of the matching process between sensory-based input pattern and the stored object representation in memory (and less with invariance operations) also presuppose that a unique pattern of the "to-be-recognized" object should be present at the input level of the object recognition system (e.g. Carpenter & Grossberg, 1993; Grossberg, 1980; Mumford, 1992). Otherwise no adequate stored "object template" (no matter whether whole- or feature-based) can be selected by the input pattern, and no successful match between the "back-projected" template (hypothesis) and the input pattern is possible. To sum up, existing theories that attempt to model the process of visual object recognition presuppose a mechanism that selects information from one object (or a few objects) as an input to the recognition system.

[3]Allport (1987) and Neumann (1987) formulate the "selection-for-action" problem as a general problem of "parameter specification" (Neumann, 1987) for action control. I will restrict myself to space-based motor actions like saccades or grasping movements.

In addition to these computational considerations that were central to the visual attention models of LaBerge & Brown (1989), Goebel (1993), and Olshausen et al. (1993), there is a further and still neglected reason to assume "capacity limitations" in object recognition. It concerns the structure and the information flow within the "pre-recognitional" primate visual system, which creates two "binding problems". (a) Based on evidence from animal lesion studies, from human neuropsychology, and from neurophysiological work (see Desimone & Ungerleider, 1989; Goodale & Milner, 1992; Mishkin, Ungerleider, & Macko, 1983), it is commonly assumed that two main information-processing pathways of the visual system can be distinguished. It is the ventral "what"-pathway whose function is mainly object recognition and the dorsal "where"-pathway whose task it is to compute spatial parameters for motor actions. This parallel and distributed way of processing information is even more evident when the brain areas within both pathways are considered. The first cortical processing stage, area V1 (striate cortex or area 17), computes spatially and temporally in parallel "local low-level features" of objects within diverse retinotopic sub-maps or modules,[4] for instance, the local orientation of luminance contrasts, local colour values, or local movements (e.g. DeYoe & Yan Essen, 1988; Livingstone & Hubel, 1988; Zeki, 1992). This type of information in V1 is not sufficient for object recognition—that is, it is not sufficient for a match with the stored object representation. More invariance transformations are needed, which occur in "later" higher-level stages of the what-pathway, such as within the brain areas V4, IT, and STS of the macaque (see Desimone & Ungerleider, 1989; Harries & Perrett, 1991; Rolls, 1992). Only information at these cortical levels is invariant enough with regard to colour or shape information and can therefore be matched against memory representations. Recent single-cell studies on the macaque brain (Fujita, Tanaka, Ito, & Cheng, 1992; Tanaka, 1993) support this conjecture and suggest how information for object recognition is represented. The results show that the IT cortex (more precisely, its anterior part, TE) contains columns that code an "alphabet of visual primitives"—that is, "object features of moderate complexity", for instance, certain shape primitives (see Tanaka, 1993). Other single-cell studies also revealed that the receptive fields of IT neurons are relatively large and cover several degrees of visual angle (e.g. Desimone & Ungerleider, 1989). In other words, information about the retinal location of visual primitives is lost or at least represented with very low spatial precision (e.g. Desimone & Ungerleider, 1989; Felleman & Van Essen, 1991, p. 6.). Figure 1 gives a highly simplified and selective sketch of elements of the primate visual system that are important to VAM's architecture. For simplicity, certain areas, e.g. V2, are left out, and only a few modules are presented (absent modules are symbolized by empty boxes).

[4]The term "module" should be understood in the sense of "weak modularity" (e.g. Kosslyn & Koenig, 1992), which excludes some attributes of the "strong version" (Fodor, 1983).

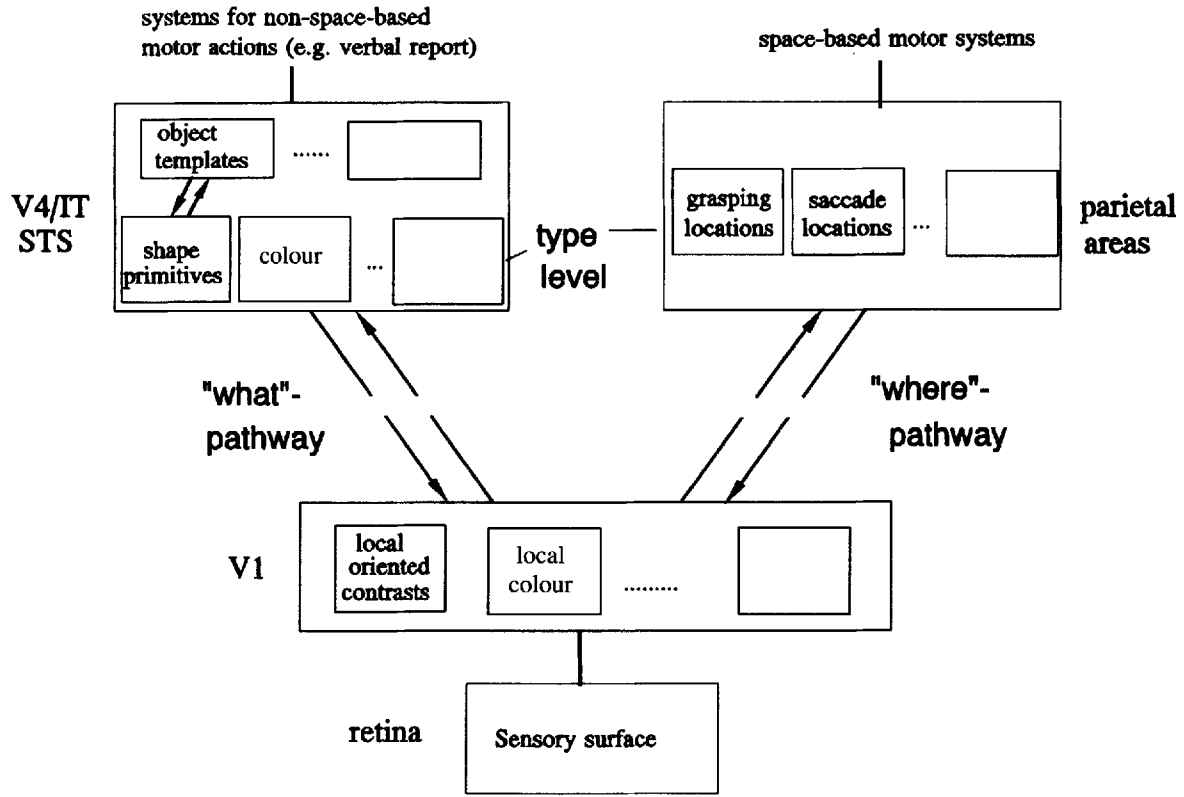

FIG. 1. A simplified and selective sketch of the primate visual system.

These largely location-independent and highly invariant representations of visual primitives in IT are examples of what will be called a "type" representation. Type information has to be distinguished from "token" information, which refers only to a single object (e.g. Kahneman & Treisman, 1984; Kanwisher, 1987; Marr, 1982). That is, *type* information refers to representations of largely invariant higher-level visual features such as to the alphabet of visual primitives in IT, whereas a *token* contains all type representations that relate to one particular object. The object recognition mechanism, therefore, faces the problem that it can only rely on type information as an input, but that type information should correspond exactly to an object token. If only one object is present within the visual field, no problem arises. In a situation with several objects, however, such as in natural viewing, object recognition is not possible without prior selection. At the input level of the recognition system, type level representations of visual primitives (e.g. certain shape primitives) from several objects are active, and the system has to "know" which of these individual type representations—which might be called "type-level features"—belongs to which object. As stated above, implicit location coding in the form of retinotopic structures (as in V1) is not sufficiently precise at these stages of the cortical machinery and can therefore not be used to solve what I call the *inter-object binding problem* (see also Goebel, 1991, 1993; Milner, 1974; von der Malsburg, 1981). The important point is that the object recognition system needs information that refers to an object token. This is a non-trivial problem, as there is no simple way to select all visual type-level feature representations of one object token and to ignore or keep separate information from other tokens. How does the brain give those type-level representations of visual primitives that correspond to one object token access to the object recognition system while representations of other tokens are temporarily excluded or held separate? Somehow, type information from one object has to be distinguished from information from other objects, and this distinction has to be made available for the object recognition system. One could also say that the problem is to "mark", "tag", or "segment" the type information of single objects. The resulting state of such attentionally based tagging or segmentation implements an object token. I will argue that the main function of visual attention is to solve this problem of "selective token implementation".

This problem refers not only to object recognition, but also to the second function of visual attention, "selection-for-space-based-motor-action". Let us consider a classical type of task from experimental psychology, the "filtering task" (e.g. Kahneman & Treisman, 1984). For instance, a subject has to saccade to a red object surrounded by green and blue objects. To fulfil this task, the primate brain has to use the information "colour red"—represented with sufficient constancy at the type level within the ventral what-pathway (e.g. V4)—in order to select the corresponding information about the location of the saccade object—represented within the dorsal where-pathway (e.g. area LIP). Again, a binding or "cross-referencing" problem is created, because colour and location

information reside within modules of different pathways and because a simple direct referencing from the colour type module to the location type module(s) is not possible. Retinotopy is largely lost at the V4/IT type level and cannot be used. Not only representations of different objects have to be kept separate within each module (such as shape primitives)—inter-object-binding-problem—but it has also to be determined which information of one module (set of shape primitives) belongs to which information of the other modules (location representations); this will be called the *intra-object binding problem.* The implementation of an object token solves both binding problems. Such a token, therefore, refers not only to type-level representations of the what-pathway, but also to type-level location representations within parietal areas of the where-pathway, which is, for instance, used for the space-based motor actions such as saccadic eye movements or grasping actions. Therefore, the two functions of visual attention—selection-for-object-recognition and selection-for-space-based-motor-action—are claimed to be solved by one set of mechanisms that implements object tokens.

MECHANISMS OF VISUAL ATTENTION

How are these two attentional functions realized? The mechanisms VAM suggests will be presented in three steps. The first step introduces the basic architecture of attentional control and its processing dynamics. Stimulus-driven and attentional segmentation, attentional selection via V1, what- and where-based attentional control are introduced as VAM's central elements. They are further explicated by describing the information-processing events in a simple experimental task. The second step justifies and further specifies the central elements by referring to data patterns from visual search (similarity effects, local feature contrast, conjunction search), spatial precueing, lateral masking, and the interaction of visual attention and saccades. Step three attempts a further neurophysiological specification of VAM's control dynamics. Temporal modulation of the neuronal firing ("neuronal synchrony") is suggested as a "common coding" scheme for stimulus-driven segmentation and attentional segmentation control. Finally, the role of the pulvinar and the results of recent single-cell studies are discussed.

Basic Architecture and Processing Dynamics

Overview

Recognizing objects and carrying out space-based motor actions requires information about the visual scene that is segmented. The result of these stimulus-driven processes will be called "visual chunks". These chunks are object-based "pieces of information" that are segmented locally at the level of V1 but not globally at the higher type-level. Attentional control signals try to achieve

the required global segmentation for one of these V1 chunks in order to solve the inter- and intra-object binding problems. One form of attentional control is "what-based". It depends on the task instruction (for example, "Name the red object") and originates at the type level of the what-pathway. This control signal propagates to V1 in order to segment the corresponding visual chunk. This attentionally mediated global segmentation is then further transmitted to all higher type-level modules of the what- and where-pathway. The final result of this attentional control signal flow should be the globally segmented information about one object. This information is adequate for recognition and space-based actions and implements an object token. If the what-based attentional control was not successful, then where-based attentional control is required. For instance, if the what-based signal supplied not only one but several chunks, then these chunks are resegmented as a new "supra-chunk". Where-based control relies on regions to solve this problem. It selects in a serial search one region after another in V1. Each region is thus scanned serially, the corresponding chunk is segmented (from the supra-chunk), and it is checked by the recognition process as to whether it matches the target object. If the match is successful, the token implementation process is completed by making available information about the identity of an object and about the parameters for space-based motor action control.

Stimulus-driven Perceptual Grouping and Segregation in the Primate Visual System

When a primate looks at a natural scene that includes many different objects, the retina transforms light energy into neural patterns that after several levels of "pre-processing" (for example, within the LGN) propagate to the first cortical stage, V1. As stated before, different sub-maps (modules) of V1 compute different low-level visual features, such as local oriented contrasts or local colour. The information is further processed within the dorsal where- and the ventral what-pathway. Higher-level areas within these two main pathways compute increasingly invariant object information while losing information about the retinal location. At the highest level of this pre-object-recognition flow, for example, at V4/IT and STS within the what-pathway, diverse type-level representations ("stimulus dimensions") are computed within specialized modules, such as shape primitives, colour constancy, and so forth. Furthermore, there is also feedback flow of information from the "non-retinotopic" higher type-level modules, such as the V4/IT areas, back to lower-level retinotopically organized modules (sub-maps) in V1 (e.g. Damasio, 1989; Felleman & Van Essen, 1991; Finkel & Edelman, 1989; Zeki, 1992). One function of this feedback flow is to perform perceptual grouping and segregation processes that *segment*[5] visual information in an object-based way according to various stim-

[5]I use the term "segmentation" as generic term for perceptual grouping and segregation.

ulus-driven Gestalt principles (e.g. Rock & Palmer, 1990; Wertheimer, 1923). For instance, neuronal representations of local oriented contrasts that belong to the same shape contour are grouped together and segregated from other shape contours (e.g. Grossberg & Mingolla, 1985). The result of these stimulus-driven segmentation processes that represent potential objects are *visual chunks*. In other words, a visual chunk consists of "pre-recognitional" segmented visual object information.

Up to now, the impression might be that the neural activation flow stops at the type-level and does not enter the level where "object templates" and space-based motor patterns are stored. This is not assumed. The point is that information flow indeed enters the object template level and may also be able to "prime" certain templates, but this is not sufficient for the successful recognition of objects. Globally segmented token information has to be available for this purpose.

Attentional Segmentation

How do attentional processes relate to these segmentation processes? I suggest that attentional processes are not fundamentally different from stimulus-driven grouping and segregation processes. Instead, attentional processes convert the *local* segmentation of all visual chunks into the *global* segmentation of one chunk—that is, an object token. In other words, attentional and stimulus-driven segmentation are two successive forms of structuring visual information for further processing, namely object recognition and space-based motor action control.

As stated before, *stimulus-driven* perceptual grouping processes determine which low-level features in V1 belong together, forming a common chunk, whereas segregation processes determine which features belong to different chunks. This segmentation in V1 is local in the sense that it does not imply the segmentation at the type-level (inter- and intra-object binding problems). A second form of parsing visual information is required. It is *attentional* and tries to achieve global segmentation of visual information about one object. "Global" means that visual information is grouped and segregated (from information about other objects) at all levels of the visual system ranging from V1 to the type-level. The final result of this global grouping and segregation is an object token. If the token contains information about one object, then immediate recognition of this object is possible. If, however, the token contains information from several objects and only one of these objects should be recognized, then the attentional control signal flow has to be revised. Too much information is present within the token; it has to be restricted to information from one object—the sub-section on "what- and where-based attentional control" will explicate this point.

How are stimulus-driven and attentional segmentation processes and their results coded? I assume that both processes modulate the neuronal activation of content-specific representations at all levels of the pre-recognitional part of the

visual system. "Grouping" means a modulation of elements of content-specific representations (such as low-level features in V1, type-level features in V41/IT, and so on), making explicit that these elements belong together. "Segregation" is expressed by another aspect of modulation, which makes explicit that elements do not belong together and should be treated as separate. Therefore, perceptual grouping and segregation processes are considered as competitive: Grouping can override segregation, and vice versa. The final result of this competition determines the segmentation. How both processes might be implemented at the neurophysiological description level is discussed in the section "the Neurophysiological Specification of VAM's mechanisms". In short, the modulation of representations—either attentional or stimulus-driven—does not consist of changing the neuronal firing rate but of a temporal modulation of the firing. Grouping modulation means that representations of elements (such as neurons that code low-level features) fire in a highly synchronized manner, for example, by firing in the same time slice, whereas segregation modulates the temporal separation of elements, for example, forcing them to fire in different time slices.

A further common feature of stimulus-driven and attentional perceptual grouping and segregation processes is that both are implemented via top-down modulatory feedback from higher to lower levels.[6] Stimulus-driven segmentation processes make sure that elements (such as low-level features) that are close in space but belong to different chunks are segregated and can be distinguished. Attentional segmentation is more ambitious with the aim that a visual chunk is segregated not only from the local neighbour chunk but from all other chunks as well. How is this global attentional segmentation coded and distinguished from local stimulus-driven segmentation? I suggest that attentional signals strongly increase the strength of the stimulus-driven segmentation signal. This difference in signal strength (modulation of neuronal activation) allows global segmentation to be implemented (see the section on "The Neurophysiological Specification of VAM's mechanisms).

In summary, visual attentional as well as stimulus-driven segmentation processes rely on the same coding scheme—that is, modulation of neural activation. The difference between the two processes is that the goal of the stimulus-driven segmentation process is to achieve local grouping and segregation of several object representations (visual chunks in V1), whereas the goal of attentional segmentation is to achieve global grouping and segregation of one visual chunk (object token). Object recognition and space-based motor control processes require the latter type of segmentation.

[6]To emphasize the role of feedback-based information flow in segmentation does not exclude that intra- and inter-module connections (e.g. in V1) contribute to grouping and segregation processes.

Attentional Selection via V1

How does visual attention achieve this global segmentation of information from one object? The answer is given by referring to a simple experimental task that requires the red object among other coloured objects to be named. First, in order to do such a simple task, the "task instruction"[7] has to have access to the attentional control system. It generates an attentional signal in the corresponding modules of the type-level. Information at this level of the processing hierarchy is sufficiently invariant for reliable selection. For the naming task, this means that the colour representation "red" of the colour module (at the type level) receives the task-instruction-based attentional signal. This signal groups and segregates all neurons of the colour representation "red", segmenting it from representations of other colours. However, this is not sufficient for the naming task, which presupposes the shape-based recognition of the red object. How, then, can the attentionally segmented information "red" also segment the corresponding representation within the shape-primitive module, so that the object recognition mechanism can distinguish and access this shape information (here the shape information of the one red object)? As stated above, a direct reference from the colour module to the module for representing shape primitives—the intra-object-binding problem—is not possible because the modules do not contain a sufficiently precise location-based "binding code" like V1. V4/IT modules of the what-pathway have sacrificed the spatial reference for invariant type-level representations.[8] The colour representation "red" within V4/IT could be derived from any location within its large receptive fields. So how could such an attentionally segmented colour representation of a red target object (globally) segment the representation of shape primitives of the same object and not of other objects with other colours? More generally expressed, how are these inter- and intra-object binding problems solved?

My suggestion is that *back-referencing to the first cortical stage* of the visual processing, V1, is used for this job (see also Damasio, 1989; Zeki, 1992). Why V1? It is organized in retinotopic sub-maps (such as modules with representations of local oriented luminance contrasts, or representations of colours without constancy) that allow for a location-based solution[9] to the binding problems.

[7]To my knowledge, it is currently not known from which neural structures the task instruction signal comes and how it is generated. Possible candidates are frontal lobe areas in combination with the limbic system.

[8]There are also cross-connections at the higher levels (e.g. Felleman & Van Essen, 1991; Goodale & Milner, 1992). The point is that these higher-level cross-connections with relatively large receptive fields do not allow sufficiently precise (location-based) referencing of the object-based representations from one module to the corresponding representations in the other module.

[9]That visual attentional selection is realized in a location-based way is also supported by a large amount of experimental data from psychology (e.g. Eriksen & Hoffman, 1973; Hoffman & Nelson, 1981; Nissen, 1985; Posner, 1980; Tsal & Lavie, 1988; cf., above all, Van der Heijden, 1992, 1993, for a forceful argument on that point).

This solution works as follows: After applying the instruction-based attentional signal to the type-level colour module (which implies the global segmentation of feedforward delivered colour information), the attentional signal propogates back to the V1 sub-map for representing colours. The colour representation of the red object—the locally segmented visual chunk—is then globally segmented by this top-down signal [see (1) in Fig. 2]. Because the retinotopic sub-maps of V1 are connected via common locations,[10] the segmented representation "red" of the colour sub-map propagates its attentional signal to the representations of the other sub-maps of the same location (region). Therefore, also the sub-map that represents oriented luminance contrasts—the V1 basis for computing shape primitives at the type level—receives the location-specific attentional signal.

After this region-based attentional signal distribution within the sub-maps of V1 is completed, the signal propagates to all connected type-level modules [see (2) in Figure 2] and continues its global segmentation task. One could say that V1 acts as a location-based distributor of attentional signals. When the signal arrives at the type-level, it segments all those patterns that belong to the attended and segmented visual chunk in V1. The result of this global grouping and segregation process is an object token. This token solves the inter- and intra-object binding problem and is a prerequisite for object recognition and selective parameter delivery for space-based motor actions.[11] Finally, the attentional selection in the sense of global grouping and segmentation is not a matter of purely "cortical information flow" within the what- and where-pathway but needs the interaction with a further subcortical structure, the pulvinar (e.g. LaBerge & Brown, 1989; Posner & Petersen, 1990) (see the section on "The Neurophysiological Specification of VAM's mechanisms).

"What- and Where-based Attentional Control"

The above-mentioned task of naming a red object is an example of "what-based attentional control". This means that a task-instruction-based attentional signal originates from type-level modules of the what-pathway. What-based attentional control is, however, not always sufficient—for instance, if a task requires all red letters among other non-red letters to be named. According to VAM, the task-instruction-based attentional signal segments the type-level representation of "red" within the colour module, and propagates to V1. In V1, not only one but several stimulus-driven segmented chunks (that contain infor-

[10] Additional to or instead of lateral connections in V1, feedback connections from V2 could fulfil this location-based attentional signal distribution. Feedback connections from a higher- to a lower-level module of one dimension (e.g. colour) seems to contact not only the feedforward neurons of the same module, but also neurons of other modules that code other dimensions (see Zeki, 1992)

[11] "Reflexive" or "exogeneously based" motor action control (e.g. a saccade to abruptly appearing object) does not require such a token of type level representations – see further on.

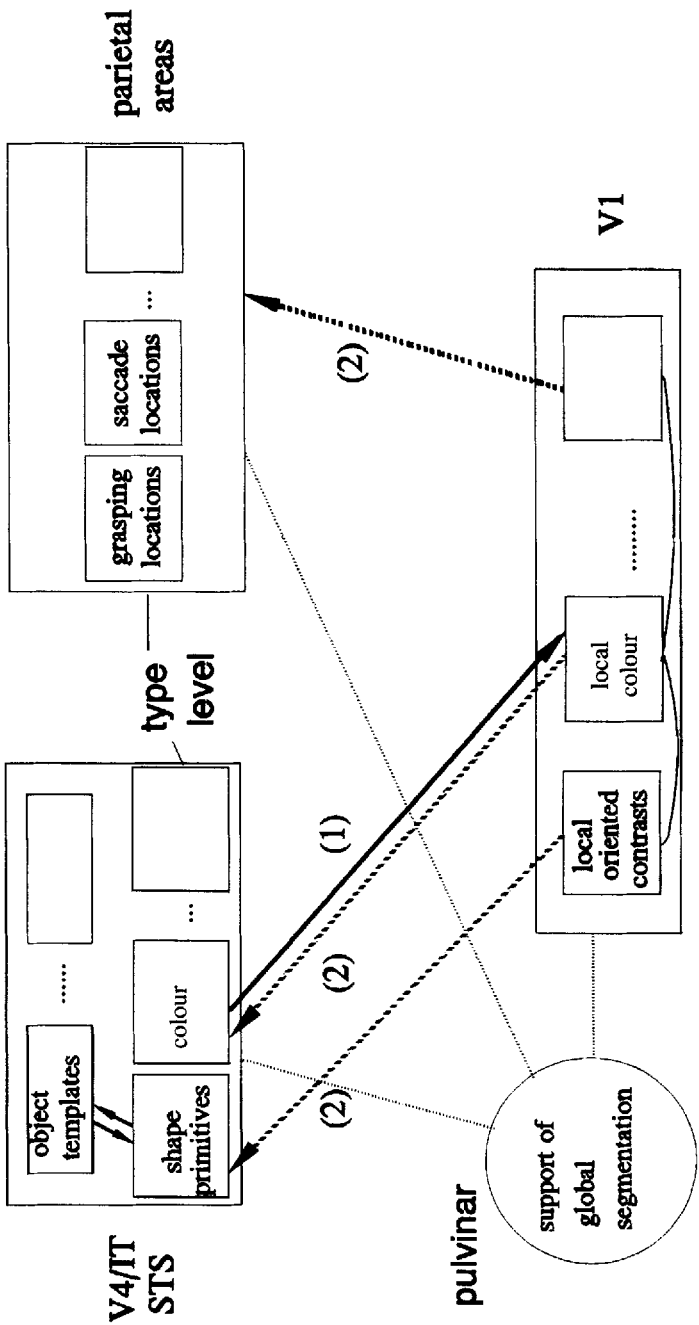

FIG. 2. Attention selection in V1.

mation about red letters) are supplied with the attentional signal. This causes the resegmentation of the "attended" chunks into a new single "supra-chunk", which contains information about all red letters. The feedforward flow of this attentional signal from V1 to the type-level is therefore based on this supra-chunk [see (2) in Fig. 3]. Recognition is still not possible because information from several red letters is simultaneously present at the type-level of the letter recognition module (inter-object binding problem!).

How is this problem solved? For this purpose, a second form of attentional control is assumed: "where-based attentional control". It resegments the supra-chunk (which contains all red letters) – that is, it decomposes and segregates it into one chunk that corresponds to a single object (single red letter) and one chunk that contains information about all the other objects (other red letters). The resegmentation of where-based attentional control is based on regions. Regions are different from chunks, and they probably rely on relatively primitive segmentation principles that are different from the relatively sophisticated what-based segmentation. A region can correspond to a sub-part of a chunk. In our example, several chunks were regrouped by an attentional what-based signal into a new supra-chunk (all red letters). This does not prevent regions within the dorsal pathway from being computed (corresponding to single red letters). The where-based attentional control selects one of these regions and sends an attentional signal to V1 [see (3) in Fig. 3]. The "hidden" chunk of this region is globally resegmented in V1. This chunk then propagates its attentional signal to the type-level modules [see (4) in Fig. 3], and the corresponding object token is implemented. After successful recognition, the next region is selected by the where-based attentional control, and the same resegmentation processes as for the first region (chunk) are initiated. Therefore, several letters can be serially scanned in this way and recognized.

The where-based attentional mechanism needs little "intelligence". The requirements it has to fulfil are the following. (a) It must be able to scan one region after another. (b) It must register the regions and store them for a certain amount of time. Intervening eye movements should not disturb this storage. (c) It must "eliminate" (or at least tag) those regions from the search list that have already been selected by the where-based attentional signal to avoid the same region to be selected again. (d) Where-based attentional control should be able to generate location expectations. In other words, task-dependent selection-by-location should be possible. Knowing where objects could appear is an important information for guiding attention. Especially in multi-object scenes, knowledge about the overall scene could be used to determine potential locations of target objects. If the location of a possible target object is known in advance, then the where-based attentional signal can already be applied to a certain region of V1 before retinal input information has arrived. The form of this advanced where-based signal might be adjustable like a "zoom-lens" (e.g. Eriksen & St. James, 1986).

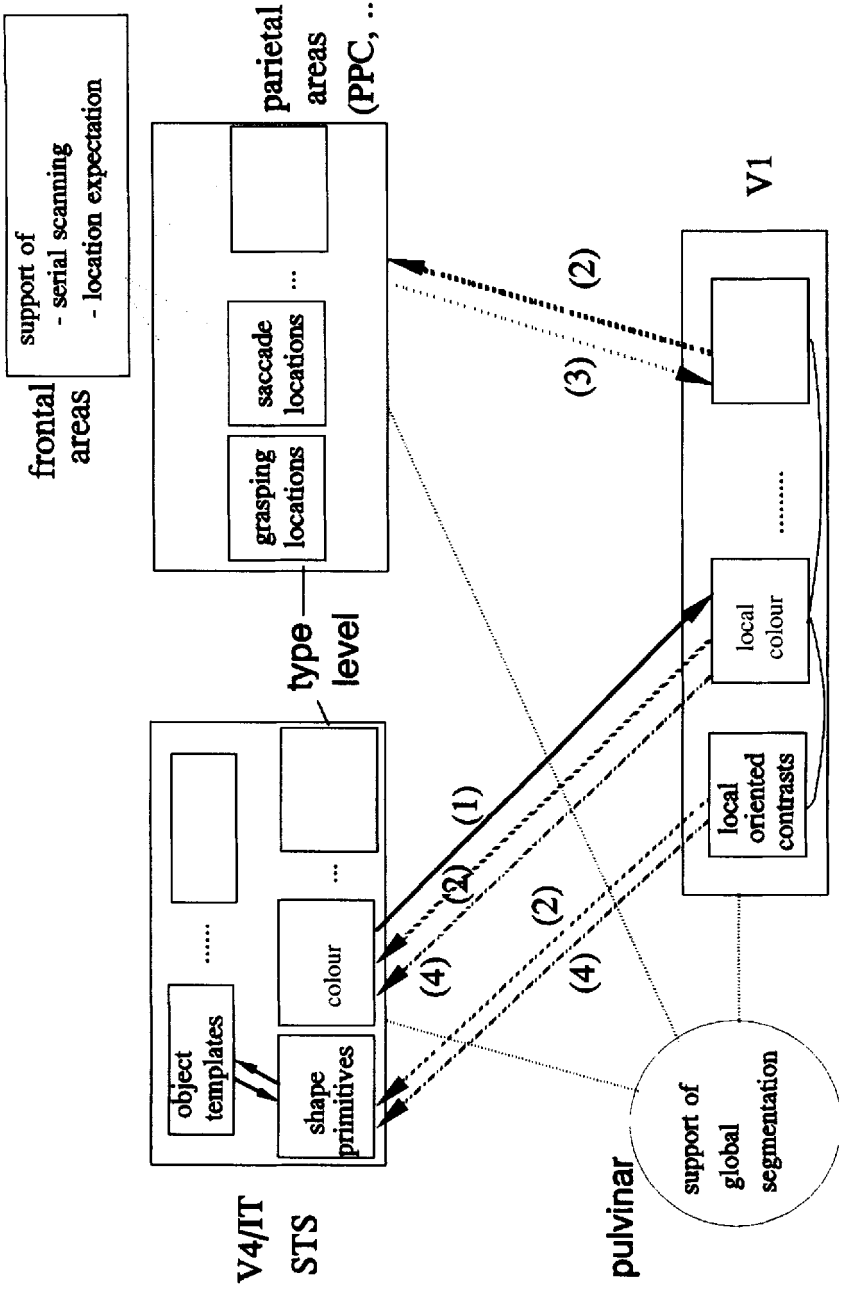

FIG. 3. What- and where-based endogeneous attentional control.

The posterior parietal cortex (PPC) seems to be the cortical site of where-based attentional control. (a) The PPC is part of the dorsal where-pathway. (b) The PPC contains representations of locations that can be considered to be stable across eye movements (e.g. Andersen, Essick, & Siegel, 1985). However, where-based control can presumably not be handled exclusively by the higher-level location modules of the posterior parietal cortex but needs the frontal lobe where-areas that are connected with the PPC (e.g. Wilson, Scalaidhe, & Goldman-Rakic, 1993).

Up to now, attentional control has only been exerted by *endogeneous*, "intentional", or "goal-directed" factors, such as by the instruction to attend to a certain object. Endogeneous control refers to cortical "top-down" what- and where-based attentional signals. There is a second type of control, called the *exogenous*, "reflexive", or "stimulus-driven" attentional control (see Posner, 1980; Yantis, 1993). It relates to situations where objects tend to attract visual attention by certain stimulus attributes (such as abrupt onset or a pop-out stimulus), contrary to the current endogenous control via task demands.

Two different pathways of exogenous control can be distinguished. (a) One object can be especially "salient"—for instance, a bright object among dim objects. The corresponding control signals are probably subcortical in origin—for instance, originating from the superior colliculus (SC).[12] Furthermore, objects with abrupt onset might also transmit their attentional control signal via this subcortical route. (b) Exogeneous control can be based on more complex object representations (than on mere brightness differences or abrupt onsets), such as "pop-out" displays in visual search tasks (e.g. Treisman & Gelade, 1980), for instance when one vertical line is presented among many horizontal lines. V1[13] itself seems to be involved in generating this type of exogenous cortical attentional signal (see Knierim & Van Essen, 1992).

Space-based Motor Actions, Object Recognition, and a "When" System: Making a Goal-directed Saccade

The problem of selective delivery of a spatial parameter (end position) for space-based motor-actions, either saccadic eye movements or grasping actions, is also solved with an implemented object token because tokens contain not only ventral what- but also dorsal where-information. The modules of the dorsal pathway that code locations for different motor actions contain as part of the token a unique representation of the location of the attended object. The location

[12]The superior colliculus (SC) is a reasonable candidate, because it is sensitive to such relatively primitive stimulus differences and because it can be accessed directly and quickly from the retina. Empirical support for a role of the SC in attentional control comes from the micro-lesion study by Desimone et al. (1990).

[13]The attentional effect in V1 could be mediated by V2 (V4?) feedback (see Knierim & Van Essen, 1992, p. 978).

representation for a saccade landing point is probably different from that for a grasping movements (e.g. Rizzolatti, Gentilucci, & Matelli, 1985). "Later" high-level motor-related areas use this location information and send it down to the low-level motor structures.

How visual attention, object recognition, and space-based motor control interact according to VAM is illustrated by means of a further simple example. A subject is instructed to saccade to a red object among objects of other (sufficiently different) colours. At the beginning of the experimental trial, the computer screen is blank. Then three objects are shown: a red circle, a blue square, and a green triangle. Stimulus-driven perceptual grouping and segregation processes segment the scene information into three chunks. Simultaneously, their type representations are computed—for instance, representation of shape primitives, of the three colours, and the three potential saccade locations. The task instruction is converted into an attentional signal that segments the type-level representation of the colour red (V4/IT), which propagates to the V1 chunk of the red object. From V1 the attentional signal goes back to the type-level modules, where the corresponding patterns (of the attended red object) are globally segmented within each module. The object recognition system uses this information and registers a red circle, while the segmented type-level location representation of the red object is propagated further to later motor-related areas. In the case of the saccade system, the high-level motor-related area might be the "frontal eye field" (FEF), which, in turn, sends a signal to the low-level motor structure "brain stem"—either directly or via the superior colliculus (e.g. Goldberg, Eggers, & Gouras, 1991). The result of this signal flow is an overt saccade.

Can selection-for-space-based motor action rely on one selection mechanism only—that is, visual attention? This is certainly not the case. Evidently, not every generated object token should immediately lead to a space-based motor action. In the above-described task it was required to saccade as fast as possible, but obviously primates can decouple the decision to initiate a space-based motor action towards a target object from the corresponding object token implementation process that delivers the unique location information. Therefore, the decision about the overt execution of motor actions is not a matter of visual attention. Metaphorically expressed, what visual attention does instead is to generate candidates for potential actions—that is, to make sure that objects are recognized and that their various spatial parameters are computed. A further mechanism for the *when-aspect* of action control is therefore necessary. Its function is to deliver the go-signal for the overt initiation of motor actions (e.g. Bullock & Grossberg, 1988; Dominey & Arbib, 1992). For instance, in the above-described saccade task, the go-signal might be supplied by FEF (see Dominey & Arbib, 1992).

Further Specifications of VAM's Central Assumptions: Discussing Experimental Data on Visual Search, Spatial Precueing, Lateral Masking, and the Relationship between Object Recognition and Saccades

The following sub-sections apply VAM to well-known paradigms and data patterns from experimental psychology—namely, to visual search (similarity effects, local feature contrasts, and conjunction search), spatial precueing, lateral masking, and the relationship between object recognition and saccades. My intention is to explain these data patterns and to specify further VAM's central assumptions.

What-based Attentional Control: "Similarity Effects" and "Local Feature Contrasts" in Visual Search

The top-down what-based attentional signal flow implies divergence from non-localized type-level representations with large receptive fields to localized V1 representations with small receptive fields. Therefore, all those V1 chunks receive the top-down attentional signal that had sent output to the attentionally segmented type level representations. In our example, all chunks of red objects that had sent activation up to the type level representation (of the colour red) also receive the top-down attentional signal. This implies that a higher-level neuron sends activation back to those lower neurons that are part of its "receptive field". Such a conception of a top-down attentional signal flow that runs parallel to the already used feedforward connections explains the *similarity principles* that Duncan and Humphreys (1989) postulated for predicting visual search data. The authors distinguish two factors that determine the efficiencu of visual search—namely, "target-distractor-similarity" and "distractor-distractor-similarity". The data that motivated this distinction show firstly that the more similar target (T) and distractors (Ds) are, the less efficient the search—the *T–D-similarity* principle. Secondly, with increasing similarity between D elements, the efficiency of the search increases, too—the D–D-similarity principle (see Duncan & Humphreys, 1989). VAM's explanation of high T–D-similarity is that the T type-level representation sends an attentional signal (and also a stimulus-driven grouping signal) not only to the T chunk, but also to the D chunks. T and D chunks have both sent feedforward activation to the T type-level representation due to common type-level "features" that T and Ds share. The attentional signal from the type-level is, in turn, distributed to the T chunk and the D chunks, which are resegmented as a common chunk. Serial where-based attention is then required for segmenting the T only (see the next subsection on "conjunction search").

Let us consider a visual search task for explication. The subject is instructed to search for the T-letter *R* in a display that also contains D-letters, e.g. a *P*, an

X, and a *Q*. If the T-letter *R* is present, a "yes"-button has to be pressed as fast as possible; if it is not present, a "no"-button has to be pressed. The result shows slow, inefficient search—that is, a strong reaction-time increase with the number of Ds (e.g. see Duncan & Humphreys, 1989, Experiment 5; Treisman & Gelade, 1980, Experiment 4). According to VAM, the pre-recognitional type-level representation of the letter *R* (e.g. a certain line configuration)—the "search template"—is supplied with the attentional signal and thus segmented from other representations. Because the type-level representation of an *R* consists of several elements of a distributed representation, such as line configurations, it is reasonable to assume that some elements of the search template of the T-letter *R* overlaps with elements of D-letter representations of *P* and *Q*[14] (see Figure 4 for a highly simplified graphical illustration of this idea without the claim of being neurophysilogically realistic). Considering the feedforward flow, both D letters (*P* and *Q*) have sent activation to type-level elements that are shared with the T type-level search template. The *R* template, therefore, propagates an attentional signal not only to the *R*, but also to the *P*- and *Q*-V1-representations and achieves, therefore, a segmentation of the T-letter *R* and of the D-letters *Q* and *P* into a new common "supra-chunk". Where-based serial scanning is then required to segment the T globally (from the supra-chunk). The letter *X* is not so much affected by the top-down attentional signal, because there is less overlap with the T type-level representation of *R*.

Because T–D-similarity is determined by the featural overlap at the type level, recent data by Wolfe, Friedman-Hill, Stewart, & O'Connell (1992) receive a simple explanation. They found visual search for a single orientation T (e.g. a line tilted to the right) to be efficient if the Ds fall into a different "orientation category" (e.g. tilted left) from the T. If T and Ds are close in orientation and fall into the same orientation *category*, search becomes inefficient. Neuronal representations of orientation at the type-level (e.g. V4) are more broadly tuned than are V1 representations and could fit into the categories Wolfe et al. (1992) determined. Therefore, if T and Ds access non-overlapping representations—that is, different categories—then what-based attentional control can select the T chunk only, and search should be efficient. With an overlap between T and D type-level representations (same category), search should have to rely on serial where-based attentional control and should therefore become inefficient.

How can the second principle for determining visual search efficiency, the *D–D-similarity*, be explained? D–D-similarity refers to stimulus-driven perceptual grouping and segmentation (Duncan & Humphreys, 1989; see, also, Humphreys, Quinlan, & Riddoch, 1989). The V1 representations of highly

[14]If there were a single type-level feature that distinguishes the T from the Ds, such as searching for a *Q* among *O*, then the overlap could be avoided by restricting the search template to the non-overlapping feature. Search asymmetries (e.g. Treisman & Souther, 1985) can be explained in this way.

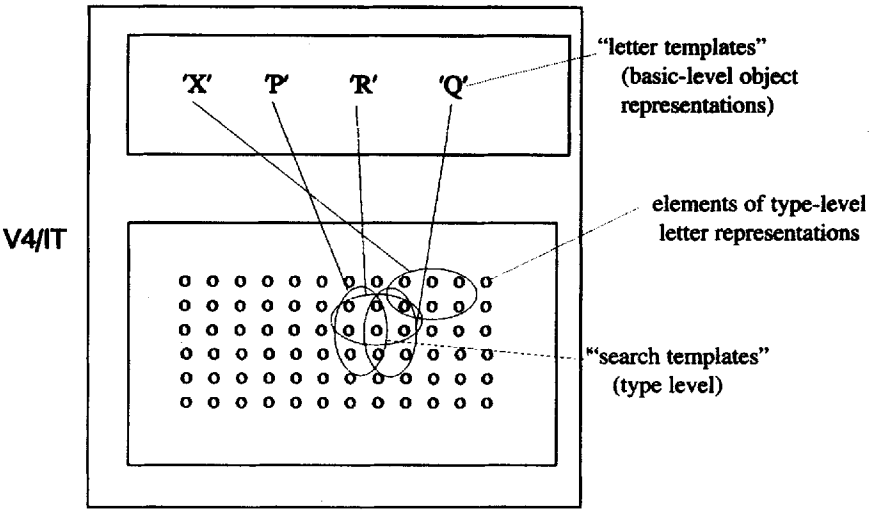

FIG. 4. Overlapping type-level letter representations.

similar Ds (such as only *X*s as D-letters) are grouped and resegmented as one chunk according to the Gestalt principle of similarity (e.g. Wertheimer, 1923; Rock & Palmer, 1990). Therefore, two chunks—a T chunk and a chunk for all Ds—are segmented. Because there are only two chunks, segmentation is global, and the object recognition system can immediately recognize the T chunk. This is also the explanation for the pop-out effect mentioned earlier (e.g. Treisman & Gelade, 1980).

A further interesting finding of the visual search literature that is related to the similarity principles are effects of *local feature contrasts*. Nothdurft (1993; see also Nothdurft, 1985) has shown that efficient search with a high average T–D-similarity is possible when D elements close to the T have sufficient "featural contrast" (low T–D-similarity), even when other D elements that are further away have low featural contrast (high T–D-similarity). For instance, Nothdurft (1993) has shown that an orientation-defined T, a 90° (upright) line, can be detected very efficiently—that is, reaction time was almost independent of the number of Ds—when the immediate D neighbours are 45° lines (dissimilar elements) and the other, more distant Ds consist of lines with an orientation very similar to the T (e.g. 70° or 100°). But if the local featural contrast around the T was changed from low to high T–D-similarity (low local contrast), then visual search was inefficient.

VAM's explanation of these findings relies on stimulus-driven segmentation. In Nothdurft's search there is only a strong featural contrast between the T and close Ds. Other, more distant Ds are always similar to their neighbours and do not differ from each other by strong local contrasts. The segmentation process,

therefore, segregates the T from its close D neighbours and segments it as a single chunk, whereas the other Ds are similar and are therefore segmented as one chunk. Consequently, again, a large background chunk consisting of the Ds and a single T chunk result from the segmentation process.

Where-based Attentional Control Signal: "Conjunction Search", "Spatial Precueing", and Brief Remarks on "Unilateral Neglect"

Another classical data pattern in visual search concerns the "conjunction search" task (Treisman & Gelade, 1980). In a typical conjunction search task, the subject is instructed to search for a T defined by two attributes—for example, by the form *X* and the colour "green". The Ds share one attribute with the T, for example, are either a "red *X*" or a "green *O*". The results usually show that search is not very efficient (but see Wolfe, Cave, & Franzel, 1989), and the reaction time increases linearly with the number of Ds in the display—the so-called "display size effect". The slopes of "Yes"- and "No"-responses for the display size function normally have a ratio of 2:1.

VAM's explanation again relies on an interplay between what- and where-based visual attention. Due to the task instruction, two type-level feature representations of the T—that is, "green" in the colour module and the pre-recognitional letter representation for *X*, are supplied with the attentional signal. The "back propagation" of this attentional signal to V1 reaches all chunks of the T and Ds and initiates a resegmentation. The two features deliver competing segmentation suggestions. The information in V1 could either be segmented according to the colour—that is, green items are segregated from red items—or it could be segmented according to the letter shape—that is, *X* items are segregated from *O* items. Whether one of these two segmentation possibilities is realized depends on the relative segmentation strength of the attentional signals of the two type-level features. Task instruction, strategy or stimulus factors determine which of both alternatives is selected. A third possibility is that neither feature determines the segmentation and all T and D chunks are only locally segregated.[15] For each of the three alternatives where-based attentional control is required, which scans one region (chunk) after another. The where-based top-down signal is stronger than the what-based signals and is able to override competing segmentation tendencies. In each case of an "attended region", the object recognition checks whether the search template of the T matches the segmented input. If the match is successful, the process of shifting the where-

[15]The T chunk receives a top-down attentional signal from both features and the D chunks receive only a signal from one feature. The observation that the T chunk does not "pop-out" ad is not immediately found suggests that the two attentional signals from different sources (different features) do not simply add in V1. If "neuronal synchrony" is the code for the strength of the attentional signal—see the section on "The Neurophysiological Specification of VAM's Mechanisms"—and competitive segmentation is the attentional control principal then this non-additivity is plausible.

based signal stops. Otherwise, "serial scanning" continues until all locations of objects are "attended". This leads to the usual display size effect and a 2:1 slope. However, if one feature is much stronger in its segmentation tendency than the other feature—either due to task-instruction or stimulus-driven factors—then this feature determines the segmentation. For instance, if the T-colour "green" is dominant in its segmentation tendency then all green items are segmented as one chunk and only its regions (green items) are serially scanned.

As mentioned in the previous section, attention can also be controlled by location-based expectations. Experimental evidence for this form of attentional control was collected within the framework of the *spatial precueing* paradigm (e.g. Posner, 1980; Posner & Cohen, 1984; see also Eriksen & Hoffman, 1973; Jonides, 1981). In such a paradigm, subjects usually work through a reaction-time task—for instance pressing a button as fast as possible in response to a target stimulus. The spatial precueing manipulation consists either in not informing the subjects about the stimulus location (neutral precueing condition), informing them correctly (valid precueing), or informing them incorrectly (invalid precueing). The basic and often-replicated result of this paradigm is that valid precueing leads to the best performance (such as fastest reaction times), neutral precues to medium values, and invalid precues to the worst performance (such as slowest reaction times).

In a spatial precueing task, the attentional where-based signal is allocated in advance to the region of V1 where the T will probably appear. When the sensory input arrives at V1, global segmentation can immediately begin and continue until it reaches the type level in order to implement an object token. "Benefits" in reaction time in the valid precueing condition (compared to the neutral condition) are due to the time that is saved by the advance allocation compared to stimulus-induced allocation. "Costs" of invalid precueing reflect the reallocation time. If only one T object appears, benefits and costs of precueing are small because the abrupt onset of the object itself attracts the attentional signal in a purely stimulus-driven manner (see also Van der Heijden, 1992). Furthermore, precueing the location of an object by "peripheral" cues (e.g. Jonides, 1981; Posner, 1980) is also done via this stimulus-driven exogeneous attentional control pathway.

This subsection should be closed by briefly discussing neuropsychological evidence on where-based attentional control. As stated before, the PPC (in connection with frontal lobe where-areas) is probably the site for the control of where-based attention. What happens if the PPC is damaged? Patients with right parietal lesions who suffer from so-called *unilateral neglect* deliver some information on this question. In short, they tend to ignore objects that are presented within the visual field contralateral to the lesion. The neglect is often considered to be an attentional disturbance (e.g. Allport, 1993; Humphreys & Bruce, 1989) and, more specifically, a disturbance of the ability to "shift" or "disengage" attention (e.g. Posner, Walker, Friedrich, & Rafal, 1984; Riddoch & Humphreys,

1987). This interpretation fits nicely to the role ascribed to PPC by VAM—namely, where-based attentional control. Damage to the PPC should therefore affect all those tasks for which what-based attentional control is not sufficient and where-based control is needed. For instance, conjunction search tasks that require serial search by the where-based system should be disrupted, in contrast to pop-out tasks that can be handled by the what-based system—see, for example, Riddoch and Humphreys (1987) for supporting data. Furthermore, if there is no (or less) PPC feedback to certain parts of the visual field representation in V1—whether attentional or stimulus-driven in source—then the corresponding visual chunks without such PPC feedback have a disadvantage in that they get less top-down modulation as compared to chunks with PPC feedback.

Attentional Segmentation and Task-relevance: Categorial Effects in "Lateral Masking"

An experimental finding, termed *lateral masking* (e.g. Bouma, 1970; Wolford & Hollingsworth, 1974), gives further information about the interaction of stimulus-driven and attentional segmentation. The task requires identification of a briefly presented T-letter that is surrounded by one or more D-letters. An independent variable is, for example, the distance between T- and D-letters. The dependent variable is the percentage of correctly reported letters. The basic finding is that with decreasing distance between T and D, the accuracy of reporting the T-letter decreases as well. This is not unexpected if the Gestalt principles of distance and similarity (e.g. Rock & Palmer, 1990; Wertheimer, 1923) are recalled. The closer to each other similar elements are, the stronger the grouping effect between them, and the more difficult their segregation. Therefore, if letters that share many type-level features are relatively close and only one of them has to be reported (T-letter), then the system has to "break up" the grouping between target and distractors in order to segregate them.

There are a number of further data patterns on lateral masking, but one of the most interesting and in our context relevant findings comes from Styles and Allport (1986) and can be termed the *categorial effect* in lateral masking. In one of their experiments, the task consisted of the brief presentation of a linear five-item string followed by a pattern mask. Subjects had to report the red target letter that could appear at any one of the five positions and that was surrounded either by other black letters or by black digits. Two main results were observed. (a) Report accuracy is higher for both outer-end letters as compared to the three inner positions. In other words, outer T-letters experience less masking from D-letters than do inner T-letters. (b) If a T-letter is surrounded by digit Ds, then inner and outer letter Ts are reported with the same high degree of accuracy. The disadvantage of the inner items—the masking effect—vanishes for the letter-digit-combination. The label "categorial effect" owes to the fact that the lateral masking effect of Ds on a T seems to be restricted to D members of the

same category (letters) as the T. The masking effect does not show up—or is, at least, strongly reduced—for D members (digits) that do not belong to the T category.

The first finding, that outer T letters allow for better performance than do inner letters, can be explained by differences in the segmentation difficulties. Outer items have just one close neighbour, and their segregation against the counteracting grouping influences is therefore easier and faster compared to inner items that have two neighbours. However, why does a T-letter surrounded by digits fail to show this advantage for inner positions? When a T-letter is embedded among D-letters, top-down what-based segmentation from the letter module is not possible. The T-letter is not known in advance; thus feedback from the letter modules spreads out to all V1 chunks and attempts to resegment them as a common chunk. Although attentional feedback from the colour module is available, it has to struggle with strong feedback from the letter module. A where-based attentional signal is required for a fast segmentation. The situation is different in the presentation of one T-letter among digits. Attentional feedback from the type-level letter module goes only to the one visual chunk (T) that is the same that receives the colour feedback. Feedback from the digit module is not task-relevant, thus much weaker in strength, and is restricted to Ds.[16]

There are general lessons to be learned from this categorial effect in lateral masking. (a) The ease of attentional segmentation depends on the *task-relevance* of the to-be-segmented visual information. Two sources of task-relevant information can be distinguished for such "filtering tasks" (e.g. Kahneman & Treisman, 1984). The first source, the "criterion attribute" information, determines the relevant object; the second source, the "response attribute" information, determines the kind of action (e.g. Bundesen, 1990; Van der Heijden, 1992; Van der Heijden, LaHeij, Phaf, Buijs, & Van Vliet, 1988). In the lateral masking task of Styles and Allport (1986), letter identity is the response attribute—it has to be reported—whereas the colour "red" is the selection attribute—it determines the relevant T-object. Both attribute representations (letters and colours) are supplied at the type-level with the attentional signal. If there is only one chunk at V1 that receives attentional feedback from both attribute representations—a red letter among black digits—attentional top-down segmentation can be efficiently realized, and the masking effect is strongly reduced. With several chunks that receive attentional feedback—a red letter among black letters—however, the top-down what-based segmentation is difficult and has to be supplemented by where-based attentional control. (b) Stimulus-driven perceptual grouping effects between T and Ds can be overcome if the Ds are not task-relevant and do not share type-level representations with the T. For the lateral masking paradigm, the good performance in the letter-digit-condition is made

[16]This explanation implies two separate modules for letters and digits. This means that letter representations share elements with each other but not with digit representations.

possible by the top-down attentional segmentation from response and criterion attribute representations. It supplies only one chunk and overcomes the grouping effect of spatially close items (Gestalt principle of proximity).

Object Recognition and Space-based Motor Actions: An Obligatory Attentional Coupling

According to VAM's basic architecture, object recognition and space-based motor control (such as saccading, reaching, and so on) are strictly coupled and both depend on the allocation of visual attention. Only information about one object at a time can be "attended"—that is, globally segmented and converted into an object token. Object recognition and parameter delivery for space-based motor actions presupposes this object token. For instance, a subject should not be able to make an immediate visually guided saccade to a target within a multi-object scene and—during the same time slice of *saccade programming*[17]—to *identify* another *object* in the scene. During the allocation of attention to an object, only this object allows recognition and parameter specification. Relevant psychological data were collected at the beginning of the 1980s (e.g. Klein, 1980; Remington, 1980). Some of these early studies seem to cast doubt on this claim of a strict coupling (e.g. Klein, 1980), but they had serious flaws (see Rizzolati, Riggio, Dascola & Umiltá, 1987; Shepherd, Findlay, & Hockey, 1986).

This issue is not quite decided yet, but data that support the claim of a strict coupling have been collected very recently by a colleague and myself (see Deubel & Schneider, 1994; Schneider & Deubel, 1995). Our paradigm consists of a saccade task combined with a letter-discrimination task—the object-recognition measure. Our subjects had to saccade to positions within a horizontal string (forward masks) to the left or right of a fixation cross. The saccade target (ST) location was designated by a peripheral cue or a central cue. After ST onset and well before the saccade initiation, the forward masks were replaced for a limited time by a discrimination target (DT)—an *E* or a mirror *E*—and surrounding distractors (*S* or mirror-*S*). The discrimination task required the subject to decide whether an *E* or a mirror-*E* was present during the trial. ST and DT were always presented in the same hemifield, but their locations were independently varied at Positions 2, 3, and 4 of the letter string. Dependent measures were discrimination performance, saccadic latencies, and saccadic landing positions. What were the results? Let us consider the case in which ST was shown at Position 2. DT performance was good if DT was also at Position 2 but was drastically reduced when DT was located at the other two possible positions, 3 or 4. Here, performance was close to chance level (50%). The same pattern held for the other ST positions. The data clearly show that successful discrimination prior to a saccade is restricted to the location of the intended ST. In other words, "programming" a saccade to a location implies that during this process—a

certain time slice before the saccade—object-recognition abilities are "focused" at the intended saccadic landing location.

A Neurophysiological Specification of VAM's Mechanisms: Stimulus-driven and Attentional Segmentation by Synchronized Neuronal Activation, Pulvinar Functions, and Single-Cell Data on Visual Attention Effects

Most of the central assumptions of VAM were also motivated by neurophysiological and neuroanatomical data that concern the structuring of the primate visual system—for instance, the "what"- and "where"-pathway distinction, and the multi-level "hierarchy" of processing stages. However, questions were left open about the coding of stimulus-driven perceptual grouping and segregation signals, of the corresponding attentional signals, and of the implementation of an object token. The following neurophysiological specification of VAM attempts to answer these questions. Yet if the specification would turn out to be wrong, the central assumptions might nevertheless be valid.

"Common Coding" of Stimulus-driven and Attentional Segmentation by Neuronal Synchrony and the Role of the Pulvinar

Those readers familiar with neurophysiology and/or neural network modelling may already have answered these questions, probably by assuming that differences in firing rates are the "common coding"[18] scheme for stimulus-driven and attentional segmentation. My suggestion is a different one. The coding scheme is not the firing rate, but a modulation of the temporal fine structure of neuronal activation. *Temporal modulation* offers a parameter in addition to firing rate for neuronal coding, namely the "synchrony" of the firing of neuron populations (e.g. Milner, 1974; von der Malsburg, 1991).

Applied to stimulus-driven perceptual grouping and segregation processes, this means that information of a visual chunk is coded by highly *synchronized firing* of all those neurons that represent the features of the chunk (within and across feature modules of V1). Segregation, on the other hand, refers to locally desynchronized firing of neurons from different chunks (e.g. Goebel, 1993; von der Malsburg & Buhman, 1992).Synchronized firing means that the neurons generate their spikes (action potentials) almost simultaneously, providing the additional advantage of reliable signal transmission through the cortical stages (e.g. Abeles, 1991). Desynchronized firing could be implemented by making neurons of different chunks fire at different time slices. For instance, the neurons of one chunk fire simultaneously at time slice t1 while the neurons of the close neighbour will fire at a later time t2 (e.g. Goebel, 1991). The temporal modula-

[18]This "common coding" notion should not be confused with the "common coding" theory of perception and action (Prinz, 1990).

tion scheme offers an advantage if two chunks are located spatially close—for example, in the case of partly overlapping objects. The chunks are then segregated by firing in different time slices without necessarily having difficult firing rates. Nevertheless, in natural scenes with many objects this local temporal segregation scheme does not prevent more distant chunks from firing in accidental synchrony with each other. Only a limited number of objects can be temporally segregated (e.g. Crick & Koch, 1990).

The idea of grouping and segregation—that is, "tagging" neural patterns by modulating the temporal fine structure—was first formulated by Milner (1974) in his object recognition model and by von der Malsburg (1981) in his general attack on purely firing-rate-coded information processing models. Both authors have stated that the temporal structure of the neural firing might be used to solve the object-related binding problems. Again, this means that all those neurons that code information about one object should fire in a highly synchronized and coherent manner, whereas the firing of those neurons that relate to information from other objects should not be synchronized. The acceptance of this "temporal tagging" idea was greatly improved after the publication of neurophysiological data by Singer and colleagues (Engel, König, Kreiter, Schillen, & Singer, 1992; Gray & Singer, 1989) and Eckhorn and colleagues (e.g. Eckhorn et al., 1988). They have shown, for instance, that neurons in the cat's primary visual cortex fire in a highly synchronized and oscillating manner within the 40-Hz range when they code the movement of one stimulus across the visual field. Neurons that code opposite movements of two stimuli are much less correlated in their firing (e.g. Gray, König, Engel, & Singer, 1989). Other laboratories could not replicate these findings with primates (Young, Tanaka, & Yamane, 1992) or with static stimuli (Tovee & Rolls, 1992) (but see Kreiter & Singer, 1992). However, this failure concerned the absence of a periodic oscillating structure of neural activation, and not synchrony in the sense of simultaneous spiking of neurons, nor aperiodic repetitive firing. Therefore, as already stressed by Singer and colleagues, temporal coding by synchrony has to be distinguished from oscillations: "Cells can synchronize their responses without engaging in regular oscillatory discharges, and, conversely, responses may be oscillatory without being synchronized" (Engel, König, & Singer, 1992, p. 388). For my neurophysiological interpretation, it is mainly synchrony that is significant.[19]

How is the attentional control signal coded? One possibility is the degree of synchrony. If the task-instruction-dependent attentional signal is applied to the type-level representations, then the synchrony in the firing of these representations should be enhanced and global segmentation should be achieved. This enhanced synchrony is then propagated via the feedback connections to V1,

[19]In a system like VAM, where feedforward and feedback signal flow occurs simultaneously, repetitive but not necessarily strictly periodic firing is probably required for achieving sufficient coordination between the signals flows.

where the neuronal representations of the corresponding chunk(s) increase the synchrony of their firing. If the degree of synchrony is sufficiently different from other chunk representations, then segmentation also changes from local to global. This means that not only the local neighbouring chunks are desynchronized in relation to the attended chunk, but also all other chunks of V1. If desynchronization is implemented via firing in different time slices, then there should be one time slice where only the neurons of the attended chunk fire and other neurons are relatively silent. The global temporal segmentation of the V1 patterns is in the next step transported to the higher-level modules [see (2) in Figure 2]. The result is an object token containing a globally segmented pattern of object information at several levels of the visual system, from V1 to the type-level areas. Recognition processes and parameter delivery for space-based motor actions are then possible. If this neurophysiological interpretation is correct, then attention-dependent changes in neuronal synchrony of firing should be found in V1 and in higher-level areas. Data of such experiments that measure attentional effects in terms of synchrony are currently not available.

The subcortical thalamic nucleus pulvinar is a reasonable candidate for carrying out or at least supporting attentional control—that is, for achieving global segmentation. Neurobiological experimental evidence that suggests a role of the pulvinar for visual attention has indeed been collected—for instance, in lesion studies by Petersen, Robinson & Morrris (1987) and Desimone et al. (1990) and a PET study by LaBerge and Buchsbaum (1990). Moreover, the pulvinar has the required neuronal connections. Two of its parts—that is, the lateral (PL) and inferior pulvinar (PI: see Robinson & Petersen, 1992)—are connected to V1 and to the higher type-level areas, such as IT and the PPC (see Baleydier & Morel, 1992; Robinson et al., 1992).

Single-Cell Studies on Visual Attention

What can the neurophysiological evidence tell us about visual attention effects at the single-cell level (e.g. Bushnell, Goldberg, & Robinson, 1981; Chelazzi, Miller, Duncan, & Desimone, 1993; Moran & Desimone, 1985; Motter, 1993)? The emerging picture of these studies looks complicated (see also Allport, 1993). (a) Some studies on spatial visual attention obtained partly contradictory results (Moran & Desimone, 1985; Motter, 1993). (b) Attentional effects on firing rates seem to occur relatively late in processing (Chelazzi et al., 1993). With regard to the first point, the contradiction is that the study by Moran and Desimone (1985) did not reveal spatial attention (where-based attention) dependent firing rate changes in V1 and V2 (but changes in V4), whereas Motter (1993) found effects in V1, V2, and V4. In the Moran and Desimone (1985) single-cell study, an alert monkey had to release a bar in response to a target stimulus appearing at a certain location and to ignore another simultaneously presented distractor stimulus. The authors report that neurons in V4 and IT

representing the distractor greatly reduced their firing rate if the target stimulus was within their receptive fields. The target itself maintained its firing rate. Furthermore, V1 neurons showed no "attentional effects"—neither enhancement nor inhibition. In contrast to Moran et al. (1985), the study by Motter (1993) revealed firing rate changes in V1 and V2. Alert macaque monkeys had to make a bar orientation discrimination task (two-choice) where a spatial marker indicated the T among Ds that were placed in non-overlapping receptive fields. The results showed enhancement effects for most "attended" neurons in V1 and V2 (compared to "non-attended" neurons). Concerning V4 neurons with an attentional effect, roughly 50% showed an enhancement effect, whereas the other 50% showed a reduction of firing rate as compared to non-attended ones. How can this contradiction be resolved? As suggested by Chelazzi (personal communication), there is one feature of Motter's experiment that might induce some caveats. Spatial attention was directed to its location by a small dot. The pure sensory input of this dot, which was located inside the receptive field for the attended condition and outside the field for the non-attended condition, might have caused the firing rate changes in V1 and V2. Directing attention by a central arrow would be a way to avoid this problem.

However, the single-cell study by Chelazzi et al (1993) revealed an illuminating finding: Firing-rate changes of "non-attended" IT neurns occur relatively late in the processing sequence. The visual task required from an alert monkey to saccade to a target (such as a square) among distractors (such as triangles). Firing rates of neurons in IT were recorded which corresponded either to a target or a distractor stimulus. The data show that the neural activation of the T and Ds increased at the same rate after stimulus presentation, until 90–120 msec before saccade onset D-neurons began to reduce their firing rate while T-neurons increased further. Interestingly, when more D-stimuli were added to the search display and the saccade latency was increased, the firing-rate change was still locked to saccade onset (90–120 msec in advance). My interpretation of this action-locked firing change is that it does not reflect visual attentional selection (token implementation) but "response selection". If the system decides to act on a T-object (object token) and to select the "corresponding response", then a reduction of the firing rate of D-Tokens is required. Temporal coding allows the "simultaneous" implementation of up to 4 tokens (e.g. Crick & Koch, 1990), that is, T- and D-Tokens.

EXTENDING VAM'S EXPLANATORY RANGE AND MAKING PREDICTIONS: A NEW LOOK AT "ERIKSEN"-INTERFERENCE AND SPACE-BASED MOTOR ACTION CONTROL

The primary goal of this section is to show that VAM's main assumptions can be used for explaining experimental tasks beyond those mentioned up to now and that new predictions can be made.

"Eriksen-interference": Task-Relevance, Perceptual Grouping, and Attentional Segmentation

The "Eriksen"-interference paradigm (Eriksen & Eriksen, 1974) usually consists of a two-choice reaction-time task where subjects are required to react as fast as possible to the middle target item within a string of several distractor items. The decisive experimental manipulation consists of varying the identity of the distractor letters. They can be either compatible (for example, identical to the target letter), neutral (a non-response-related letter), or incompatible (letter that requires the opposite reaction). The basic finding is that despite the advance knowledge of the target position, incompatible distractor items slow down the reaction time and produce a higher error rate than do neutral or compatible items. This performance decrement can be called the "Eriksen-interference".

What causes the Eriksen interference? In contrast to the standard explanation, which assumes solely "response competition", VAM suggests a segmentation problem as the cause for at least part of the reaction time delay.[20] My argument is analogous to my treatment of the categorial effect in lateral masking. Top-down attentional segmentation from the letter module is possible for the neutral distractor condition (ND) but not for the incompatible distractor condition (ID). For ND, only the type-level representations of the T-letter send an attentional signal to the T-letter representation in V1, and what-based segmentation can be used to determine the T response.[21] However, for ID, all activated letter representations (response attribute) send attentional signals to V1, reaching T and D chunks. Therefore, a where-based attentional signal is required to segment the T chunk and to verify its identity as T.

Because top-down what-based attentional control depends on T–D similarity, it is predicted that the reaction-time difference between ID and ND, the Eriksen-interference, should vary with T–D-similarity.[22]

Furthermore, Kramer and Jacobson (1991) and Baylis and Driver (1992) investigated the *effects of perceptual grouping* on the Eriksen-interference. Their results show that several manipulations that emphasize the grouping between T and Ds (e.g. common colour, etc.) strongly enhance the Eriksen-interference measured as reaction time and error difference between incompatible and

[20] I do not claim that later "response selection processes"—see Pashler (1991) for data that show the independence of visual attention and response selection—do not contribute to the Eriksen-interference (see e.g. Gratton, Coles, Sirevaag, Eriksen, & Donchin, 1988)

[21] As T is usually defined by its "relative" position (in the middle position of the string, such information can only be computed after T and Ds have been attended. Therefore, even in the ND case, where what-based attention would be sufficient, where-based attention might be added to segment the Ds. This would guarantee that the response is really T-based.

[22] Eriksen and Eriksen (1974) varied the similarity between letter features. The results are inclusive (e.g. depending on the T-D-distance). The problem is that type-level letter features have to be known. One way of solving this problem is to determine T-D-featural similarity by a visual search task and then to do an Eriksen experiment with the same stimuli.

compatible conditions. This finding is especially interesting because it casts doubt on the usual explanation of interference that relies on "response competition" alone. As the word suggests, response competition is usually located at a "late" level—the "response level". But how could "early" grouping factors influence "late" response competition? According to VAM, the fact that perceptual grouping manipulations influence the performance in the ID condition is caused by the same factors that increase the lateral masking effect when the distance between letters is decreased.[23] In both cases, segmentation of the T-item is more difficult and takes a longer time for strongly grouped T–D-configurations than for non-grouped configurations. However, VAM is able to make a further and less expected prediction. The perceptual grouping manipulations of Baylis et al. (1992) and Kramer et al. (1991) should produce, for the ND condition, only a minor increase (or no increase at all) for the reaction time and the error values. In this case—analogous to the digit condition of the categorial effect in lateral masking—direct what-based top-down attentional segmentation within the letter domain (or other type feature domains) is possible, provided there is sufficient T–D-dissimilarity.

Visual Attention, Object Recognition, and Space-based Motor Actions

The second domain will show that VAM is also able to produce new research questions and predictions about the control of space-based motor actions. (a) A central assumption is that the selection of location parameters (end points of trajectories) for different space-based motor actions (such as saccades, grasping) is coupled to one common object. This predicts that subjects should not be able to program a goal-directed grasping movement to one object and to program, in the same time slice, another type of motor action, such as a saccade, to a different object. In other words, the tight coupling that has been shown for saccade programming and *object recognition* (Deubel & Schneider, 1994; Schneider & Deubel, 1995) should be found for *saccade* and *grasping programming*, too. (b) The *spatial precision* of space-based motor actions, such as saccades, should depend on *speed requirements*—that is, in VAM's terms, it should depend on the time of the go-signal. An early go-signal—a fast saccade to an abruptly appearing object—should direct the saccade according to the early available segmentation, that is, the exogenously controlled attentional segmentation (mediated by the abrupt onset pathway, possibly via SC). This segmentation is relatively "primitive" in nature and might be based on a "rough" luminance-

[23]This allows for a further prediction for the Eriksen-interference that concerns the distance between T and Ds. The usual finding is that the larger the distance, the smaller the interference (e.g. Eriksen & Eriksen, 1974). But this effect could be due to retinal acuity disadvantages of more distant Ds. Even in the case of controlled acuity effects, a distance effect due to varying grouping strengths should be found.

based location representation. A later go-signal—a slow saccade to the same object—should direct the saccade according to an endogeneously controlled attentional signal. This signal reflects a more sophisticated cortical segmentation (for example, Gestalt principles) of the target object (visual chunk). For fast saccades, Findlay (1982) and Deubel, Wolf, and Hauske (1984) have shown that the saccade to a T-object in combination with a spatially close D lands at the "centre-of-gravity" of the T and the D. This centre-of-gravity should reflect the exogenously based attentional segmentation. Coeffe (1987) has shown for the same experimental situation that slower saccades avoid the centre-of-gravity and go mainly to the T. The endogeneous object-based attentional signal overrides the exogeneous signal, and the location for the saccade system is therefore more precise and T-dependent.

This idea of an attentional segmentation process that changes in time and gets more precise can be combined with VAM's emphasis on the relationship between segmentation and *task-relevance*. For instance, imagine a task[24] in which a *saccade* should be made to a conjunction T (such as a green circle) next to one D. If the D does not share an attribute with the T (such as a blue square) then segmentation is easy. If, however, the D shares one attribute (for example, with a green square), then segmentation cannot be purely what-based, because the T and the D chunk both receive a task-dependent attentional signal (see the section on conjunction search). Therefore, the where-based signal is also needed. In this case, the final attentional segmentation of the T takes more time than it does in the case of a D with no task-relevance. Therefore, saccades with the same latency should land closer to the D in the case of one shared attribute (task-relevance) than in the case of no shared attribute (no task relevance). Segmentation has proceeded further in the latter case.

RELATIONSHIPS TO OTHER THEORIES OF VISUAL ATTENTION, CHALLENGING ISSUES, AND OPEN QUESTIONS

This final section discusses VAM's relation to other recent major theories and models of visual attention—that is to the work by Treisman (1988; Treisman & Gelade, 1980), LaBerge and Brown (1989), Kosslyn, Flynn, Amsterdam, & Wang (1990), Olshausen et al. (1993), Goebel (1993), Wolfe and Cave (1990), Van der Heijden (1992), Bundesen (1990), Duncan and Humphreys (1989), and others. The content of these theories will not be described—it would require more than a further article—but the theories will be selectively and comparatively evaluated. Furthermore, VAM's specific contributions will be stated. The paper will close with challenging issues and open questions with which VAM has to cope—namely, selective report under conditions of brief and masked

[24]This experimental idea was developed in cooperation with Heiner Deubel.

stimulus presentation, the question of exogenous vs. endogenous control, the large and very specific data base on visual search and on perceptual grouping, neuropsychological phenomena, and, finally, "response selection effects"

Relationship to Other Theories of Visual Attention

This section is restricted to a discussion of recent major work that is mainly concerned with offering specific models of visual attention. Important "older" but nevertheless essential theories (e.g. Broadbent, 1958; Norman, 1968; Posner, 1978; Schneider & Shiffrin, 1977) that deal with attentional processes of the whole sensory-cognitive system at a more general level (and are not restricted to vision) are, therefore, not analysed.[25] The discussion is started by describing *VAM's specific contribution* to the theoretical literature on visual attention. (a) VAM deals in a single model with the two main functions of visual attention, selection-for-object-recognition, and selection-for-space-based-motor-action, and it offers one set of mechanisms for both functions. (b) The solution to the inter- and intra-object-bindings is unique: A feedback flow of endogeneous attentional control signals from higher type-level modules to the location-reference area V1 and back to the type level modules again. (c) VAM explicitly distinguishes between what- and where-based attentional control; also, their interactions are specified. (d) A common mechanistic basis of stimulus-driven and attentional segmentation is postulated. It is claimed that stimulus-driven grouping and segregation achieve local segmentation, whereas attentional control is concerned with global segmentation. What is VAM's relationship to other theories?

Treisman (1988; Treisman & Gelade, 1980) published a very influential and stimulating theory of visual attention—the "feature-integration theory"—that generated a large number of research questions and experimental results. Furthermore, it has integrated a number of important concepts—for instance, the binding problem, the location-based nature of visual attention, serial scanning of locations—into a new theoretical framework. The current version of feature integration (e.g. Treisman, 1988) is in many respects different from the original theory by Treisman and Gelade (1980). One major difference from VAM is that the binding problem was positioned at a different level of visual architecture—that is, V1, where location-coding is available for solving it. Furthermore, the question of how attention is controlled is not answered (see, also, Van der Heijden, 1992, p. 251), and the way of explaining visual search is different, except for the scanning idea for conjunction search (e.g. Treisman, 1988).

LaBerge and Brown (1989) presented a theory of visual attention that is concerned with shape-based object recognition. It deserves credit for a number of features that distinguish it from former theories or models. For instance, it is

[25]However, VAM is in general more in line with the classical two-stage theories of "early" visual attention (e.g. Neisser, 1967), than with "late" selection theories (e.g. Deutsch and Deutsch, 1963).

one of the first models that specifies attention in terms of computationally defined structures and operations, and it discusses the relationship of these structures to neurophysiologically based knowledge about the primate visual system. Furthermore, it specifies explicitly the location-based nature of visual selection. Main differences from VAM—in addition to specific elements mentioned earlier—are a more restricted range of data that can be explained (for example, no visual search data), an analysis of the object recognition problem without reference to the binding problems, and an emphasis on attentional exclusion of information instead of attentional modulation.

Kosslyn et al, (1990) introduced a theory that brought data from neurophysiology, neuroanatomy, and neuropsychology to the focus of theories on visual attention. Furthermore, they introduced neurocomputational concepts such as the what- and where-pathway distinction, a visual buffer (V1), and an attentional window, which are now the basis for many attentional models. One main difference from VAM is that Kosslyn et al.'s model deals with hardly data from experimental psychology, such as visual search, but only with neuropsychological phenomena (neglect, and so on). Furthermore, as for LaBerge and Brown (1989), the binding problems are not analysed and not seen as central for attentional control.

Olshausen et al. (1993) recently presented a neurobiologically plausible neural network model of how visual attention solves the object recognition problems of position and scale invariance. It makes explicit—in contrast to former models—how this solution (in short, a spatial relationship preserving attentional window—see also Goebel, 1993) works by simulating attentional operations with neuron-like elements. Major differences from VAM are the different attentional control concept (for example, a direct gating of the information flow by dynamically modifying receptive fields), a lack of specifying the binding problems and of explaining data from experimental psychology (for example, visual search).

Goebel (1993) introduced the first visual attention model that explicitly specifies the relationship between segmentation, object recognition, and visual attention. Furthermore, the binding problems and its oscillatory solution, as well as the solution to the invariance problems in object recognition (see also Olshausen et al., 1993), are described clearly and simulated successfully with a neural network model. These solutions differ from VAM mainly in the specific elements described above, especially in the treatment of how segmentation and attention are related. Location-based attention precedes complex segmentation and object-based attention in Goebel's model.

Wolfe (1992); Wolfe & Cave, 1990) presented a "guided search model" of visual attention. The basic idea is that the serial visual attention stage can be guided by the parallel feature-computation stage. This delivers very detailed explanations for a large amount of data and allows precise predictions. Its control concept is in some respects very different from VAM, which does not

assume a guidance of visual attention by a priority list of locations. VAM's where-based control mechanism is relatively simple and works in a way that is different from the what-based control.

Van der Heijden (1992) published an elaborate attentional model that explains a large amount of data patterns from experimental psychology. It explicitly specifies the task-dependent control processes and pathways for location-based attentional selection. Furthermore, it incorporates the idea of "post-categorial filtering" (Van der Heijden, 1981)—that is, the idea of a feedback-based attentional control signal from higher-level modules back to lower-level modules (see also Phaf, Van der Heijden, & Hudson, 1990). The main differences from VAM are that object recognition is not considered as a "capacity-limited" operation that requires attentional selection, and that what-based attentional control can only be exerted via accessing where-based control.

Bundesen (1990) introduced a mathematically specified "theory of visual attention (TVA)" that is large in scope. It covers data from very different domains (such as selective report in brief-duration multielement displays, spatial precueing, visual search, and so on), and it is able to generate quantitative explanations. VAM is difficult to compare to TVA because it uses a more abstract level of description. For instance, the "parallel capacity-limited processing" assumption of TVA might be implemented at the mechanistic level by the combination of what- and where-based attentional processes of VAM.

Duncan and Humphreys' (1989) theory of visual attention is concerned with visual search. It has introduced the T–D- and D–D-similarity principles described before that have large explanatory power and are an essential element of VAM. The main differences from VAM are a lack of a where-based attentional control mechanism and no dynamic attentional descriptions at the mechanistic level. Furthermore, it is a "late-selection" theory that assumes parallel capacity-unlimited computations of non-visual object properties—that is, it ascribes visual attention no role for object recognition (like Van der Heijden, 1992).

Before an overall evaluation of the relationship of VAM to other theories is given, two more models that treat specific aspects of visual attention and whose ideas are to some extent incorporated in VAM need to be discussed. (a) *Crick and Koch (1990)*, based in part on the earlier work by Crick (1984) and Koch and Ullman (1985), have suggested that visual attention operates at the neurophysiological level by changing the temporal structure of firing. Attention is conceptualized as a "spotlight" that can be shifted in V1 and boosts the synchronized oscillating neuronal activation. This idea has been further specified by Niebur et al (1993) in the form of a neuronal network model that simulates the single-cell data patterns found by Moran and Desimone (1985). (b) *Rizzolati et al. (1987)* have presented what they call a "premotor theory of attention". The basic idea is to reduce attentional processes to the programming of eye movements. The only differences between overt and covert orienting is that in the latter case the go-signal for initiating an eye movement is not given. VAM

claims that a reduction of visual attention to saccade programming is inadequate—attention also has a central function for object-recognition. However, the tight link between visual attention and motor programming is also postulated but more specified than in the case of the "premotor theory".

There are several main advantages of VAM compared to other theories. (a) VAM's new architecture and processing dynamics allow seemingly incompatible theoretical ideas to be integrated—for example, Duncan and Humphreys' similarity principles with Treisman's serial scanning idea. Additionally, Van der Heijden's postcategorial filtering principle is incorporated. (b) VAM addresses a range of experimental data patterns (such as visual search, spatial precueing, Eriksen-interference, object recognition and saccade control, segmentation phenomena, and so on) that is larger than for most theories (an exception is Bundesen, 1990). However, there is a trade-off between the specificity of models and their explanatory range. Therefore, it is not surprising that models that are more specific and precise than VAM (Goebel, 1993; Olshausen et al., 1993) try to explain a smaller group of data sets. (c) VAM allows an empirical validation not only at the behavioural level but also at the neurophysiological level. These results may, in turn, help to specify and modify the architecture and processing dynamics further, which, in turn, might allow further behavioural testing.

Challenging Issues and Open Questions

1. A large data domain that was not covered here and that delivers important insights about attentional processes comes from *selective report tasks for multiple-item displays* of very brief duration. Results by Sperling (1963, 1967), Shibuya and Bundesen (1988), and Shibuya (1993) show that when stimulus information is available for a very short time interval (such as 150 msec in the case of backward masks) several (up to 3–4) items can be selectively reported (e.g. Ts based on colour or alphanumeric identity). This fact puts a heavy time constraint on any attentional mechanism and probably requires the assumption of a visual short-term or working memory. Furthermore, even semantic content seems to be able influence the attentional selection performance in such tasks (Allport, 1977). These data may show a way towards the discovery of a further visual attention mechanism (see Bundesen, 1990).

2. The large and very elaborate data base on *visual search* (e.g. Cheal & Lyon, 1992; Treisman & Gormican, 1988) is a serious challenge for any attentional theory. VAM has only covered the well-known effects and has not dealt with the full richness and complexities of search results. Conjunction search tasks where one feature is more salient (for example, due to task instruction) are ideally suited to test VAM's mechanisms (see the section on "Further Specifications of VAM's Central Assumptions"). Furthermore, a number of interesting data on perceptual grouping and attentional selection are omitted in this treatment of VAM (e.g. Prinzmetal, 1981; Kahneman & Henik, 1981) and deserve further analysis.

3. VAM does not specify explicitly the relationship between *exogenous vs. endogenous* attentional control. Again, there are many (often puzzling) results from experimental psychology available for specifying this relationship (e.g. Folk, Remington, & Johnston, 1992; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Theeuwes, 1991; Yantis, 1993; Yantis & Jonides, 1984).

4. The *neuropsychological literature* on attention-related effects (e.g. Allport, 1993; Farah, 1989; Humphreys & Bruce, 1989; Kosslyn & Koenig, 1992) offers many unexpected data patterns whose understanding will surely refine any theory or model of visual attention.

5. In the previous section, firing rate reductions at the single-cell level (e.g. Chelazzi et al., 1993) were explained as a consequence of "response selection processes". Extending this idea might allow a fresh look at the so-called "psychological refractory period" (e.g. Bertelson, 1966; Pashler, 1993; Welford, 1959) and other possible "after-effects of response selection"—for instance, "negative priming" (Allport, Tipper, & Chmiel, 1985; Tipper, 1985).

In taking all these *challenges* seriously, we might end up admitting that "attentional functions are of very many different kinds, serving a great range of different computational purposes. There can be no simple theory of attention, any more than there can be a simple theory of thought" (Allport, 1993, p. 206). Whether VAM will contribute something to this enterprise of an adequate theory is an empirical question.

REFERENCES

Abeles, M. (1991). *Corticonics: Neural Circuits of the cerebral cortex.* Cambridge: Cambridge University Press.

Allport, D.A. (1977). On knowing the meaning of words we are unable to report: The effects of visual masking. In S. Dornic (Ed.), *Attention and performance VI* (pp. 505–533). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Allport, D.A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In H. Heuer & A.F. Sanders (Eds.), *Perspectives on perception and action* (pp. 395–419). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Allport, A. (1993). Attention and control. Have we been asking the wrong questions? A critical review of twenty-five years. In D.E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV. Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 183–218). Cambridge, MA: MIT Press.

Allport, D.A., Tipper, S.P., & Chmiel, N.R. (1985). Perceptual integration and postcategorical filtering. In M.I. Posner & O.S. Marin (Eds.), *Attention and performance XI* (pp. 107-132). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Anderson, R.A., Essick, G.K., & Siegel, R.M. (1985). Encoding of spatial location by posterior parietal neurons. *Science, 230,* 456–458.

Baleydier, C., & Morel, A. (1992). Segregated thalamocortical pathways to inferior parietal and inferotemporal cortex in macaque monkey. *Visual Neuroscience, 8,* 391–405.

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

Bertelson, P. (1966). Central intermittency twenty years later. *Quarterly Journal of Experimental Psychology, 18,* 153–163.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychology Review, 94,* 115–147.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226,* 177–178.

Broadbent, D.E. (1958). *Perception and communication.* New York: Pergamon Press.

Bullock, D., & Grossberg, S. (1988). Neural dynamics of planned arm movements: Emergent invariants and speed-accuracy properties during trajectory formation. *Psychological Review, 95,* 49–90.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97,* 523–547.

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.

Carpenter, G.A., & Grossberg, S. (1993). Normal and amnesic learning, recognition and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences, 16,* 131–137.

Cheal, M., & Lyon, D.R. (1992). Attention in visual search: Multiple search classes. *Perception & Psychophysics, 52,* 113–138.

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

Coeffe, C. (1987). Two ways of improving saccade accuracy. In J.K. O'Regan & A. Levy-Schoen (Eds.), *Eye movements: From physiology to cognition* (pp. 105–113). Amsterdam: Elsevier.

Crick, F. (1984). Function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of Sciences, 81,* 4586–4590.

Crick, F., & Koch, C. (1990). Some reflections on visual awareness. *Cold Spring Harbor Symposia on Quantitative Biology, 55,* 953–962.

Damasio, A.R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation, 1,* 123–132.

Desimone, R., & Ungerleider, L.G. (1989). Neural mechanisms of visual processing in monkeys. In F. Boller, & J. Grafman (Eds.), *Handbook of neuropsychology, Vol. 2* (pp. 267–299). New York: Elsevier Science.

Desimone, R., Wessinger, M., Thomas, L., & Schneider, W. (1990). Attentional control of visual perception: Cortical and subcortical mechanisms. *Cold Spring Harbor Symposia on Quantitative Biology, 55,* 963–971.

Deubel, H., & Schneider, W.X. (1994). *The coupling of visual attention, object recognition and saccade target selection.* MPI-paper 4/1995, München: Max-Plank-Institut für psychologische Forschung.

Deubel, H., Wolf, W., & Hauske G. (1984). The evaluation of the oculomotor error signal. In A.G. Gale & F. Johnson (Eds.), *Theoretical and applied aspects of eye movement research* (pp. 55–61). Amsterdam: Elsevier.

Deutsch, J.A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70,* 80–90.

DeYoe, E.A., & Van Essen, D.C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences, 11,* 219–226.

Dominey, P.F., & Arbib, A. (1992). A cortico-subcortical model for generation of spatially accurate sequential saccades. *Cerebral Cortex, 2,* 155–172.

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

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., & Reitboeck, H.J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybernetics, 60,* 121–130.

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

Engel, A.K., König, P., & Singer, W. (1992). Reply to "The functional nature of neuronal oscillations". *Trends in Neurosciences, 15,* 387–388.

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

Eriksen, C.W. (1990). Attentional search of the visual field. In D. Brogan (Ed.), *Visual search.* London: Taylor & Francis.

Eriksen, C.W., & Hoffman, J.E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics, 1,* 155–160.

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.

Farah, M.J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision.* Cambridge, MA: MIT Press.

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

Findlay, J.M. (1982). Global processing for saccadic eye movements. *Vision Research, 22,* 1033–1045.

Finkel, L.H., & Edelmann, G.M. (1989). Integration of distributed cortical systems by reentry: A computer simulation of interactive functionally segregated visual areas. *The Journal of Neuroscience, 9,* 3188–3208.

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Folk, Ch.L., Remington, R.W., & Johnston, J.C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1030–1044.

Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature, 360,* 343–346.

Goebel, R. (1991). Binding, Episodic Short-Term Memory, and Selective Attention, Or Why are PDP Models Poor at Symbol Manipulation? In Touretzy, D.S., Elman, J.L., Sejnowski, T.J., & Hinton, J.E. (Eds.), *Connectionists models: Proceedings of the 1990 summer school* (pp. 253–264). San Mateo, CA: Morgan Kaufman.

Goebel, R. (1993). Perceiving complex visual scenes: An oscillator neural network model that integrates selective attention, perceptual organisation, and invariant recognition. In C.L. Giles, S.J. Hanson, & D.J. Cowan (Eds.), *Advances in neural information processing systems.* San Mateo, CA: Morgan Kaufman.

Goldberg, M.E., Eggers, H.M., & Gouras, P. (1991). The ocular motor system. In E.R. Kandel, J.H. Schwartz, & Th. M. Jessell (Eds.), *Principles of neural science* (pp. 660–676). New York: Elsevier Science.

Goodale, M.A., & Milner, A.D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15,* 20–25.

Gratton, G., Coles, M.G., Sirevaag, E.J., Eriksen, C.W., & Donchin, E. (1988). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology, Human Perception and Performance, 14,* 331–344.

Gray, C.M., König, P., Engel, A.K., & Singer, W. (1989). Oscillatory responses in the cat's visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature, 338,* 334–337.

Gray, C.M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences, 86,* 1698–1702.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review, 87,* 1–51.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries, an emergent segmentations. *Perception & Psychophysics, 38,* 141–171.

Harries, M.H., & Perrett, D.I. (1991). Visual processing of faces in temporal cortex: Physiological evidence for a modular organization and possible anatomical correlates. *Journal of Cognitive Neuroscience, 3,* 9–24.

Hillyard, S.A., Munte, T.F., & Neville, H.J. (1985). Visual-spatial attention, orienting, and brain physiology. In M.I. Posner & O.S.M. Marin (Eds.), *Attention and performance XI* (pp. 63–84). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hoffman, J.E., & Nelson, B. (1981). Spatial selectivity in visual search. *Perception & Psychophysics, 30,* 283–290.

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

Humphreys, G.W., & Bruce, V. (1989). *Visual cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Humphreys, G.W., Quinlan, P.T., & Riddoch, M.J. (1989). Grouping processes in visual search: Effects with single- and combined-feature targets. *Journal of Experimental Psychology: General, 118,* 258–279.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. Long and A. Baddeley (Eds.), *Attention and performance IX* (pp. 187–302). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Kahneman, D., & Treisman, A.M. (1984). Changing views of attention and automaticity. In R. Parasuraman & D.R. Davies (Eds.), *Varieties of attention* (pp. 28–61). New York: Academic Press.

Kanwisher, N.G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition, 27,* 117–143.

Keele, S.W. (1990). Motor programs: Concepts and issues. In M. Jeannerod (Ed.), *Attention and performance XIII: Motor representation and control* (pp. 77–110). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Klein, R. (1980). Does oculomotor readiness mediate cognitive control of visual attention? In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 259–276). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4,* 219–227.

Kosslyn, S., Flynn, R., Amsterdam, R., & Wang, G. (1990). Components of high-level vision: A cognitive neuroscience analysis and accounts of neurological syndromes. *Cognition, 34,* 203–277.

Kosslyn, S.M., & Koenig, O. (1992). *Wet mind: The new cognitive neuroscience.* New York, NY: Free Press.

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.

Kreiter, A.K., & Singer, W. (1992). Short communication: Oscillatory neuronal responses in the visual cortex of the awake macaque monkey. *European Journal of Neuroscience, 4,* 369–375.

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

LaBerge, D., & Buchsbaum, M.S. (1990). Positron emission tomography measurements of pulvinar activity during an attention task. *Journal of Neuroscience, 10,* 613–619.

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

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

Milner, P. (1974). A model for visual shape recognition. *Psychological Review, 81,* 521–535.

Mishkin, M., Ungerleider, L.G., & Macko, K.A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6,* 414–417.

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.

Müller, H.J., & Rabbitt, P.M. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 315–330.

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological Cybernetics, 66,* 241–251.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research, 29,* 1631–1647.

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

Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer & A.F. Sanders (Eds.), *Perspectives on perception and action* (pp. 361–394). Hillsdale,

NJ: Lawrence Erlbaum Associates, Inc.

Niebur, E., Koch, C., & Rosin, C. (1993). An oscillation-based model for the neuronal basis of attention. *Vision Research, 33,* 2789–2802.

Nissen, M.J. (1985). Accessing features and objects: Is location special? In M.I. Posner & O.S.M. Marin (Eds.), *Attention and performance XI* (pp. 205–219). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Norman, D.A. (1968). Toward a theory of memory and attention. *Psychological Review, 75,* 522–536.

Nothdurft, H.C. (1985). Sensitivity for structure gradient in texture discrimination tasks. *Vision Research, 25,* 1957–1968.

Nothdurft, H.C. (1993). Saliency effects across dimensions in visual search. *Vision Research, 33,* 839–844.

Olshausen, B., Anderson, Ch., & Van Essen, D. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience, 13,* 4700–4719.

Pashler, H. (1993). Dual-task interference and elementary mental mechanisms. In D.E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV* (pp. 245–264). Cambridge, MA: MIT Press.

Petersen, S.E., Robinson, D.L., & Morris, J.D. (1987). Contributions of the pulvinar to visual spatial attention. *Neuropsychologia, 25,* 97–105.

Phaf, R.H., Van der Heijden, A.H., & Hudson, P.T. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22,* 273–341.

Posner, M.I. (1978). *Chronometric explorations of mind.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Posner, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32,* 3–25.

Posner, M.I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and performance X* (pp. 531–556). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Posner, M.I., & Petersen, S.E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13,* 25–42.

Posner, M.I., Walker J.A., Friedrich, F.J., & Rafal, R.D. (1984). Effects of parietal injury on covert orienting of attention. *Journal of Neuroscience, 4,* 1863–1874.

Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann & W. Prinz (Eds.), *Relationships between perception and action* (pp. 167–201). Berlin: Springer-Verlag.

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

Remington, R.W. (1980). Attention and saccadic eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 726–744.

Riddoch, J.M., & Humphreys, G.W. (1987). Perceptual and action systems in unilateral visual neglect. In M. Jeannerod (Ed.), *Neurophysiological and neuropsychological aspects of spatial neglect* (pp. 151–181). Amsterdam: Elsevier Science.

Rizzolatti, G., Gentilucci, M., & Matelli, M. (1985). Selective spatial attention: One center, one circuit, or many circuits? In M.I. Posner & O.S.M. Marin (Eds.), *Attention and performance XI* (pp. 251–265). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia, 25,* 31–40.

Robinson, D. L., & Petersen, St.E. (1992). The pulvinar and visual salience. *Trends in Neurosciences, 15,* 127–132.

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

Rolls, E.T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philosophical Transactions of the Royal Society of London, B335,* 11–21.

Rosenbaum, D.A. (1991). *Human motor control.* San Diego, CA: Academic Press.

Schneider, W.X. (1993). Space-based visual attention models and object selection: Constraints, problems, and possible solutions. *Psychological Research, 56,* 35–43.

Schneider, W.X., & Deubel, H. (in press). Visual attention and saccadic eye movements: Evidence for obligatory and selective spatial coupling. In J.M. Findlay, R.W. Kentridge, & R. Walker (Eds.), *Eye movement research: Mechanisms, processes and applications.* North-Holland: Elsevier.

Schneider, W., & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,* 1–66.

Shepherd, M., Findlay, J.M., & Hockey, R.J. (1986). The relationship between eye movements and spatial attention. *Quarterly Journal of Experimental Psychology, 38A,* 475–491.

Shibuya, H. (1993). Efficiency of visual selection in duplex and conjunction conditions in partial report. *Perception & Psychophysics, 54,* 716–732.

Shibuya, H., & Bundesen, C. (1988). Visual selection from multielement displays: Measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 591–600.

Sperling, G. (1963). A model for visual memory tasks. *Human Factors, 5,* 19–31.

Sperling, G. (1967). Successive approximations to a model for short term memory. *Acta Psychologica, 27,* 285–292.

Styles, E.A., & Allport, D.A. (1986). Perceptual integration of identity, location and colour. *Psychological Research, 48,* 189–200.

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

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics, 49,* 83–90.

Tipper, S.P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *The Quarterly Journal of Experimental Psychology, 37A,* 571–590.

Tovee, M.J., & Rolls, E.T. (1992). Oscillatory activity is not evident in the primate temporal visual cortex with static stimuli. *NeuroReport, 3,* 369–372.

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

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

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

Treisman, A., & Souther, J. (1985). Search asymmetries: A diagnostic for preattentive

processing of separable features. *Journal of Experimental Psychology: General, 114,* 285–310.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition, 32,* 193–254.

Van der Heijden, A.H.C. (1981). *Short-term visual information forgetting.* London: Routledge & Kegan.

Van der Heijden, A.H.C. (1992). *Selective attention in vision.* London: Routledge & Kegan.

Van der Heijden, A.H.C. (1993). The role of position in object selection in vision. *Psychological Research, 56,* 44–58.

Van der Heijden, A. H., Heij, W.L., Phaf, R.H., Buijs, D.A., Van Vliet, E.C. (1988). Response competition and condition competition in visual selective attention. *Acta Psychologica, 67,* 259–277.

von der Malsburg, Ch. (1981). *The correlation theory of brain function.* Internal Report. Göttingen, Germany: Max-Planck-Institute for Biophysical Chemistry, 1–38.

von der Malsburg, Ch., & Buhmann, J. (1992). Sensory segmentation with coupled neural oscillators. *Biolgical Cybernetics, 67,* 233–242.

Welford, A.T. (1959). Evidence of a single-channel decision mechanism limiting performance in a serial reaction task. *The Quarterly Journal of Experimental Psychology, 11,* 193–210.

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

Wilson, F.A.W., 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* (4), 124–128.

Wolfe, J.M., & Cave, K.R. (1990). Deploying visual attention: The guided search model. In A. Blake & T. Troscianko (Eds.), *AI and the eye* (pp. 79–103). New York: John Wiley.

Wolfe, J.M., Cave, K.R., & Franzel, S.L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 419–433.

Wolfe, J.M., Friedman-Hill, St.R., Stewart, M.I., & O'Connell, K.M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 34–49.

Wolford, G., & Hollingsworth, S. (1974). Lateral masking in visual information processing. *Perception & Psychophysics, 16,* 315–320.

Yantis, S. (1993). Stimulus-driven attentional capture. *Current Directions in Psychological Science, 2,* 156–161.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Psychophysics, 10,* 601–620.

Young, M.P., Tanaka, K., & Yamane, S. (1992). On oscillating neuronal responses in the visual cortex of the monkey. *Journal of Neurophysiology, 67,* 1464–1474.

Zeki, S. (1992). The visual image in mind and brain. *Scientific American, 267,* 42–50.

Manuscript received 5 January 1994
Revised manuscript received 21 June 1994