Abstract
The proposal of the present work is that the brain selects stimuli for further processing based on their saliency, defined as the strength of their neural representation. A saliency-based model of attentional selection is proposed consisting of two hierarchically arranged processing stages. At the first stage basic features of stimuli are represented in distinct feature maps, which are subject to top-down modulation from working memory. Feature maps feed into the saliency map, which represents the second stage of processing, and create distinct saliency differences in this map. An important feature of the model is that saliency is affected by physical characteristics of stimuli such as brightness or colour as well as by top-down influences such as behavioural goals, and that at the level of the saliency map these distinct influences are indistinguishable. The model proposes that attention selects the most salient stimulus for further processing.

Reference

DOI : 10.13097/archive-ouverte/unige:34620
Saliency, Selection, and Spatial Neglect

Radek Ptak

Division of Neurorehabilitation
Department of Clinical Neurosciences
Geneva University Hospitals

Habilitation thesis for the qualification as a Privatdocent
Medical school, University of Geneva
August 2008
# CONTENTS

**ABSTRACT**  4

1 – THE PURPOSE OF ATTENTION  5
The criterion problem  6
Preview  8

2 – A SALIENCY-BASED MODEL OF ATTENTIONAL SELECTION  9
The saliency map  9
Feature enhancement  11
Location enhancement  15
Interactions within and between saliency maps  17
Summary  18

3 – NEURAL MECHANISMS OF SALIENCY REPRESENTATION  19
Attentional modulation of neuronal responses to visual features  19
The posterior parietal cortex and spatial attention  24
Top-down modulation by spatial attention and action goals  28
A posterior parietal saliency map  30
Summary  33

4 – COGNITIVE MECHANISMS OF ATTENTIONAL SELECTION  35
Visual search and stimulus saliency  35
Attentional set and the capture of covert attention  40
Attention and oculomotor performance  49
Summary  54

5 – SALIENCY, SELECTION, AND SPATIAL NEGLECT  55
Spatial neglect: clinical picture and anatomy  55
Spatial extinction  60
Pre-attentive and attentive processing in neglect and extinction  61
Surface-based and location-based enhancement of saliency  69
A bias favouring foveated stimuli  74
Attentional capture and attentional set  78
Summary  84

6 - CONCLUSIONS  86

REFERENCES  89
Acknowledgements

I am grateful to my friends and collaborators who directly or indirectly supported my career and without whose contribution the present work would have hardly ever seen the light of day:

Armin Schnider and Klemens Gutbrod, who taught me neuropsychological thinking, both from a cognitive and a neurological perspective.

René Müri, who was involved in the studies examining oculomotor performance in spatial neglect.

Christian Camen, Marie Di Pietro, Carole Greber, Laetitia Golay, Claude-Alain Hauert, Stéphanie Morand, Estelle Robert and Nathalie Valenza, who all contributed to the studies performed at the Division of neurorehabilitation.

The patients participating in these studies, whose patience and interest in scientific research always filled me with amazement.

Suzanne, Noé, and Elie, who suffered from my lack of attention to some fundamental Activities of Daily Living while I was writing this thesis.

The studies summarized here were supported by grants from the Swiss National Science Foundation, the Foundation Carlos and Elsie de Reuter, the Schmidheiny Foundation, and the Commission Informatique de l’Université de Genève.
ABSTRACT

One of the basic problems of attentional selection is the definition of a global measure allowing the comparison of stimuli that have very dissimilar physical qualities. The proposal of the present work is that the brain solves this criterion problem by selecting stimuli for further processing based on their saliency, which is defined as the strength of their neural representation. A saliency-based model of attentional selection is proposed consisting of two hierarchically arranged processing stages. At the first stage basic features of stimuli are represented in distinct feature maps, which are subject to top-down modulation from working memory as well as mutual inhibition. Feature maps feed into the saliency map, which represents the second stage of processing, and create distinct saliency differences in this map. An important feature of the model is that saliency is affected by physical characteristics of stimuli such as brightness or colour as well as by top-down influences such as behavioural goals, and that at the level of the saliency map these distinct influences are indistinguishable. The model proposes that attention selects the most salient stimulus for further processing.

Single-cell studies suggest that the posterior parietal cortex is the most likely location for the putative saliency map, because this region has a multimodal representation of the environment, strong inputs from extrastriate visual cortex and is interconnected with the prefrontal cortex, which actively maintains information in working memory. Studies with healthy participants suggest that phenomena such as visual search, surface-based selection, or attentional capture may be understood in terms of a saliency-based selection process. Further, the pathological attentional bias observed in brain-injured patients with deficits of spatial attention may be explained in terms of a lack of balance between bottom-up and top-down modulation of saliency.

List of abbreviations
AIP: Anterior intraparietal area; CT: Computerized tomography; DLPFC: Dorsolateral prefrontal cortex; FEF: Frontal eye fields; FIT: Feature integration theory; IOR: Inhibition of return; IPL: Inferior parietal lobule; IT: Infero-temporal; LIP: Lateral intraparietal area; LRP: Lateralized readiness potential; LVF: Left visual field; MIP: Medial intraparietal area; MRI: Magnetic resonance imaging; MT: Middle temporal area; PPC: Posterior parietal cortex; PSP: Progressive supranuclear palsy; RT: Reaction time; RVF: Right visual field; SBAS: Saliency-based Attentional Selection Model; SC: Superior colliculus; SEF: Supplementary eye fields; SOA: Stimulus-onset asynchrony; STG: Superior temporal gyrus; TPJ: Temporo-parietal junction; V4: Extrastriate visual area 4; VIP: Ventral intraparietal area.
2 – THE PURPOSE OF ATTENTION

Perceptual systems provide the organism with an enormous amount of information. As an example, consider the human visual system. It is capable of processing lines, angles, curves, colours and movement in parallel and to considerable detail. In a single glance, thousands of stimuli reach our eyes and are transferred to the visual cortex. The arrival of such an overwhelming amount of information obviously presents a serious problem for our brain. How can the brain differentiate between relevant and irrelevant stimuli? What is the mechanism that makes us disregard some stimuli and focus our interest on others? What is the fate of unattended information? What variables affect the selection process?

Why are these questions of importance? If our brains were just highly sophisticated information-processing machines, there would be no need for the selection of information. We would simply feed them with whatever information meets our senses. However, this is not the purpose of our brains. Neural systems evolved to provide organisms with the possibility to exert control over their environment, that is the ability to find food or a mating partner, to build a nest, to flee an enemy. In other words, brains primarily evolved for action, not for information storage. However, while sensory resources are nearly unlimited, action resources are severely restricted. This is why there is a biological need for a mechanism that differentiates among all incoming stimuli those that are behaviourally relevant from those that are not. This mechanism is what we call attention.

The effect of attention has been described in varying terms. For example, in one of the first cognitive studies, attention has been conceived as a ‘bottle-neck’ that blocks a large amount of sensory information, leaving only a small part for further processing (Broadbent, 1954). According to this idea attention acts as a relatively static filter, essentially reflecting the natural capacity restrictions of the brain. According to other authors attention enhances processing of selected information and attenuates processing of ignored information, eventually leading to improved perception and altered appearance of stimuli selected for further processing (Pestilli and Carrasco, 2005; Yeshurun and Carrasco, 1998). Such enhancement of attended information might be achieved by the reduction of noise that is necessarily present in complex scenes (Pashler, 1998). We will see in Chapter 3 that the notion of enhancement and attenuation describes well the effect of attention at the neural level. A further characteristic of attention is its spatial nature. Akin to a ‘spotlight’, attention is directed to regions in space and enhances processing of stimuli that are present in these regions (Cave and Bichot, 1999; Eriksen and Hoffman, 1972; Posner and Petersen, 1990). In some situations, attention increases the spatial resolution of focused information by ‘zooming-in’ (Eriksen and St. James, 1986; Yeshurun and Carrasco, 1998).
Finally, attention results in awareness, and since we are aware of what we are attending to people often use both terms interchangeably. However, the primary purpose of attention is to focus sensory resources on relevant stimuli and to select information for action. Therefore, awareness is rather a by-product of attention rather than its primary aim.

**The criterion problem**

It is introspectively easy to describe what attention *does* (James, 1890/1950), but it is much more difficult to determine *how* attention achieves its effects. Any flexible selection mechanism that is based in the brain faces a fundamental problem: if we assume that attention is a unique mental faculty (that is, there aren’t several selection functions proceeding in parallel), how is this faculty able to select the most relevant stimulus among several candidates originating in different modalities, stimuli that a priori cannot be compared? What criterion does attention use to select stimuli for further processing, when these stimuli belong to physically very distinct categories? I call this fundamental challenge for attentional selection the **criterion problem**. Consider for example, the following situation. While sitting in a bar that faces a busy street you are absorbed with reading an interesting magazine story. Though there are many potential distracters – cars, people, dogs and a policeman – you don’t find it difficult to continue reading your article. Then, suddenly, a small black animal that slowly creeps toward your hand captures your attention. This example demonstrates the core of the criterion problem. How did your brain determine that the black shape was *at that moment* more relevant than the text you were reading? Your brain somehow elaborated representations of the conspicuity of the two visual stimuli, and, after comparing these representations, judged to shift attention toward the small creeping shape. Any system able to make such a comparison between different visual inputs needs a measure of conspicuity that is applicable to all stimuli irrespective of their visual characteristics, and the definition of such a measure provides a solution of the criterion problem. I will use the term *saliency* to denote this measure (Ptak, 2006). This term is often used to refer exclusively to the physical quality of signals in relation to surrounding stimuli (Fecteau and Munoz, 2006; Itti and Koch, 2001; Knudsen, 2007). A stimulus is said to be salient if it stands out among or ‘pops up’ among other stimuli, like a fly sitting on a white table. According to this definition, saliency is purely determined by the physical, bottom-up characteristics of the display. It is sometimes contrasted with the term relevance, a term expressing whether a particular stimulus fits to the momentary goals. However, there are many reasons to believe that at the neural level, saliency and relevance are two sides of the same coin (see Chapter 3). I agree with the above authors that there are situations when saliency is nearly exclusively determined by the perceptual characteristics of the display. An example for such ‘bottom-up’ biases is shown in Figure 1,
which exemplifies how a single letter ‘pops up’ among homogenous distracter letters. While there are several factors that determine the saliency of stimuli (Figure 1 demonstrates the effect of similarity of a target stimulus to surrounding stimuli), situations when attention is influenced exclusively by perceptual biases are rare. I will argue in this paper that ‘top-down’ biases such as conceptions, motivation, or expectations may significantly affect saliency and appearance of stimuli (see Chapter 4). The neural response to the very same stimulus changes dramatically as a function of whether this stimulus is an action goal or whether it is irrelevant. My position therefore is that the term saliency should not denote the absolute physical quality (such as size, lightness, or speed), but the relative or perceived quality of the signal. This quality is represented in terms of an entity that scales the strength of neural response of stimuli of different sensory origins. In order to permit selection among stimuli with different spatial coordinates, this scaling is performed in a spatially selective manner, resulting in the coding of saliency in a saliency map (Fecteau and Munoz, 2006; Koch and Ullman, 1985). I will specify the different properties of saliency maps and the factors affecting saliency in Chapter 2.

If we accept that the criterion problem is solvable with a mechanism that selects stimuli based on their saliency contrast, we must also conceive where in the brain saliency is represented. A possible response is that saliency is coded by those regions that represent sensory characteristics of stimuli, that is primary sensory regions. However, this possibility leaves us with a multitude of different and widespread regions. For example, since colour and motion of a stimulus are coded in distinct cortical areas, the respective saliency values for colour and motion would be represented in spatially disparate regions. This solution faces the problem that every single stimulus would receive several saliency values, one for each of its perceptual features, resulting in an important number of saliency values that would have to be represented by the perceptual system. The main
advantage of the saliency concept – the representation of the conspicuity of different stimuli in a single measure – would be lost. A more plausible and computationally simpler solution is that saliency is represented by specialized regions in the brain. In order to correspond to the demands of a system that selectively represents saliency, these regions should fulfill the following criteria: First, to be able to represent the saliency of different perceptual characteristics, these regions should receive information from different sensory sites. Second, in order to preserve the advantage of selecting stimuli on the basis of saliency alone, neurons in these regions should code the world in a featureless manner (Fecteau and Munoz, 2006); that is they should represent only the saliency of objects without representing their perceptual features. Finally, these regions should be capable of spatial coding, since attention is always directed to a specified region in space. A few brain structures correspond to these criteria. At the cortical level, probably the best candidate is the posterior parietal lobe. This region (in particular the lateral intraparietal area, LIP) contains neurons that are spatially selective, receive inputs from the occipital lobe that convey bottom-up information and the frontal lobes carrying top-down signals, and are sensitive to the target-status of a stimulus.

Preview

I have outlined several concepts that are of importance for the understanding of the basic problems and the functioning of attention. The aim of this work is to present converging and complementary evidence supporting the view that the criterion problem is solvable with a neural mechanism that selects stimuli based on their saliency. I will first illustrate some basic properties of saliency maps before describing a model of attentional selection based on the saliency of stimuli (Chapter 2). Support for this model will be presented in Chapters 3 (arguments from neurophysiology), 4 (arguments from cognitive studies), and 5 (arguments from neuropsychological studies). Finally, a conclusion will be made and some limitations of the model will be discussed in Chapter 6.
2 – A Saliency-Based Model of Attentional Selection

This chapter describes the Saliency-based Attentional Selection Model (SBAS, see Figure 2) of attention, in which attentional selection is based on a winner-take-all mechanism that at any given moment in time picks up the stimulus with the highest value on a putative “saliency map”. The model integrates findings from neurophysiological, behavioural, and computational studies. Similarly to several previous models of visual attention (Koch and Ullman, 1985; Treisman and Gelade, 1980; Wolfe, 1994), the present model is based on a two-stage process. In a first stage, primary features are extracted from the environment by specialized feature maps. Feature extraction is parallel across the visual field, and is independent of attention. Feature maps activate retinotopically corresponding locations in the saliency map, which provides a featureless representation of the environment that is accessible to spatial attention. Based on the considerations that were outlined in the preceding chapter, I begin with describing the core of the model, the saliency map.

The saliency map

Several neurophysiological, cognitive, or computational models of visual attention assume that visual space is represented in retinotopically organised maps (Gottlieb, 2007; Itti and Koch, 2001; Koch and Ullman, 1985; Treisman and Gelade, 1980). The central part of the model proposed here is the saliency map, that is a two-dimensional representation of space that encodes saliency of perceived objects. Saliency is not an absolute property, and activity in the saliency map reflects the perceptual distinctiveness of an object relative to environing objects in the scene. Figure 2 shows that the degree of activity in the saliency map is the result of a combination of multiple retinotopic low-level feature maps representing basic perceptual features (such as length, orientation, curvature, or colour), which are computed by primary and secondary visual areas. Since feature maps and the saliency map are retinotopically organised, every point on a feature map has its corresponding point in the saliency map. However, spatial coding at the level of the saliency map is not only retinotopic, but also head- and body-centred. This triple representation of space assures that target representations remain constant across eye shifts with respect to the head and body.

An important question is how the neural coding of saliency is computationally related to activity in the different feature maps. The simplest possibility is that saliency is the average of all individual feature values. While computationally simple, such an averaging process is not compatible with some classical findings from visual attention studies. For example, when in a display one feature (say, colour) dominates other features, this feature also generally captures attention (Treisman and Gelade, 1980). This greater contribution of
colour to the computation of saliency would be obscured if colour information were averaged with several other features (Itti, Koch and Niebur, 1998). Feature averaging also fails to explain the superiority of partial report over whole report, which is the ability to focus attention on a specific feature (e.g. shape) and to report observed changes in that feature better than changes in another feature (e.g. size, Bundesen, 1998). Effects such as the superiority of partial report are therefore only possible if the input of feature maps into the saliency map is weighted according to its relative importance. In the SBAS model weighting has three possible sources: mutual inhibition, location enhancement and feature enhancement. Mutual inhibition is a result of competitive interactions between distinct feature maps (Figure 2, blue arrows) that overemphasizes (higher weighting) maps containing highly distinct features and whose input into the saliency map will therefore be most important. Competition between feature maps is organized retinotopically, that is, activity at a specific position in one feature map inhibits activity at the same position in other feature maps. Location enhancement results from top-down attentional signals from the saliency map to feature maps, which leads to enhancement of specific locations in all feature maps, independently of the content of these locations. A third source of feature weighting is ‘top-down’ feature enhancement by behavioural goals represented in working memory. The model also contains a direct route between working memory and the saliency map that represents task-selective signals leading to additional enhancement or inhibition of saliency peaks by task-demands. Instructions such as ‘focus on red, discard white items’ bias the competition
between features in favour of the target feature and thus increase the relative contribution of this feature when saliency is computed.

Attentional selection based on saliency follows a “winner-take-all” rule (Koch and Ullman, 1985). Thus, the region with the greatest saliency values captures attention and is selected for further processing. The speed of attentional selection depends on the difference between single activation peaks of the saliency map. For example, when there is an isolated peak attentional selection is much faster than when there are several peaks of comparable activity.

**Feature enhancement**

The activity of the saliency map is subject to highly dynamic and rapid changes that are triggered by perceptual, conceptual, emotional, or motivational factors. For example several purely perceptual factors modulate activity through perceptual organisation of the display (see Figure 3). Thus, edges and intersections are more salient than straight lines even when contrast and intensity are equal. A powerful perceptual factor affecting saliency is luminance contrast, since contours, edges and shapes are only perceived when stimuli are presented with enough contrast. Luminance contrast often interacts with colour contrast, though both factors might affect saliency independently (visual stimuli purely defined by colour contrast necessitate an equiluminant scene, which may only be obtained in the laboratory). Another important factor is perceptual similarity of objects in the display. Highly similar objects (for example, regular patterns) are less salient than single irregularities. A singleton (a single object defined by the presence of a unique perceptual feature that differentiates it from other objects in the scene), be it defined by shape, colour, or other characteristics, is always more salient than several similar objects. Singletons defined by the *absence* of a feature (e.g. the letter P presented among letters R; see Figure 1) are more difficult to find than singletons defined by its presence (Treisman and Gelade, 1980). In addition to the saliency-modulating effect of a single feature, the homogeneity of distracters significantly affects their saliency, because homogenous objects appear as a repetitive perceptual pattern or a texture (Figure 3). Visual search for a target defined by the presence of a feature that is shown together with highly homogenous distracters is almost independent of the number of distracters in the display (Duncan and Humphreys, 1989). This effect can be understood in terms of the fact that representations in feature maps are subject to lateral interactions, which result in enhanced representation of discontinuities (e.g. intersections, borders, colour changes) and attenuated representation of regularities (e.g. regular patterns and textures). Note that since feature maps are exclusively retinotopically organized and therefore are unable to represent orientation-independent views of objects, distracters are only homogenous if they are similar and have the same orientation (Figure 3).
A second group of perceptual factors affecting saliency are factors that differentiate stimuli on the basis of their spatial arrangement. The distance from other objects increases the saliency of an isolated object, even if it is highly similar to other objects in the scene (Figure 4). Artists often use this principle to underline the importance of a specific motive in their composition. A second important spatial factor is retinal position. Due to the greater spatial resolution of the central compared to the peripheral visual field stimuli closer to the fovea will appear more salient than stimuli presented in the visual periphery.

Third, some stimuli enhance saliency primarily due to their emotional valence. Animal teeth, small hairy insects, blood as well as other emotionally arousing stimuli belong to this category (Figure 5). Still other perceptual factors determine the species-specific behavioural significance of visual stimuli. For
example, human faces characteristically capture the gaze of observers. Similarly, text is only salient to human observers, not to animals. Species-specific factors might be subject to individual preferences (e.g. an architect will be more sensitive than the average observer for buildings). However, though emotional and species-specific stimuli appear to capture attention reflexively, it is not always clear whether this effect is due to their affective content or to a specific perceptual component. For example, all three emotional stimuli shown in Figure 5 are characterised by high figure/ground contrast.

All factors influencing saliency have in common that they bias the representation of individual features present in a scene in favour of a specific feature, increase the saliency of that feature and thus influence the attentional selection process. A central assumption of the SBAS model is that all factors affecting saliency act at the level of feature maps. This is also the case for conceptual factors, behavioural goals, expectations and other ‘top-down’ biases. Behavioural goals constitute one of the most important components of attentional selection: attentional set. Attentional set may be defined as the behavioural predisposition to attend to some stimuli more than to others, an operation that strongly influences the filtering mechanism of attention. For example, if you search for a book with a green cover, you very efficiently ignore all books with a red, blue, or yellow cover, and your search will only be affected by the number of green distracters. Figure 6 exemplifies the impact of attentional set on the computation of saliency. The display shows two singletons, one defined by its shape (white letter T), the other by its colour (red letter L). The figure shows schematically the activity of two feature maps representing shape and colour and the resulting activity in the saliency map. Saliency of the four stimuli is represented as Gaussian ‘bumps’ of different heights. If the observer searches for the letter T, the computation of saliency will primarily be affected by the feature map representing letter shape, and the contribution of colour will be limited. In contrast, if the target is the red letter, the map representing colour will be enhanced and consequently saliency computation will essentially be based on colour. Note that saliency of the ignored singleton (e.g. red L when the target is a
T) is greater than the saliency of the other distracters (white Ls) due to perceptual (bottom-up) influences on saliency. Visual search for a target defined by a single feature is not influenced by the number of distracters, because the bottom-up effects on saliency are much weaker than the top-down effect of attentional set. In this situation, attentional set enhances saliency of the target.

Consider now a search task in which the target is defined by two features (e.g. shape and colour) and half of the distracters share one feature while the other half share the other feature (Figure 7). In this situation attentional set leads to strong activation of both feature maps, which in turn results in strong mutual inhibition. Since only the target letter has comparable activity in the colour map and the feature map, at target position competition between feature maps is particularly important and the contribution of both features to the computation of saliency is limited. Therefore, when the search target is defined by a conjunction of two features, it has only slightly greater saliency than the distracters sharing a feature and visual search is strongly affected by the number of distracters. Thus, top-down feature enhancement by attentional set does not always result in strong enhancement of targets at the level of the saliency map.

Figure 6. Effect of attentional set on saliency of the letters shown in the display when the target is defined by a simple feature. Schematic representation of the feature maps for colour and shape and the saliency map when the observer is instructed to search for the letter T (left) or for the red letter (right).
Location enhancement

Representation of saliency is not only the result of feature-based inputs, but is also affected by spatial factors. For example, due to cortical magnification and the ensuing overrepresentation of central space, even in the absence of visual stimulation our visual system has a bias favoring central locations (Figure 8 A). Adding a surface that acts as reference frame for a given task (e.g. a TV-screen when we watch a film, a book when we read, a sheet of paper when we write) leads to enhancement of those locations of the saliency map that code for the portion of space delimited by the surface (Figure 8 B). Contrasting with the feature-based enhancement discussed in the previous section, such reference-frame effects are more adequately described as space-based or surface-based. Surface-based enhancement and the central bias eventually act together and...
result in greater saliency of central compared to more peripheral regions of the surface (Figure 8 C).

In contrast to the enhancement of feature maps by behavioural goals, enhancement of spatial locations affects directly representations at the level of the saliency map. The attentional ‘spotlight’ highlights locations in the saliency map and feeds back into the feature maps, resulting in enhanced processing of any object falling within the spotlight of attention. An example of this dynamic surface-based enhancement of saliency provides the spatial cueing effect. In a spatial cueing experiment a visual target is typically presented at slightly peripheral positions in the left or right hemifield. Prior to the target a brief cue summons attention either to the hemifield of the upcoming target (valid cue) or to the opposite hemifield (invalid cue). If the cue precedes the target by 200-300 ms subjects are faster to detect the target in the valid cueing condition compared to the invalid cueing condition (Posner, 1980; Posner, Cohen and Rafal, 1982). However, these spatial effects are of short duration and the activity in the saliency map that is elicited by a sudden peripheral onset rapidly decays. At the behavioural level, this rapid decay of saliency results in a phenomenon termed inhibition of return (Posner, Rafal, Choate and Vaughan, 1985). In visual search, inhibition of return ensures that distracters that have already been visited (that is: ‘old’ items) cease capturing attention and interfering with attentional selection of items that have not yet been processed (‘new’ items). When eye movements are measured during visual search, participants are less likely to direct their gaze toward a previously fixated distracter for a period lasting 3-4 saccades (McCarley, Wang, Kramer, Irwin and Peterson, 2003). After that time, the probability that attention revisits previously fixated distracters approximates the probability to select distracters that were never fixated before, suggesting that the inhibitory effect lasts 3-4 saccades (or one second). The spatial cueing effect and inhibition of return suggest that activity in the saliency map is subject to highly dynamic spatial and temporal changes.

Though these spatial dynamics are an important feature of the saliency map, another component is even more important: representations in the saliency map are not coded exclusively in retinocentred coordinates, but have access to head-centred and body-centred coding. Consider the advantage of a system having access to different spatial coordinates. In a purely retinocentred representation, the focus of attention does not keep a stable position with respect to the head or the body, but shifts with every eye movement. Consequently, an object will only remain in the focus of attention as long as the eyes are stable, but will leave this focus as soon as the eyes move. In order to correct for this error, the attentional ‘spotlight’ must shift correspondingly, but in direction opposite to the eye shift. Similarly, a system that is purely head-centred shifts its representation with every movement of the head. Only a system that codes the environment in body-centred coordinates retains a stable representation across shifts of the eyes and the head (note that such a system is also best suited for showing crossmodal
attention effects, such as spatial cueing effects with auditory cues and visual targets or vice versa). For these reasons, one of the most important properties of the saliency map is that it has access to body-centred representations of the environment.

**Interactions within and between saliency maps**

Spatial enhancement of saliency and inhibition of return are direct properties of the saliency map rather than the consequence of mechanisms operating at the level of feature maps. They are the results of activating or inhibitory lateral interactions between neural units constituting the saliency map. Activating inputs exist between nearby locations in the saliency map while inhibitory inputs arrive from more distant sites. Activation and inhibition are proportional to the degree of activity in the saliency map: highly salient locations strongly activate nearby locations and strongly inhibit distant locations. This principle guarantees that activity differences will be accentuated in order to facilitate attentional selection.

Neurophysiological studies suggest the presence of more than one saliency map in the primate visual system (see Chapter 3). If several brain regions represent saliency these regions must somehow interact to permit efficient attentional selection. A possibility is that activity differences in distinct saliency maps reflect the affordance of visual stimuli for different actions. For example, a stimulus that is farther than reaching distance will activate saliency maps more that are mainly involved in oculomotor programming than to the programming of hand movements. Conversely, the selection of a target for a directed hand
movement relies mainly on a ‘hand movement map’. Thus, the interactions between saliency maps are mainly affected by output factors such as whether an oculomotor response or a goal-directed manual response is planned.

**Summary**

The present chapter described a model of attentional selection that is characterized by the following features:

a) The strength of the neural response evoked by sensory stimulation results in the representation of stimulus saliency that is the primary determinant for the attentional selection of the stimulus for further processing. The term here is explicitly used to denote the strength of the neural representation of the stimulus, though this use deviates from the more common meaning of saliency as a psychological, not a neural quality.

b) Saliency is represented in different spatial coordinates at the level of a saliency map. The saliency map contains the saliency values of each region of represented space.

c) Saliency is affected by different factors. While many of the factors may be represented on a continuum between purely stimulus-driven (bottom-up) and controlled (top-down) factors, at the neural level they all have the same effects: they either strengthen or attenuate the neural signal that constitutes perceptual saliency.

d) Attentional selection is based on a ‘winner-take-all’ principle: the region of space that at a given moment in time has the greatest saliency will momentarily receive all attentional resources and be selected for further processing.

e) The brain contains regions that have specialised in representing saliency. These regions are characterised by several features, such as multimodal input, featureless coding, and spatially-selective responses. The primary cortical region specialized in saliency representation is the posterior parietal cortex (PPC). The PPC has convergent inputs from prefrontal areas involved in the selection of action-goals as well as perceptual inputs from occipital areas, and communicates with premotor areas involved in action programming.
3 – NEURAL MECHANISMS OF SALIENCY REPRESENTATION

Substantial evidence from neurophysiological studies supports the basic components of the SBAS model of attentional selection. The present chapter reviews neurophysiological evidence for top-down modulation of responses of neurons in extrastriate visual areas to basic visual features, and shows how saliency of stimuli is computed, supporting the contention that specialized brain areas establish an abstract representation of the environment in a saliency map.

Attentional modulation of neuronal responses to visual features

Neural responses to visual stimuli are distributed across a network of at least 30 cortical visual areas (Felleman and Van Essen, 1991, see Figure 9). Functional mapping of the human visual cortex suggests a similar organisation of the human visual system (Grill-Spector and Malach, 2004). Neural representations in this highly elaborate and specialized system are characterized by a high degree of redundancy and interact with each other in a complex way. Attention provides a biasing signal that reduces competition by increasing the saliency of the attended stimulus in expense of the unattended stimulus (Desimone and Duncan, 1995; Kastner and Ungerleider, 2001; Reynolds and Chelazzi, 2004). In 1985 Moran and Desimone (1985) published a seminal study in which they described the attenuation of neural responses when attention was directed away from a visual stimulus. The authors examined the response of neurons in area V4 and the inferior temporal cortex (IT), both of which are part of the ventral visual stream critical for object recognition (Mishkin, Ungerleider and Macko, 1982; Ungerleider and Mishkin, 1982). Neurons in V4 have receptive fields up to 4° wide, and IT cells have large receptive fields that often span more than 25°, making it highly likely that in real-life situations more than one object is present at the same time in the receptive field (Gross, 2008). Moran and Desimone (1985) argued that receptive fields of that size must create competition between visual objects for the cells’ response, and that this competition could be subject to top-down modulation by attention. They determined which stimulus most effectively elicited a response of V4 and IT cells and presented this stimulus simultaneously with a non-preferred stimulus in the receptive field of the cell. When the monkey attended to the preferred stimulus, the cell showed a strong response. However, when the monkey attended to the non-preferred stimulus (which was equivalent to ignoring the preferred stimulus), the response of the cell was greatly attenuated. When the non-preferred stimulus was placed outside the cells receptive field, no such attenuation of response was observed. This inhibition of responses to non-attended stimuli was confirmed by other studies. Thus, Luck and collaborators (Luck, Chelazzi, Hillyard and Desimone, 1997) found strong attenuation of activity of V2 and V4 cells when attention was directed to the non-preferred of
two simultaneously presented stimuli. When the two stimuli were presented sequentially, the size of this top-down attentional modulation was reduced. Sato (1988) examined cells with receptive fields spanning across both visual fields and reported stronger suppression of responses when the neutral stimulus and the attended stimulus were located in the contralateral visual field than when the neutral stimulus was presented ipsilaterally, though in both cases the neutral stimulus was within the receptive field of the cell. Motter (1994), in some contrast to the results of Moran and Desimone (1985), found suppression of the response to a preferred stimulus in V4 even when attention was directed outside of the cell’s receptive field.

The implication of the suppressive effect of attention on cell’s responses in a simple visual search task was examined by Chelazzi and collaborators (Chelazzi, Duncan, Miller and Desimone, 1998; Chelazzi, Miller, Duncan and Desimone, 1993). The monkeys were required to make a saccade to one of two peripherally presented visual choice stimuli, one of which was a preferred stimulus for the cell, the other a non-preferred stimulus. The target stimulus was indicated by a central cue presented prior to appearance of the target display. If the cue indicated that the monkey had to make a saccade to the preferred stimulus, cells in the IT cortex responded vigorously upon appearance of the target display. If, however, the target was the non-preferred stimulus, the cell’s response to the now irrelevant preferred stimulus initially increased, but was suppressed about ~200 ms following the onset of the target display. Interestingly, this suppression of cell activity started well before the monkey made a saccade. The behavioural response of the monkey was thus predicted by the activity of IT cells.

In a recent study Bichot, Rossi and Desimone (2005) demonstrated that attention-based enhancement of activity might account for the differences
observed between parallel feature search and serial conjunction search. In one
task the authors requested monkeys to search for a target defined by its colour or
shape and measured responses to stimuli falling in the receptive field of V4 cells
while the monkey freely moved its gaze across the display (Figure 10). The
critical responses were those when a target stimulus fell in the receptive field of
the cell, but was not detected by the monkey. When this occurred and the target
stimulus had the preferred feature of the cell the activity of the cell was
enhanced. In contrast, response of the cell was suppressed when the stimulus was
not the search target, even though it had the cell’s preferred feature. In addition,
V4 cells were also strongly activated by features that were similar, but not
identical to the target feature. Thus, during feature search V4 cells signalled the
presence of the target feature or a visually similar feature when this feature fell in
their receptive field, in parallel throughout the visual field.

When the authors examined responses of V4 cells to distracters in a
conjunction search task, they found parallel activation of neurons that shared a
single feature with the target across the visual fields. For example, when the
target was a red X, activity of cells whose preferred stimulus was red was
enhanced compared to their activity when the target was green. Thus, whether a
single feature or a feature conjunction defined the target, activity linked to
distracters sharing a feature with the target was enhanced. However, while this
parallel activation of visual filters is beneficial for feature search, it slows down
visual search for targets defined by a conjunction of features.

One question arising from these findings is whether the suppression of cell
activity by attention is mainly space-based or feature-based (Maunsell and Treue,
Recall that in the study by Moran and Desimone (1985) the suppressive effect was observed when the monkey directed its attention to the non-preferred stimulus. Thus, suppression of activity was the consequence of a spatial shift of attention. In area MT (which is part of the dorsal stream), spatial attention leads to a shift of the receptive field of neurons in direction of the focus of attention (Womelsdorf, Anton-Erxleben, Pieper and Treue, 2006). In contrast, in the study by Chelazzi and collaborators (Chelazzi et al., 1998; Chelazzi et al., 1993) the monkey did not know in advance the position of the target stimulus. The observed suppression of cell activity upon onset of the target display was therefore a result of attentional selection based on object features. More direct evidence for feature-based enhancement was reported by Fallah and collaborators (Fallah, Stoner and Reynolds, 2007), who found that feature-based attention modulates the activity of V4 cells independently of spatial attention. These authors presented virtual surfaces defined by rotating dot patterns. On each trial two such dot patterns were presented that were of different colours and rotated in opposite directions. One pattern was presented first and could have the preferred colour of the neuron or the non-preferred colour. The second pattern was added after 500 ms and could again have the preferred or the non-preferred colour. The authors found that responses of colour-selective neurons in area V4 were mainly driven by the delayed surface, but only if this surface was of the preferred colour. Thus, cell activity was driven by the pattern with abrupt onset, which captured attention in the absence of a spatial modulation.

Strong evidence for feature-based selection was also found in area MT, which is involved in visual motion processing. Thus, neural responses of MT cells to a dot pattern that moves in the preferred direction are suppressed when the monkey directs attention to a superimposed pattern that moves in the opposite direction, even though in the two conditions the spatial location of attention is identical (Wannig, Rodriguez and Freiwald, 2007). Treue and Martinez-Trujillo (1999) presented a random dot pattern in the receptive field of MT cells that moved in the preferred direction and thus activated the cell maximally. At the same time, a second dot pattern that moved in the preferred direction of the cell or the opposite direction was presented outside the receptive field. When the pattern outside the receptive field moved in the preferred direction the cell significantly increased its response compared to when this pattern moved in the opposite direction. This effect was observed even though, in both conditions, the monkey attended to this second pattern shown outside of the receptive field. Thus, in this study the effect of attention interacted with the effect of the preferred feature presented outside the receptive field of MT cells. Treue and Martinez-Trujillo (1999) proposed that the enhancement and suppression of neural activity is a function of the similarity of the features of the behaviourally relevant stimulus and the distracter stimulus, the spatial selectivity and the sensory selectivity of the neuron. Thus, if a cell’s preferred feature is attended, the response of the cell is enhanced even though attention is directed outside its receptive field. A similar
enhancement is observed when attention is directed inside the receptive field of the cell independent of its response selectivity.

In sum, the findings of attentional suppression and enhancement of cell activity provide converging support for the SBAS model. Neural responses in area V4 and the inferior temporal cortex suggest that feature maps are elaborated in the ventral visual stream, and that activity of these maps is subject to top-down suppression or enhancement. These attentional effects are very similar to purely perceptual modulations of visual stimuli like increasing stimulus contrast (Pestilli and Carrasco, 2005; Reynolds and Chelazzi, 2004), suggesting that bottom-up sensory and top-down attentional influences have identical effects on target visual areas (Treue, 2001). These results also support the notion that attention gates visual processing in the ventral stream by inhibiting or ‘filtering out’ irrelevant information and thus attenuates the competition between representations of stimuli in the ventral visual stream (Desimone, 1999; Desimone and Duncan, 1995; Reynolds and Desimone, 1999). Such competitive interactions provide the physiological basis of the mutual inhibition between feature maps postulated in the SBAS model. In addition, the results of Luck and collaborators (1997) support the surface-based enhancement by spatial attention postulated in our model. These authors found enhancement of spontaneous cell activity when attention was directed inside the receptive field of V2 or V4 cells compared to when attention was directed elsewhere, even when no stimulus was present in the receptive field.

If V4 and IT elaborate maps of visual features, what is the effect of selective damage of these regions on visual attention? Schiller and Lee (1991) found that following selective damage to V4 monkeys showed specific retinotopic deficits in colour, pattern, size, or brightness discrimination that were particularly striking when the target stimulus had lower visual saliency than the comparison stimuli from which it had to be discriminated. This latter finding suggests that visual processing with a V4 lesion might be particularly affected by distracter stimuli. In accord with this prediction, De Weerd and collaborators (De Weerd, Desimone and Ungerleider, 2003; De Weerd, Peralta, Desimone and Ungerleider, 1999) required monkeys with selective damage to V4, TEO, or combined damage of these two regions to make perceptual discriminations of targets flanked with neutral distracters or distracters that shared a feature with the target. The authors found that distracters sharing a feature with the target impaired discrimination performance more than neutral distracters. In addition, discrimination performance varied as a function of distracter contrast, such that increasing distracter contrast above the target contrast resulted in much stronger impairments than when distracter contrast was below the target contrast. These effects were strongest after combined V4 and TEO damage, but remained significant even when damage was restricted to V4 or TEO. Very similar impairments were observed in a rare patient who suffered isolated damage to the lingual gyrus in the right hemisphere (Gallant, Shoup and Mazer, 2000). In
addition to achromatopsia affecting the left upper quadrant the patient showed strong influence of flanking distracters on orientation discrimination of sinusoidal gratings. In the absence of distracters discrimination in the affected quadrant was comparable to discrimination in the rest of the visual fields. The presence of task-irrelevant flanking distracters resulted in pathologically increased discrimination thresholds in the impaired quadrant. Thus, the very similar results of De Weerd et al. (1999) and Gallant et al. (Gallant et al., 2000) suggest that human V4 is located in the inferior-posterior temporal lobe. Together, the monkey and human lesion studies suggest that damage to V4 and TEO leads to an inability to filter out the distracting effect of peripheral information on the processing of visual stimuli. Such a deficit is compatible with the proposal that stimuli compete for attentional resources in the ventral visual pathway (Desimone, 1999; Duncan, Humphreys and Ward, 1997; Kastner and Ungerleider, 2000), though it does not permit to draw conclusions about the origin of the signal that biases competition in favour of one stimulus or the other. Given the claimed role of inferior temporal regions for the elaboration of feature maps, the fact that an isolated damage to V4 mainly impairs visual discrimination under conditions of sensory competition is surprising. It is possible that other temporal visual regions are sufficient for visual discrimination, but that the top-down biasing signal affecting competition in favour of the discrimination stimulus passes through V4 before reaching these regions. This conclusion is supported by the finding that the seemingly attentional impairment after V4 damage is strictly retinotopic.

The posterior parietal cortex and spatial attention

Several results presented in the previous section raise the question what brain regions are the origin of top-down modulation of feature maps by attention. We have seen that modulation can be space-based or feature-based. Space-based selection is the result of the attentional spotlight - a metaphor used to describe the two-dimensional extent spatial attention (Cave and Bichot, 1999). Spotlight-theories assume that attention highlights locations in space to which it has covertly been shifted (Eriksen and Hoffman, 1973; Posner and Petersen, 1990). However, these theories are founded on results that have been obtained with paradigms that require participants to fixate a central spot and to inhibit eye movements. Consequently, they do not specify whether the space in which attention operates is coded in retinocentred, head-centred, or body-centred coordinates. Since human observers make approximately 3-5 saccades per second (Findlay and Gilliehrist, 2003; Henderson and Hollingworth, 1999), the retinotopic location of objects selected by attention constantly changes though their spatial location with respect to head and body may remain constant. Attention would seriously lack functionality if it used an exclusively retinocentred representation of space, because highlighted locations would shift
with every movement of the eyes. For all tasks requiring a goal-directed manual response it is therefore essential to represent the environment in coordinates that remain stable with respect to the body irrespective of eye and head movements. Spatial attention is necessary when a visual stimulus is the target of an eye movement or a manual response (e.g. pointing or reaching movements) and therefore entails access to space representations in multiple coordinates. The brain region that has access to multiple representations of space is the posterior parietal cortex. The monkey posterior parietal cortex (PPC) is composed of several functionally distinct subregions that are involved in sensory-motor integration and the preparation of actions. The best-studied areas of the PPC are the ventral intraparietal area (VIP), the medial intraparietal area (MIP), the anterior intraparietal area (AIP), and the lateral intraparietal area (LIP) located within the intraparietal sulcus, as well as area 7a of the inferior parietal lobule. Cells in VIP, MIP and AIP respond to visual and tactile stimulation and are involved in the representation of perioral and head-centred space (VIP), reaching space (MIP) and space used for the guidance of grasping with the hand (Colby and Goldberg, 1999). While these three areas establish a representation of space that is relevant for sensori-motor transformations, activity of LIP and 7a is directly involved in the preparation of eye movements and spatial attention. Andersen, Essick and Siegel (1985) examined the visual responses of neurons of area 7a and found that these were strongly modulated by eye position. They first mapped the receptive field of individual neurons with the animal looking straight ahead. They then had the animal look at different locations on the screen and always stimulated the same retinotopic location, thus ensuring that the position of the stimulus was constant in retinotopic space, but varied with respect to head position. The authors observed that the response of the cells was significantly influenced by the current eye position, a finding indicating that these cells are able to compute a representation of visual stimuli in head-centred coordinates. In this early experiment the monkey’s head was immobilized, precluding a distinction between head- or body-centred representations in the PPC. However, later experiments in which monkeys were allowed to move their heads horizontally, revealed that approximately half of the sampled cells encoded space as a function of the absolute gaze position (Andersen, 1995; Snyder, Grieve, Brotchie and Andersen, 1998), whether this was achieved by moving the eyes (head fixed, eyes moving freely) or by moving the head (eyes fixed, head moving freely). These results suggest that neurons in the PPC encode space in head-centred and body-centred coordinates.

In addition to its ability to code space in different coordinates, several findings suggest that the PPC plays an essential role in spatial attention. For example, Lynch, Mountcastle, Talbot and Yin (1977) found that neurons in area 7 are active when the monkey fixated an object that it desired (e.g. food) or become active before and during visually-evoked saccades, but not during spontaneous eye movements in the dark. Bushnell, Goldberg and Robinson
(1981) observed characteristic enhancement of responses of area 7 cells in several tasks requiring a shift of attention to peripheral visual stimuli. Thus, in comparison to an irrelevant peripheral stimulus, cell activity was enhanced prior to a saccade to a peripheral stimulus, when the monkey attended to a peripheral stimulus while maintaining fixation, or when it made a reaching response to a peripheral stimulus without leaving the fixation point. Importantly, cells that showed enhanced responses in the saccade task also exhibited very similar enhancement by peripheral attention in the absence of a saccade. Such task-independent enhancement of cell activity prior to a response might be the neural basis of the finding that visual processing of targets presented at the landing position of an upcoming saccade is enhanced (Deubel and Schneider, 1996; Kowler, Anderson, Dosher and Blaser, 1995).

Further results supporting the attentional role of the PPC come from studies that examined the effect of spatial cues and of abrupt onsets on neuronal activity in the LIP. Both manipulations have been extensively examined in healthy
Saliency, Selection, and Spatial Neglect

participants and are well known for their strong impact on spatial attention (Cave and Bichot, 1999; Yantis and Jonides, 1990). Cohen, Cohen and Gifford (2004) examined the effect of visual and auditory spatial cues on LIP activity preceding a saccade to a visual target. Monkeys showed strong cueing effects in both conditions, that is, they made saccades of shorter latency when visual and auditory cues correctly predicted the position of the target. At the neural level, these cueing effects resulted in enhanced responses of LIP neurons to both kinds of spatial cues, though the modulation of activity by visual cues was stronger.

Gottlieb, Kusunoki and Goldberg (1998) required monkeys to make a saccade to the centre of a display composed of eight irrelevant visual stimuli arranged on an imaginary circle (Figure 11). The saccade brought one of the eight stimuli in the receptive field of a LIP neuron. In one condition this stimulus was present from the beginning of the trial, while in the recent-onset condition it appeared 500-2000 ms prior to the onset of the saccade. Neural activity triggered by the recent-onset stimulus was much stronger than activity in the stable condition, consistent with the interpretation that its abrupt onset increases the saliency of a stimulus and therefore captures attention in a reflexive way. Also, the neural activity triggered by an abrupt onset is clearly dissociated from the more sustained activity of the same neurons prior to a saccade (Gottlieb, Kusunoki and Goldberg, 2005), suggesting that it is elicited by stimulus-driven factors and not by the motor intention. Finally, Bisley and Goldberg (2003) found that the population activity of LIP neurons predicted the locus of attention, be it attention captured by a task-irrelevant, abrupt-onset distractor or a visual stimulus that was the saccade target. It is unlikely that this effect only reflects the intention to make an eye movement (Snyder, Batista and Andersen, 1997), since neural activity reflected the attentional priority of a location and did not change when the monkey made no eye movement or when it directed the response elsewhere.

Together, these results strongly suggest that the PPC (in particular area LIP) is an important component of a network involved in directing attention. This conclusion is further supported by findings of impairments of attention in monkeys following damage to the PPC. Lesions of monkey IPL lead to contralateral spatial extinction (Lynch and McLaren, 1989), an attentional disorder that is also found in human patients with parietal damage. Functional inactivation of LIP leads to neglect-related deficits in saccadic responses, with difficulty to initiate saccades to unilateral contralesional stimuli and absence of saccades to these stimuli when they are presented simultaneously with an ipsilesional item (Wardak, Olivier and Duhamel, 2002).

When considering the role of the PPC for attention it is important to recall the distinction between bottom-up attentional capture and top-down controlled attention (Egeth and Yantis, 1997). Some of the results presented above suggest that the primary role of the PPC is to direct attention in space and thus to highlight the position of possible action targets in head-centred and body-centred
coordinates. Isn’t this role of the PPC in conflict with the finding that activity of LIP neurons is influenced by bottom-up variables such as abrupt onset? There is no contradiction if we accept the idea that bottom-up processes interfere with the controlled deployment of the attentional spotlight. For example, the neural computation of stimulus saliency might be strongly biased in favour of features of an abrupt-onset stimulus, resulting in a corresponding bias of neural activity in the PPC. Indeed, Hayden and Gallant (2005) found that the abrupt onset of a stimulus in the receptive field of V4 cells results in a transient increase of activity 100-150 ms after target onset that overrides the modulating effect of spatial attention directed in the RF. Such temporal elimination of top-down by bottom-up signals might temporarily increase the saliency of distracter stimuli, resulting in reflexive attentional capture reflected in neural activity of the PPC.

In sum, this overview of PPC function suggests that the PPC transforms spatial representations of the environment from retinotopic in head-centred and body-centred coordinates. Further, the PPC (in particular area LIP) codes the allocation of attentional resources to visual (and to a significant, but lesser extent, auditory) stimuli, whether attention has been captured by the stimulus in a bottom-up manner, or whether attention has been directed voluntarily to that stimulus.

Top-down modulation by spatial attention and action goals

The SBAS model posits that the selection of locations by attention feeds down into the feature maps and enhances features of the selected location in comparison to unattended features. The preceding section has shown that activity of cells in the PPC strongly correlates with the focus of attention, suggesting that it might be the source of such top-down modulation. However, comparable attentional enhancement is observed in many areas of the visual system, including ventral areas V2, V4, and IT. How can we be sure that the origin of these attentional effects is the PPC and not some other brain region? The only way to answer this question is to measure attention-related activity of many brain areas and to determine which of these is the first to show attentional enhancement. In addition, we need to know whether the area that has been identified in this way directly influences activity of other, more specialized areas. Presently, these questions cannot be answered satisfactorily. However, in a recent study Saalmann, Pigarev and Vidyasagar (2007) measured simultaneously the activity of areas LIP and MT while monkeys performed a delayed match-to-sample task requiring them to match the orientation and location of two successively presented stimulus gratings. The second (target) grating was always a preferred stimulus for the neuron. The authors were interested in the modulation of cell activity by different attentional states. Therefore, before the target grating appeared, attention of the monkey was directed to the receptive field of the cell, or elsewhere. Attention significantly modulated activity in LIP
and MT, but the modulation of activity began earlier in LIP than MT neurons. In addition, spikes from MT neurons closely followed spikes generated by LIP neurons, suggesting that neural activity of both areas was highly correlated. These results suggest that LIP is the origin of attentional enhancement of MT neurons, though they do neither prove that it is the only source of enhancement, nor that it is the first area that exhibits attentional enhancement. Indeed, at least two areas located in the frontal lobes are similarly involved in top-down enhancement. Johnston and Everling (2006) examined simultaneous activity of neurons in the dorsolateral prefrontal cortex (DLPFC) and neurons of the superior colliculus. They identified DLPFC neurons that send direct projections to the intermediate layer of the superior colliculus, a midbrain structure involved in preparation and execution of saccades. The authors used a prosaccade/antisaccade task in which on some trials monkeys were required to make a saccade toward a flashed peripheral stimulus, on other trials to make the saccade in the direction opposite the target. Activity of many DLPFC neurons was modulated by stimulus location (i.e. whether the stimulus was in the left or right hemifield) and saccade direction (i.e. whether the saccade was to be made toward or away from the stimulus), suggesting that these neurons can potentially send task-relevant information directly to the superior colliculus. A further study examining the time-course of neural interactions between areas involved in visual or visuo-motor processing was conducted by Moore and Armstrong (2003). These authors recorded activity of neurons in area V4 while visual stimuli were presented in their receptive field and frontal eye field (FEF) cells were simultaneously stimulated with weak electric current. Electrical stimulation of FEF cells normally leads to the spontaneous execution of a saccade in a specific direction (Schall and Thompson, 1999), but the authors prevented this by using low current stimulation. They found that stimulation of retinotopically corresponding sites within the FEF enhanced V4 responses to the stimulus, while stimulation of non-corresponding sites attenuated V4 activity. Importantly, the authors were able to exclude a direct connection between stimulated FEF sites and V4 neurons, suggesting that activity of the latter was modulated indirectly, perhaps through the ‘top-down’ activation of a network that controls visual responses in the ventral stream. Given the strong connections between the FEF and the PPC, and the strong implication of these areas in saccadic eye movements, it is probable that this network includes the PPC.

In sum, these findings suggest that the PPC sends direct top-down attentional signals to extrastriate visual areas and thus modulates activity of these areas while the prefrontal cortex sends task-related signals to other structures involved in visual and saccadic processing. This role of the prefrontal cortex is consistent with the role of this area in action planning. Thus, the prefrontal cortex transforms external (e.g. experimental instructions) or internal (e.g. motivations) requirements on the organism into a representation of target stimuli, maintains an internal representation of the action target in working memory until an adequate
response has been formulated, and protects this representation against distracting influences (Fuster, 1997, 2001; Miller and Cohen, 2001).

**A posterior parietal saliency map**

The previous sections presented evidence for the modulation of feature representations by spatial attention and the role of the PPC and prefrontal cortex in this modulation. Several findings discussed above also suggest that various visual areas might in some way represent visual saliency. This finding has led to two divergent views of saliency representation in the brain. On the one hand, saliency is seen as an emergent property from distributed representations within the visual system (Treue, 2003). On the other hand, some dedicated brain areas might have specialized in representing saliency (Maunsell and Treue, 2006). Thus, where exactly is the saliency map (or maps) located? Before attempting to formulate an answer to this question, we should first consider the properties that characterize a saliency map. Unfortunately, there seems to be no consensus regarding this question. For example, Li (2002) defines a saliency map by its ability to represent saliency irrespective of the actual feature (e.g. the saliency of a red spot at one location could be compared with that of a black vertical bar at another location) and its dependence on the organization of a visual scene. The author concludes that such a saliency map is located in V1. According to Mazer and Gallant (2003) the saliency map is retinotopically organized, receives input from visual areas capable of representing fine spatial details, and necessarily has strong projections to the oculomotor system. Since V4 fulfils these requirements, the authors conclude that V4 represents saliency of visual stimuli. Similarly to Li (2002), Fecteau and Munoz (2006) propose that a saliency map represents the distinctiveness of stimuli independently of the actual feature. In addition, since saliency is computed from feature inputs the saliency map should receive inputs from the ventral visual pathway. Finally, lesions to the saliency map should produce deficits in attentional selection. According to Fecteau and Munoz (2003, 2006) structures of the oculomotor network (in particular: the FEF and superior colliculus) fulfill these criteria.

In my opinion, none of these definitions define a saliency map that is capable of resolving the criterion problem satisfactorily. I therefore propose seven criteria that specify the saliency map: first, in accord with the above authors saliency is coded independently of the actual feature represented. However, in order to be capable of representing saliency of stimuli presented in different modalities, the map is also able to code the environment independently of modality. Third, in contrast to Mazer and Gallant (2003), spatial coding in the saliency map is not be exclusively retinotopic, but also has access to head-centred and body-centred reference frames. These different spatial frames allow for a constant highlighting of locations by spatial attention irrespective of eye and head movements. Fourth, the saliency map receives input from feature maps computed in the ventral
stream as well as from frontal regions representing task-constraints. Fifth, the saliency map has strong outputs to these ventral regions and is therefore able to bias competition between single features. Sixth, representations in the map are subject to modulation by attention. Finally, the map codes stimuli independently of the required response (eye movement, hand movement, or other). This more precise definition of the saliency map discards several regions as possible candidates for the primary saliency map. Thus, the striate cortex (V1: Li, 2002) is not a likely candidate for a saliency map defined in this way: it exclusively represents visual information (violation of criterion 2) in retinocentred coordinates (violation of criterion 3), and it is not the origin of attentional modulation in ventral visual cortex (violation of criterion 5). Similarly, the extrastriate cortex (V4: Mazer and Gallant, 2003) does not meet criteria 2, 3 and 5.

It is more interesting to follow the suggestion by Fecteau and Munoz (2003, 2006) that the saliency map is localized in brain regions implicated in oculomotor processing. These regions involve (among other) the FEF, LIP and the superior colliculus. It is not too surprising that some authors consider these regions as the neural substrate of the saliency map. They have a very close relationship with the attentional network, and their underlying neural circuitry has been well identified. The FEF are directly involved in computing motor commands necessary to produce visually-guided eye movements. Many FEF cells have visual responses, and these responses are enhanced if the stimulus is the target of a saccade (Bruce and Goldberg, 1985; Schall and Thompson, 1999; Wurtz and Mohler, 1976a). The FEF contain distinct populations of neurons that are active during fixation (fixation neurons) and neurons that become active shortly prior to a saccade (saccade neurons, Munoz and Everling, 2004). These two cell types represent a neural balance through mutual inhibitory interactions: when fixation cells are more active than saccade cells, fixation is maintained. Conversely, eye movements are released when the activity of saccade cells is higher than the activity of fixation cells. In addition to these characteristics, FEF cell activity is subject to modulation by attention (Buschman and Miller, 2007; Fecteau, Bell and Munoz, 2004) and is sensitive to action goals. For example, in a conjunction search task, FEF neurons exhibited higher activity when a distracter that shared a feature with the target fell in their receptive field compared to when the distracter shared no feature (Bichot and Schall, 1999). These results suggest that the FEF contain a map in which visual stimuli are coded with respect to their behavioural relevance. However, in contrast to area LIP, FEF neurons only show activity related to visual stimuli and their activity is modulated only by oculomotor responses (Colby and Goldberg, 1999). The same is true for the superior colliculus (SC), which is an important component of the saccadic system. The SC has a laminar organization with the superficial layer mainly showing visual responses while the intermediate layer contains cells that have combined visuomotor responses (Mohler and Wurtz, 1976). Activity of
intermediate layer cells reflects capture of attention by spatial cues that bias attention toward one hemifield (Fecteau et al., 2004). A characteristic of neurons in the superficial layer is that they respond more strongly to a visual stimulus in their receptive field when this stimulus is used as target for a saccade (M. E. Goldberg and Wurtz, 1972; Wurtz and Mohler, 1976b). McPeek and Keller (2004) reported that following functional inactivation of the SC, monkeys had difficulty differentiating between irrelevant distracters and the saccade target. A study by Johnston and Everling (2006) has shown that the DLPFC sends task-relevant signals directly to the SC, providing evidence that the modulation of SC activity is due to top-down influence from the prefrontal cortex. However, this enhancement by task relevance is dependent on the type of response and does not occur prior to a hand movement. Thus, a possible role of the SC is in selecting visual targets for foveation based on top-down task-relevant signals (Ramat, Leigh, Zee and Optican, 2007). Interestingly, a significant proportion of SC cells respond to combinations of visual, auditory, or somatosensory inputs (Meredith and Stein, 1986; Stein and Meredith, 1993), suggesting a role of this structure in multimodal integration. These findings indicate that the SC has some, though not all, functional characteristics that define a saliency map: spatial coding in the SC is retinotopic, activity of neurons is exclusively saccade-related, and due to the lack of anatomical connections, the SC is definitely not the origin of attentional modulation in the ventral visual stream.

None of these limitations apply to the PPC, in particular area LIP. We have seen in the preceding sections that LIP neurons have access to spatial representations in retinotopic, head-centred and body-centred reference frames. Activity of these cells shows a high degree of independence from the specific feature input, and a significant proportion of them react to auditory and visual input. LIP receives strong input from extrastriate visual areas and the prefrontal cortex (including the FEF) and sends projections back to these areas. I have summarized evidence that activity in area LIP is modulated by attention, and that this area might be the origin of attentional modulations of ventral visual areas. Finally, LIP activity appears to be modulated independently of the required response.

In addition, even more direct evidence exists for the representation of saliency in the PPC. Gottlieb et al. (1998) examined responses of LIP neurons in a task requiring monkeys to make a saccade toward a stimulus matching a visual cue. The monkeys first fixated a point outside a circular array containing eight stimuli (Figure 11). Then a cue appeared that matched one member of the array, and the monkey made first a saccade to the centre of the display (which brought one of the eight stimuli in the receptive field of the cell), and then to the cued stimulus. When the first saccade brought the saccade target in the receptive field, the cell responded strongly to the stimulus. In contrast, a much weaker response was registered when a stimulus fell in the receptive field that was not the saccade target. In control experiments, the authors could exclude that the effect was due
to activity related to the saccade toward the target stimulus. Thus, only the target status of the stimulus decided whether the cell would respond to it, suggesting that LIP neurons selectively code the behavioural relevance of stimuli. Similarly, responses to behaviourally relevant stimuli were registered in area 7a. Neurons in this area respond strongly when a target stimulus, but not when a neutral stimulus is presented inside their receptive field (Constantinidis and Steinmetz, 2001). The build-up of activity suggests that under some circumstances PPC neurons actively represent action targets and do not merely reflect a selection process completed in the prefrontal cortex (Gottlieb, 2002). A recent study has cast light on the respective contributions of the prefrontal cortex and the PPC to attentional selection and representation of the target. Buschman and Miller (2007) have had monkeys perform a search task in which the target either ‘popped up’ and was quickly found or had to be searched for serially because of its similarity to the distracters. The authors simultaneously recorded activity of LIP, FEF and lateral prefrontal cortex (LPFC) neurons and found that in the pop-out task LIP activity predicted the location of the target ~80 ms prior to the LPFC and ~120 ms prior to FEF activity. In contrast, in the serial search task this pattern was reversed, with LIP being the last area to indicate target location. These results show that, when search must be organized and performed serially in order to detect tiny features that differentiate the target from distracters, target selection is indexed first by neural activity in frontal areas. In contrast, in the pop-out task, in which performance is determined by large saliency differences between the target and the distracters, target selection occurs first in LIP. This role of the LIP was confirmed in a study examining LIP responses to overtly ignored popup stimuli. Ipata and collaborators (Ipata, Gee, Gottlieb, Bisley and Goldberg, 2006) found that neural responses of LIP neurons to pop-out distracters was suppressed in comparison to ignored non-pop-out distracters, though in both cases the response of the monkey was directed toward a target stimulus. The active ignorance of a salient stimulus that automatically captures attention thus necessitates suppression of activity in the LIP, supporting the importance of this area for bottom-up attentional responses.

Summary

This review of the relevant neurophysiological literature presents several arguments favouring the SBAS selection model. Studies of the ventral visual stream show that representations of visual stimuli in areas V2, V4, TEO and IT are subject to important modulation by spatial attention. These areas constitute the neural site where a representation of the environment in terms of multiple feature maps is elaborated. Attentional modulation of these representations is space-based or feature-based, and recent studies suggest that a dorsal network is responsible for these modulations. This network comprises strongly interconnected regions of the prefrontal cortex (including the DLPFC and the
FEF) and the PPC (in particular areas LIP and 7a). In comparison to the FEF and SC, which represent the saliency of visual stimuli that may act as targets for eye movements only, the PPC represents saliency independently of whether an eye movement, a covert attentional shift or a hand movement is planned. The PPC also fulfills several other criteria – such as featureless and modality independent coding, strong connections with prefrontal areas important for task-setting and connections with visual areas in the ventral stream, strong modulations of activity by spatial attention, and coding of space in several coordinate frames – demonstrating that this area is the best candidate for the representation of stimulus saliency.
4 - COGNITIVE MECHANISMS OF ATTENTIONAL SELECTION

The scientific study of attention makes a fundamental distinction between two modes of attention. William James (1890/1950) called these modes ‘active’ and ‘passive’ attention, other scientists called them ‘stimulus-driven’ and ‘goal-directed’, ‘automatic’ and ‘controlled’, ‘reflexive’ and ‘voluntary’, or ‘bottom-up’ and ‘top-down’ attention. Though this basic distinction has early been made, the first systematic studies of attention focused on the question whether attention selects stimuli prior to the processing of their semantic features (early selection) or whether attention interferes later, after semantic characterisation of the stimulus (late selection, Broadbent, 1954). The first studies examining bottom-up and top-down attentional control were spatial cueing studies performed by Eriksen and Hoffman (1972, 1973). Since then, more than thirty years of research have been spent on the study of attentional control processes. Many studies using paradigms such as visual search, spatial cueing, abrupt onsets, or oculomotor exploration are relevant for the SBAS selection model, and these studies will be discussed in this chapter. It starts with a discussion of results of the vast visual search literature.

Visual search and stimulus saliency

Visual search is one of the most studied paradigms in the cognitive study of spatial attention. The most important work in this domain has been accomplished by Anne Treisman, who performed several influential studies and developed a theory that proposes a solution to the binding problem. The binding problem can be described as follows: the visual cortex contains many areas that are highly specialized in representing basic shape information, colour, motion, or other features. However, we do not perceive these single features in isolation, but

Figure 12. Ideal search slopes in parallel (feature) and serial (conjunction) search.
always as parts of an object. In other words: our brain must have developed a mechanism that binds the distinct features together to form coherent representations of objects (Luck, 1994; Luck and Ford, 1998; Reynolds and Desimone, 1999; L.C. Robertson, 2003; von der Malsburg, 1995). To give an account of this mechanism, Treisman (Treisman and Gelade, 1980; Treisman and Schmidt, 1982) developed the feature-integration theory (FIT). She and her collaborators proposed that there is an early stage of perceptual processing at which separate features are coded independently of each other and in distinct brain regions. Features are encoded automatically, outside of attention (pre-attentively), and in parallel in distinct ‘feature maps’. However, there is no conscious access to these feature maps, otherwise the visual environment would appear as a collection of disembodied features. Therefore, at a second stage, features are integrated into meaningful objects that can be consciously perceived.

According to Treisman and collaborators, feature-integration necessitates focal attention. At any given moment in time attention selects a location within a ‘master map of locations’ that contains information regarding the spatial position of a subset of features. The master map represents the stage at which spatial attention (Treisman refers to the spotlight metaphor) can be deployed, and that therefore contains a representation of space that is accessible to consciousness.

Based on the assumption that attention is only required when features are conjoined, this model makes two easily testable predictions: the first is that processing of single features is not affected by attentional limitations, hence that features are processed in parallel across the visual field. The second is that perception of feature conjunctions is severely limited by attentional restrictions, hence that conjunctions are processed serially. Parallel and serial processes in visual search predict distinct patterns of reaction times when the number of distracter items is systematically increased (Figure 12). Because each item is examined individually, serial search is characterized by a linear increase of detection times with increasing numbers of distracters. The degree of increase in detection time is often expressed as the slope of a linear regression calculated as a function of the number of distracters. A positive slope that significantly deviates from zero suggests a significant relationship between search times and the number of distracters. A second characteristic of serial search is the search slope for target-absent trials, that is the time that subjects need to say that the target is not present. This search slope is approximately the double of the target-present search slope, because a present target is on average found after half the items have been scanned while all items have to be examined for a correct target-absent decision. In contrast to serial search, parallel search results in a search slope that is not significantly different from zero, indicating that the number of distracters does not affect the search times. In addition, the search times for target-present and target-absent trials are comparable.

Treisman and Gelade (1980) tested the theory using different visual search tasks and showed that these basic predictions were correct. For example, when
Subjects were asked to search for a blue letter among brown Ts and green Xs, the slope of the regression line was only 4 ms per item, indicating that detection times did not depend on the number of distracters. In contrast, when the target was defined by a conjunction of features (green T among brown Ts and green Xs), the slope was 29 ms when the target was present, and 67 ms when it was absent. Thus, when feature search and conjunction search are compared, distracters have a very different effect on search performance, though in both conditions the same distracters are present. Similar search asymmetries were observed when the target was a Q among Os compared to when the target was an O among Qs (Treisman, 1988; 1998, see also Figures 1 and 3). In the first case search is fast, because the task can be solved by detecting the presence of the disjunctive feature (the tail in the Q) that characterizes the target. This search task can be realized without the need to bind the single tail in the display to any of the items. On the other hand, finding an O among Qs requires locating all the tails and binding them to the Qs. According to FIT, a target defined by the presence of a specific feature (i.e. a feature that is only present in the target, but not the distractors) will quickly be detected, because specific detectors of that feature will only become active at the location where the feature is present. In contrast, when the target is defined by the absence of a feature (i.e. the feature is only present in the distractors), feature detectors will be activated throughout the visual field, and the activity generated by the target will be drowned in the resulting noise. Comparing search for the presence with search for the absence of a feature, Treisman and Gormican (1988) presented evidence that certain feature primitives are processed pre-attentively, while others necessitate serial processing (Figure 13). Thus, line curvature, deviation from a prototypical colour or shape (a magenta target in red distracters; an elliptical target in circle
distracters), or closure may result in fast search, suggesting that these features may participate in pre-attentive perceptual processes such as texture segregation and figure-ground grouping. FIT does not only make predictions regarding the speed of visual search for feature and conjunction targets, but also postulates specific deficits that appear when binding by attention is hindered, for example if a densely populated display is presented for very short time (Treisman, 1996). According to FIT, in such a situation, binding errors in the form of illusory conjunctions should be observed. For example, in an experiment of Treisman and Schmidt (1982) subjects were shown for less than 200 ms displays consisting of two black peripheral digits and three coloured central letters (Figure 14). On each trial, participants reported the two peripheral digits (primary task) and thereafter the locations and colours of the central letters (secondary task). The aim of the primary task was to bind attention to the peripheral items and to prevent that the central letters were processed attentively. The fact that attention was withdrawn from the central items resulted in many identification errors and recombinations of features on the secondary task (see Figure 14). Subjects made many illusory conjunctions with mislocalization of one or both reported features. Importantly, illusory conjunctions were more frequent than the report of features that were not present in the display, supporting the view that illusory conjunctions result from a failure to bind features to locations.

FIT is probably one of the most influential theories in cognitive psychology, and an impressive amount of experimental findings has provided support for the theory. Nevertheless, several questions remain unclear. The first is whether feature maps represent spatial information. Clearly, even at the level of simple features, at least a rudimentary representation of spatial coordinates of features must be elaborated, otherwise texture segregation, pre-attentive grouping and the perception of optic illusions (Driver and Baylis, 2000; Pomerantz and Lockhead, 1991; Rock, 1986) could only be achieved in the presence of attention. According to Treisman and Schmidt (1982), features are organized in own spatial maps, but these spatial representations remain pre-attentive, and are therefore not accessible to conscious awareness.

<table>
<thead>
<tr>
<th>Error type</th>
<th>Example</th>
<th>Answers per trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>Middle, brown X</td>
<td>0.44</td>
</tr>
<tr>
<td>Illusory conjunction</td>
<td>Left, green N</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>Right, brown S</td>
<td>0.06</td>
</tr>
<tr>
<td>Single feature correct</td>
<td>Right pink</td>
<td>0.49</td>
</tr>
<tr>
<td>Single feature wrong</td>
<td>Something blue</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Figure 14. Example display and results of an experiment examining illusory conjunctions between shape and colour (adapted from Treisman and Schmidt, 1982).
An interesting aspect of the theory is that parts of it are not only relevant to visual attention, but also make predictions about perceptual processes. Thus, many results from the visual search literature suggest that perceptual features are processed in parallel, but that more complex selection processes are subject to attentional restrictions. This dichotomy has been implemented in several two-stage models of visual attention (Desimone, 1999; Duncan and Humphreys, 1989; Wolfe, 1994), including the SBAS model. Among these, FIT is the only model predicting that activity in feature maps is not accessible to consciousness. Single features can only be accessed consciously when they are bound to a specific location. In fact, this assumption appears to be correct only for some features (such as shape), while under conditions of attentional restriction some ‘pre-attentive’ properties may be accessed (such as redness) without the possibility to assign this property to a specific object (Wolfe and Bennett, 1997).

However, a major critique of FIT concerns the assumption that feature search and conjunction search are respectively accomplished with parallel and serial mechanisms. Sagi and Julesz (1985) found that the time to detect and to localize target lines that differed from the distracter lines only by their orientation was independent of the number of targets in the display. However, when observers were asked to make a discrimination judgement (whether all target lines had the same orientation), their decision time increased linearly as a function of the number of targets. Thus, in contrast to the prediction of FIT, a localization judgement did not depend on a serial process. In the same vein, Duncan and Humphreys (1989) argued that under some conditions, FIT predicts serial search, yet search is independent of the number of distracters. For example, according to FIT searching for a target L among Ts should require serial attention. Since both letters contain the same strokes, and only the different arrangement of these strokes differentiates them, individual feature maps for horizontal and vertical strokes cannot make a distinction. However, the authors found that search slopes were compatible with parallel search (with a maximum of 6 ms per item), independently of whether distracters were homogenous (all distracters were either upright or 90° rotated Ts) or heterogenous (a mixture of upright and rotated Ts). When distracters were rotated 180° and 270°, and thus the similarity of target and distracters was increased, a strong effect of distracter heterogeneity emerged. Similarly, Quinlan and Humphreys (1987) found that search slopes for a conjunction were much steeper when the target shared two features with each of the distracters than when it only shared one feature. Duncan and Humphreys (1989) concluded that there was no clear indication for a dichotomy between serial and parallel search modes. Instead, they proposed that the ease of visual search was determined by an interaction between two factors: the similarity between target and distracters and the similarity between distracters. According to this model, visual search is hardest when the target and distracters are highly similar and distracters are highly dissimilar. Conversely, search is easy when that target and distracters are dissimilar and distracter similarity is high. These simple
principles explain many of the effects found in the visual search literature without referring to qualitatively different search processes (note, however, that very difficult search tasks might require a serial component, Bricolo, Gianesini, Fanini, Bundesen and Chelazzi, 2002).

From this description of FIT, an important question arises: what differentiates the SBAS model from FIT? Clearly, both models have many points in common. Both are two-stage models, comprising a feature stage at which features are extracted in parallel and outside of attention, and an integration stage at which feature information is integrated. However, the saliency map in the SBAS model has several features that are different from the master map of FIT. For example, space is coded in retinotopic, head-centred and body-centred coordinates in the saliency map, while spatial coding in the master map has not been specified. A more important difference is that saliency map and feature maps in the SBAS model have important task-selective inputs (comparable to the Guided Search model of Wolfe, 1994), whereas FIT is a pure bottom-up model. Computation of saliency in SBAS depends on the similarity between stimulus inputs, with similar items leading to similar degrees of activation and mutual inhibition at the level of feature maps (compare Figure 7). Difficult search (that is, search characterized by steep search slopes) is explained by the fact that activity of the target stimulus at the level of the saliency map is similar to activity evoked by distracters. According to this account, differences between feature search and conjunction search are explained by the amount of noise produced by the distracters, and there is no need to refer to parallel and serial search processes. Finally, the SBAS model has a sound basis in neurophysiological findings of attentional mechanisms in the brain.

Attentional set and the capture of covert attention

The results of the visual search literature support the two stages of processing of the SBAS model. I now turn to the cognitive psychological literature that supports another idea implemented in the model. It is the idea that behavioural goals, task-constraints and expectations constituting the attentional set of the observer influence the saliency of stimuli similarly to purely bottom-up factors such as perceptual conspicuity, and that the most salient stimulus captures spatial attention. I start with a description of the important results of studies examining the covert shifting attention.

The study of attention has traditionally distinguished between overt (shifts of attention accompanied by eye movements) and covert (shifts of the spotlight of attention in the absence of eye movements) aspects of attention. The classical paradigm used to examine covert attentional processes is the spatial cueing paradigm, a variant of which is shown in Figure 15. For example, Eriksen and Hoffman (1973) studied the dynamics of spatial attention by asking participants to identify letters briefly displayed on an imaginary circle. On each trial a cue
Saliency, Selection, and Spatial Neglect 41

was presented shortly prior to the onset of the target display, pointing to the letter that had to be reported. The authors varied the stimulus-onset asynchrony (SOA; the time between onset of a first and onset of a second stimulus) between cue and target and found that identification of the target letter was better with increasing SOA up to about 200 ms, where it reached asymptote. Similar cueing effects were observed by Posner (1980), who presented a cue at fixation either in the form of an arrow pointing left or right, or in the form of the letter X (neutral cue). Participants were instructed to react to a peripheral square, which was either at the location to which the arrow pointed (valid condition) or at the opposite location (invalid condition). Reaction times (RTs) were faster following a valid cue compared to a neutral cue (suggesting a benefit from valid cues), and shorter following a neutral cue compared to an invalid cue (suggesting a cost from invalid cues). When peripheral cues (boxes, one of which brightened prior to appearance of the target inside of it or inside of another box) were used, the cueing effects appeared to be even greater (Posner et al., 1982), and benefits of valid cues started at an SOA of about 50 ms between cue and target. Another property of the attentional spotlight was suggested by the findings of Hoffman and Nelson (1981). These authors showed participants target displays consisting of four letters presented on an imaginary circle. On each trial a small U-shaped.

Figure 15. Two classical variants of the spatial cueing task: peripheral cues (upper row) and central cues (lower row). Each variant shows a valid and invalid trial.
probe was flashed close to the target letter or close to one of three distracter letters, and participants were asked both to indicate whether the display contained the target letter and to judge the orientation of the probe. In both tasks performance was highest when target and probe were at adjacent positions, but fell close to chance when they were spatially separated. This study suggests that the attentional spotlight cannot simultaneously highlight two different positions in space.

The studies discussed so far suggested that covert attention is akin to a mental spotlight that highlights selected locations. For the present, let us accept this metaphor and consider some further characteristics of the attentional spotlight (Cave and Bichot, 1999). Attention does not only highlight locations, but also enhances processing of objects. The first indication for this conclusion comes from studies that examined whether cueing benefits could be observed in object discrimination (in addition to detection) tasks. Müller and Findlay (1988) reported that valid peripheral cues improved the ability to discriminate between different orientations of the target letter T. Duncan (1984; see also Vecera and Farah, 1994) showed participants two closely adjacent objects, each varying on two dimensions and asked them to make a simultaneous discrimination concerning two dimensions of the same object or two dimensions of different objects. Performance did not differ when participants made a discrimination of one or two dimensions of the same object, but it fell significantly when they judged dimensions of two different objects. These results suggest that attention did not select relevant locations, but relevant objects. A similar conclusion was reached by a study that directly compared the effects of within- and between-object spatial cueing (Egly, Driver and Rafal, 1994). The task in this study was to detect and react to a target presented close to the end of one of two rectangles. On every trial one end of a rectangle was precued, and the target stimulus appeared at this end (valid location and valid object), at the uncued end of the precued rectangle (invalid location and valid object), or within the uncued rectangle (invalid location and invalid object). The critical comparison was whether costs by invalid cueing were greater in the invalid location/invalid object compared to the invalid location/valid object condition. This was found to be the case, indicating that attention was not only influenced by spatial factors, but was also modulated by the presence of objects. An object-based component of attentional selection was also confirmed in a study by Kahneman, Treisman and Gibbs (1992), showing that attention was bound to an object by virtue of this objects history (see also Maljkovic and Nakayama, 1994).

A further finding of spatial cueing studies is that cueing benefits may be observed when the modalities of cues and targets are crossed. For example, auditory and tactile cues influence RTs to visual targets similarly to visual cues (Buchtel and Butter, 1988; Butter, Buchtel and Santucci, 1989), and these influences are reflected in physiological measures (Eimer, 2001; Eimer and
These results suggest a common mechanism for spatial attention that is independent of modality.

Overall, compelling evidence exists that attention enhances performance at cued locations relative to uncued (or invalidly cued) locations. However, it is less clear whether this enhancement is due to better perceptual processing (e.g. faster object identification), facilitation of target selection (e.g. through decrease of noise in the system), accelerated preparation of the response, decreased uncertainty, or some other variable. Several recent studies have suggested that attention directly affects perceptual processing. For example, attending to a specific colour in an apparent motion paradigm increases the perceived saliency of this colour in a manner similar to increasing its saturation (Blaser, Sperling and Lu, 1999). This finding is very important for the basic assumption of the SBAS model: it shows that the influence of top-down signals (attention) on the perceived saliency of visual stimuli is indistinguishable from a bottom-up (perceptual) modulation of the same stimuli. Yeshurun and Carrasco (1998) used a texture segregation task to examine whether spatial attention has an effect on spatial resolution. Participants decided whether a texture consisting of oriented lines contained a texture-target whose orientation differed from that of the background. Prior to appearance of the target display a peripheral cue was presented at the position of the upcoming texture-target or at a non-target position. Unlike in other visual tasks, performance in this task is better when the target appears in the visual periphery than when it appears at the fovea, where spatial resolution is highest. Because high spatial resolution is detrimental to performance in this task, it allows for testing whether attention enhances spatial resolution: the expectation is that performance is worse at cued compared to uncued locations, and this is exactly the result that the authors found. In another study, Yeshurun and Carrasco (1999) confirmed the resolution hypothesis using spatial resolution tasks that are better solved at visual field positions where spatial resolution is high (e.g. Landolt square or Vernier acuity). In all tasks valid cues increased spatial resolution in a manner similar to shifting the experimental display nearer toward the fovea. A newer study by the same group showed that spatial cues increase contrast sensitivity in a task requiring observers to make a forced discrimination about the orientation of a Gabor target (Pestilli and Carrasco, 2005). Together, these results indicate that spatial attention may enhance perceptual processing by tuning visual channels that are selectively responsive to different spatial frequencies (Blakemore and Campbell, 1969). They also suggest that covert attention has a similar function to overt eye movements, which bring an interesting object located in the retinal periphery onto the fovea: to enhance visual processing of peripheral objects prior to an eye movement.

The results of spatial attention studies have stimulated the development of several theories, two of which have been particularly influential. The first is the ‘zoom lens’ model of selective attention by Eriksen and St. James (1986). In one
experiment that was performed with the aim to seek support for this theory, the authors showed participants circular displays of eight letters. On each trial, one, two, or three adjacent letters were precued and participants were asked to indicate whether the cued letters contained the letter S or C. Most of the uncued letters of the display were neutral, but one was either compatible with the target letter (an S when the target was an S) or incompatible (a C). The distance of the compatible/incompatible letter to the target letter varied. Participants were slower to identify the target letter when the distracter letter was incompatible compared to when it was compatible. Interestingly, this perturbing effect of the incompatible letter decreased with increasing distance from the target letter, and this effect interacted with the SOA between cue and target onset. Thus, at an SOA of 50 ms an incompatible letter located 3 positions away from the target disrupted target processing, but had no effect at longer SOAs. This result is compatible with the idea that within about 100 ms following a cue attention constricts its focus around the cued area very similarly to a zoom lens. This model of spatial attention is particularly interesting in view of the proposal that attentional effects on responses of inferior temporal cells are akin to a decrease of the size of the receptive field and thus resemble a zooming-in around the relevant stimulus (Moran and Desimone, 1985). The zoom lens model is also supported by newer findings noted above that attention increases the resolution of information within its focus.

The second influential model is the three-process model elaborated by Posner and collaborators (Posner, 1980; Posner et al., 1982; Posner and Petersen, 1990). Posner distinguishes between three fundamental operations necessary for the covert orienting of attention to visual stimuli. When attention moves from one stimulus to another, it first has to be disengaged from the current location before moving to the new location and finally being engaged on the new stimulus.

Figure 16. Idealized patterns of performance in a spatial cueing task predicted for healthy participants, patients with a unilateral deficit in attentional disengagement, patients with an impaired move operation, and patients with impaired engagement of attention. The blue data line shows performance for the intact visual hemifield while the red line shows performance for the impaired side.
According to the theory, an impairment in any of these basic mechanisms results in a specific pattern of RTs in the spatial cueing paradigm shown in Figure 15. Prototypical patterns of normal and impaired performance are shown in Figure 16. In order to understand the rationale behind the idealized effects shown in Figure 16 it is important to note that the difference in performance between valid and invalid cueing is due to facilitatory influences of the valid cue and inhibitory influences of the invalid cue, relative to a ‘neutral’ cue. The neutral cue is a cue that either highlights both of the positions that can contain the target, or none of them. In the peripheral variant of the cueing task normal performance is characterized by comparable facilitation and inhibition effects. A deficit in the disengagement operation results in a difficulty to detect targets following a cue presented in the opposite hemifield and thus affects the inhibition effect by invalid cues. A difficulty to move attention is characterized by very small facilitation and inhibition effects. Finally, a deficit in the engagement of attention is manifested by very slow RTs to the target, be it cued validly or invalidly. Posner and collaborators showed that patients with Progressive supranuclear palsy (PSP), a neurological syndrome characterized by Parkinson-like motor deficits and characteristic oculomotor impairments, were much less influenced by valid or invalid cues than healthy participants (Posner et al., 1982). The authors therefore concluded that the brain structures associated with PSP (in particular, oculomotor centres located in the midbrain) are important for the movement of attention. A different pattern was found in patients with focal parietal damage: these patients showed a difficulty to detect and react to contralesional targets following the presentation of a ipsilesional cue – a pattern compatible with a disengagement deficit (Posner, Walker, Friedrich and Rafal, 1987; Posner, Walker, Friedrich and Rafal, 1984). Finally, patients with thalamic lesions were found to have similar facilitation by valid cues and inhibition by invalid cues as healthy participants, but overall their RTs to contralesional targets were severely slowed, suggesting a deficit in the engagement of attention (Rafal and Posner, 1987). However, only the disengagement deficit has been replicated by other researchers and appears to be characteristic of patients with spatial neglect. These results will be discussed in detail in the following chapter.

One of the most important findings of studies of spatial attention is that cueing benefits vary as a function of the SOA separating cue and target. Posner (1980) found that benefits of valid cueing were largest at SOAs of 100-150 ms, approached zero at SOAs of 250-300 ms and were negative at longer SOAs. This cost of valid cues at long SOAs reflects a difficulty to return attention to the location indexed by the cue, and therefore has been termed inhibition of return (IOR, Klein, 2000; Posner et al., 1985). Müller and collaborators (Müller and Findlay, 1988; Müller and Rabbitt, 1989) examined the time-course of cueing effects with peripheral and central cues in a target discrimination task. The cues were predictive, that is a target was three times as likely to appear at the cued position than at uncued positions. The authors found that with peripheral cues
performance builds up rapidly, but decreases with SOAs greater than 300 ms whereupon it is sustained at the same level. In contrast, with central cues performance builds up slowly and is then maintained at the same level. To explain these results, Müller and Findlay (1988) proposed a two-component model of spatial attention (Figure 17). The first component is automatic capture of attention by physical properties of peripheral cues, such as their sudden onset. This automatic orienting of attention toward a peripheral stimulus cannot be interrupted by competing processes (such as controlled processes or another abrupt onset stimulus), shows a fast build-up and a fast decay. Since in the automatic mode attention is driven by perceptual characteristics, it has also been termed *exogenous attention*. Automatic capture is followed by a period of controlled facilitation and active maintenance of attention at the cued position. This second component shows a slow build-up and may only be observed when valid cues provide a potential performance benefit (e.g. when the target is more likely to appear at the cued position than at an uncued position). Since in the controlled mode attention is actively deployed by the observer, it has also been termed *endogenous attention*.

Though the two-component model accounts well for the findings from cueing studies, it has provoked a theoretical controversy, which was mainly concerned with the proposal that the first attentional component reflects fully automatic capture of attention (Egeth and Yantis, 1997; Gibson, 1996; Gibson and Kelsey, 1998). Some studies have suggested that any singleton may capture attention (Killingsbaek, Schneider and Bundesen, 2001; Theeuwes, Kramere, Hahn, Irwin and Zelinsky, 1999), but it appears that abrupt onsets are the most potent attention capturing stimuli (Egeth and Yantis, 1997; Theeuwes *et al.*, 1999).
Jonides and Yantis (1988) examined whether abrupt onsets capture attention differently than other physical features such as contrast or colour of the stimulus. Participants viewed circular arrays of letters, one of which was characterized by abrupt onset, unique brightness or a unique colour. They then decided whether the array contained a specific target letter, which sometimes was identical with the unique letter. The results showed that with increasing numbers of distracters the detection time increased linearly in the brightness and colour condition, but not in the abrupt onset condition, where the number of distracters did not affect decision times. Thus, abrupt onsets appeared to capture attention automatically independently of the degree of noise in the display. However, a potential methodological confound precludes acceptance of this conclusion: the abrupt onset letter could potentially be the target and was therefore relevant for the task. If abrupt-onset stimuli capture attention automatically, they should do it regardless of the subjects’ intention. Yantis and Jonides (1990) examined this prediction by asking participants to react to a letter presented left or right of the fixation point, which was precued by an arrow pointing toward or away from the letter. The authors argued that if abrupt onset targets capture attention automatically, subjects should be as quick at identifying them with valid and invalid cues. However, they found that identification of abrupt onset letters was similarly affected by the orientation of attention as the identification of no-onset letters. Further, Folk, Remington and Johnston (1992) presented strong evidence that attentional capture depends on the attentional set of the observer. They asked participants to react to targets defined by abrupt onset (much like the peripheral cueing paradigm shown in Figure 15) or a single colour that were either presented without precueing, or precued with valid or invalid abrupt onset cues. Participants showed a significant interference by invalid cues only when the onset cue preceded an onset target, not when it preceded a colour target. Based on these results the authors proposed the contingent capture hypothesis, which states that attentional capture will be contingent on whether a stimulus shares a feature property that is critical for the performance of the task at hand. The findings of Folk et al. (1992) raise a question that challenges our intuitive knowledge about the functioning of attention: does any stimulus, whatever physical characteristic define it, capture our attention if we are not prepared to deal with a stimulus of its sort? In its extreme form the contingent capture hypothesis says no, and thus questions the postulated independence of pure bottom-up processes from attentional set. Though intuitively this conclusion seems implausible, several convincing examples have been presented in the literature on attentional capture that support this view. For example, Gibson and Jiang (1998) showed that an unexpected appearance of a coloured distractor after prolonged practice with monochromatic displays does not interfere with the task at hand. On the other hand, very rare stimuli fail to capture attention in visual search tasks, though they are defined as targets (Wolfe, Horowitz and Kenner, 2005). The most dramatic evidence for the accuracy of the contingent capture
hypothesis comes from the literature on inattentional blindness (Simons, 2000). In a now classic experiment, Simons and Chabris (1999) showed students the video tape of two basketball teams wearing black or white dresses, and asked them to count the number of passes. While the participants counted passes, an unexpected event occurred: in one condition, a woman holding an umbrella walked left to right in front of the basketball players. In a second condition, the unexpected event was an actor in a gorilla dress that stopped in the middle of the display, turned to face the camera, thumped its chest, and continued walking across the field of view. The results were amazing: 46% and 50% of all observers failed to notice the umbrella women and the gorilla, respectively. Clearly, this result is very counterintuitive and contrary to the common opinion of most people that unexpected events capture attention (Simons, 2000).

Despite these convincing findings favouring the contingent capture hypothesis, the debate concerning the existence of fully automatic capture of attention remains open. For example, some studies have shown that the presence of a singleton distracter defined by colour interferes with detection of a target defined by orientation, though the singleton is not contingent with the task demands (Theeuwes, 2004). Similarly, abrupt onsets capture the gaze of observers in a task requiring participants to orient attention toward a colour target (Theeuwes et al., 1999). Thus, attention may sometimes be captured by stimuli that bear no contingent relation with the task at hand, though the precise
conditions leading to such automatic capture have not yet been elucidated. Nevertheless, these results do not present an argument against the fact that many stimuli interfere with online attentive processing only because they have a feature that is relevant for the task at hand. For example, our recent results question the contention that abrupt irrelevant onsets capture gaze in a reflexive manner. We examined oculomotor capture with a paradigm similar to the one used by Theeuwes et al. (1999). Participants saw four coloured dots arranged on an imaginary circle around a central fixation cross (Figure 18). After 1200 ms three of the dots changed colour, while at the same time the fixation cross disappeared and a fifth dot appeared. The fifth dot (the abrupt onset distracter) was always presented in the same hemifield as the target, but in the opposite quadrant (i.e. when the target was in the upper left hemifield, the distracter was in the lower left hemifield and vice versa). 18 participants were asked to make a saccade from the centre toward the dot that had not changed its colour. Theeuwes et al. (1999) have shown that in this task participants make an important number of saccades to the irrelevant abrupt onset distracter. However, the contingent capture hypothesis predicts that the abrupt onset should only capture attention (and the gaze) when it has a relation with the task at hand. We therefore manipulated the relationship between the distracter and the target by presenting on some trials an abrupt onset that had the same colour as the target (contingent distracter), while on other trials it had the same colour as the distracters (non-contingent distracter). Figure 18 shows that participants indeed produced only about 70% saccades correctly directed to the target in the contingent distracter condition. However, this proportion rose to nearly 90% correct saccades when the colour of the abrupt onset was not contingent with the target colour. There was no difference between the visual hemifields. These results clearly support the contingent capture hypothesis and suggest that this hypothesis applies to tasks that require a manual response as well as tasks requiring saccadic responses.

**Attention and oculomotor performance**

Classic textbooks dealing with cognitive studies of attention do not systematically discuss literature that deals with eye movements. One reason for this is that many studies of attention examined covert aspects of attention and thus required participants to inhibit eye movements. The second reason is that in the early years of attention study, only few commercial eye tracking devices with sufficient spatial and temporal resolution necessary for the examination of fast shifts of the gaze were available. Given the strong functional and neuroanatomical link between attention and eye movements (Corbetta et al., 1998; Kustov and Robinson, 1996; Perry and Zeki, 2000), this negligence of the contribution of eye movements to spatial attention is regrettable (Findlay and Gilchrist, 2003).
The relationship between attention and eye movements may be described by one of the following rules: one possibility is that attention and eye movements co-occur, but are functionally entirely independent. This *independence hypothesis* predicts that in the extreme case, overt and covert attention are dissociable, that is attention may be shifted without shifting the eyes and the eyes may be shifted without shifting attention. The second alternative is the functional dependence hypothesis, which predicts that attention and eye movements are fully dependent, hence that every eye movement is accompanied by a movement of attention in the same direction, and vice versa. Finally, the *interdependence hypothesis* postulates that attention might be shifted without shifting the eyes (though in real-life situations this is rarely the case), but that eye movements and attention cannot dissociate (e.g. the eyes cannot move in one direction and attention at the same time in another direction). Remington (1980) asked healthy participants to execute saccades to a peripheral location indicated by the onset of a peripheral cue. On half of the trials and at different intervals following cue onset a barely perceptible probe stimulus was shown at fixation position or at position of the saccade target. Prior to the onset of the saccade subjects showed a significant advantage in detection when the probe was presented at the cued position. Remington concluded that attention precedes eye movements even if there is no incentive to move attention with the eyes. However, the problem with this interpretation is that the peripheral cue might have captured attention automatically. In order to avoid this difficulty, Deubel and Schneider (1996) asked participants to make saccades in the direction indicated by a central arrow cue to one of three peripheral targets. Prior to the onset of the saccade, a discrimination target was presented at the position of the saccade target or at a nearby position. Participants showed better performance when the discrimination stimulus was shown at the landing position of the upcoming saccade compared to nearby positions, suggesting that attention is allocated to the end position of an eye movement prior to commencement of the movement. Similarly, Shepherd, Findlay and Hockey (1986) observed enhanced detection of a peripheral stimulus when it was presented at the landing position of the saccade. In another study, Hoffman and Subramaniam (1995) established optimal conditions allowing subjects to separate attentional allocation from eye movements. Participants saw an array of four boxes centred on the edges of an imaginary diamond. On each trial an arrow indicated the most likely location of an upcoming target letter, which the participant was required to identify. Throughout a block of trials participants made saccades in the same direction, which was completely independent of the location of the target letter. Thus, the most sensible strategy was to attend to the precued location though this location was different from the location targeted by the eye movement. However, results showed that participants were not able to follow this strategy: their level of target identification was predicted by the direction of the saccade, not by the directional cue, suggesting that participants were unable to dissociate attention from the
concurrent eye movement. A further confirmation of the interaction between attention and eye movements was presented by Kowler et al. (1995). These authors showed participants circular arrays containing a single numeral among seven letters and asked them to execute a saccade toward the numeral or to the letter that was located opposite the numeral. They argued that if the singleton numeral captured attention and saccade planning was dependent on attention, executing saccades away from the numeral should be more difficult than making saccades toward the numeral. Indeed, participants made longer latency and less accurate saccades when they directed their gaze away from the numeral, confirming this prediction. Finally, several studies have reported that saccade scan paths measured in visual search tasks paralleled the deployment of attention measured as manual RTs in these same tasks (Findlay, 1997). For example, participants require only few (often none) saccades when searching for a feature target independently of the number of distracters, but an increasing number of
saccades in conjunction search (Williams, Reingold, Moscovitch and Behrmann, 1997; Zelinsky and Sheinberg, 1997). Thus, the number of saccades to find the target strongly correlate with measures of reaction time in feature and conjunction search tasks. In sum, together with the findings regarding covert attention presented in the preceding section, these results strongly argue in favour of the interdependence hypothesis: attention may be shifted without an eye movement, but eye movements are necessarily preceded by a shift of attention to the landing location of the upcoming gaze shift.

An important consequence of this functional interdependence between attention and saccadic eye movements is the conclusion that studying the scanpaths of saccades tells us something about the attentional processes involved in complex scene perception. When we explore a visual scene, our gaze shifts on average three times per second to a new location (Henderson and Hollingworth, 1999) and thus samples our environment in a highly dynamic way. Several studies attempted to determine the perceptual characteristics leading to the selection of an image region for fixation. Early studies showed that observers direct their gaze toward particularly informative regions. For example, participants preferentially fixate faces when asked to estimate the age of people (Yarbus, 1967), and fixation density during visual exploration is highest for regions rated as particularly informative (Antes, 1974). However, though observers appear to be largely consistent when evaluating the perceptual conspicuity of image regions (Cerf, Cleary, Peters, Einhäuser and Koch, 2007), it is difficult to define local informativeness independently of a subjective assessment. A more objective strategy is to study the perceptual distinctiveness of fixated regions by comparing local statistics of these regions with randomly defined image patches. This approach shows that healthy participants preferentially look at image regions characterized by high local contrast (Parkhurst and Niebur, 2003; Reinagel and Zador, 1999) and edge content (Tatler, Baddeley and Gilchrist, 2005; Tatler, Baddeley and Vincent, 2006), suggesting that the guidance of eye movements during active vision is driven to a significant degree by local perceptual characteristics of the scene.

In order to examine the local image content that is selected for fixation we asked 18 healthy participants to freely explore a series of images (Ptak, Golay, Müri and Schneider, submitted). Stimuli were 16 colour photographs showing natural scenes, architecture, man-made objects, or regular patterns. The photographs were presented on a 20” CRT and subtended 32° x 26° of visual angle. The horizontal and vertical position of the right eye was recorded with 240 data points per second and a spatial resolution of less than 0.3°. We computed local intensity, chromatic contrast, luminance contrast, and edge content of 1° x 1° image patches drawn around each fixation and compared these to randomly chosen patches. Figure 19 shows examples of fixation distributions of one participant and four different stimulus images. There was substantial variability
between the numbers of fixations and their distribution across the image. Despite this variability, fixated regions were characterized by significantly higher luminance and contrast, and contained more edges than randomly selected regions (Figure 20). This finding extends findings from similar studies (Parkhurst and Niebur, 2003; Reinagel and Zador, 1999; Tatler et al., 2005) and suggests that oculomotor exploration during scene viewing is significantly guided by local image content.

Results similar to these provide support for the bottom-up model of attentional selection elaborated by Koch and collaborators (Itti and Koch, 2000, 2001; Itti et al., 1998; Koch and Ullman, 1985; Parkhurst, Law and Niebur, 2002). The model attempts to reproduce attentional selection in a series of computational steps performed in a hierarchically organized system. At the lowest level of the hierarchy input images are decomposed in a set of early visual features (colour, intensity, and orientations), represented at different spatial scales. These spatial scales represent information at different spatial frequencies, akin to the operation of neurons of the visual system. Through the calculation of centre-surround differences between finer and coarser maps, a set of feature maps is obtained from the initial colour, intensity and orientation maps. The final step is the calculation of a saliency map by linear combination of the different feature maps. A normalization operator is applied to the saliency map in order to enhance the most salient location and thus to decrease spatial noise from distracters with similar, though slightly lower saliency. Attentional selection in the model is achieved by selecting the location with highest saliency. This model successfully reproduces human performance in search tasks where the target is defined by a single feature (Itti et al., 1998) or a feature conjunction (Itti and Koch, 2000). Further, it explains a significant amount of the variance that is inherent in ocular fixation patterns when participants observe natural scenes (Parkhurst et al., 2002). However, since the model is based on purely stimulus-
driven processes it has limited predictability for performance in tasks requiring
top-down guidance of attention (Torralba, Oliva, Castelhano and Henderson,
2006). Nevertheless, the combination of bottom-up (feature information) and
top-down guidance (task-related selection of stimuli) might provide a valuable
predictor of human performance in future computational models (Itti and Koch,
2001).

**Summary**

This review discussed the findings of three basic topics of research on
attentional mechanisms in healthy participants: visual search, spatial attention
and overt attention. The results of the rich visual search literature suggest that for
a variety of tasks, attentional selection operates at two stages: at the first stage,
attention processes features of objects in parallel and independently of their
number across space. Processing at this stage is fast, independent of the number
of distracting elements in the scene, and has nearly no limitation of capacity. At
the second stage stimuli are selected one-by-one, serially, with a limited-capacity
process. The first stage selects features, while the second stage selects feature-
compounds building visual objects. This two-stage architecture has been
implemented in several cognitive models of attentional selection, and it is also
the basis of the SBAS model proposed in Chapter 2.

The function of the second processing stage has been characterized in studies
examining covert attention. These studies have shown that the attentional
spotlight may select locations or objects, that it cannot be divided, and that it
affects perceptual processing by increasing spatial resolution, or enhancing
perceptual differences. Studies of attentional capture have provided strong
evidence favouring the importance of attentional set in selection processes. These
studies strongly suggest that top-down (task-induced) signals determine whether
an object captures attention or not, suggesting that such task-induced signals
affect the perceived saliency of visual stimuli.

The last section of this chapter reviewed relevant studies showing that
attention and eye movements are interdependent: attention may be shifted
without shifting the gaze, but any gaze shift is necessarily preceded by a shift of
attention in direction of the upcoming eye movement. Computational studies
similar to the SBAS model have been proposed to explain fixation patterns of
healthy participants observing visual scenes. These studies have shown that
purely bottom-up models may account for a substantial amount of variance in
exploratory eye movements, though they underestimate the impact of task-
related factors and exploratory strategies.
Studies with brain-injured participants may provide information about a specific function that cannot be gained by examining healthy participants. Neuropsychology is a young discipline, and it has gone (and is still going) through several important struggles about its methodological foundations (Caramazza, 1986; Caramazza and Badecker, 1991; Farah, 1994; L.C Robertson, Knight, Rafal and Shimamura, 1993; Shallice, 1988). Probably the major source of disagreement concerns cognitive architecture. How does the brain create the mind? Theoreticians have proposed at least two possible answers, both strongly inspired by computer science. One model proposes that information is represented and processed in distributed networks allowing the processing of large quantities of information in parallel (Bibby, 1992; Crick, 1989; Sutherland, 1986). Cognitive neuropsychologists have always been more concerned with the second model. It proposes that cognitive functions are represented in several, functionally and anatomically distinct processing units: modules. The classic description of modules has been formulated by Fodor (1983, 1985), who defined several characteristics of modules such as their mandatory operation (modules necessarily put out an answer when they are fed with information), informational encapsulation (information enclosed within modules cannot be accessed or processed by other systems), domain specificity (modules process only one type of information), speed (modules are fast), hardwiredness and common phylogeny (modules have a neural basis that is the same across individuals). The idea that cognitive functions are represented in distinct modules (though sometimes slightly differently defined than Fodor did) has always been central to cognitive neuropsychology (E. Goldberg, 1995). Probably the most important consequence of this thinking is that cognitive neuropsychology seeks to find associations and dissociations between functions (Coltheart, 2001). This approach allows describing dependencies between functions and specifying the cognitive architecture of mental faculties.

Any theory of a cognitive function should be able to explain cognitive impairments. The aim of the present chapter is therefore to examine whether the evidence from studies of deficits of visual-spatial attention are in agreement with the basic predictions of the SBAS model.

Spatial neglect: clinical picture and anatomy

Spatial neglect is a frequent result of large ischemic damage most often affecting the right cerebral hemisphere (Driver and Mattingley, 1998; Halligan and Marshall, 1993; Heilman, Watson and Valenstein, 1993; Kerkhoff, 2001; Rafal, 1998). Left neglect is characterized by the following clinical features: a) strongly diminished attentional orienting and exploration behaviour leading to the failure to report or to act to stimuli presented on the side contralateral to the
damaged hemisphere; b) a tendency to direct attention toward the more ipsilesional of two stimuli; c) a difficulty to detach attention from ipsilesional stimuli when it has to be moved contralesionally; d) omission of contralesional stimuli when asked to imagine objects or scenes; e) a difficulty to move the ipsilesional arm toward the contralesional space; f) generalized attentional deficits of alertness, and g) a deficit of spatial working memory. In addition, many neglect patients have associated deficits affecting reading, and constructive function: spatial dyslexia is a reading disorder characterized by the tendency to omit or transform the contralesional part of words (Arduino, Vallar and Burani, 2006; Làdavas, Umiltà and Mapelli, 1997; Miceli and Capasso, 2001). Writing and written calculation are often affected by the spatial deficits characterizing neglect. Constructive failures are characterized by the difficulty to respect the size, position, spatial arrangement, and two- or three-dimensional structure of different elements when asked to reconstruct complex drawings or 3D-models (Halligan and Marshall, 2001; Kerkhoff and Marquardt, 1998).

Neglect patients exhibit striking spatially lateralized deficits in everyday activities. They have difficulty with dressing (e.g. they fail to turn the T-shirt correctly), grooming (they fail to shave the left part of their face or to adjust correctly their glasses), eating (they fail to eat food located in the left half of the plate), or navigation (they bump into objects located on their left). Patients with spatial neglect have worse motor outcome and have greater impairment in activities of daily living compared to non-neglect patients (Denes, Semenza, Stoppa and Lis, 1982; Ptak, Bonvallat and Schnider, 2004; I. H. Robertson, Halligan and Marshall, 1993).

The exact locus of damage leading to this striking and debilitating disorder has been intensely discussed (Karnath, Fruhmann Berger, Zopf and Küké, 2004; Karnath and Himmelbach, 2002; Marshall, Fink, Halligan and Vallar, 2002; Mort et al., 2004). Early investigations, based on visual inspection of computerized tomography (CT) scans, suggested the IPL and the temporo-parietal junction (TPJ) as the brain regions most often damaged in patients with spatial neglect (Heilman, Watson, Valenstein and Damasio, 1983; Vallar and Perani, 1986). However, these studies may be criticized regarding inclusion criteria – in particular, neglect may have been confounded with visual field loss – as well as the absence of a comparison between patients with and patients without neglect. Nevertheless, many authors affirm that among the cortical structures involved in spatial orienting, the IPL plays a crucial role (Driver and Mattingley, 1998; Mattingley, Husain, Rorden, Kennard and Driver, 1998; L.C. Robertson, 2000; Vallar, 2001). It was therefore surprising, when in a recent study comparing 25 neglect to 25 control patients, Karnath et al. (Karnath, Ferber and Himmelbach, 2001) reported that the region best predicting the occurrence of spatial neglect was the middle part of the superior temporal gyrus (STG). However, these results were contested by Mort et al. (Mort et al., 2003), who objected that the authors had excluded patients with visual field loss, which
might have biased their analysis towards more anterior damage. However, Karnath et al. (Karnath, Fruhmann Berger, Küker and Rorden, 2004) confirmed their results in a second study in which patients with visual field loss were not excluded. These new results were based on a comparison between 78 patients with spatial neglect and 62 control patients. Mort et al. (Mort et al., 2003) performed high-resolution MRI on 14 neglect patients with vascular damage in the territory of the middle cerebral artery. In significant contrast to Karnath et al. (2001; 2004), they found that the region predicting best the occurrence of neglect was the angular gyrus. Thus, the controversy about the anatomical bases of spatial neglect remains open, and more studies are needed to clarify the question.

In order to examine neglect anatomy in an independent sample, we (Golay, Schneider and Ptak, in press) compared brain lesions of 28 neglect patients with 19 control patients with right hemisphere damage using the same approach as Karnath et al (2001; 2004).

| Table 1. Demographic and clinical description of the neglect and control patients. |
|-----------------------------------|-----------------|-----------------|
|                                   | Neglect         | Control         |
| Number                           | 28              | 19              |
| Sex (female / male)              | 16/12           | 8/11            |
| Age (mean)                       | 64.9 ± 13.9     | 69.3 ± 12.4     |
| Etiology                         | 18 infarction   | 15 infarction   |
|                                  | 9 haemorrhage   | 4 haemorrhage   |
|                                  | 1 tumour        |                 |
| Time lesion-exam (days, mean)    | 74.9 ± 77.9     | 79.2 ± 82.6     |
| Time lesion-scan (days, mean)    | 59.9 ± 112.9    | 56.3 ± 122.5    |
| Contralateral paresis            | 89.3%           | 84.2%           |
| Visual field defect              | 17.9%           | 5.3%            |
| Bells test (contralesional omissions) | 11.5 ± 4.3 | 0.9 ± 1         |
| Line bisection (% deviation)     | 11.7 ± 9.2      | 1.9 ± 1.9       |
| Drawing neglect (mean)           | 0.8 ± 0.9       | 2 ± 0           |

We included all patients with a recent unilateral brain injury, the only exclusion criteria being the presence of previous brain damage or absence of a visible brain lesion on MRI or CT. The presence of motor, somatosensory or visual field defects was assessed using standardized neurological examination. Patients were attributed to the neglect group if they a) showed at least some signs of contralesional unawareness in everyday actions (e.g. ipsilesional head and gaze deviation, difficulty with dressing or grooming due to personal neglect, unawareness of objects or people placed contralesionally etc.) and b) had scores indicative of spatial neglect in at least two of three formal neglect tests: a cancellation test (Gauthier, Dehaut and Joanette, 1989), line bisection
Saliency, Selection, and Spatial Neglect

Schenkenberg, Bradford and Ajax, 1980), and drawing from memory (Wilson, Cockburn and Halligan, 1987). Lesion mappings were based on manual transposition of lesions from individual MRI-scans or CT-scans into a template brain using MRICro software (Rorden and Brett, 2000). MRICro was used to generate images of superposed lesions, lesion subtractions, and voxel-wise comparisons. Table 1 presents the clinical characteristics of both groups and Figure 21 shows the results of the anatomical comparisons.

The region that was most frequently damaged in neglect patients involved the insula and its underlying white matter, as well as the posterior temporal and inferior parietal cortex (Figure 21). However, these regions were also frequently damaged in control patients. In order to identify regions that were more frequently damaged in neglect patients in relation to the control group, we subtracted the superimposed lesions of the control group from the lesions of the neglect group. This subtraction identified a region reaching from the insula into the white matter of the frontal lobe as predicting best the incidence of spatial neglect. The advantage of the subtraction technique is that only regions for which frequency of involvement is different in the neglect compared to the control group are highlighted. However, the technique does not provide a statistical test of frequency of lesion involvement. In order to determine statistically which specific regions were more frequently damaged in neglect than control patients, we performed a voxel-wise analysis on each voxel involved in at least one patient. In this analysis, for every voxel the hypothesis was tested whether it was involved significantly more often in the neglect group compared to the control group. There were essentially two clusters of voxels for which the test was statistically significant, an anterior and a posterior cluster. The anterior cluster reached from the inferior to the superior insular cortex and further into the white matter of the frontal lobe anterior to the horn of the lateral ventricle. Within this cluster a group of voxels differentiating best between both groups was situated slightly anterior to the head of the caudate nucleus (Talairach-coordinates: 21, 22, 14). The posterior cluster was located at the TPJ and reached into the supramarginal gyrus and the inferior postcentral gyrus. The STG was also involved, but only with its most posterior part neighboring at the TPJ. Within the posterior cluster, a group of voxels in the white matter beneath the supramarginal gyrus (Talairach-coordinates: 36, -43, 24) differentiated best between neglect and control patients.

Thus, our comparison between neglect and non-neglect patients revealed an anterior and a posterior cluster differentiating best between these two groups. Both, the IPL and the STG were damaged significantly more often in neglect than in control patients, confirming both the results of Karnath et al. (2001; 2004) and Mort et al. (2003). Nevertheless, these results should be considered with some caution. The lesion subtraction technique represents a form of statistical abstraction that may miss a critical feature of spatial neglect, namely the necessity of large lesions that affect several right-hemisphere regions
involved in spatial attention. Indeed, clinically severe neglect almost invariably necessitates large lesions affecting substantial parts of the territory of the middle cerebral artery, while isolated damage to TPJ does not result in severe forms of neglect.
Spatial extinction

The second attentional disorder that will be discussed in this chapter is spatial extinction. Patients with spatial extinction report single stimuli briefly presented in the ipsilesional or contralesional visual field. However, they fail to report a contralesional stimulus presented simultaneously with an ipsilesional stimulus (Denny-Brown, Meyer and Horenstein, 1952; Driver and Vuilleumier, 2001; Karnath, 1988). Spatial extinction may be observed within the visual, auditory, and tactile modality. Furthermore, extinction may occur when the ipsilesional stimulus is presented in a modality different from the contralesional stimulus (di Pellegrino, Lâdavas and Farnè, 1997; Lâdavas and Pavani, 1998). Thus, extinction represents a pathological selection bias in favour of the ipsilesional stimulus due to a diminished generalized processing capacity of the damaged hemisphere for the contralateral space (Duncan et al., 1997). While many extinction patients have slight difficulties to detect unilateral contralesional stimuli, some patients show an exceptionally pure form of the disorder with entirely intact detection of contralesional stimuli when they are presented alone, but impaired detection under bilaterally simultaneous presentation. Such a patient is SP, a 63 year-old woman, who suffered a stroke of the right middle cerebral artery (Figure 22). Extinction was tested with coloured circles presented for 150 ms at four different eccentricities (3, 6, 9 and 12 degrees). SP detected all unilateral targets in the left and right hemifield, but showed between 30 and 50% extinction of left targets under bilaterally simultaneous presentation (Figure 22). Such extraordinarily pure forms of extinction strongly support the view that extinction reflects attentional failure, rather than a low-level sensory impairment.
Some authors have suggested that extinction and neglect are two independent disorders (Smania et al., 1998), and that their anatomy is different (Karnath, Himmelbach and Küker, 2003). However, these studies cannot exclude that the failure to find extinction in some neglect patients is a result of the examination technique rather than the confirmation that extinction and neglect doubly dissociate. Indeed, the degree of extinction varies with the physical characteristics of the stimulus, and choosing a less salient stimulus might reveal a lateralized bias that wasn’t found with larger or brighter stimuli (Geeraerts, Michiels, Lafosse, Vandenbussche and Verfaillie, 2005).

**Pre-attentive and attentive processing in neglect and extinction**

Similarly to other two-stage attentional models the SBAS model has a hierarchical structure, with a first stage responsible for parallel processing of simple features and a resource-limited second stage that is involved in selection of spatial ‘objects’. Spatial neglect and spatial extinction provide strong support for this two-stage hierarchy. Let us start with the findings suggesting that neglect and extinction patients process stimuli to a considerable degree outside of awareness. Several studies have examined whether visual extinction is diminished when the ipsilesional and contralesional item are grouped by perceptual factors (Driver, Mattingley, Rorden and Davis, 1997; Driver and Vuilleumier, 2001). Probably the first attempt was reported by Luria (1959), who found that connecting two simple shapes with a line substantially diminished the degree of spatial extinction in a patient with simultanagnosia. Ward and collaborators (1994) observed better performance when patients were shown two brackets compared to when a contralesional bracket was presented paired with a dot. Gilchrist et al. (1996) observed that shape identity as well as stimulus polarity affected grouping both across and within a visual field. Further variables found to affect grouping are the axis of alignment (Boutsen and Humphreys, 2000), edge collinearity (Cooper and Humphreys, 2000; Humphreys, Romani, Olson, Riddoch and Duncan, 1994; Pavlovskaya, Sagi, Soroker and Ring, 1997) or embeddedness in a larger figure (Farah, Wallace and Vecera, 1993), even if the latter is produced by illusory contours spanning over the visual fields (Mattingley, Davis and Driver, 1997).

We had the occasion to test the effect of grouping by symmetry in a patient with severe visuo-spatial deficits and spatial extinction following bilateral parietal-temporal damage. The patient PS suffered from an ischemic attack affecting the superior parietal lobe bilaterally as well as the left temporal lobe (Figure 23). At the time of the study he exhibited severe visual confusion by irrelevant visual information (e.g. he was unable to identify the picture of a key that was masked with a crosshatched pattern), moderate spatial neglect and severe left extinction. We asked PS to indicate the number of square brackets shown for 50 ms left and right of fixation, or bilaterally. In the bilateral condition
Saliency, Selection, and Spatial Neglect

brackets were either both facing left or right (repetition condition), or both facing inwards or outwards (symmetry condition). PS showed significantly less extinction of symmetrical compared to repeated items (Figure 23). This finding is in accord with the role of symmetry in perceptual grouping processes. Symmetry is an almost exclusive property of figures, and is known as a powerful perceptual factor involved in figure-ground segmentation (Rock, 1986). Its importance in figure parsing predestines symmetry as a factor reducing contralesional extinction in neglect patients. Further observations support this suggestion. Driver et al. (1992) described a neglect patient who, from a pattern of alternating symmetrical and asymmetrical columns consistently chose the symmetrical columns as “figures”, identifying the asymmetrical columns as “ground”. The fact that the patient neglected the left side of the figures when trying to copy them suggests that structural parsing by symmetry occurred at a pre-attentional processing stage. Together with the results of patient PS, these findings suggest that perceptual grouping factors such as symmetry enhance pre-attentive visual processing by binding the contralesional to the ipsilesional item or by increasing the contrast between figure and ground.

In addition to these examples of bottom-up grouping other studies found evidence for semantic processing of extinguished stimuli. Volpe, LeDoux and Gazzaniga (1979) reported that extinction patients were able to indicate whether two simultaneously presented stimuli were identical or different, despite being unable to name the contralesional stimulus. Though this result might reflect residual visual processing sufficient for a perceptual comparison, but insufficient for identification (Farah, Monheit and Wallace, 1991), other evidence supports the claim of the authors that some semantic features of extinguished stimuli are perceived and may influence semantic judgements. For example, an extinction

**Figure 23.** MRI of extinction patient PS showing bilateral parietal and left inferior temporal damage. Conditions of the extinction experiment are shown in the middle and the results of item detection in the left hemifield are shown on the right.
patient examined by Berti et al. (1992) was able to decide whether two simultaneously presented objects had the same name even when they were shown from different viewpoints or when different exemplars of the same object were presented. Similarly, several studies have reported semantic processing of extinguished verbal stimuli (letters and words, Kumada and Humphreys, 2001, Kanne, 2002 #1548). Another set of studies showed that semantic factors may determine whether a stimulus reaches awareness or not. Ward and Goodrich (1996) showed that pictures of intact objects are less subject to extinction than scrambled images of the same objects. Similar effects were observed with stimuli such as arrows (Ward et al., 1994), faces (Vuilleumier, 2000) or fear-inducing stimuli (Vuilleumier and Schwartz, 2001), suggesting grouping between the ipsilesional and contralesional stimulus by high-level semantic processes. Finally, Riddoch et al. (2003) reported diminished extinction when the ipsilesional and contralesional stimulus grouped together by common utilization (e.g. a corkscrew and a bottle) and were displayed in the configuration in which they are generally used (compare also di Pellegrino, Rafal and Tipper, 2005).

Similar to extinction studies, studies with neglect patients have revealed a substantial degree of preservation of pre-attentive processing (Driver and Vuilleumier, 2001). In a classic study Marshall and Halligan (1988) presented to a patient with severe left neglect drawings of houses, in some of which bright red flames emerged from the left or right side. On each trial the patient was shown a pair of two houses and was asked in which she would prefer to live. Though she indicated seeing no difference between the burning and the intact house, the patient consistently preferred the intact house. Similar results were observed by Doricchi and Galati (2000) using a larger set of stimuli (e.g. a broken cup). As noted above, Driver et al. (1992) reported that symmetry positively influenced the distinction between figure and ground in a neglect patient. Finally, several studies examining patients with spatial dyslexia have shown that patients have access to semantic information of words they are unable to identify explicitly (Làdavas, Paladini and Cubelli, 1993; Làdavas, Shallice and Zanella, 1997; Làdavas, Umiltà et al., 1997).

Given these converging findings suggesting an important degree of preservation of pre-attentive processing in extinction and neglect the question emerges whether this processing may be related to the recruitment of brain regions that are also involved in attentive (conscious) processing. To answer this question, Rees et al. (2000) used event-related fMRI to measure brain activity in an extinction patient to whom drawings of different objects were shown in conditions leading to correct detection or extinction. The authors found that extinguished stimuli and consciously perceived stimuli activated similar striate and extrastriate regions in the damaged right hemisphere. Similarly, Vuilleumier et al. (2001) found that extinguished faces activated striate and inferior temporal cortex. However, these authors also observed that conscious perception activated
a larger network including V1, bilateral temporal and parietal cortex suggesting that the recruitment of additional regions is necessary for conscious perception.

In sum, these experimental and functional imaging findings strongly suggest that pre-attentive processing in neglect and extinction is preserved to a significant degree. However, does this imply that pre-attentive processing is sufficient to allow parallel processing of visual features across the visual fields, similarly to what has been shown in healthy participants searching for a feature target (Chapter 4)? This question cannot be answered using the extinction paradigm, because in this paradigm almost always only one stimulus is presented in the impaired hemispace. Studies using the visual search task are therefore better suited for the evaluation of parallel processing of distributed features. Indeed, several studies reported preserved parallel processing in patients with spatial neglect asked to search for a feature target. Eglin, Robertson and Knight (1989) found that visual search of neglect patients for a target defined by a single feature was unaffected by the number of distracters and that, as in healthy participants, conjunction search was characterized by search slopes compatible with serial search. However, the most interesting finding was an exaggerated influence of ipsilesional distracters on conjunction search of neglect patients: whereas search for ipsilesional targets was unaffected by the number of contralesional distracters, the search for a contralesional target progressively slowed down with increasing numbers of ipsilesional distracters. Comparable results were reported when neglect patients searched for a conjunction of colour and orientation (Pavlovskaya, Ring, Groswasser and Hochstein, 2002), a target defined by the absence of a particular feature (the letter O among Qs, Eglin, Robertson, Knight and Brugger, 1994; Esterman, McGlinchey-Berroth and Milberg, 2000) or an inverted among upright items (Humphreys and Riddoch, 1993). Together, these studies suggest two important conclusions: on the one hand, the fact that speed of feature search is unaffected by the number of distracters confirms the hypothesis that single features are processed by parallel search mechanisms. On the other hand, pathologically enhanced search slopes for conjunction targets, suggest an impairment of serial, attentive search in neglect (Esterman et al., 2000; Grabowecky, Robertson and Treisman, 1993; L.C. Robertson and Eglin, 1993). These findings are in line with a biased-competition account of spatial neglect, which proposes that impaired detection of contralateral stimuli in spatial neglect reflects competitive interactions between target and distracters (Desimone, 1999; Desimone and Duncan, 1995; Duncan et al., 1997). The biased-competition account is based on the claim that different brain systems activated by multiple visual stimuli enter into competition for neural activity. Perceptual factors such as abrupt onset or perceptual conspicuity, as well as top-down influences such as behavioural goals may bias this competition, favouring a stimulus that will dominate further processing. Unilateral brain damage biases attentional competition in favour of stimuli presented in ipsilesional space. Thus, the biased-competition account considers
visual objects as the anchor points of attention upon which orienting movements in visual search are organized. One important alternative is the gradient account of spatial neglect, which assumes that visual search of neglect patients mainly reflects a biased gradient of attention for extrapersonal space (Kinsbourne, 1987, 1993; Pouget and Driver, 2000). The gradient account is based on the finding that detection times of neglect patients gradually increase with increasing retinal eccentricity or distance of the target from body midline (Behrmann, Watt, Black and Barton, 1997; Natale, Posteraro, Prior and Marzi, 2005; Smania et al., 1998).

In a recent formulation of the gradient account, Driver and Pouget (2000) stated that detection performance of neglect patients is not only determined by the absolute position of the target with respect to an egocentric reference frame, but also by its relative position respective to other stimuli. The authors further proposed that the attentional gradient of neglect patients is not strictly linear, but has a steeper fall-off for targets located further to the egocentric left, which is consistent with studies examining spontaneous exploratory eye and head movements of neglect patients searching for a visual target in the dark (Karnath, 1999; Karnath, Niemeier and Dichgans, 1998).

It would be tempting to interpret increased search slopes of neglect patients in conjunction search as evidence for the biased-competition account. However,
several methodological problems preclude this conclusion. Given that most of the studies discussed above evaluated search slopes, they varied task difficulty by increasing the number of distracters (Eglin et al., 1989; Eglin et al., 1994; Esterman et al., 2000; Grabowecky et al., 1993). One difficulty with this approach is a multiplication of confounding variables such as head or eye movements or repeated scanning of the same items that prevent evaluating the precise role of ipsilesional distracters. A second, perhaps more important problem is that by increasing the number of distracters, researchers are forced to decide between increasing the absolute size of the display and keeping item density constant or increasing item density and keeping absolute size constant. Further, an attentional gradient whose steepness depends on the number of distracters might contribute to the results. The more objects are present on the ipsilesional side, the more patients may allocate attention ipsilesionally, which would lead to an accelerated attentional gradient in the contralesional visual field.

In order to examine the possibility that neglect is a result of a combination of an attentional gradient and biased competition, we tested the performance of neglect patients in a search task that manipulated the quality of ipsilateral distracters while keeping target position and the number and position of distracters constant (Ptak and Valenza, 2005). This approach offered the possibility to study search performance in conditions with identical target and distracter positions, but different degrees of similarity between distracters and the target. Patients were asked to indicate whether a search display consisting of four elements contained the search target (which was always defined by the combination of two features: colour and filledness). There were three search conditions, which only differed by the quality of the three distracters (Figure 24A). In the dissimilar-distracter condition, all distracters were the same, with both features different from the target. For example, when the target was red/filled, all three distracters were green/unfilled. In the similar-distracter condition, all distracters were the same and shared one feature with the target. Thus, for a red/filled target, all three distracters were either green/filled or red/unfilled. In the mixed-distracter condition two distracters shared one feature (colour) and one distracter shared the other feature (filledness) with the target. A red/filled target was thus presented together with two red/unfilled distracters and one green/filled distracter or one red/unfilled distracter and two green/filled distracters. The position of these different distracters varied randomly across the three distracter positions. We studied 14 patients with left-sided neglect following an isolated right-hemispheric vascular lesion and 12 healthy controls. Figure 24B shows detection times and error rates of both groups, as a function of distracter condition. There was no significant effect of condition or visual hemifield on healthy control performance. In contrast, neglect patients showed two distinctive effects in their RT and error data. The first effect was a highly significant difference in performance between the left and the right hemifield,
which was found whether distracters were homogenous or heterogenous. This
generalized left-right difference suggests the presence of an attentional gradient.
The second effect was increased RT and error rate in the mixed distracter
condition compared to the other conditions, when the target was in the left
hemifield. This increased influence of heterogenous distracters is compatible
with the hypothesis that neglect is characterized by increased competition
between distracters and the target for attention.

How can these results be understood in terms of the SBAS model? Note that
in the mixed condition all distracters shared one feature with the target. In this
condition the SBAS model predicts only slight saliency differences between the
target and the distracters (compare Figure 7, p. 12), while greater saliency
differences are expected in the other two conditions, where only one feature
defines the target (compare Figure 6, p. 11). Therefore, only in the mixed
condition might a biased competition mechanism increase the saliency of right-
sided relative to left-sided items, and thus render the search task particularly
difficult for neglect patients. However, the hemifield difference that is found in

Figure 25. Lesion comparison between distractible (reddish colours) and non-
distractible (bluish colours) neglect patients. Non-distractible search (mean
RTs shown on the left) was associated with posterior parietal damage while
distractible search (mean RTs shown on the right) were associated with
inferior temporal damage.
the similar and dissimilar condition suggests the additional contribution of an attentional gradient favouring all ipsilesional over contralesional items. In sum, these results suggest that neglect patients suffer from a combination of biased competition for attention and a biased attentional gradient, which both contribute to their lateralized bias in visual search.

Do biased competition and the biased attentional gradient afford the same brain regions? In order to examine this question we separated our neglect group into those patients who showed highly distractible search (defined as a search slope, calculated across the three conditions, that was significantly steeper than the search slope of healthy participants) and patients whose search slopes were within the range of healthy controls. We then compared the localization of lesions in these two groups. Both neglect groups showed a significant left-right difference in visual search affecting all three conditions and reflecting a biased attentional gradient (Figure 25). In addition, neglect patients exhibiting distractible search showed particular impairment in the mixed condition, reflecting biased competition. The lesion analysis revealed that the region that was most often associated with distractible search was the inferior temporal lobe. These results suggest that biased competition for attention may be exacerbated when lesions of neglect patients extend into the inferior temporal lobe, while isolated parietal lesions only lead to a biased attentional gradient.

It is noteworthy that even in the easiest condition neglect patients showed a significant slowing of search for targets in the left hemifield. This condition required the search for a feature target, and should therefore depend on pre-attentive processing. How do these findings relate to the conclusion reached by previous studies that pre-attentive processing is intact in spatial neglect? A possible answer to this question is that pre-attentive processing remains functional, but is qualitatively altered in spatial neglect. This hypothesis is difficult to verify, yet we found support for it in a recent study examining the pre-attentive guidance of gaze by local scene characteristics (Ptak et al., submitted). Participants (healthy controls, right-hemisphere damaged patients

\[ \text{Figure 26. Quadratic polynomial functions based on the local features of fixed image patches. Horizontal position denotes the position of the patch within the picture. The green area shows the 99\% confidence interval of control data.} \]
without neglect, and neglect patients) freely explored a sample of images, and local content of image patches drawn around each fixation was analyzed (see page 48 for a description of the method). The influence of spatial neglect on local content selected by gaze was examined by calculating a quadratic trend for luminance, chromatic contrast, luminance contrast, and edge content as a function of the horizontal position of the fixation. Figure 26 shows for the three groups and for each local feature the quadratic polynomial trends, together with the 99% confidence interval of healthy participants. The figure shows that the relation between feature content and the horizontal position of a fixated patch was often nonlinear, indicating that the content of image regions sampled by gaze either increased or decreased toward the edges. Interestingly, while the data of healthy controls and RH-damaged participants closely overlapped, neglect patients showed significant deviations from both control groups for luminance, chromatic contrast and edge content. Though the selection of image regions in the right hemifield was based on the same local features as in healthy participants, neglect patients were particularly biased toward selecting regions of high local luminance and low edge content in the left hemifield. The shift away from edge sampling and toward privileged sampling of luminance information suggests that the processing of very local image features might be qualitatively altered in spatial neglect.

What relation do these findings have to the SBAS model? Many of the studies discussed above showed that processing at the level of feature maps is relatively spared in patients with spatial neglect or spatial extinction. Do these studies suggest that the first processing stage of the SBAS model (feature maps) is fully intact in patients with spatial neglect? Several findings suggest that this is not the case, at least not for all neglect patients. First, neglect patients exhibit a biased attentional gradient that results in slowed processing of all contralesional stimuli. This slowing not only affects complex search tasks requiring feature binding, but may also be observed in simple detection tasks that do not require specific search or feature integration processes (Natale et al., 2005; Smania et al., 1998). Second, it is probable that feature processing by neglect patients differs qualitatively from healthy participants, as our results of a shift away from contour and edge processing toward preferred luminance processing demonstrate. Third, depending on the extent of their lesion into the temporal lobe, neglect patients are differently affected by distracters in visual search. Visual search of patients with a combination of parietal and inferior temporal damage appears to be particularly vulnerable to a competition bias favouring ipsilesional distracters over a contralesional target.

**Surface-based and location-based enhancement of saliency**

An important component of the SBAS model is enhancement of the saliency of specific regions in a surface-based manner. Because of the spatially lateralized
attentional failures in spatial neglect, it is particularly interesting to study such surface-based enhancement in this disorder. One of the most controversial subjects in the literature dealing with spatial neglect is object-based neglect (Driver, 1999). Consider the drawing of one of our neglect patients (AO) shown in Figure 27. At the first sight, the drawing does not reflect important signs of neglect. However, a closer look at the different objects depicted reveals that the right side of each object is drawn with more care than the left. This is true for the tree, the woman and the sun shown on the left as well as the windows of the house right of the middle and the bird in the upper right corner. There appears to be an attentional gradient that affects the distribution of local elements within individual objects.

That object-based neglect of this kind may occur in drawings is not contested. Rather, the controversy about object-based neglect has to do with the question whether object-based effects may be understood with regard to a reference system that is entirely independent of egocentric coordinates or whether they represent a special form of egocentric coding. The first possibility follows the hypothesis of Marr (1982), who proposed that the brain constructs a 3D-sketch of objects consisting of a 3-dimensional representation whose coordinates are referenced to the centre of the object. In such a representation the left and right side of objects are coded unambiguously despite changing orientations of the object in space (for example, your left hand remains your left hand even if you

Figure 27. Object-based neglect in the drawing of a landscape.
are head-down). Applied to spatial neglect, such object-centred coding predicts neglect of the left side of the object irrespective of its orientation; that is, when the object is in its canonical (upright) orientation, the left side of the object is in left space, and object-centred neglect cannot be dissociated from left spatial neglect. However, when the object is head-down, its left side is on the right side of space. In this condition, pure object-centred neglect should result in omissions on the right side of space. Several studies have reported that rotating different objects away from their canonical position results in failures suggesting object-centred neglect. Thus, Behrmann and Moscovitch (1994) presented to neglect patients letters drawn with several colours and shown upright or rotated 90° to the left or to the right and asked them to indicate the presence of specific colours. They found that patients neglected the upper part (corresponding to the left side in object coordinates) more often than the lower part when the letter was rotated to the right, but showed the inverse pattern when the letter was rotated to the left. Driver and Halligan (1991) reported a patient who neglected the upper left side of objects more than their lower right side when these were rotated 45° to the right, in which case their upper left fell relatively more to the right than their lower right from the viewpoint of the patient. Caramazza and Hillis (1990) described a patient with right neglect who, when asked to read, consistently omitted or transformed the end of words irrespective of the orientation in which the word was presented. A particularly interesting form of object-centred neglect was observed by Behrmann and Tipper (1999). The authors showed barbell
stimuli that rotated 180° around their own axis before a target stimulus appeared at the left or the right end of the barbell to which the patients were asked to react. When the barbell rotated to the right, neglect patients were slower to react to stimuli on the right (defined in a reference frame centred on the body of the patient) than stimuli on the left. When two barbells were shown, but only one of them rotated to the right, patients showed left neglect for the barbell that was stationary, and right neglect for the rotated barbell, suggesting that in this condition patients accessed simultaneously two independent reference frames. This finding argues against an explanation of object-centred effects as a result of mental rotation of the object plane (Buxbaum, Branch Coslett, Montgomery and Farah, 1996). Nevertheless, many instances of apparent object-centred neglect (e.g. Pavlovskaya, Glass, Soroker, Blum and Grosswasser, 1997) may be explained with a relative gradient model that supposes that the biased attentional gradient observed in spatial neglect is significantly influenced by the boundaries defining an object (Driver and Pouget, 2000; Pouget and Sejnowski, 2001). According to the model the attentional gradient is steeper within objects than between objects, which results in strong neglect of the left part of objects whatever their absolute position. However, as Figure 28A shows this model not only predicts neglect within objects, but also strong neglect of the scene, which is not the case of the drawing shown in Figure 27. As an alternative, the SBAS model proposes a surface-based selection process (Figure 28B), which predicts that whole surfaces are selected as reference frames for attention, and that spatial neglect will be observed in the momentarily selected surface. This mechanism offers an explanation of the observation that similarly severe neglect may be observed for the left side of objects at different absolute positions.

Surface-based selection of reference frames also predicts significant variations of neglect for surfaces located at the same absolute position, but whose horizontal extent varies. This prediction was confirmed in a visual search study conducted by Karnath and Niemeier (2002). The authors asked neglect patients to search for a target letter in a search field populated with a large number of distracter letters. In one condition, all stimuli were the same colour and patients were instructed that the target could appear anywhere in the search field. In a second condition, the search field was segmented into six areas that differed in the colour of the stimuli. In both conditions, neglect patients explored a region between 20° and 120° to the right of the middle, and there was no difference between attentional gradients across this region. However, in the third condition patients were instructed that the target would appear in the orange stripe, which covered an area ranging from 40° to 80° degrees to the right of the middle. When the search field was restricted in this manner, neglect patients only explored the orange stripe. However, the most important result was that the attentional gradient in this stripe was much steeper than when neglect patients attempted to explore the whole search field. These findings suggest that whether search is distributed over a wider surface or spatially restricted determines the degree of
neglect for the same portion of space. A similar conclusion may be reached when comparing the results of two studies examining the distribution of gaze fixations across different search surfaces. In a study by Karnath, Niemeier and Dichgans (1998) patients were free to move their eyes and head in a surface covering 360°. The authors found that patients produced a bell-curved distribution of fixations whose mean was shifted 40° to the right of the middle. In a recent study, we evaluated the distribution of fixations of healthy participants, right-hemisphere damaged patients without neglect, and neglect patients within images of different contents (natural scenes, architecture, man-made objects etc.) and found a rightward shift of the distribution by only 6° (Ptak et al., submitted, see Figure 29). There was no indication that the patients of Karnath et al. (1998) had more severe neglect than our patients. Rather, these important differences between the two studies suggest that the size of the search surface significantly affects the extent of the rightward shift during visual exploration of space.

Figure 29. A) Colour-coded representation of the 2D-density function based on all fixations made during free exploration of 16 images. Warm colours show regions with high fixation density. B) and C) Fixation density functions resulting from free exploration of two sample images.
A bias favouring foveated stimuli

In addition to these demonstrations of surface-based modulations of spatial neglect by the size of a visual scene and its segmentation, the visual system exhibits a strong bias favouring central regions. The classic paradigm used to demonstrate this central bias is the overlap task. In this task, participants are required to make saccades from the fixation point toward peripheral targets. On some trials the central fixation stimulus disappears shortly prior to appearance of the target (gap condition) while on other trials the fixation stimulus remains on the screen (overlap condition). Compared to the overlap condition the gap condition significantly reduces saccade latency (Saslow, 1967), eventually leading to saccades with extremely short latencies (Fischer and Breitmeyer, 1987). In order to examine whether the effect of a distracter stimulus presented at fixation affects saccade latency similarly in healthy participants and neglect patients, we asked neglect patients to execute saccades to peripheral targets presented at 6° or 12° in the LVF or RVF (Ptak, Schnider, Golay and Müri, 2007). 150 ms prior to appearance of the target the central fixation cross was extinguished. The target (a small, coloured square) was shown either alone or simultaneously with a coloured distracter that appeared in the same hemifield as the target, in the opposite hemifield, or at fixation. Healthy participants and right-hemisphere damaged patients without neglect showed an increase of saccadic latency of approximately 20% when the distracter appeared at fixation (overlap task) compared to when there was no distracter (gap task). In contrast, the increase averaged more than 60% in patients with spatial neglect (Figure 30). Interestingly, a foveal distracter had no influence on performance when patients were asked to respond manually (by pressing a response button) upon the onset of the peripheral target stimulus. The observation that the effect is bilateral and restricted to oculomotor responses is coherent with a model of spatial neglect that posits mutual inhibitory interactions between the parieto-occipital cortex and the superior colliculus (Sprague, 1966), a structure involved in saccade initiation and maintenance of active fixation. In animals, impaired orienting responses toward stimuli contralateral to a parieto-occipital lesion are restored if the contralateral colliculus is functionally inactivated (Payne, Lomber, Geeraerts, van der Gucht and Vandenbussche, 1996; Sprague, 1966). This finding suggests that parieto-occipital cortex normally exerts a facilitatory influence on the ipsilateral and an inhibitory influence on the contralateral colliculus. In spatial neglect following a cortical lesion, the ipsilesional colliculus has a lower (inhibition) and the contralesional colliculus a higher glucose metabolism (disinhibition, Rushmore, Valero-Cabre, Lomber, Hilgetag and Payne, 2006), which could contribute to the directional bias in favour of ipsilesional items. An important feature of the superior colliculus is its subdivision in two functionally distinct regions. The rostral colliculus contains neurons that discharge when a stimulus in the central ~2° of the visual field is actively fixated (Gandhi and
Saliency, Selection, and Spatial Neglect

Keller, 1999; Munoz and Wurtz, 1992, 1993b), while neurons in the caudal colliculus show activity related to the preparation and execution of saccades (Dorris, Paré and Munoz, 1997). According to the model, in patients with neglect following a right cortical lesion, a functionally disinhibited left colliculus would exhibit increased fixational activity in its rostral region and increased saccade-related activity in its caudal region. An important finding of our study was that neglect patients made many directional errors toward distracters shown in the right hemifield when the saccade had to be made toward the left. This high number of right directional errors of our patients can be explained by a functional disinhibition of the left caudal colliculus. On the other hand, the bilateral increase of saccade latency following a foveal distracter can be understood in terms of functional interactions between the two colliculi. Fixation neurons pause before saccades in all directions (Munoz and Wurtz, 1992, 1993b) and stimulation of these neurons in one colliculus activates fixation neurons and deactivates saccade-related neurons in the other colliculus (Munoz and Istvan, 1998), suggesting that rostral colliculi work as a unit that maintains active fixation. Disinhibition of one rostral colliculus delays saccades in all directions (Munoz and Wurtz, 1993a), which would explain the bilateral increase of saccade latency in our neglect patients. Such disinhibition would produce an increase of saccade latency via a mechanism that amplifies the neural response to fixated items relative to peripheral items. Though animal studies support the Sprague-model of spatial neglect, there is yet no direct evidence that the model also applies to spatial neglect in humans. Nevertheless, our findings strongly suggest that spatial neglect is characterized by a strong bias favouring stimuli at the centre of vision.

**Figure 30.** Mean increase of saccade latency (A) and manual reaction time (B) when a distracter appeared at fixation relative to when no distracter was presented. Relative increase is shown for each group as a function of target position.

Keller, 1999; Munoz and Wurtz, 1992, 1993b), while neurons in the caudal colliculus show activity related to the preparation and execution of saccades (Dorris, Paré and Munoz, 1997). According to the model, in patients with neglect following a right cortical lesion, a functionally disinhibited left colliculus would exhibit increased fixational activity in its rostral region and increased saccade-related activity in its caudal region. An important finding of our study was that neglect patients made many directional errors toward distracters shown in the right hemifield when the saccade had to be made toward the left. This high number of right directional errors of our patients can be explained by a functional disinhibition of the left caudal colliculus. On the other hand, the bilateral increase of saccade latency following a foveal distracter can be understood in terms of functional interactions between the two colliculi. Fixation neurons pause before saccades in all directions (Munoz and Wurtz, 1992, 1993b) and stimulation of these neurons in one colliculus activates fixation neurons and deactivates saccade-related neurons in the other colliculus (Munoz and Istvan, 1998), suggesting that rostral colliculi work as a unit that maintains active fixation. Disinhibition of one rostral colliculus delays saccades in all directions (Munoz and Wurtz, 1993a), which would explain the bilateral increase of saccade latency in our neglect patients. Such disinhibition would produce an increase of saccade latency via a mechanism that amplifies the neural response to fixated items relative to peripheral items. Though animal studies support the Sprague-model of spatial neglect, there is yet no direct evidence that the model also applies to spatial neglect in humans. Nevertheless, our findings strongly suggest that spatial neglect is characterized by a strong bias favouring stimuli at the centre of vision.
In our study of the central bias in spatial neglect, distracters at fixation appeared simultaneously with the target. In the overlap condition the central distracter remained on screen, while in the gap condition it disappeared 200 ms prior to the onset of the target. In all other conditions (blue squares) the distracter disappeared 200 ms prior to target onset, but reappeared at varying time intervals prior to (negative asynchronies), simultaneously with (asynchrony = 0), or after (positive asynchronies) onset of the target. Figure 31 shows that saccade latencies were longest in the overlap condition. Among the off-on conditions, saccade latency was longest when the distracter reappeared simultaneously with the target, while reappearance before and after target onset resulted in significantly shorter latencies. When the distracter reappeared 100 ms or more after target onset, its effect was equivalent to the gap-condition. These results confirm that a central distracter systematically increases saccade latencies to peripheral targets, and that the presence of the distracter is particularly perturbing when it appears simultaneously with the target. These effects of a central distracter contrast with
the effect of peripheral distracters, which influence saccade amplitude, but have a much lesser impact on saccade latency (Findlay, 1982; Ptak et al., 2007; Walker, Deubel, Schneider and Findlay, 1997). Neurophysiological studies suggest a generalized overrepresentation of fovea in the visual system, though this overrepresentation probably results from different mechanisms in the cortex and the midbrain (Gross, 2008; Munoz and Wurtz, 1993a). A recent study by Beck and Lavie (2005) has shown that distracters at fixation not only interfere with saccade programming, but also the detection of peripheral targets. These findings support the idea that processing of visual information presented at the fovea is prioritized compared to more peripheral information. This priority of the central visual field has been implemented in the SBAS model in terms of slightly increased saliency of foveal relative to peripheral regions (see Figure 8).

However, this relative increase of saliency of central regions does not imply that peripheral regions are not subject to surface-based or location-based enhancement. In fact, the most important demonstration of location-based enhancement of saliency in spatial neglect comes from studies of peripheral cueing. Posner, Walker, Friedrich and Rafal (1984) examined patients with unilateral damage to the parietal lobe with the spatial cueing paradigm (see Figure 15). The patients had comparable reaction times to contralesional and ipsilesional targets in the valid condition, but disproportionately slowed RTs to contralesional targets in the invalid condition. Subsequent studies reproduced and extended this finding of an increased validity effect, showing that attentional

![Figure 32](image-url)

**Figure 32.** A) Percent correct identifications of features of the left (LVF) or right (RVF) stimulus after valid or invalid cueing. The dashed horizontal line represents chance performance. B) Percentage of contralesional letters identified following an auditory stimulus that was coherent with the contralesional letter (white bar), the ipsilesional letter (black bar) or neither (grey bar).
capture by ipsilesional cues strongly correlated with the degree of spatial neglect (Losier and Klein, 2001; Morrow and Ratcliff, 1988; Petersen, Robinson and Currie, 1989; Posner et al., 1987). Similar findings were reported for long cue-target intervals (Bartolomeo, Siéroff, Decaix and Chokron, 2001), and even when cues and targets were presented in different modalities (Farah, Wong, Monheit and Morrow, 1989; Golay, Hauert, Greber, Schnider and Ptak, 2005). In addition, peripheral cues generally produced stronger effects of cue validity than central (symbolic) cues (Losier and Klein, 2001). These results are consistent with the interpretation that neglect patients have a deficit of attentional disengagement (compare Figure 16) because ipsilesional stimuli capture and bind their attention much stronger than in the healthy brain (Morrow and Ratcliff, 1988; Posner and Petersen, 1990; Rafal, 1998; L.C. Robertson and Eglin, 1993). The SBAS model proposes a relatively simple mechanism that accounts for the capture of attention by peripheral stimuli. However, since this mechanism critically depends on the attentional set of the patient it is important to present evidence for attentional capture and attentional set on spatial neglect before giving an explanation for these effects.

**Attentional capture and attentional set**

The attentional failures of patients with spatial neglect and extinction described in the preceding sections suggest that both syndromes reflect an impairment of automatic aspects of attention and therefore may not be affected by expectations, action goals, or task instructions. However, in the last years an increasing number of studies have been published demonstrating that neglect and extinction may to a significant degree be modulated by top-down influences. Thus, contralateral extinction decreases when the location of the contralesional stimulus is predetermined compared to when it is uncertain (Geng and Behrmann, 2006), or when patients are instructed to disregard stimuli appearing

![Figure 33](image)

**Figure 33.** Mean reaction times for experiment 1 depending on cue validity, cue identity and target position for younger (circles) and older (squares) healthy controls, right-hemisphere damaged patients without neglect, and neglect patients. Note that the data of healthy participants and patients are presented in different scales.
in the ipsilesional visual field (Karnath, 1988). We observed a similar effect in extinction patient AG (Ptak, Valenza and Schneider, 2002). The patient was shown unilateral or bilateral displays of red or green squares and diamonds. Similarity of the bilateral displays was systematically varied and the displays were coherent with respect to form (e.g. both squares) or colour (e.g. both red), or incoherent. On single trials, the patient identified one feature (e.g. colour) of the contralesional or ipsilesional stimulus. We explicitly instructed the patient only to identify the contralesional stimulus regardless of the presence of an ipsilesional stimulus. AG identified a contralesional feature significantly better when it was identical with the ipsilesional feature (e.g. colour when both stimuli were red) than when it was different. This effect reflected a real benefit of stimulus repetition and could not be explained by a bias to simply report the ipsilesional feature, since such a tendency was not evident in identification of dissimilar displays. However, it contrasts with the observation made by Baylis, Driver and Rafal (1993) that, when extinction patients are asked to identify both stimuli, they show increased extinction when the stimuli are similar. The authors presented bilateral red or green letters and asked patients to report their colour or form. The interesting finding was that when the identified feature (e.g. colour) was shared by both stimuli (e.g. both letters were red) the extent of extinction was significantly lower than when the features were different.

![Figure 34. Mean reaction times for experiment 2 depending on cue validity, cue identity and target position for right-hemisphere damaged control patients and neglect patients. Targets were presented in the same or in a different modality than the cue.](image-url)
increased in comparison to when the feature was different (e.g. red and green) and irrespective of the similarity of the irrelevant feature. Thus, the instruction to report only one specific dimension of the bilateral stimuli significantly affected the degree of extinction. Another example of the effect of task constraints on spatial extinction is the observation that spatial cueing significantly affected extinction of patient AG (Ptak et al., 2002). The patient was shown bilateral stimuli that either shared only form, only colour, form and colour, or no feature. Prior to presentation of the stimuli, a cue word appeared at fixation indicating the dimension the patient was requested to identify (e.g. ‘colour’). After presentation of the stimuli, a question appeared that the patient was to answer (e.g. ‘left colour’). On most trials the question was coherent with the cue, but on some trials the cue was misleading and the patient was asked to identify the uncued dimension. As Figure 32A shows, AG showed a cueing effect only for stimuli presented in the left (contralesional) visual field. He was well above chance when identifying the validly cued dimension, but at chance when identifying the invalidly cued dimension. This finding shows that extinction may be modulated by the expectation of the patient, but that such modulation only affects some features of the stimulus.

A further example of the effect of attentional set on spatial extinction is the observation that, similarly to the influence of lateralized sounds on detection of visual targets by neglect patients (Farah et al., 1989; Golay et al., 2005), auditory cues significantly affect the degree of extinction. We asked TG, a neglect and extinction patient with focal right parietal damage to identify letters presented for 150 ms bilaterally or left or right of fixation. Each presentation was preceded by a letter cue presented auditorily through a central speaker. The cue was either coherent with the contralesional letter or the ipsilesional letter, or it was incoherent with both letters. As Figure 32B shows TG showed most extinction when the cue was coherent with the ipsilesional letter, and less extinction when it correctly indicated the contralesional letter. This finding suggests that the cue attracted attention of the patient ipsilesionally when it was coherent with the ipsilesional cue, and contralesionally when it was coherent with the contralesional cue. The studies reviewed here provide converging evidence that attentional set significantly modulates the competition between the ipsilesional and the contralesional stimulus.

Attentional set also plays a similar role when other measures of spatial neglect are evaluated. Danckert and collaborators (Danckert, Maruff, Kinsella, de Graaff and Currie, 1999) used a flanker task to examine whether ignored information affects processing of a central target. They asked participants to identify the shape or the colour of a central letter flanked on its left or right by an irrelevant distracter letter. The flanker shared the feature with the target that had to be identified or the irrelevant feature. Healthy participants and a patient with left neglect showed facilitation of target identification by flankers, but only if flankers shared the relevant target feature. In a group study using the same
paradigm, Snow and Mattingley (2005) confirmed the facilitation by task-relevant feature information, but they also found that the task-irrelevant feature of ipsilesional flankers interfered with target identification in patients with spatial neglect. The authors therefore concluded that neglect patients have an impaired ability to selectively inhibit task-irrelevant information within the ipsilesional field. However, this interpretation is not entirely supported by findings of studies using the spatial cueing paradigm. Thus, Bartolomeo and collaborators (Bartolomeo et al., 2001) examined performance of healthy participants and neglect patients in the peripheral cueing task and systematically varied the time interval between cue and target as well as cue predictability. Remember that healthy participants show a benefit of valid cues at short cue-target intervals, but a cost (inhibition of return) at long intervals. Cue predictability is the factor that determines the probability to which the target is presented at the position of the cue. The authors used cue predictabilities of 50% (unpredictable cues), 80% (cues predict that the target is most likely to appear on the same side), and 20% (cues predict that the target is most likely to appear on the opposite side). The aim of the manipulation of predictability was to observe interactions between automatic capture of attention by the cues and controlled aspects of attention. When predictability is 50%, the most sensible strategy is to disregard the cue and to keep attention at the central location. When predictability is 80%, it is more sensible to shift attention to the location of the cue, because the target is most likely to appear at this position. Finally, when predictability is 20% the best strategy is to shift attention in the direction opposite the cue. In healthy participants, this expected pattern was found with cue-target intervals of 550 ms or 1000 ms, while some capture of attention was only observed at the 150 ms interval in the unpredictable cue (50%) condition. Neglect patients showed greater and prolonged capture of attention in the 50% and the 20% condition. However, contrary to the proposal of Snow and Mattingley (2005), they did not show a generalized inability to inhibit the influence of cues in these conditions, since at longer cue-target intervals they were able to control attention similarly to healthy participants.

In order to test the contention that neglect patients lack inhibition of task-irrelevant information in the ipsilesional field we examined whether the ipsilesional orienting bias of neglect patients in a spatial cueing task would be modulated by induced behavioural biases in patients (Ptak and Schneider, 2006). In contrast to previous studies of spatial cueing, we manipulated the identity of the cues to make them similar or dissimilar to the target. Similar cues possessed features that also defined the target, and were thus of high behavioural salience, but patients had to withhold reactions to them. If spatial cues capture attention of neglect patients automatically, contralesional RTs in the invalid cueing condition should not differ between cues that share some task-relevant property with the target (behaviourally relevant cues) and cues that don’t. In experiment 1 the patients were instructed to react to a coloured form presented left or right of
fixation simultaneously with a coloured distracter. The target display was precued with a form that had the same colour as the target or a different colour, presented at the position of the upcoming target or in the opposite hemifield. The critical question was whether an ipsilesional cue that was dissimilar to the target would capture attention similarly to a cue that shared the behaviourally relevant colour with the target. In healthy participants and RH-damaged patients without spatial neglect similar and dissimilar cues had no differential effect on performance when they were presented at the invalid location (Figure 33). In contrast, neglect patients showed significantly slowed reactions to contralesional targets following ipsilesional cues, but only when cues were similar to the target (LVF-invalid condition, see Figure 33). Clearly, task-irrelevant cues did not capture attention of neglect patients.

In experiment 2 we asked whether the interference by similar cues was due to their physical similarity or some higher-level relationship with the target. We repeated experiment 1 with another group of neglect patients, but added a condition. On some trials the target was a red circle while on others it was the word ‘RED’. The cue was either visually related to the target (e.g. red circle when the target was a red circle; the word ‘RED’ when the target was the word ‘RED’) or only semantically related (RED - red circle; red circle - RED). Interestingly, we found that ipsilesional cues captured attention of neglect patients when they were relevant for the task, but irrespective of whether cue and target were presented in the same modality or in a different modality (Figure 34). Thus, these results suggest that capture of attention by ipsilesional cues did not depend on visual similarity, but on a non-visual relation. One possibility is that cues captured attention because of their semantic relationship with the target, since the red circle and the word ‘RED’ both refer to the same concept. The other possibility is that cues captured attention because the red circle and the word RED were both defined as targets and were therefore behaviourally relevant. In
order to distinguish between these possibilities we repeated experiment 2, but this time instructed patients only to react to the colour cue and to disregard the word cue. As Figure 35 shows, when the word cue was irrelevant it ceased to capture attention of neglect patients. In sum, contrary to the suggestion by Snow and Mattingley (2005) that neglect patients fail to inhibit task-irrelevant information presented in the ipsilesional hemifield, these results show that only behaviourally relevant ipsilesional stimuli capture attention of neglect patients.

We further evaluated whether the capacity of neglect patients to disregard irrelevant ipsilesional stimuli is immediate or whether it takes some time to develop (Ptak and Golay, 2006). Again, we tested neglect patients with a spatial cueing task. Patients reacted to a target letter presented in the LVF or RVF simultaneously with a distracter letter in the other hemifield. A letter cue identical to or different from the target appeared at the same position as the upcoming target or at the opposite position 100 ms or 1000 ms prior to target appearance. At 1000 ms, we observed capture of attention by ipsilesional letter cues only if these were identical to the target. However, at 100 ms all ipsilesional cues captured attention irrespective of their similarity to the target, suggesting that the process that allows neglect patients to disregard irrelevant stimuli is not instantaneous.

Together, these results suggest that attentional capture by ipsilesional stimuli may be represented in a modified two-component model of spatial attention. The original model (Müller and Findlay, 1988, see Figure 17) suggested that attentional capture is automatic and characterized by fast onset and fast decay. The second component is active maintenance of attention by controlled processes. The difficulty of neglect patients to disengage attention from ipsilesional stimuli suggests that attentional capture is prolonged compared to healthy participants (Bartolomeo and Chokron, 2002). However, the critical factor responsible for this extended influence of the first process had not been
specified. Our results show that the variable determining whether a stimulus continues to capture attention after 200-300 ms is the behavioural relevance of the stimulus. This modified two-component model is shown in Figure 37. According to the model attentional capture is determined by an automatic component whose action is extensible in time as a function of the behavioural relevance of the stimulus, and that directly influences the activation that the stimulus generates at the level of the saliency map. In the SBAS model, this modulation of the influence of the automatic attentional component on the saliency map is a result of saliency enhancement by behavioural goals represented in working memory.

**Summary**

In this chapter neuropsychological evidence has been presented that supports several components of the SBAS model. A considerable number of studies have reported that patients with spatial extinction or neglect are capable of extracting a significant amount of information from stimuli they are unable to perceive consciously. Thus, extinguished stimuli affect processing of stimuli shown in the intact hemifield, and information presented in neglected space may influence judgements of patients about stimuli in intact space. In agreement with these capacities, neglect patients’ visual search is not strongly affected by the number of distracters when they search for a target defined by a single feature, indicating that they are capable of parallel processing of single features (but not combinations of features) across the visual field. Together, these findings suggest that pre-attentive processing at the level of feature maps is preserved to a considerable degree in spatial neglect, though it is probably not entirely intact. Recent studies with neglect patients also show that these patients show a non-spatial attentional bias favouring foveal stimuli when oculomotor responses are measured.

Yet another important feature of spatial neglect is revealed when enhancement of attention by spatial cues is examined: neglect patients show strongly impaired detection and identification of contralesional stimuli when their attention is drawn ipsilesionally toward a distracter stimulus. Systematically varying the time between cue onset and target onset as well as the relevance of the distracter stimulus further reveals that the detection of targets presented in the impaired hemifield is a function of the interaction between these two factors. Finally, evidence is accumulating that the lateral attentional bias characteristic of spatial neglect is significantly modulated by the attentional set of the patient. For example, stimuli that are relevant for the task at hand capture attention in a seemingly automatic manner, while irrelevant stimuli presented at the same position do not affect performance.

In sum these results suggest that spatial neglect can only be fully understood if spatial and temporal dynamics of attentional displacement as well as the
demands of the current task are examined. Whether a stimulus captures attention of neglect patients depends on a) whether it is presented in the left or the right hemifield, b) whether it is presented shortly prior to the target or some time ago, and c) whether it has some relevance for the task at hand. Thus, stimuli shown in the intact hemifield only capture attention of neglect patients if they appear shortly prior to the onset of the target or if they are relevant for the current task. These findings also suggest that attentional failures in spatial neglect may be understood in terms of a damaged saliency map. As a consequence of such a damage, the saliency of stimuli presented on the contralesional side of space is reduced, making contralesional information more difficult to detect and more vulnerable in comparison to ipsilesional information. Since attention selects the most salient stimulus, ipsilesional stimuli are consistently privileged by attention, though the feature input into the saliency map remains unaffected.

Figure 37. Modified two-component model, in which automatic allocation of attention (red line) is extensible as a function of the behavioural relevance of the stimulus. Note that controlled attention (blue line) has a tendency to decrease the saliency of the stimulus, when it is not defined as action target.
6 - CONCLUSIONS

The present work presented evidence from neurophysiological, cognitive, and neuropsychological studies for a model of attention, which postulates that stimuli are selected for further processing based on their saliency. I have argued that, in order to solve the criterion problem the brain needs a representation of stimuli which allows the comparison of different features (e.g. colours, shapes, sizes), modalities (images, sounds, touches), or histories (previously encountered or new stimuli). Such a model has several advantages, such as its relative simplicity (at any given moment in time the most salient stimulus is selected by attention), economy (in the simplest form of the model, only one saliency map is needed), neurophysiological foundation (its components and function can be localized in the brain), and possibility to test it with computational methods. Clearly, there are some limitations that need to be addressed.

One problem concerns the exact role of working memory for attentional selection. In the SBAS model working memory is conceived as a cognitive component that is external to the model. The aim of working memory is the online representation and activation of action plans that feed into the saliency map or feature maps to enhance representations of the environment. However, the model does not specify how action plans are selected, how they are transmitted, and how enhancement is achieved. Further, the model suggests a unilateral connection between working memory (prefrontal cortex) and the saliency map (posterior parietal cortex). However, neurophysiological studies suggest strong feed-forward and feed-back projections between both regions. An important modification of the model would be to implement a connection between the saliency map and working memory allowing information about the selected stimuli to become available for processes involved in action preparation.

A second limitation has to do with the differences in performance that depend on the type of response. Some newer findings with neurological patients suggest that the saliency of stimuli is a function of the requested response. For example, performance varies according to whether an oculomotor or a hand action is planned. In its present form the SBAS model is unable to explain such differences, though several alternatives exist: according to one possibility, different saliency maps might independently and in parallel code the saliency of environmental stimuli for hand action and oculomotor action. Another possibility to explain the effect of the type of response is by keeping one saliency map, but by adding saliency ‘modulators’ in each output channel that would account for saliency differences of stimuli selected as saccade targets and stimuli selected as reach targets. For example, an oculomotor modulator would increase the saliency of foveal regions, thus accounting for the strong capture of gaze by distracters presented at fixation.
A further problem with the model is the representation of hemispace and hemispheric differences. Functional brain imaging studies and the greater frequency of occurrence of spatial neglect following right compared to left hemispheric damage suggest a dominance of the right hemisphere for spatial attention. This right-hemispheric dominance has been implemented in models of neglect in a bilateral representation of space in the right hemisphere and an exclusively contralateral representation of space in the left hemisphere (Heilman et al., 1993; Mesulam, 1981). The SBAS model does not make reference to such hemispheric differences, but they could be implemented by specifying different contributions of the two hemispheres to the elaboration of the saliency map.

Despite these shortcomings, the present work has provided many arguments for a model of attention taking the saliency of stimuli as criterion for attentional selection. The model is supported by converging findings of the neurophysiological, cognitive, and neuropsychological literature. Thus, single cell studies have shown that the activity of cells to environmental stimuli is significantly modulated by bottom-up (e.g. perceptual stimulus characteristics or lateral suppression due to competing stimuli in a crowded display) as well as top-down factors (behavioural goals, the target status of a stimulus etc.). Presently, there is no evidence for a distinction between these two kinds of inputs at the level of single cells; that is, a cell is not able to differentiate between bottom-up or top-down afferents. Consequently, bottom-up and top-down inputs have very similar influences on cell activity. The second fundamental finding of neurophysiological studies is that several brain regions including the FEF, the superior colliculus, and the PPC represent the strength of environmental stimuli, but only the PPC fulfils several criteria that are necessary for the representation of saliency. Thus, the PPC has access to multiple reference frames for the representation of the environment, codes the environment independently of single visual features and independently of input modality. These neurophysiological findings are completed by studies of healthy participants strongly supporting a two-stage architecture of spatial attention. The main support for this architecture comes from visual search studies showing that the processing of single features is independent of attentional limitations, while more complex search requires a one-by-one, serial search mechanism. The latter may represent the operation of the attentional ‘spotlight’, a spatially delimited focus of covert attention. Early studies have suggested that objects may capture covert attention in an automatic manner. However, more recent findings show that attentional capture is often contingent on top-down signals such as behavioural goals and expectations. This finding fits well with the idea that modulation of saliency by bottom-up and top-down processes is indistinguishable. A further argument is the finding that capture of attention by stimuli presented to the preserved visual field of patients with spatial neglect is significantly modulated by the behavioural relevance of the stimulus. Studies with neglect patients also
support the idea that attention may select surfaces, and that central stimuli are systematically privileged over stimuli presented in the visual periphery.

In conclusion, though some of aspects of the model of attentional selection presented in this manuscript may be criticized, the main idea that the brain solves the criterion problem by relying on a single measure reflecting the strength of neural response is supported by convergent findings from neurophysiology, cognitive psychology, and neuropsychology.
REFERENCES


Saliency, Selection, and Spatial Neglect

Eglin, M, Robertson, LC, Knight, RT and Brugger, P (1994). Search deficits in neglect patients are dependent on size of the visual scene. *Neuropsychology, 8*(3), 451-463.


Felleman, DJ and Van Essen, DC (1991). Distributed hierarchical processing in the primate cortex. *Cerebral Cortex, 1*, 1-47.


Golay, L, Schneider, A and Ptak, R. Cortical and subcortical anatomy of chronic spatial neglect following vascular damage. Behavioral and Brain Functions (in press).


Saliency, Selection, and Spatial Neglect

Saliency, Selection, and Spatial Neglect


Ungerleider, LG and Mishkin, M (1982). Two cortical visual systems. In Ingle, DJ, Goodale, MA and Mansfield, RJW (Eds.), Analysis of visual behavior (pp. 549-586), Cambridge: MIT Press.


