

# Computational Models of Visual Selective Attention: A Review

Dietmar Heinke, Glyn W. Humphreys  
Behavioural and Brain Science Centre  
University of Birmingham  
Birmingham B15 2TT  
United Kingdom

## Abstract

We review some of the major computational models of visual selective attention in terms of how they apply to psychological data and to theoretical concepts derived from experiments concerned with stimulus filtering, visual search and spatial cueing. We also review attempts to model neuropsychological disorders of visual object recognition and attention, including visual agnosia and neglect. We highlight how computational models can unite several dichotomies in psychological models: space-based versus object-based selection; early versus late selection; representational versus attentional accounts for neurological deficits. A comparison of the models exposes weaknesses and advantages in different accounts, but it also highlights that most models suggest that competitive interactions in visual selection are the basis for attentional effects in human data. There is also agreement across models on explaining neurological disorders as due to imbalances in competition following damage to certain areas of the brain. We propose that computational models clarify theoretical accounts of visual selective attention and integrate concepts across areas. The models provide a useful contribution to psychological research.

## 1 Introduction

This paper reviews some of the major computational models of visual attention and highlights, how they help to clarify concepts in psychological theories. To do this we focus on some of the major issues and experimental paradigms that have influenced psychologists, including: space vs. object-based attention, stimulus filtering tasks and visual search. We also discuss how the models can be used to simulate neuropsychological disorders of visual cognition, including agnosia and neglect – disorders that themselves highlight particular theoretical issues such as the role of parallel grouping processes in vision, the relations between attention and spatial representation, and the interaction between bottom-up and top-down factors in selection. We conclude by evaluating the contribution of the modeling process to understanding cognition.

To provide the framework for our analysis, we first consider the issues and paradigms that have dictated much of the psychological research on visual attention.

### **Space vs. object-based attention**

Psychological theories traditionally hold that visual selection is spatially-based, assuming that selection involves some form of internal spotlight that activates attended regions or

leads to the inhibition of unattended regions of visual field (Posner et al. 1980, Treisman 1988). Strong evidence for spatial selection comes from studies of cueing which show that detection of a target is improved when it is preceded by a spatially valid cue (Posner et al. 1980). On the other hand experiments have also demonstrated a more object-based style of selection. For instance, the detrimental effects of invalid cues are reduced if cues and targets fall within the same perceptual object (Egley et al. 1994), and the costs in detecting two stimuli presented simultaneously relative to when they are presented successively in time are reduced if the stimuli are attributes of single object (e.g. Duncan 1984). To capture these results, we need theories that specify not only how both space- and object-based selection operates, but importantly, how these different forms of selection may interact.

### **Stimulus-filtering tasks**

One procedure that has been used extensively to investigate attentional selection in humans has been stimulus-filtering, where participants are given a stimulus attribute to use as a cue to select other properties of a target in a display. For instance, subjects may be asked to name the colour of a shape (e.g. red square). Experimenters may then be interested in the impact on selection of stimuli carrying non-target attributes (e.g., blue circle or red circle), impact on responses to targets, and whether any impact is related to factors such as the response associated with the relevant and irrelevant items. Generally, the effect of the distractors is dependent on their similarity to the relevant target attributes (e.g. van der Heijden 1981). On the other hand, the experiments showed that the interference depends on the spatial distance between relevant and irrelevant items (e.g. Eriksen & Eriksen 1974). Eriksen & Yeh (1985) took these experimental findings as evidence for an imperfect spatial selection process in the shape of a "zoom lens" stimuli falling in the penumbra of this zoom lens may still be processed to some degree, to generate interference even at the response level. Hence, theories need to explain these aspects of non-spatial (attribute-based) as well as spatial selection.

### **Visual search**

One special kind of filtering task, in which the number of non-target elements are varied systematically, is visual search. Studies of visual search have been used particularly to assess the question of which visual processes are spatially parallel and which necessitate that selection becomes spatially restricted, so that a target is only detected following some form of serial search (operating across one restricted spatial region at a time) (see Wolfe 1998, for a recent review). Typically, search efficiency is measured using the slope of the function relating reaction time (RT) to the number of distractors present. The slope of the search function then depends on the exact nature of the target and the distractors in the display. For instance, if the target is distinguished by a salient feature (feature search), such as a horizontal bar amongst vertical bars, search slopes are shallow or even flat (a result termed 'pop-out' in search). In contrast, if the target is defined by a conjunction of features (e.g., a blue vertical bar amongst green vertical and blue horizontal bars) search is slow and typically linearly related to the number of distractors present. Usually these findings are interpreted as indicating different modes of processing during search. Efficient (flat) search functions are said to reflect spatially parallel processes whereas linear and steep functions are attributed to spatially serial processes (e.g. Treisman & Gelade 1980). There are numerous theories based on this dichotomy. Here, we focus on two theories each of which have formed the basis for computational models which we review subsequently. Wolfe & Cave (1990) suggested that the visual system computes for each item in the display its saliency relative to surrounding items. A "saliency map" is then assembled in parallel across a display. The saliency map is also modulated by the observer's "set"

for the target, which acts in a top-down manner to raise activation values for relevant attributes. Values in the saliency map are used to direct attention to the most salient item which is then compared with a memory representation of the target. If the item is not the target, the next most salient item is selected and tested for being a target and so forth. This serial search process, guided by saliency, is the hallmark of the Guided Search model.

An alternative account has been proposed by Duncan & Humphreys (1989). They suggested that items are grouped according to their similarity in an early parallel stage of processing. These groups compete for access to a limited capacity visual short term memory on the basis of the similarity between the stimuli and the memory template of the target. If the group that has won the competition contains the target, the search is terminated, if not, the selection process starts again. In the case of search for a target defined by a salient feature (feature search) the distractor items differ from the memory template for the target and so tend to create little competition for selection with the target. In addition, if the distractors are homogeneous they can be grouped and rejected together. As a consequence of both factors, search is efficient and little affected by the number of distractors present (search is spatially parallel). For targets defined only by a conjunction of features, however, search is disrupted by a variety of factors: (i) there can be grouping between target and the distractors sharing a critical feature, reducing the match between the template of the target and the representation of the display; (ii) there is little grouping between distractors that do not share features, disrupting the efficient rejection of distractors; and (iii) distractors sharing features with targets will partially activate the target's template, creating competition for selection. The net result is that conjunction search will be relatively inefficient and even spatially serial – selecting each item in the display in turn. Computational versions of these theories will be discussed below.

## **Neuropsychological disorders:**

### **Visual agnosia**

In conjunction with studies of visual attention in normal observers, one other important constraint on understanding the mechanisms of visual selection comes from neuropsychology – the study of disorders of cognition. One example of such a disorder is visual agnosia, a term used to apply to patients with a selective impairment in the visual recognition of common objects (e.g. Humphreys 1999, Humphreys & Riddoch 1993, Farah 1990). Patients with visual agnosia can have difficulties in selecting information from visual displays due to problems in grouping together relevant visual features and in segmenting apart irrelevant features. This becomes particularly apparent in visual search tasks in which target selection is aided by the efficient grouping and rejection of distractors (e.g., when distractors are homogeneous). Normal subjects show relatively flat search functions under these conditions, and absent responses can even be faster than present (due to fast rejection of distractors) (Duncan & Humphreys 1989, Humphreys et al. 1989, Quinlan & Humphreys 1987). In contrast to this agnosic patients can show inefficient search and they may fail to demonstrate any evidence of fast absent responses (Humphreys et al. 1992b). This is not due to a breakdown in visual attention per se, since such patients can conduct quite efficient serial search, when grouping processes do not act to facilitate search (Humphreys et al. 1992b). Thus the study of such patients can inform us about the contribution of grouping processes to selection. Theories hoping to provide a full account of how visual attention relates to the processes involving in grouping and object recognition need to be able to account for deficits of this sort.

## **Visual neglect**

An example of a more particular disturbance in visual attention per se is the neuropsychological disorder of visual neglect. This term is applied to patients who, following damage to temporo-parietal and fronto-parietal regions of the brain, may fail to respond to stimuli presented on the side of space opposite to their lesion (see Humphreys & Heinke 1998, Husain & Kennard 1996, Walker 1995, Robertson & Marshall 1993). Performance can be affected by the position of an object relative to the patient's body (e.g. Karnath et al. 1991, Riddoch & Humphreys 1983). This can be interpreted as an impairment of space-based selection, reflecting the coding of objects in a body-centred co-ordinate schema. In contrast, other patients neglect parts relative to an object, consistent with a spatial deficit in object-based selection. For instance, in neglect dyslexia the impaired identification of letters on the contralesional side of a word can occur even when the word is presented in the ipsilesional visual field (e.g. Young et al. 1991), suggesting that neglect is then determined by the positions of the letters within the word, rather than in a co-ordinate system centred on the body. Studies of neglect, then, enable us to explore the relations between different forms of spatial coding in visual selection. There are also other interesting aspects of the syndrome. For example, studies of word reading demonstrate that knowledge can modulate the severity of neglect, because neglect can be mitigated in reading words compared to reading non-words (e.g. Sieroff et al. 1988). Hence neglect provides a means for studying the interaction of bottom-up and top-down processes in selection. Visual extinction is typically a milder problem in visual selection and refers to effects where patients can detect a single stimulus presented on the contralesional side of space, but fail to detect the same item when it is exposed simultaneously with an item on the ipsilesional side (e.g. Karnath 1988). Recent studies of extinction have demonstrated that effects can vary as a function of grouping between elements in the ipsi- and contralesional fields. Extinction is reduced when elements group relative to when they fail to group and so are treated as separate objects (e.g. Gilchrist et al. 1996, Ward et al. 1994). Here again models need to be able to account for an interaction between grouping and selection, and how selection can be subject to a spatial bias after lesioning.

## **Attentional vs representational accounts of neglect**

Psychological theories of neglect have typically characterized the disorder as either being due to damaged visual attention or to an impaired internal representation of space (see Riddoch & Humphreys 1987, for one discussion). Arguments for an impairment in spatial representation come from studies of so-called object-centred neglect, where patients fail to report features on one side of an object irrespective of its lateral position in space (Caramazza & Hillis 1990, Young et al. 1991). In contrast, arguments for a more specific impairment in visual attention comes from studies showing that neglect can be ameliorated by cueing attention to the affected side (Riddoch & Humphreys 1983). Furthermore, deficits can be exacerbated when attention is paid to the ipsilesional side from where patients seem to have problems in disengaging attention (Posner et al. 1984). As we shall show, computational models can demonstrate that the contrast between "attentional" and "representational" aspects of neglect is false; both aspects of neglect can emerge from a spatially-specific lesion in models in which attention modulates the elaboration of particular forms of spatial representation.

## **Organization of this chapter**

The computational models we will review here have attempted to account for subsets of the properties and paradigms used to study attention. In this chapter we will introduce the models and elaborate on the explanations they give to the different aspects of attention. In the discussion section we will compare the different models. It should be noted that our

focus is on the major models used to simulate behavioural data. Hence, we exclude models which have preliminary focused on applying the idea of selective attention to computer vision problems (e.g. Tsotsos et al. 1995) or on modeling the underlying neurobiology of attention (e.g. Niebur et al. 1993). We begin by considering models applied to two main paradigms used to study attention in humans – stimulus-filtering and visual search tasks. Typically simulations of such tasks do not contain very elaborate mechanisms for object recognition. Models that incorporate procedures for object recognition as well as attention, then, provide a better account of the way in which object-based and space-based selection interact and of how top-down knowledge can modulate bottom-up biases in selection. Consequently, we proceed to discuss these models before proceeding to outline how models can be applied to neuropsychological disorders such as visual agnosia and neglect.

## 2 Computational models of visual selection

### 2.1 The Selective Attention Model (SLAM)

SLAM was proposed by Phaf et al. (1990) to model experimental findings from filtering tasks. In filtering tasks participants are asked to name a specific attribute of one of two objects in the display, with the target item being specified by another attribute (e.g. van der Heijden 1981). For instance, an experiment may contain three elementary attributes: color (red, blue), form (disk, square) and position (left, right). A typical display may consist of a red square on the left and a blue circle on the right, and participants are asked to name the color of the square. The general finding in such filtering tasks is, if items in the display possess the same response attribute (congruent condition, e.g. a red disk and a red square with task is name the colour of a square), then participants perform better than if items have different response attributes (e.g. a blue disk and red square – the incongruent condition). Phaf et al. (1990) examined whether these congruence effects could emerge in their computational model, incorporating processes that select visual information in parallel over time in a competitive fashion. SLAM’s architecture is shown in Fig. 1. Input to the model was provided by units that stood for combinations of particular attributes – form and colour, colour and position, and form and position and position. These input units fed through to feature modules, in which the units responded only to colour, form, or position. Finally, the feature modules fed activation through to motor programme modules, used for verbal responses (e.g., red, blue, circle, square, left, and right). Units with each module were inhibitory. Thus one colour (blue) inhibited other colours (red), one shape other shapes (square to circle), and activation in one location inhibited that in other locations. There were recurrent connections between the modules, forcing SLAM to converge into a consistent state, where all active units fit with the response. For instance, if the answer to the colour of the square is red, then units which stand for red, circle and left should be active whereas all other units are suppressed. Excitatory connections generated mutual support between items with shared attributes, whilst inhibitory connections generate competition between items with different attributes. As a result of the mutual support between elements shared by different stimuli, there could be facilitation when separate items processed a common response attribute (concurrent condition). In contrast, in the incongruent condition (blue disk and red square), the time taken for SAIM to converge was lengthened, due to competition between differing attributes in the same domain. Hence, SLAM simulated the basic findings of filtering tasks, where concurrence of attributes speed up reaction times and incongruent attributes delays responses.

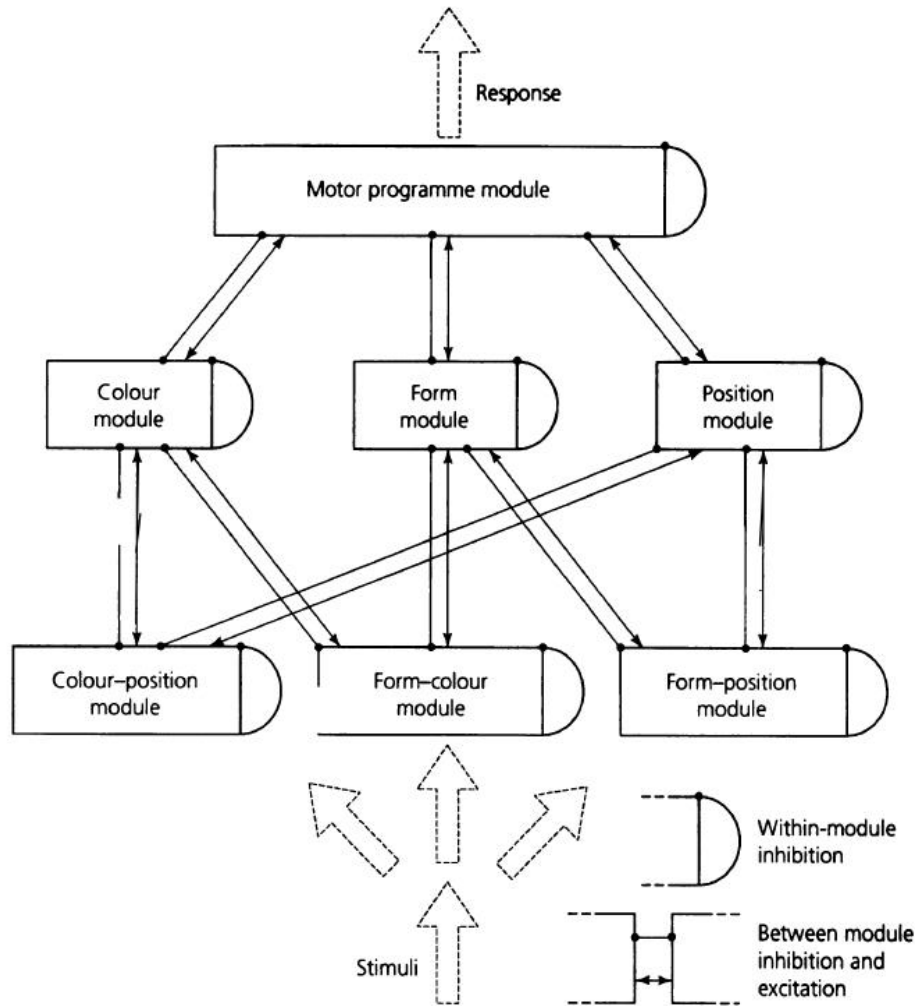


Figure 1: SLAM’s architecture consists of seven modules. All modules behave in a ”winner take all” fashion. The units in the input module encode conjunctions of color-position, form-colour and form-position, for each item. The modules at the next feature stage encode separately colour, form and position. The output stage was used to simulate motor responses in humans. There were recurrent connections between the modules, forcing SLAM to converge into a consistent state.

In a second set of simulations Phaf et al. (1990) attempted to model data from the Stroop task (Stroop 1935). In this task colour names written in different hues are presented to participants who are asked to name either the word or the hue. If asked to name the hue, RTs are delayed if the word and hue are different compared when they are the same. In contrast, if asked to name the word, participants show little effect of colour incompatibility. In order to accommodate these findings, Phaf et al. (1990) modified SLAM in two ways. First they added an input attribute, word, so that the input to SLAM comprised of three additional input modules: word-colour, word-position, word-shape. Second, in order to account for the asymmetry between colour naming and word naming, direct connections between the input stage and the motor stage were added, by-passing the feature module. For the colour naming task, however, the route via the colour feature module still determined the response of the motor programme module. This response is delayed by the strong influence from the direct routes, especially when the word name is incompatible with the colour name. In contrast, for the word naming task, this direct access process

produced responses were little affected by the colour feature module. Hence, SLAM could mimic the basic Stroop effect. In addition, Phaf et al. (1990) simulated the experimental findings when the ink colour and the word are spatially separated (e.g. Dyer 1973). In this procedure the Stroop task is turned into a classical filtering task, with subjects being required to attend to the ink colour and to filter out the word. Generally Stroop interference is reduced in these circumstances. The same result emerged in SLAM. When the colour and the word appeared at the same location, it was harder to suppress the irrelevant attribute than when the stimuli fell in separate locations, when the feedback from the position module could selectively suppress the word response. According to Phaf et al. (1990), the position-based feedback implements a type of object-based selection, because the suppression of attributes that are part of the same object is more difficult than the suppression of attributes belonging to different objects. Phaf et al. (1990) suggest, that SLAM illustrates how object-based (or attribute-based) and space-based attention can co-exist in one model, as both were simulated using the same architecture.

SLAM is a very simple model, but it nevertheless shows how cooperative and competitive interactions can generate apparent "attentional effects" in simulations. Phaf et al. (1990) argue that the model can easily be extended to process more than two items, e.g. by allowing more than two positions being encoded in input modules. However, it is unclear whether such an extended version of the model could simulate experimental results on selection with multiple items, as in visual search tasks, where intricate patterns of search functions appear. In the following section we will first discuss a serial account of such search functions and then, at the end of this section, show that also a parallel, competitive approach to selection can account for search functions. In addition, objects in SLAM have no explicit spatial coding of their parts. Nevertheless, neuropsychological evidence on neglect indicates that spatial coding within objects is important for human object recognition. As we will show in the rest of the paper, different architectures have to be used to cope with spatially extended objects.

## 2.2 Guided Search (GS) and other models of visual search

The Guided-Search model by Wolfe (1994) is an example of a class of computational models used to account for search in visual scenes using the idea of a "saliency map" which dictates the order of a subsequent serial search process. Most models within this class focus on physiological aspects of search and/or applications to computer vision (see Koch & Itti 2001, for a recent review). The Guided Search model, however, focuses on simulating behavioural data. The model assumes a two-stage model of visual selection, as has long been proposed in psychological theories (e.g. James 1890, Neisser 1967, Treisman & Gelade 1980). The first, pre-attentive stage of processing is held to be spatially parallel and to involve the computation of simple visual features. The second stage, in contrast, is held to be attentional in nature, to be spatially serial and it enables more complex visual representations (involving combinations of features) to be computed. The architecture of Guided Search is illustrated in Fig. 2. In Guided Search, the pre-attentive stage is comprised of a series of retinotopic maps that are activated in parallel by the presence of appropriate visual features (colours, lines at particular orientations and so forth). Each feature map encodes the presence of a particular feature across the display. Additionally, the activation of the maps is enhanced by the differences in features between surrounding items. For instance, if a vertical target line is surrounded by horizontal lines, then activation in the vertical feature map at the target location is higher than if the vertical line was surrounded by vertical distractors. This modulation of activation decreased with the spatial distance between any features. Apart from this bottom-up activation, knowledge of the target in visual search tasks leads to the top-down activation of feature-maps for

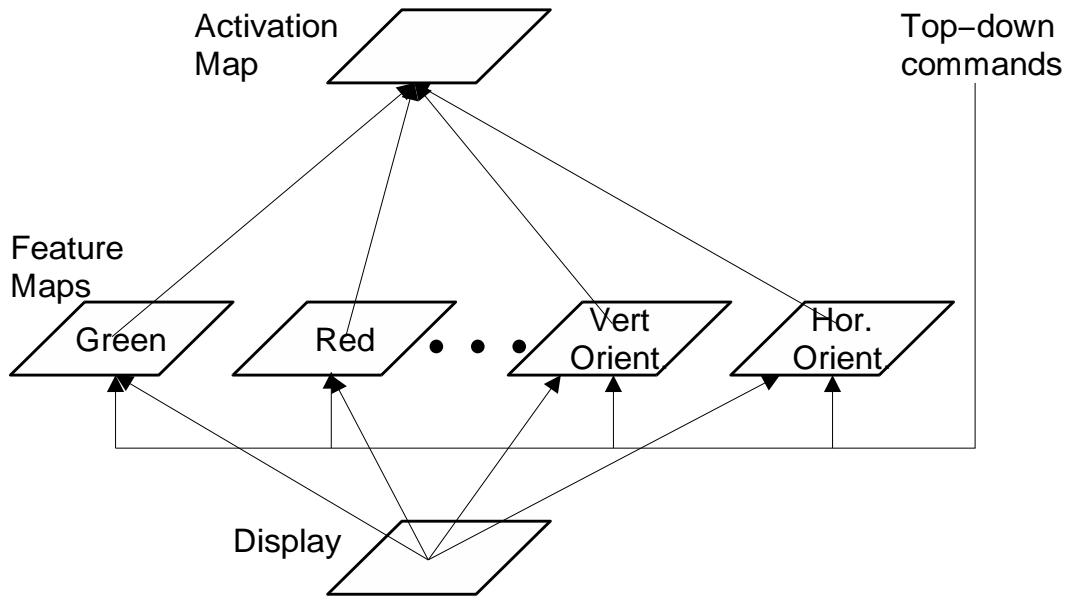


Figure 2: Illustration of the Guided Search model. In the first stage the output of feature maps results from a combination of top-down and bottom-up influences. Bottom-up activation is computed from local feature differences (see text for details). Top-down activation is based on the features of the target in the visual search. The figure gives four examples of possible feature maps: green, red, vertical line and horizontal line orientations. In the activation map the output of the feature maps is additively overlaid. Search through the display takes place in a serial fashion and follows the activation in the activation map in descending order, either until the activation falls under a threshold or the target is found.

expected target features (e.g. the maps for 'red' and 'vertical', if the target is a red vertical line). The value that features acquire, then, is a combination of bottom-up and top-down knowledge. Activation values for elements in each feature map are used to signal locations that are worthy of further attention. The greater the activation at a location, the more likely it is that attention will be directed to that location. The activation from all feature maps is topologically summed within an "activation map" which represents the overall saliency of locations in the display. Attention is directed to the location in the field where saliency is highest. If this location contains the target of the search task, the search is terminated. If not, further deployment of attention and subsequent tests for target follow the activation of the saliency map in descending order. This serial search terminates if either the target is found or if the activation of the unattended locations in the saliency map falls under a threshold (see Wolfe 1994, Chun & Wolfe 1996, for details).

Guided Search can be applied to data from visual search in the following manner. Targets that are dissimilar from distractors will produce high activation in their own feature map and within the overall saliency map. Hence, search is then efficient. This allows Guided Search to account for flat search functions from feature search with dissimilar target-distractor pairs (e.g. a horizontal line target amongst vertical distractors). However, if the targets and distractors are similar to each other as in so-called conjunction search tasks, when the critical features that specify the target are also borne by distractors (e.g. when the target is a red and vertical line and the distractors are green vertical and red horizontal lines), the corresponding feature maps do not indicate the location of the target, since the features of the target are similar to the distractor features in the same feature maps. Consequently, the summed activation from the feature maps in the saliency map does not highlight the target in the saliency map. Indeed, Wolfe (1994) suggested that



neural noise in GS would then lead to distractors being selected before targets, since their saliency value may be higher. As a consequence, serial search functions result. Top-down activation, for an expected target can modify this to some degree, since it can activate the features belonging to the target (red and vertical, in our example). The target, having both these features, can still achieve a higher value in the saliency map than distractors which each have only one of the critical features. This can explain why conjunction targets that have features quite different from those of each half set of distractors can be detected efficiently (Wolfe et al. 1989). However, if the feature values of the target and each half set of distractors become close, then distractor feature maps may be activated along with the target feature map, losing the target's advantage within the saliency map. This is lost and serial searches are made.

Guided Search is able to account for the major contrast between "single feature" and conjunction search, and by means of its top-down activation, it is able to accommodate findings of efficient conjunction search when stimuli have quite different values along their feature dimensions. However there are weaknesses. One is that the mechanism for serial search is not implemented in a manner consistent with the connectionist architecture of the earlier stages of the model. Another is that the model does not have procedures that enable visual features to be grouped, to allow object representations to be constructed. Yet there is considerable psychological evidence that grouping affects early parallel stages of vision – for instance, targets and distractors that are distinguished only by how their features group can still yield parallel search functions (e.g. Enns & Rensink 1991). Some targets can only be distinguished from distractors when the stimuli group to form discriminably different objects (Humphreys & Donnelly 2000, Donnelly et al. 1991), and targets can be difficult to detect, if they group with distractors (Rensink & Enns 1995).

Two connectionist models similar in flavour to Guided Search are the Feature Gate model of Cave (1999) and the dynamic search model of Deco & Zihl (2001).

As outlined by Cave (1999), the Feature Gate model uses a pyramidal architecture of feature processing to model visual selection. At each level of the pyramid items are selected and via an "attentional" gating mechanism passed on to the next level. This selection mechanism led to a successively reduced number of items and finally to one selected item. Similar to Guided Search the selection process at each level was guided in a bottom-up and top-down way. As Cave (1999) pointed out, this architecture is not only capable of simulating the same experiments as Guided Search, but several additional experimental findings on divided attention and Eriksen flanker interference. The effects of divided attention are reproduced by gating items through the pyramidal structure from spatially separate locations. The Eriksen flanker interference results from the fact that the attentional gates also pass on information from adjacent locations. Consequently, the flankers will somewhat influence the activation on the top level of the pyramid and delay or speed-up the response of Feature Gate, depending on the compatibility of the flankers. The model by Deco & Zihl (2001), like Guided Search, uses feature maps which can be modulated in a top-down fashion and sums output from feature maps in a "high-level" map. In addition, each feature map implements competitive interactions. This competition leads to a convergence of activation with one winner in the "high-level" map, marking the location of the target. Reaction times are expressed in terms of the number of iterations that it takes the model to produce a winner in the high-level map. Deco & Zihl (2001) reported that the speed of convergence matches the common finding that conjunction searches are slower than single feature search. In a feature search task the top down influence suppresses all distractors, making the convergence of the model independent of the number of elements in the input. In a conjunction search the top-down influence suppresses the elements in each feature map that do not have target features, but conver-

gence within each feature map is still affected by the presence of distractor elements that have the same features as targets and so are not subject to top-down suppression. As a consequence, the convergence of the high-level map was found to depend on the number of items, producing a linear search function. In essence, Deco & Zihl (2001)’s model shows that competitive interactions are capable of simulating search functions even when there is no ”spotlight” of attention, so that the serial search mechanism as assumed in Guided Search is not essential. However, Deco & Zihl (2001)’s model remains limited to dealing with simple visual features. In the following sections we will discuss models that can cope with more complex objects.

### 2.3 Search via Recursive Rejection (SERR)

SERR (Humphreys & Müller 1993) was set up to simulate effects of grouping on visual search as suggested by the psychological model of Duncan & Humphreys (1989). The model was mainly based on the architecture of a Boltzmann machine (Hinton & Sejnowski 1988). In models of this type, units behave in a stochastic fashion with a two-state output activation, either active or inactive. The probability of assuming one of the two states is determined by the size of the input a unit receives from other units. The model used an hierarchical organization to simulate visual search for letter-like stimuli made of horizontal and vertical line elements. Formally, search for a target defined by a particular combination of such features relative to distractors (e.g., a L target vs. T distractors) can be described as a conjunction search task, since each letter is composed of the same, more primitive line elements. However, provided the distractors are homogeneous, search can be relatively efficient (Humphreys et al. 1989). Search becomes inefficient primarily when the distractors are heterogeneous (e.g., a L target vs T and  $\perp$  distractors). This suggests that search in this case is strongly influenced by grouping between distractors. The first set of units in SERR’s hierarchy coded the presence of horizontal and vertical edges on a retina (see Fig. 3). Activation from these units was transmitted through to units that coded combinations of the first sets of units (L-junctions, line terminators), and activation from these units was transmitted to further representations of T-junctions formed by combining outputs from particular L- and line terminator units. In each case, a map of units was constructed to represent elements across the visual field. Activation from the T-junctions was then fed into template units, which acted to detect the targets and distractors in the experiments. The target might be a T in one orientation and the distractor Ts in other orientations. In order to capture grouping effects, excitatory connections were established between T units within the same map and inhibitory connections between units standing for the same location but in different maps. Thus identical Ts supported one another in the search, and heterogeneous Ts competed to form separate groups. Additionally the excitatory connections within the T maps were gated by a ”location map” allowing activation only to spread to locations where there was something in the visual field. This was done in order to suppress possible ”hallucinations” of elements produced by the grouping mechanism. Activation from the letter maps was transmitted through to template units, set up according to the particular targets and distractors that might be used in search. At the template level, there was rapid accumulation of evidence for identical Ts, and slower accumulation of evidence for each type of T in a heterogeneous set. The template units themselves were also competitive, so that only one such unit was incremented on each iteration of the network. A particular target or distractor was said to be detected when its template unit reached a threshold. If the first template to reach threshold was the target, search stopped. If it was a distractor, then the template and associated T units were inhibited and search continued over the remaining elements until either a target was found or all the items had been rejected.

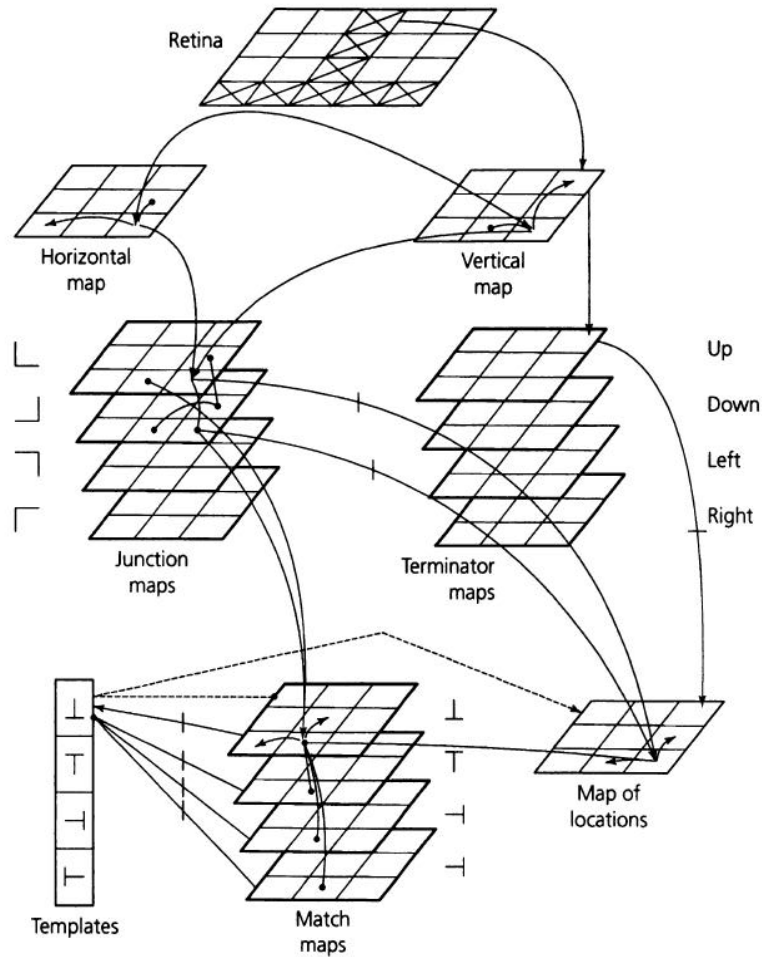


Figure 3: The architecture of Search via Recursive Rejection (SERR) model uses a hierarchical organization to simulate search for letter-like stimuli. The first two maps in SERR’s hierarchy coded the presence of horizontal and vertical line elements. Activation from these units is transmitted through to maps which encode L-junctions and line terminators, and activations from these maps are combined to activate units corresponding to T junctions in the match maps. Activation from the T-junctions is then fed into template units. The template units accumulate the evidence for each type of T, signaling the presence of a particular item when the relevant template reaches threshold. In order to capture grouping excitatory connections were established between T units within the same match map and inhibitory connections between same locations across different matched maps. Additionally, the excitatory connections within the match maps were gated by the location map allowing activation only to spread to locations where there was a stimulus on the retina.

SERR was able to simulate patterns of visual search found in human data with identical (homogeneous) and heterogeneous distractors. With identical and non-identical distractors, there was rapid grouping and rejection of the background items, enabling targets to be detected efficiently, leading to a efficient search. In contrast, with heterogeneous stimuli several groups of distractors could be selected before the target might be found. Hence, the time to detect the target depended on the number of groups that had to be selected serially, before the target was detected. In addition, given the stochastic behavior of the model, there was some probability that a target present in the field may not be found. This was particularly the case with heterogeneous distractors, which competed strongly

for selection. With heterogeneous distractor groups a target could be temporally inhibited (its corresponding units placed in an inactive state). If the distractors were subsequently selected (and rejected as being nontargets), the model could decide (incorrectly) that no target was present. To prevent errors (false rejections) of this type, search would have to be repeated. When such repetitions (re-checks) were conducted to match the error rates found in human data, then search with heterogeneous distractors was a linear function of the number of distractor present.

In subsequent work Müller et al. (1994) showed that SERR generates a strong reduction in search time when there is more than one target present (e.g. there is a redundancy gain). Müller et al. (1994) found that when two targets were present, search became faster than was ever the case even on the fastest search trials when a single target was present. This "super additive" redundancy gain is difficult to explain in models assuming strict serial search of the items present, but it is predicted by SERR (see Miller 1982). (However, see Townsend & Nozawa (1997) for a different view which argues that also serial models can produce super additivity). In SERR search benefits from the presence of two targets in two ways: (i) multiple targets form mutually-supportive groups and (ii) there is a faster accumulation of activation within template units from each grouped target. The "super additive" results generated by SERR occur because the model has a strong parallel component, accumulating activation simultaneously across the field. In addition, though, SERR incorporates a serial mechanism of successive rejection of selected items, and it can employ recursive re-checking. Hence, SERR can be classified as a hybrid model of visual search. On one hand, it incorporates parallel, competitive selection and grouping processes and on the other it applies search serially through groups and it applies serial re-checking. The ability of the model to simulate linear search functions provides an existence proof that such functions are not necessarily an indication of a strict serial (item-by-item) search process as, for instance, in Guided Search.

One interesting aspect of the work carried out with SERR is the link between computational modeling and experimentation. For example, SERR's predictions concerning redundancy gains from multiple search targets were verified in empirical studies of human search conducted by Müller et al. (1994). Other predictions derived from the model were examined by Humphreys & Müller (1993). One such prediction concerns search performance when re-checking processes are prevented by making people respond to a fast deadline. In SERR, this re-checking process takes the form of repeating search until the error rate declines (see above). When such repetitions are decreased (e.g. because a fast response deadline has to be met), then the error rates rise, particularly when large numbers of distractors are present. Humphreys & Müller (1993) found a similar pattern of errors at larger display sizes in human search under fast deadline conditions. Thus, the model served not only to simulate data, but also as a source of novel empirical predictions. Despite SERR's success both in capturing prior results and in predicting novel findings, the model is limited in several aspects. For example, it employs a fixed processing architecture useful only for searching through T-like targets and distractors. Also, although SERR groups elements to form large perceptual units, these groups are influenced by the local identity of features; they do not serve the process of putting parts together for object recognition. Only identical features are grouping, and there is no grouping between different elements that would together form a shape (e.g. four corner brackets making square). This constrains the application of the model to a broader set of data on the inter-relations between object recognition and attention. The model also encounters problems in "scaling up", since each feature combination is "bound together" by activating a local unit that represents the presence of that combination at a given location in the field. Clearly, this would require a large number of units to represent all possible feature combinations at all possible locations. This issue is essentially concerned with how feature combination can

be "bound" together. SERR accomplishes binding through allocating units in its hardware. We consider next some models that accomplish binding in a different manner and that attempt to address the broader issues concerning the relations between recognition and attention.

## 2.4 Multiple Object Recognition and attentional selection (MORSEL)

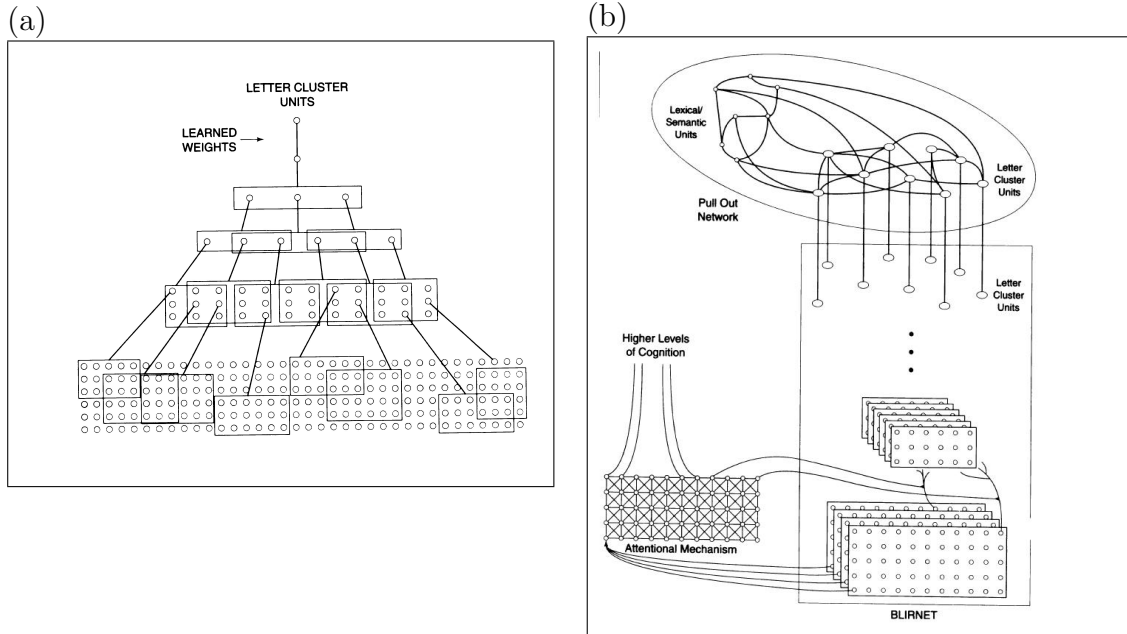


Figure 4: (a) The object recognition module (or termed BLIRNET) in MORSEL. Object recognition was achieved using a hierarchical coding schema, progressing from position-bound detectors for line orientation and colour, to units at higher levels of the hierarchy which respond to combinations of features, and which are less bound to the locations of features on the retina.

(b) The pull-out network (top) and the attentional mechanism (AM) (bottom) in MORSEL. This "pull out" network uses stored (lexical) knowledge to "clean up" the distributed code that is formed in a bottom-up manner. The attentional mechanism (AM) modulates activation in the recognition system by raising activation levels for stimuli in attended parts of the visual field over activation levels for stimuli in unattended regions. This space-based, attentional modulation reduces the likelihood of distributed features from multiple stimuli being combined incorrectly to form migration errors. External input into the AM was formed by summing the output from the line-orientation detectors at the first stage of recognition process, for each location in the field. Thus the external input into the AM network was insensitive to feature properties.

The MORSEL model, developed by (Mozer 1991, Mozer & Sitton 1998) linked visual attention to object recognition, to provide an explicit account of the inter-relations between these two processes. MORSEL essentially contained two modules – one for object recognition (derived from an earlier model termed BLIRNET (Mozer 1987)) and one for visual attention. Object recognition was achieved using a hierarchical coding schema, progressing from position-bound detectors for line orientation and colour, to units at the top end of the hierarchy which respond to combinations of features in a nearly translation-invariant manner (irrespective of the lateral position of items in the field). At each level of the recognition hierarchy, features and locations were pooled so that higher-order units

respond to increasing complex features and feature-combinations in a manner that gradually becomes less location-dependent. Many of the simulations performed with the model have been concerned with word recognition. For such simulations, units at the top-end of the hierarchy respond to clusters of letters and letter features. These features are "glued" together using an independent "pull out" network, trained to classify the distributed representation formed in the recognition system. This "pull out" network can be thought of as using stored (in this case lexical) knowledge to "clean up" the distributed code that is formed in a bottom-up manner (though note that this remains a feed-forward system; activation from the pull-out network is not used in a top-down manner, to influence feature coding in the model). As Mozer (1991) pointed out, this process of using distributed features to activate stored knowledge, can introduce some computational problems in its own right. Consider the problem of word recognition, when there are several words in the field. Each word may be recognized on its own based on a unique combination of distributed features components. When there is activation from multiple words, though, components in each word may activate some new lexical representation, so that the letters could be miscombined in perception to form a word not actually present. In perceptual report, letters could "migrate" from one letter string to another. There is some evidence from psychological studies that indicated that these migration errors can occur when words are briefly presented (see Mozer 1991). To overcome this problem, a separate "attentional" mechanism (AM) was added. The goal of the attentional mechanism was to modulate activation in the recognition system by raising activation levels for stimuli in attended parts of the visual field over activation levels for stimuli in unattended parts of the field. The main effect of this space-based, attentional modulation was to reduce the likelihood of distributed features from multiple stimuli being combined incorrectly to form migration errors, since only features of the attended object would be active. This architecture is similar to the structure of the Feature Gate model, earlier introduced in this chapter, but in Feature Gate an attentional modulation operate at every level of the recognition network (Cave 1999). Within the AM network, units had local, excitatory connections, whilst more distant units had inhibitory connections. This connectivity with the network led to activation converging into an "elastic" (as opposed to fixed size) spotlight, whose size and position depended on external inputs. Input into the AM took the form of summed outputs from the detectors at the first stage of recognition process, for each location in the field. Thus the AM network was insensitive to feature properties, responding only to the presence of some "thing" at a given region of field. In addition, Mozer & Sitton (1998) assumed that the input of the feature maps into the AM network could be modulated in a top-down manner from higher cognitive levels (e.g., when a target could be in a particular colour).

Mozer & Sitton (1998) showed that MORSEL could mimic the dichotomy between feature search and conjunction search for oriented colour elements. For the feature search task, the top-down influence suppressed the features of the distractors, so that only the target provided an active input to the AM. This caused the AM to select the target within the same time period independent of the number of items present in the display. Hence, MORSEL could produce the flat search function typical for feature search. For conjunction search, the top-down signals biased the AM to the target location by (i) activating feature maps containing target features and (ii) suppressing feature maps containing distractor features. However, provided the top-down bias was not overwhelming, distractors still influenced the selection process and the search function showed the usual dependency on the number of items due to competition in the AM. The top-down processing in MORSEL is comparable to the top-down modulation in the Guided Search model.

One interesting aspect of the simulations of conjunction search in MORSEL is that contrasting effects of the distractors were found as a function of whether many or few dis-

tractors were present. With few distractors, search was efficient and little affected by the number of targets. With many distractors, search times increased as a function of each extra distractor. Mozer & Sitton (1998) attributed this rise in search times to increased competition between items in the field, and to increased crosstalk within the recognition network (since when there are more items in the field, inter-item distances will be reduced, so higher-level units are more likely to be activated by neighbouring features). Mozer & Sitton (1998) related these results to data on human visual search by Pashler (1987) who reported a "dog leg" function for conjunction search – reaction times were initially little affected by distractors but then increased steeply when more distractors were presented. The same result has been interpreted by at least some investigators as indicating effects of crowding on search for conjunction targets, due to the increase likelihood of features being mis-combined when stimuli are in close spatial proximity (see Cohen & Ivry 1991). MORSEL provides an explicit account of such feature miscombinations, and of how attention is required to reduce such miscombinations, especially with displays with closely proximal items.

In addition to simulating visual search data, Mozer & Sitton (1998) demonstrated the generality of their approach by showing, how MORSEL could be used to simulate other standard paradigms used in attention research, including: studies of spatial attention using visual cueing and distractor-interference in stimulus-filtering tasks. Visual cueing was simulated by presenting a brief peripheral pre-cue prior to the presentation of a target at either the cued or an uncued location (on valid and invalid trials). Pre-activation of the target's location in the AM facilitated target selection, since less activation was then required when the target appeared. In contrast, invalid cues created some competition for selection of the target's location. The difference between RTs on valid and invalid cue trials mimics human data (Posner et al. 1980). As we have noted earlier, in stimulus filtering tasks distractors can interfere with responses to targets, demonstrated by slowed RTs when targets and distractors require opposite responses (Eriksen & Eriksen 1974). Mozer & Sitton (1998) showed that such competition effects could emerge in MORSEL provided that attention did not have a sharp spatial impact on processing in the recognition module, otherwise no interference would occur. As Mozer & Sitton pointed out, this delivers computational support for the zoom lens model of Eriksen & Yeh (1985), where it was assumed that attention does not focus exclusively on the target and, therefore, allow interference from distractors to occur.

MORSEL is an attempt to link attention to wider aspects of object recognition. However, it conceptualizes attention as a separate mechanism modulating object recognition. The following section will introduce a computational model which combines attention with object recognition in a more interactive way.

## **2.5 Selective Attention for Identification Model (SAIM)**

One other model that has attempted to marry object recognition to visual attention is SAIM (Heinke & Humphreys, in press, Heinke & Humphreys 1999, Humphreys & Heinke 1998, Heinke & Humphreys 1997). The architecture of SAIM followed an idea of "dynamic routing circuits" put forward by Olshausen et al. (1993, 1995). They proposed that attention can help to generate translation-invariant pattern recognition by serving as a "window" through which activation from a retinal field must pass before being transmitted through to recognition systems. Provided the window can be shifted around to different spatial positions, the same units will modulate recognition, irrespective of the lateral positions of objects in the retinal field. Consequently, recognition units do not need to be coded for every location in the field, since the same input will occur across different retinal positions, normalized by the shifting attentional window. To achieve this shifting

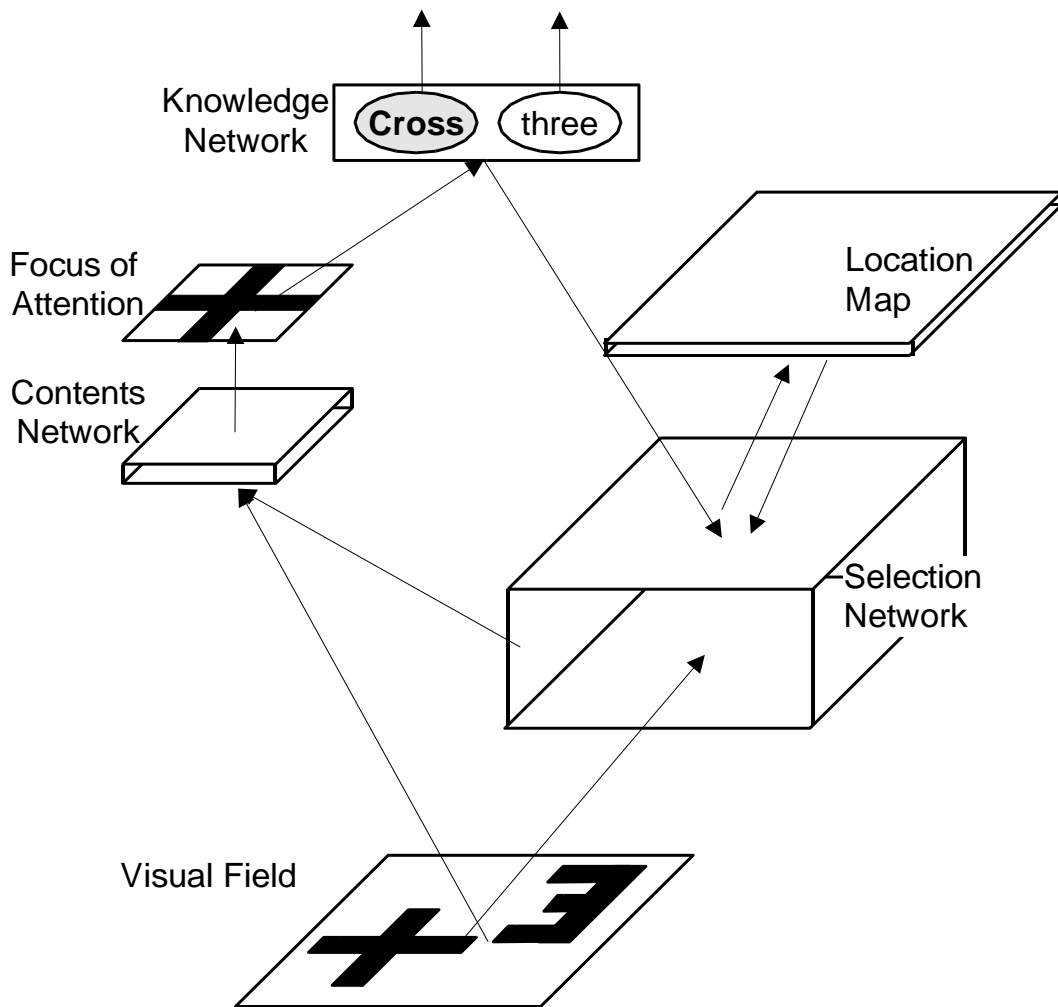


Figure 5: The Selective Attention for Identification Model (SAIM) achieves translation-invariant object recognition with three modular networks: The contents network maps the contents of the visual field into the focus of attention (FOA). This mapping is translation-invariant. The selection network determines the location in the visual field where elements in content network are mapped into the FOA. The knowledge network possesses stored-knowledge in form of template units and recognizes the contents of the FOA via simple template matching. These three component operate in parallel. The location map stores the locations of objects already attended to and prevents these locations from being selected again. This determines attention switching behaviour, with old items not being attended for a second time.

of an attentional window across space, Olshausen et al. (1993) proposed the existence of two sets of units: one set of units maps the contents in the visual field into a attentional window or "focus of attention" (FOA). The second set of units controls the shifting of the FOA. In SAIM these two sets of units were termed the "contents network", representing a partial contents of the visual field, and the "selection network", modulating the units in the contents network (see Fig. 5). In SAIM a third network, the "knowledge network", comprised of a set of template units corresponding to stored memories for objects. This was added to Olshausen's architecture to integrate a object recognition system into SAIM, and it enabled SAIM to simulate effects of stored knowledge on selection.

To achieve translation-invariant object recognition, the activation in the selection network had to fulfill several constraints: (i) The contents in the visual field should be mapped only once into the FOA. (ii) The contents of the FOA should cover only neighbouring areas in the visual field. These constraints led to a network of competitive and coopera-



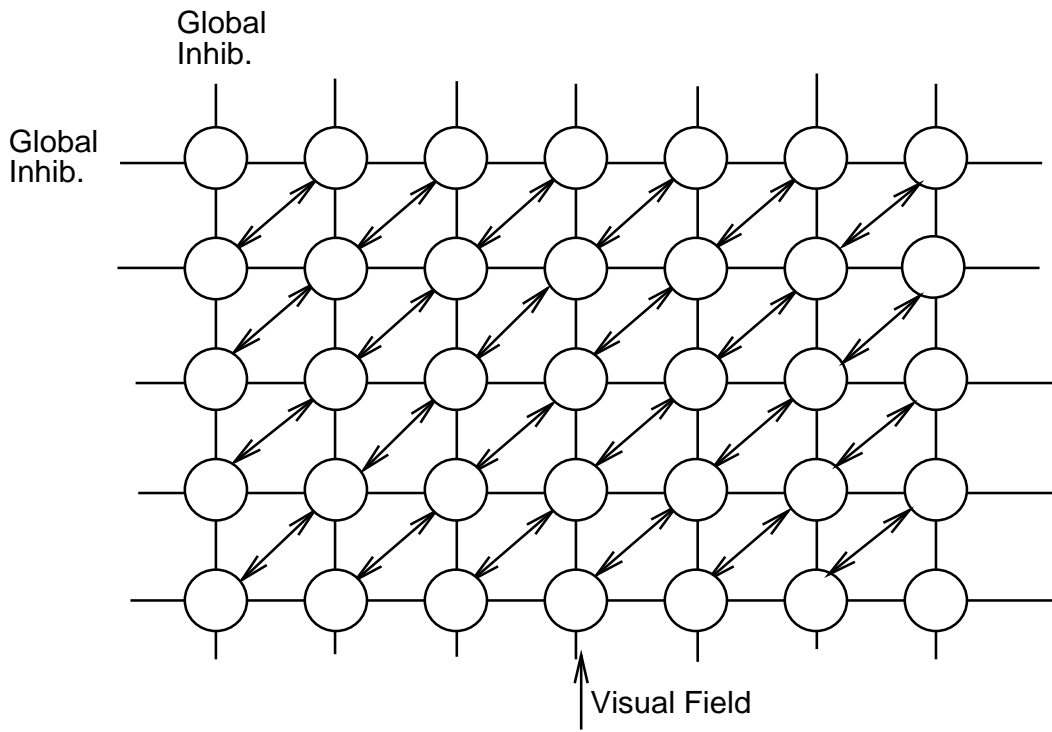


Figure 6: An illustration of the topology of the selection network, shown for a one-dimensional visual field. The horizontal and vertical lines between units stand for inhibitory connections within the rows and columns of the network. The arrows stand for excitatory connections between units on diagonals of the network. Every unit receives an input from the visual field in a topological fashion. The connections ensure that the selection network obeys a set of constraints when activating units (see text for details).

tive interactions, as depicted in Fig. 6. This step followed an idea of using connectionist approaches for soft constraint satisfaction first suggested by Hopfield (1984). The excitatory interactions ensure the activation of neighbouring units in the selection network, implementing the second constraint. In addition, the first constraint was wired into the selection network through inhibitory interactions limiting the number of units in the selection network allowed to be activated and consequently restricting the area SAIM can select. This network of inhibitory connections is illustrated in Figure 6. Each row in the network contains units that are responsible for the contents of one location in the FOA. Inhibitory connections between these units ensure that only one location in the visual field is mapped into a given location in the FOA. Similarly, competition between units in the columns of the network ensures that a given position in the visual field is only mapped into one location in the FOA. The remaining winning units determine the activation transmitted into the FOA through the contents network. Activation in the contents network is multiplied with the activation from the selection network, so that activation from the visual field is only transmitted to the FOA if it is supported by active units in the selection network. When two or more objects are presented, there is competition between the activation they create in the selection network. When top-down support is not taken into consideration (see below), the object that tends to be selected first is the one whose local parts generate most mutual support – either because the object is large or because the parts are arranged quite close to one other. Interactions in the selection network also typically lead to the FOA being aligned with the centre of mass of the selected object. In addition to operating in a bottom-up manner, SAIM differs from the other models considered here in that it used top-down object knowledge to influence selection directly, by

modulating activation within the selection network. This top-down influence originated from the knowledge network, where stored-knowledge was represented by means of a set of template units, connected to units in the FOA. The introduction of the knowledge network adds further constraints to the behaviour of SAIM. Here, the main constraint is that the template unit that best matches the contents of the FOA should be a maximally activated and all other units of the knowledge network should be inhibited. This constraint results in two types of connectivity: First a simple competitive interaction within the knowledge network and second a top-down modulation of the selection network from the knowledge network. This top-down modulation biased the selection network to favor familiar over non-familiar objects.

Having selected one of several objects present, SAIM also incorporated a mechanism for switching attention from a selected object to others in the visual field. This was accomplished by activating units in a "location map" corresponding to the position occupied by the object that was selected. Once a location unit was activated, it acted to inhibit corresponding positions in the selection network, so that these positions could no longer win the competition to control the mapping of stimuli through to the FOA. At the same time, the template for the selected object was also inhibited. Consequently, another, previously unattended, object then won the competition for selection, so that attention switched between objects. This process mimics a form of "inhibition of return", in which attention is biased against returning to locations containing objects that have already been selected (e.g. Klein 1988, Posner & Cohen 1984).

The performance of SAIM was evaluated by measuring the number of iterations that it took a template unit to reach threshold (the reaction time (RT); Heinke & Humphreys, in press). SAIM showed a bottom-up bias towards large relative to small objects and to objects in which the pixels were densely placed around their centre of mass. These effects are emergent properties of the proximity-constraints built into the selection network. Such bottom-up biases generate efficient selection of large, densely-packed stimuli, similar to the evidence of "pop out" for salient targets in human visual search (Treisman & Gelade 1980). In addition a variety of other characteristics of human selection were simulated. For example, Duncan (1980) reported a "multiple object cost", where selection was shown to suffer when multiple relative to single stimuli had to be reported. This cost was most pronounced when the reported elements come from separate objects, and it was not evident when the elements belonged to a single object. SAIM manifests similar effects. There is a cost for the second of two selected objects, when compared with when only that object is presented in the field. However, all the parts of a selected object tended to become available together in the FOA, so there were few costs for reporting multiple attributes of attended objects. In this sense, SAIM manifested a form of object-based selection, even though this was modulated through spatial attentional window. This was also demonstrated in simulations of the effects of spatial cueing. Basic cueing effects were modeled by a briefly presented a cue at a location field followed by a target at either the same or a different location. RTs were facilitated when cues and targets fell at the same location, compared with when they appeared at different locations, capturing data such as those reported by Posner et al. (1980). To examine whether object coding modulated such effects, cues were presented at locations already occupied by an object in the field. On invalid trials, the target could then appear either in the cued object, or at the same distance away but in a previously uncued object. Heinke and Humphreys (in press) showed that the spatial effect of the cue could combine with activation from the cued object, to generate an advantage for targets that fell within the cued object relative to those that fell elsewhere. This simulates the data of Egly et al. (1994), typically interpreted in favor of an object-based selection process. Other forms of object-based bias were demonstrated by varying where a cue fell within a subsequently presented shape.

Cues were most effective, if they fell at the centre of mass of a target, due to SAIM's bottom-up bias to centre its FOA at this position in a shape. Similar effects in human selection were reported by Pavlovskaya et al. (1997).

Object-based influences were apparent not only in how attention was allocated within shapes, but also in attentional biases following selection. As we have noted, inhibition of return was implemented by location-based inhibition of selected positions and inhibition of the templates for an attended object. This led not only to attention being biased against selected locations, but also against the objects that formerly occupied those locations, even if they subsequently moved to a new location. Evidence for object as well as space-based influences in inhibition of return has been documented (Posner et al. 1980, Tipper et al. 1991, Tipper & Behrmann 1996, Tipper et al. 1999). Like MORSEL, SAIM was also able to model forms of filtering by stimulus properties, using space as the medium for selection. For example, the target in an Eriksen flanker task could be selected by pre-activating a location where it would appear, though, this activation could not be so high that activation in other (distractors) positions was prevented from creating competition for selection. When a distractor matching the target was presented, then this overlaid the target's activation in the FOA, facilitating selection. In contrast, a mismatching distractor could create transient activation in the FOA consistent with its features rather than those of the target. This was sufficient to create competition too at the template level, if the the target and distractor had different identification responses. A response competition effect was consequently apparent.

As we have noted, bottom-up biases with the model could be modulated by the top-down influences from the knowledge network. If a target template had a high activation value from the outset of a trial (e.g., due to priming), or if it had high weights from the FOA relative to other templates, then selection would be biased toward this target when compared with other stimuli. This suggests that selection should be biased towards more familiar or recently primed objects. The flip side of this is that familiar objects may be difficult to ignore when the task is to select the less familiar of two stimuli; also objects similar to the primed stimulus may be misrecognized as the primed item. Again, there are psycholgocial data on (i) biases to select stimuli made familiar by consistent mapping training as targets (Shiffrin & Schneider 1977), and (ii) on false positives when targets are similar to primed representations (Schvaneveldt & McDonald 1981).

In a model such as SAIM, there is a danger that, when there are multiple stimuli in the field, features from different items could be bound together since they activate higher-order representations in parallel. This binding problem is solved by spatial attention, which limits activation from unattended items. This is similar to the view incorporated into a psychological model such as "feature integration theory" (Treisman 1998), in which spatial attention is used to filter out distractors, so that only the features of attended targets are available for binding. However, since SAIM only implements how form information may be coded in vision, it is unable to capture many of the results where feature integration theory has been applied, involving studies in which multiple stimulus dimensions have been varied (e.g. colour as well as form).

### 3 Neuropsychological deficits

We have discussed how, in addition to the paradigms used to explore visual attention in normal observers, important evidence on the nature of attention comes from the study of neuropsychological deficits. In the following section we will discuss how the models presented earlier account for neuropsychological deficits.

#### 3.1 Visual agnosia

Damage to occipital-temporal areas of the brain is associated with visual agnosia – the impaired recognition of visually presented objects (e.g. Humphreys 1987, Humphreys & Riddoch 1993). Humphreys et al. (1992b) showed that agnosic patients had impaired visual search, but only under conditions in which grouping normally generates efficient search. This has been modeled by SERR (Humphreys et al. 1992a) with the effects of the brain lesion simulated by adding noise to the Boltzmann activation functions in the model. The task was to detect a form conjunction target  $\perp$  presented amongst homogeneous distractors (T's containing the same features as the target). When the activation functions were noisy, units corresponding to the stimuli had some probability of not being activated even when the appropriate were present in the field. Similarly there was an increased probability that units corresponding to other stimuli were activated. The net result was that grouping between identical distractors was disrupted, making search inefficient. In contrast to this there was little effect of increasing the noise in the model on search for targets amongst heterogeneous distractors (e.g.,  $\perp$  vs  $\top$ ,  $\sqsubset$  and  $\sqsupset$ ). With heterogeneous distractors, grouping effects could disrupt target detection, especially if targets grouped with distractors. In this case, any loss of the beneficial effects of grouping were counteracted by positive effects of breaking-up inappropriate target-distractor groups. Consistent with these simulation results, Humphreys et al. (1992b) reported that search for form conjunction targets amongst heterogeneous distractors can be relatively little affected in agnosia (Humphreys et al. 1989). Often one might think that the effects of neural damage are simply to degrade processing, so that the more difficult of two tasks becomes even harder. The selective deficit on search with homogeneous, compared with heterogeneous, distractors contradicts this since the easier task is then affected. This indicates that rather than generating a general deterioration, the brain lesion affects a specific process: grouping based on complex form representations. The effects of lesioning SERR likewise had a selective effect with homogeneous rather than with heterogeneous distractors, due to the disruption of grouping processes.

#### 3.2 Unilateral neglect and extinction

Damage to temporo-parietal and fronto-parietal regions of the human brain can be associated with unilateral neglect (a failure to respond to stimuli contralateral to the side of lesion) and extinction (a failure to respond to contralesional stimuli when ipsilesion stimuli are simultaneously present; see our earlier discussion). These disorders have been simulated in several of the models that we have reviewed.

##### **Extinction**

Basic forms of extinction have been modeled, like neglect, by having lesions alter the spatial competition between stimuli in the visual field. In some instances, the lesions have simply been less severe to produce extinction, though in a model such as SAIM neglect and extinction can come about through lesions to different components of the model. We

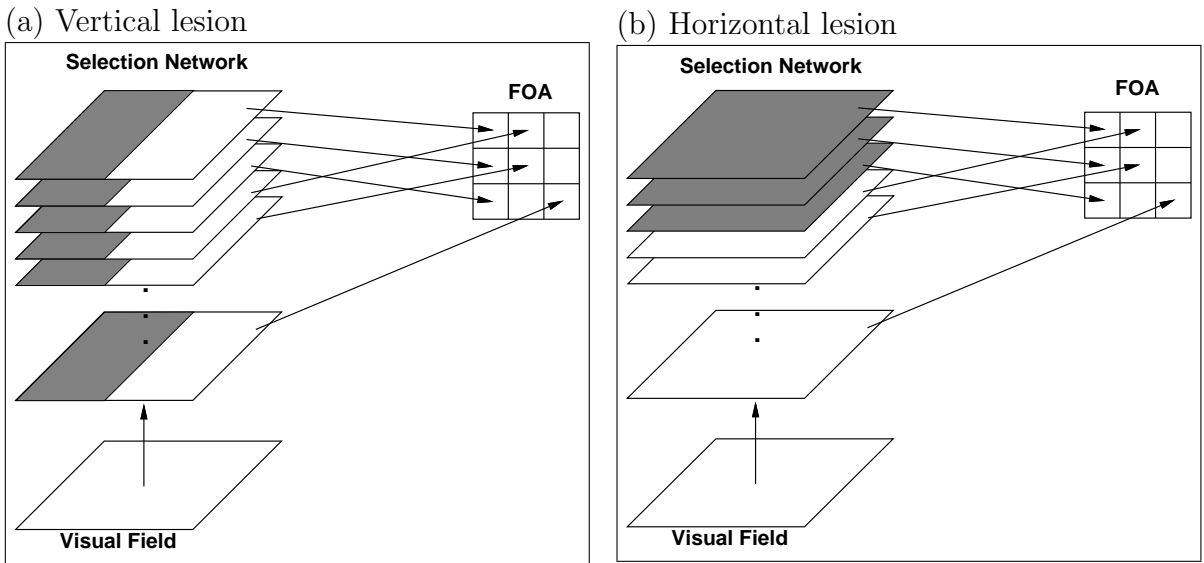


Figure 7: Illustration of the "vertical" and "horizontal" lesions used in SAIM. Each layer in the selection network in this illustration is responsible for the contents of one pixel in the FOA. The dark shaded areas illustrate the areas of lesions (reduced weights on connections to those units). The horizontal lesion affects the part of the selection network that is related to the left part of the FOA and this lesion is invariant with respect to the visual field. The vertical lesion affects units responding to one side of the visual field. For a vertical lesion, the mapping into the whole of the FOA can be affected, but only for the left part of the visual field.

return to discuss this point at the close of this section. Humphreys et al. (1996) introduced spatially-specific lesions into the map locations, or into the maps coding the form conjunctions, in SERR. In this case, the lesions reduced the probability of units going into an active state. When a single stimulus was presented on the affected part of space, the lesion was not sufficient to prevent recognition. However, when a different stimulus was placed in the unimpaired field, it exerted an inhibitory influence on the contralesional item, decreasing the likelihood that it would be detected. Extinction resulted. Humphreys et al. (1996) also noted that, when the contra- and ipsilesional stimuli were the same, extinction could be reduced because the ipsilesional item than supported detection of the contralesional stimulus, due to excitatory connections between items represented in the same form map. As noted in introduction positive effects of grouping on extinction in patients have been reported by (amongst others) Gilchrist et al. (1996), Humphreys & Heinke (1998) and Ward et al. (1994). Such grouping effects arise as a natural consequence of positive interactions in the model between elements that can combine to form larger objects.

MORSEL and SAIM also simulate extinction in quite similar ways. In MORSEL, extinction can be captured by introducing a spatial lesion into the attention module, so that units on one side of space generate less activity than those on the other side (Mozier 1991). Provided that the lesion is not too large, a single item may still be detected on the affected side whilst being extinguished by competition in the attention map when another item appears on the unaffected side. MORSEL is also able to explain why extinction effects are reduced when elements are part of a familiar object, compared with when they do not. Parts that belong to the same object can be recovered together by the pull-out network, even if one part is degraded (see Mozer 1991). In SAIM, extinction can emerge when lesions affect parts the left of the selection network correspond to input from one side the visual field. This leads to a spatial imbalance in the competitive interactions within the

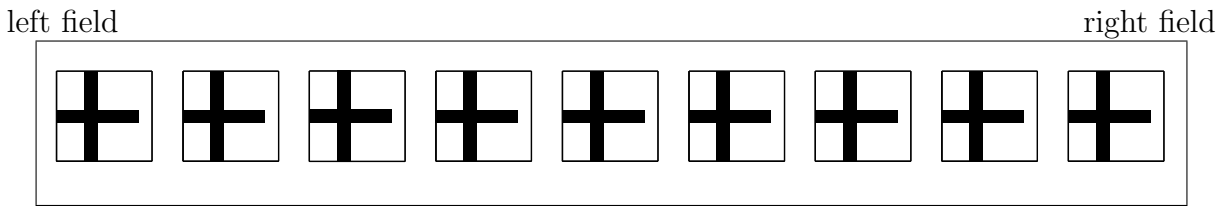


Figure 8: Object-based neglect in SAIM. Each window depicts the resulting contents of the FOA for a symmetric cross at different locations in the visual field (from left to right). Irrespective of the cross location in the visual field, the left most pixel is neglected by SAIM.

network, so that items on the unaffected side tend to win the competition for selection; these items are selected in preference to stimuli that appear on the affected side of space (Heinke & Humphreys, in press).

### Neglect

In MORSEL, neglect arises when a more severe lesion is imposed on the attention network (relative to the lesion that generates extinction). Mozer et al. (1997), for example, showed that neglect in bisection tasks could be simulated in terms of where the centre of activation in the attentional network fell with a target line. When a graded lesion was imposed across the network, attention tended to fall on the ipsilesional side of the true centre of the line. The degree of misplacement also increased in larger lines, matching the neuropsychological data (Riddoch & Humphreys 1983).

In contrast to this, neglect can be induced in SAIM in at least two different ways. One way is to increase the severity of a "vertical" lesion that can produce extinction in its stronger form. This generates a bias in which stimuli presented in one part of the visual field are favored for selection over items that fall in the lesioned area. When the lesion is sufficiently severe, there can also be a distortion in the spatial selection process, with a target being mapped into one side of the FOA. This has knock-on effects on the ability to disengage attention, once an object is identified. Normally, as we have noted, attention is disengaged from identified objects by inhibitory feedback. When the input is shifted in the FOA, the inhibitory feedback may no longer inhibit activation from the object first identified, attention becomes 'stuck' on this object. On such occasions, SAIM manifests neglect of an object presented on the contralesional side of two or more stimuli. There is space-based neglect.

A second form of neglect can be generated in SAIM by what we termed a "horizontal" lesion, affecting the units in the selection network that modulate the mapping from the visual field into one side of the FOA (Fig. 7). In this case, there can be a spatial unbalancing of mapping of information into FOA, leading to a misplaced FOA, so that elements on the contralesional side are not mapped into the FOA. Importantly, this occurs irrespective of the absolute positions of stimuli in the visual field, and it can affect each object present in the field (see Fig. 8 for an example). There is a form of "object-based" neglect, that takes place within each selected object.

SAIM is sensitive to grouping between stimuli, with there being facilitatory interactions between parts of a single perceptual group. In SAIM these interactions are based not only on bottom-up relationships (e.g., proximity) but also on top-down activity when two stimuli map onto a common template representation. For instance, in the example shown in Figure 9, SAIM was given templates either conforming to two separate letters (I and T) or to the combined pair of letters (the word IT). Following a lesion there was neglect of the letter on the affected side when the stimuli activated separate templates but not when they activated the same template (the word IT). Here, top-down activation from

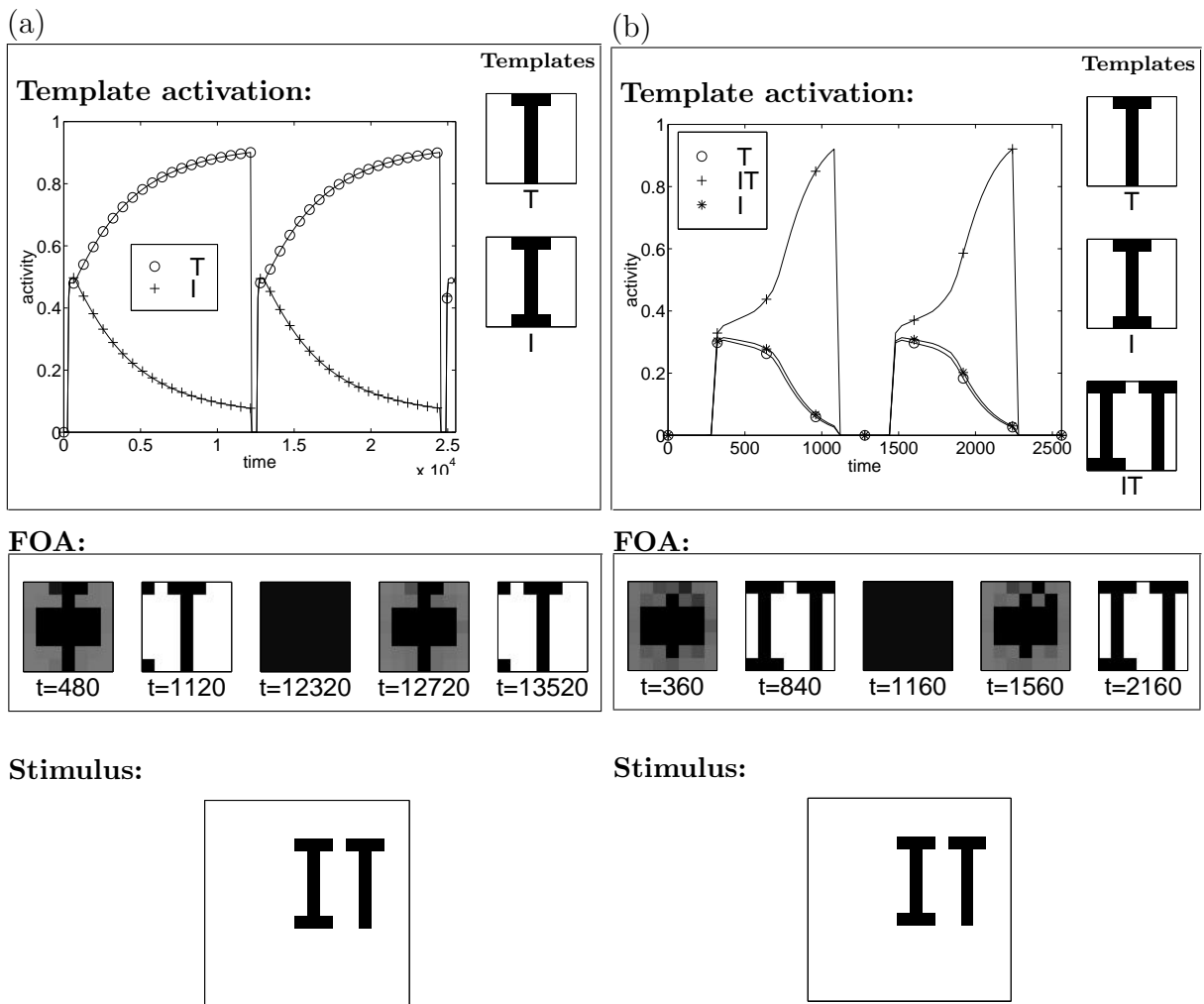


Figure 9: This simulation result with SAIM illustrates, how a top-down influence in SAIM can modulate left neglect. (a) SAIM was given templates either conforming to two separate letters (I and T); in (b) templates for the letters were used along also with a template for the combined pair of letters (the word IT). After lesioning of the selection network there was neglect of the letter on the affected side when the stimuli activated separate templates but not when they activated the same template (the word IT). Here, top-down activation from the template for the word supported units in the selection network for the letter on the affected side, enabling activation to be detected in the FOA of the model.

the template for the word supported units in the selection network for the letter on the affected side, enabling activation to be detected in the FOA of the model. This top-down activation extends a model such as SERR, where grouping is based only on bottom-up similarity effects. In SAIM grouping by activating a common template could lead to stimuli being recovered together even when there is no stronger bottom-up grouping than in other stimuli, that are subject to neglect. The facility for top-down grouping is useful, since equivalent effects have been observed in parietal patients. Sieroff et al. (1988) reported neglect between letters making up nonwords, but not when the letters combined to form words, even though bottom-up grouping should not favor the letters in words. Like SAIM, MORSEL is also able to explain why neglect effects are reduced when elements are part of a familiar object, compared with when they do not (see Mozer 1991). However, the mechanisms mediating such effects in MORSEL differ from those in SAIM. MORSEL does not employ top-down feedback from the object recognition system to affect selection

with the attentional module, but uses the "pull out" network to recover degraded stimuli even when the attentional module is lesioned. Accordingly, MORSEL, like SAIM, can account from letter recognition being improved in words relative to nonwords in neglect patients. Interestingly, there is now support for SAIM's approach of top-down influence from neurophysiological evidence. Recently McAdams & Maunsell (2000) and Motter (1994) showed that directing attention to stimulus features modulates the responses of neurons in V4 throughout the visual field. This is similar to how SAIM's top-down activation operates, where the top-down modulation influences the behaviour of the selection network throughout the visual field.

There are now also several pieces of neuropsychological evidence indicating that different forms of neglect exist and can be fractionated from one another. One double dissociation was reported by Humphreys & Heinke (1998). In their study some patients only showed neglect of the contralateral member of a pair objects, but had no neglect of each object. In contrast, other patients manifested neglect of each object, but did not neglect whole objects. Patients with this last form of "within object neglect" might be worse at identifying the left side of an object in their right field, even relative to their identification of the right side of an object in their left field (see also Driver et al. 1992). Humphreys & Riddoch (1995, 1994) even reported a dissociation between neglect within and between separate objects in a single patient, who had suffered bilateral brain lesions. This patient neglected the left side of selected objects, whilst also omitting whole objects presented on his right side. These double dissociations can be simulated in SAIM, because neglect between and with objects is caused by distinct lesions. The vertical lesion causing neglect between objects will not necessarily cause within-object neglect; similarly the (horizontal) lesion will generate object-based neglect without effects of visual field. Heinke & Humphreys (in press) even simulated the dissociation with a single patient, by introducing a "vertical" lesion affecting the right side of space and a "horizontal" lesion affecting the left.

Object-based neglect can also be simulated in MORSEL if there is a graded lesion across its attentional field (Mozer et al. 1997). With such a graded lesion, the right side of each object will be relatively more activated than the left side, even when the object is shifted laterally across the visual field. There can be neglect of left-side features even when the object falls in the right visual field. However, it is difficult to see how MORSEL can account for the double dissociation where a whole object on the right side and left parts of objects on the left side can be ignored, as it has been done with SAIM. This is a limitation of using a simple attentional mechanism confined to a two-dimensional representation of the field, where a spatial lesion can be introduced in a single way.



## 4 Discussion

### 4.1 Emergent properties and existence proofs

We have presented a number of connectionist models of visual attention, which have been applied to different psychological paradigms with both normal observers and neuropsychological patients. The models incorporate a number of common ideas – in particular that "attentional" behaviour arises out of competitive interactions between units in networks. In such networks, there is typically no single "attentional module" that imposes selectivity on perception and action; rather attentional behaviour is an emergent property of competitive interaction in networks. An exception to this approach is MORSEL which has an attentional module separate from its identification components. Another exception is Guided Search, which does not contain a competitive mechanism between stimuli. According to the "competitive" approach to attention, "attention" may infiltrate multiple levels of information processing, where competition takes place. This contrasts with traditional psychological models, which have tended to posit a single locus of attention acting either "early" or "late" in processing (see Broadbent 1958, Deutsch & Deutsch 1963, respectively). For example, in SAIM there is competition between multiple stimuli to fall within a spatial window of attention. Cueing the window to an area of space biases this competition so that unattended stimuli are filtered – a form of "early" attention. Nevertheless multiple stimuli in the field can activate their stored representations in parallel, with there being competition between templates to control the response. This implements a form of "late" selection, as evidenced by simulations of response competition in the Eriksen flanker task. The model, thus shows how traditional psychological theories are not necessarily mutually contradictory, selection can be "early" or "late", depending on a variety of conditions (e.g., the presence of cueing, the familiarity of the objects and so forth). Similarly, models such as MORSEL and SAIM demonstrate that effects of spatial cueing with networks are not necessarily a result of a module devoted to "engaging", "moving" or "disengaging" attention (e.g. Posner et al. 1980). In these models, the dynamics of processing can also be altered by lesioning, so that, for example, effects of invalid spatial cueing are much larger for targets presented in the affected rather than the unaffected field (Heinke & Humphreys, in press; see also Cohen et al., 1994, for a similar example in a small-scale connectionist model devoted purely to simulating spatial cueing effects). This interaction of invalid cueing and target field is found in studies using patients with lesions of the parietal lobe (Posner et al. 1984), and it has been attributed to a deficit in an "attentional disengagement" mechanism. In MORSEL and SAIM, however, it arises because the target on the lesioned side is slow to suppress residual activation from the cue due to the weakened impact of the target following the lesion. As a consequence, the selection of the target is delayed. The simulations indicate that we should be cautious in identifying too readily a particular experimental finding with a particular (and devoted) processing mechanism (here: for attentional disengagement).

Another dichotomy made in psychological models, that can play little role once processing is implemented in detailed computational terms, is that between "attentional" and "representational" disorders in neglect. For instance, consider both MORSEL and SAIM. In both models, neglect is produced by a bias in spatial selection so that one part of the visual world is favored for attention. In both cases, though, there is a spatial distortion in selection, so that only a part of space is represented (e.g., activates stored knowledge). For such models neglect is a disorder of both attention and spatial representation. Similarly, models such as MORSEL and SAIM incorporate mechanisms for both space- and object-based selection. In SAIM these processes interact directly, so that object-knowledge is used to guide spatial attention by top-down activation of the selection

network. In MORSEL space-based attention (in the attention module) and object-based selection (e.g., via the pull-out network) are functionally more independent, although the information that is finally selected is influenced by both factors. For these models, visual selection is not either space-based or object-based, it is both.

In these respects the models, once implemented, can act as "existence proofs" for psychological theory (in the above examples, the proofs demonstrate that object- and space-based theories are not incompatible). Another example of this concerns the implementation of visual search in models such as SERR. In SERR there is parallel processing of visual stimuli, with selection dictated by which elements form the strongest perceptual group. There is then serial selection of each perceptual group. For such a model, linear search functions do not reflect an item-by-item search process; rather they reflect serial rejection of groups plus some degree of re-checking to maintain error rates at larger display sizes. The simulations here provide an existence proof that linear search functions are not necessarily caused by spatially serial (item-by-item) search, as suggested in the Guided Search model. This point is made even more strongly in simulations of search that we have carried out with SAIM. We measured the time for a target of distractor template to reach threshold as a function of a number of distractors in the field. The time for both the target and distractor templates to reach threshold varied linearly with the size of the distractor set, even though there was not even serial selection of grouped items – the target template reached threshold first when it was present and the distractor threshold was attained first when no target was present. The effect of the distractor set size in target and distractor selection in SAIM arises because when multiple items are present, there are systematic increases in the competition for attention within the selection network and between the templates for the stimuli. Again the simulations demonstrate that linear search functions can be generated by spatially parallel search mechanisms.

## 4.2 Comparisons between models

The models we have discussed have, in some cases, been applied to rather different tasks (filtering by stimulus properties, visual search, multiple word identification, etc.). This makes it difficult to compare them directly. In particular, it is problematic to argue that one model is superior to another given the multiple differences that exist. For example, consider SAIM and SERR. One criterion on which models may be judged is the rule of Ockham's razor (William of Ockham, 1285-1347/49), that "entities should not be multiplied beyond necessity". The model that employs the fewest, simplest mechanisms should be the best. A somewhat different criterion is that the model should take into account as much psychological evidence as possible. SAIM and SERR differ in terms of these criteria. SERR is simpler in terms of its processing architecture and its processing dynamics. On the other hand, SAIM covers a broader number of tasks (object recognition, word recognition, visual search) and it is not configured to process just one particular type of stimulus (L and T-like items). Judgment about which model is better is determined by the question being asked – as a model of a particular search task, then probably SERR, since SERR unlike SAIM, incorporates form-based grouping; as a more general model applicable to many objects, and as a solution to the binding problem, then probably SAIM. Another approach to comparing between models is to ask whether one model is able to provide a better account of a particular experimental result that they both attempt to address. One example here concerns the simulations of neglect within and between perceptual objects. As we have noted these contrasting forms of neglect can be found not only in different patients, but even in the same patient. The observed double dissociation can be simulated relatively easily in SAIM, since its architecture lends itself to lesioning in different ways (the vertical vs. horizontal lesions, Fig. 7). These contrasting forms of

lesioning essentially affect the selection of stimuli in different reference frames: a retinal frame (vertical lesion) and an object-based frame (horizontal lesion). Lesioning in these different ways generates spatial deficits in the different reference frames, leading to neglect either within a single object or between separate objects. In contrast to this, MORSEL employs only a single (retinal) frame of reference for both object recognition and attention. Object-based deficits in neglect are then brought about by graded lesioning, so that there are relative biases in selection between the different parts of objects. We have argued that, in this framework, it is not clear how patterns of double dissociation could emerge especially within a single patient. We suggest that, at least as far as modeling these different forms of neglect is concerned, SAIM provides a more natural account.

Another issue relevant to the modeling enterprise concerns how parameters are used in order to capture particular sets of results. Ideally these parameters should be kept constant throughout all simulations to prove the generality of the model. However, alteration of parameters can be allowed for accounting of altered experimental set-ups. For instance, in order to account for knowledge-based biases on selection, the initial setting for SAIM's knowledge network was modified so that one template unit was favoured over others, making the selection network more likely to select the corresponding item. On the other hand, in these simulations the increased influence of the template unit had to be restricted to limit the likelihood of false alarms, especially, if stimuli had features similar to the "primed" target. In a sense this restriction represents a case of fitting parameters to the data – an example of modifying parameters in a less favourable way. However, since it seems plausible to argue that any intelligent system would have to strike a balance between useful top-down influences and a top-down influence that becomes ineffective because it overwrites bottom-up input. Hence, here the process of fitting the parameters to the data is a necessary way of constraining the model. As a next step, however, one would want to evaluate whether changes, for instance, in object familiarity can be captured by variations in the critical parameter. That is, modellers need to move beyond fitting the data to a single data to test whether qualitative changes in parameter values are appropriate. There are other examples of modifying parameters to accommodate experimental data that we have covered in this chapter, particularly the simulations of flanker interference in the "Eriksen" paradigm (Eriksen & Eriksen 1974). Flanker interference has been observed in both MORSEL and SAIM. In SAIM the spatial location of the target location was primed to reflect the knowledge of participants about the location of the target and targets would be selected in preference to distractors. This seems a plausible assumption based on the standard Eriksen flanker procedure, where the location of targets is known in advance by observers. However, the magnitude of spatial priming had to be limited, otherwise any competitive interference from distractors would be eliminated. This can be considered as a somewhat arbitrary fitting of parameters to the data, to generate interference effects of appropriate size. This argument also applies to MORSEL, where the efficiency of selection in the attentional module had to be reduced to generate the interference effect. Though it is of course important for models to demonstrate that they are capable of simulating a broad range of human data, it is clear that the resulting conclusions are more interesting and general when they emerge from intrinsic properties of models, using the same set of parameters.

Finally, an important aspect of computational modeling is that it can enable researchers to draw-in constraints that come from outside the parent discipline. For example, our focus in this chapter has been on computational models of cognitive data, and we have not discussed models that attempt to capture neurophysiological data. However, it is certainly possible for models in this area to be linked to such data (e.g Deco & Zihl 2001). Models also capture some of the motivating computational principles that presumably guided selection and object recognition (see Marr 1982), and so can link to work in ar-

tificial intelligence as well as in psychology. By incorporating such constraints, models will move beyond being "mere" implementations of a particular psychological theory to become framework for understand attention, from the single cell to the system level (see Seidenberg 1993, for an in-depth discussion of this).

### 4.3 Modeling neuropsychological data

In the field of attention, computational models have been evaluated not only by comparing their results to data from normal observers, but also by simulating neuropsychological data. In these simulations the effects of brain lesions have been modeled in a variety of ways, including: adding noise to the activation values (in SERR, Humphreys & Müller 1993) and reducing the weights on connections (e.g., in SAIM, Heinke & Humphreys, in press). At present it is not clear how best to simulate the effects of brain lesions in models and alternative procedures could include eliminating connections or adding varying degrees of noise across each iteration of a network (Olson & Humphreys 1997). One relevant issue is whether the lesion should selectively affect processing in one part of a model or whether it should exert a more pervasive influence. Humphreys et al. (1992a) changed a global parameter affecting all stages of their model, to simulate agnosia. In contrast, reduced weights on connections in one part of the field may be useful for simulations of neglect, since this deficit appears to reflect a spatial imbalance in selection. It remains for future work to establish how different forms of simulated lesion may capture the contrasting effects of brain damage on human performance.

## 5 Conclusion

In this review we have highlighted contributions of computational models to research on visual selective attention. These contributions include demonstrating that traditional dichotomies such space-based versus object-based attention, early vs. late selection, and representational vs. attentional accounts of visual neglect do not necessarily hold. In each case selective processing in the models arises from competitive mechanisms in which units interact to gain control of behaviour, and attention can be an emergent property of the network and not a distinct module in its own right. The models can provide detailed accounts of both normal and abnormal selection and some models can also do this across a range of different tasks. We conclude that computational models do provide a useful contribution to research in this area.

### Acknowledgments

This research was supported by the Medical Research Council (MRC) and the European Union.

## References

- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon.
- Caramazza, A. & Hillis, A. E. (1990). Levels of Representation, Co-ordinate Frames, and Unilateral Neglect. *Cognitive Neuropsychology*, 7(5/6):391–445.
- Cave, K. R. (1999). The FeatureGate model of visual selection. *Psychological Research*, 62:182–194.

- Chun, M. M. & Wolfe, J. M. (1996). Just Say No: How Are Visual Searches Terminated When There Is No Target Present? *Cognitive Psychology*, 30:39–78.
- Cohen, A. & Ivry, R. B. (1991). Density Effects in Conjunction Search: Evidence for a Coarse Location Mechanism of Feature Integration. *J. of Experimental Psychology Human Perception and Performance*, 17(4):891–901.
- Cohen, J. D., Romero, R. D., Servan-Schreiber, D., & Farah, M. J. (1994). Mechanisms of Spatial Attention: The Relation of Macrostructure to Microstructure in Parietal Neglect. *Journal of Cognitive Neuroscience*, 6(4):377–387.
- Deco, G. & Zihl, J. (2001). Top-down selective visual attention: A neurodynamical approach. *Visual Cognition*, 8(1):119–140.
- Deutsch, J. A. & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, 70:80–90.
- Donnelly, N., Humphreys, G. W., & Riddoch, M. J. (1991). Parallel Computation of Primitive Shape Descriptions. *J. of Experimental Psychology: Human Perception & Performance*, 17(2):561–570.
- Driver, J., Baylis, G. C., & Rafal, R. D. (1992). Preserved Figure-Ground Segmentation in Visual Matching. *Nature*, 360:73–75.
- Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, 87:272–300.
- Duncan, J. (1984). Selective Attention and the Organization of Visual Information. *Journal of Experimental Psychology: General*, 113(4):501–517.
- Duncan, J. & Humphreys, G. W. (1989). Visual Search and Stimulus Similarity. *Psychological Review*, 96(3):433–458.
- Dyer, F. N. (1973). Interference and facilitation for color naming with separate bilateral presentations of the word and color. *J. of Experimental Psychology*, 46:314–317.
- Egley, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal subjects. *J. of Experimental Psychology: Human Perception and Performance*, 123:161–177.
- Enns, J. & Rensink, R. (1991). Preattentive Recovery of Three-Dimensional Orientation From Line Drawings. *Psychological Review*, 98(3):335–351.
- Eriksen, B. A. & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 14:155–160.
- Eriksen, C. W. & Yeh, Y.-Y. (1985). Allocation of Attention in the Visual Field. *Journal of Experimental Psychology: Human Perception and Performance*, 11(5):583–597.
- Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. Cambridge, MA: MIT Press.
- Gilchrist, I., Humphreys, G. W., & Riddoch, M. J. (1996). Grouping and Extinction: Evidence for Low-Level Modulation of Selection. *Cognitive Neuropsychology*, 13:1223–1256.

- Heinke, D. & Humphreys, G. W. (1997). SAIM: A Model of Visual Attention and Neglect. In *Proc. of the 7th International Conference on Artificial Neural Networks–ICANN’97*, pages 913–918, Lausanne, Switzerland. Springer Verlag.
- Heinke, D. & Humphreys, G. W. (1999). Modelling Emergent Attentional Properties. In Heinke, D., Humphreys, G. W., & Olson, A., editors, *Connectionist Models in Cognitive Neuroscience – The 5th Neural Computation and Psychology Workshop*, pages 240–251. Springer Verlag, University of Birmingham, England.
- Heinke, D. & Humphreys, G. W. (in press). Attention, spatial representation and visual neglect: Simulating emergent attentional processes in the Selective Attention for Identification Model (SAIM). *Psychological Review*.
- Hinton, G. E. & Sejnowski, T. J. (1988). Learning and Relearning in Boltzmann Machines. In *Parallel Distributed Processing; Explorations in the Microstructure of Cognition; Volume 1: Foundations*. A Bradford Book, The MIT Press.
- Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 81:3088–3092.
- Humphreys, G. & Riddoch, M. (1995). Separate Coding of Space Within and Between Perceptual Objects: Evidence from Unilateral Visual Neglect. *Cognitive Neuropsychology*, 12(3):283–311.
- Humphreys, G. W. (1987). Objects, words, brains and computers: Framing the correspondence problem in object and word recognition. *Bulleting of the British Psychological Society*, 40:207–210.
- Humphreys, G. W. (1999). The neuropsychology of vision. In Humphreys, G. W., editor, *Case studies in the neuropsychology of vision*, pages xi–xiii. London: Psychology Press.
- Humphreys, G. W. & Donnelly, N. (2000). 3-D constraints on spatially parallel shape perception. *Perception & Psychophysics*, 62(5):1060–1085.
- Humphreys, G. W., Freeman, T. A. C., & Müller, H. M. (1992a). Lesioning a Connectionist Model of Visual Search: Selective Effects on Distractor Grouping. *Canadian Journal of Psychology*, 46:417–460.
- Humphreys, G. W. & Heinke, D. (1998). Spatial representation and selection in the brain: Neuropsychological and computational constraints. *Visual Cognition*, 5(1/2):9–47.
- Humphreys, G. W. & Müller, H. J. (1993). SEArch via Recursive Rejection (SERR): A Connectionist Model of Visual Search. *Cognitive Psychology*, 25:43–110.
- Humphreys, G. W., Olson, A., Romani, C., & Riddoch, M. J. (1996). Competitive mechanisms of selection by space and object: A neuropsychological approach. In Kramer, A. F., Coles, M. G. H., & Logan, G. D., editors, *Converging Operations in the Study of Visual Selective Attention*. American Psychological Association.
- Humphreys, G. W., Quinlan, P. T., & Riddoch, M. J. (1989). Grouping processes in visual search: Effect with single and combined-feature targets. *Journal of experimental Psychology: General*, 118:258–279.

- Humphreys, G. W. & Riddoch, M. J. (1993). Interactive Attentional Systems and Unilateral Visual Neglect. In Robertson, I. & Marshall, J., editors, *Unilateral neglect: Clinical and experimental studies*, pages 139–167. Hove: Lawrence Erlbaum Associates Inc.
- Humphreys, G. W. & Riddoch, M. J. (1994). Attention to Within-object and Between-object Spatial Representations: Multiple Side for Visual Selection. *Cognitive Neuropsychology*, 11(2):207–241.
- Humphreys, G. W., Riddoch, M. J., Quinlan, P. T., Price, C. J., & Donnelly, N. (1992b). Parallel pattern processing in visual agnosia. *Canadian Journal of Psychology*, 46:67–103.
- Husain, M. & Kennard, C. (1996). Visual neglect associated with frontal lobe infarction. *Journal of Neurology*, 243:652–657.
- James, W. (1890). *The Principles of Psychology*. New York: Dover.
- Karnath, H.-O. (1988). Deficits of attention in acute and recovered visual hemi-neglect. *Neuropsychologica*, 26:27–43.
- Karnath, H.-O., Schenkel, P., & Fischer, B. (1991). Trunk orientation as the determining factor of the 'contralesional' deficit in the neglect syndrome and as the physical anchor of the internal representation of body orientation in space. *Brain*, 114:1997–2014.
- Klein, R. (1988). Inhibitory tagging system facilitates visual search. *Nature*, 334:430–431.
- Koch, C. & Itti, L. (2001). Computational Modelling of Visual Attention. *Nature Reviews: Neuroscience*, 2:194–203.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.
- McAdams, C. J. & Maunsell, H. R. (2000). Attention to Both Space and Feature Modulates Neuronal Responses. *J. of Neurophysiology*, 46:1751–1755.
- Miller, J. (1982). Divided Attention: Evidence for Coactivation with Redundant Signals. *Cognitive Psychology*, 14:247–279.
- Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J. of Neuroscience*, 46:2178–2189.
- Mozer, M. (1991). *The perception of multiple objects: a connectionist approach*. The MIT Press.
- Mozer, M. C. (1987). Early parallel processing in reading: A connectionist approach. In Coltheart, M., editor, *Attention and Performance XII*. American Psychological Association.
- Mozer, M. C., Halligan, P. W., & Marshall, J. C. (1997). The End of the Line for a Brain-Damaged Model of Unilateral Neglect. *Journal of Cognitive Neuroscience*, 9(2):171–190.
- Mozer, M. C. & Sitton, M. (1998). Computational modeling of spatial attention. In Pashler, H., editor, *Attention*, pages 341–393. London: Psychology Press.

- Müller, H. J., Humphreys, G. W., & Donnelly, N. (1994). SEArch via Recursive Rejection (serr): Visual Search for Single and Dual Form-Conjunction Targets. *J. of Experimental Psychology: Human Perception and Performance*, 20(2):235–258.
- Neisser, U. (1967). *Cognitive psychology*. Appleton-Century-Crofts, New York.
- Niebur, E., Koch, C., & Rosin, C. (1993). An Oscillation-based Model for the Neuronal Basis of Attention. *Vision Research*, 33(18):2789–2802.
- Olshausen, B., Anderson, C. H., & Van Essen, D. C. (1995). A Multiscale Dynamic Routing Circuit for Forming Size-and Position- Invariant Object Representations. *Journal of Computational Neuroscience*, 2:45–62.
- Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A Neurobiological Model of Visual Attention and Invariant Pattern Recognition Based on Dynamic Routing of Information. *J. of Neuroscience*, 13(11):4700–4719.
- Olson, A. & Humphreys, G. W. (1997). Connectionist models of neuropsychological disorders. *Trends in Cognitive Science*, 1:222–228.
- Pashler, H. (1987). Detecting conjunctions of color and form: Reassessing the serial hypothesis. *Perception & Psychophysics*, 41(3):191–201.
- Pavlovskaya, M., Glass, I. and Soroker, N., Blum, B., & Groswasser, Z. (1997). Coordinate Frame for Pattern Recognition in Unilateral Spatial Neglect. *Journal of Cognitive Neuroscience*, 9(6):824–834.
- Phaf, H. R., Van Der Heijden, A., & Hudson., P. (1990). SLAM: A Connectionist Model for Attention in Visual Selection Tasks. *Cognitive Psychology*, 22:273–341.
- Posner, M. I. & Cohen, Y. (1984). Components of Visual Orienting. *Attention and Performance*, pages 531–556.
- Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the Detection of Signals. *Journal of Experimental Psychology: General*, 109(2):160–174.
- Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal injury on convert orienting of attention. *J. of Neuroscience*, 4(7):1863–1874.
- Quinlan, P. T. & Humphreys, G. W. (1987). Visual search for targets defined by combination of color, shape and size: An examination of the tasks constraints on feature and conjunction search. *Perception & Psychophysics*, 41:455–472.
- Rensink, R. A. & Enns, J. (1995). Pre-emption Effects in Visual Search: Evidence for Low-level Grouping. *Psychological Review*, 102:101–130.
- Riddoch, M. J. & Humphreys, G. W. (1983). The effect of cueing on unilateral neglect. *Neuropsychologia*, 21:589–599.
- Riddoch, M. J. & Humphreys, G. W. (1987). Visual Object Processing in a case of optic aphasia: A case of semantic access agnosia. *Cognitive Neuropsychology*, 4:131–185.
- Robertson, I. & Marshall, J. C. (1993). *Unilateral neglect: Clinical and experimental studies*. London; Psychology Press.



- Schvaneveldt, R. W. & McDonald, J. E. (1981). Semantic context and the encoding of words: Evidence for two modes of stimulus analysis. *J. of Experimental Psychology: Human Perception & Performance*, 7:673–687.
- Seidenberg, M. S. (1993). Connectionist Models and Cognitive Theory. *Psychological Science*, 46:228–235.
- Shiffrin, R. M. & Schneider, W. (1977). Control and Automatic human information processing: II. Perceptual learning and automatic attending and a general theory. *Psychological Review*, 84:127–190.
- Sieroff, E., Pollastek, A., & Posner, M. I. (1988). Recognition of visual letter strings following injury to the posterior visual spatial attention system. *Cognitive Neuropsychology*, 5:427–449.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. of Experimental Psychology*, 18:643–662.
- Tipper, S. P. & Behrmann, M. (1996). Object-Centered not Scene-Based Visual Neglect. *Journal of Experimental Psychology: Human Perception and Performance*, 22:1261–1278.
- Tipper, S. P., Driver, J., & Weaver, B. (1991). Object-Centered Inhibition or Return of Visual Attention. *Quarterly Journal of Experimental Psychology*, 43A:289–299.
- Tipper, S. P., Jordan, H., & Weaver, B. (1999). Scene-based and object-centred inhibition of return: Evidence for dual orienting mechanisms. *Perception & Psychophysics*, 61:50–60.
- Townsend, J. T. & Nozawa, G. (1997). Serial Exhaustive Models Can Violate the Race Model Inequality: Implications for Architecture and Capacity. *Psychological Review*, 104(104):595–602.
- Treisman, A. (1988). Features and Objects: The Fourteenth Bartlett Memorial Lecture. *The Quarterly Journal of Experimental Psychology*, 40A(2):201–237.
- Treisman, A. (1998). Feature binding, attention and object perception. In Humphreys, G. W., Duncan, J., & Treisman, A., editors, *Brain mechanisms of selective perception and action*, volume 353 of *Philosophical Transactions: Biological Sciences*, pages 1295–1306. The Royal Society.
- Treisman, A. M. & Gelade, G. (1980). A Feature-Integration Theory of Attention. *Cognitive Psychology*, 12:97–136.
- Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, 78:507–545.
- van der Heijden, A. H. C. (1981). *Short term visual information forgetting*. London: Routledge & KeganPaul.
- Walker, R. (1995). Spatial and Object-based Neglect. *Neurocase*, 1:371–383.
- Ward, R., Goodrich, S., & Driver, J. (1994). Grouping Reduces Visual Extinction: Neuropsychological Evidence for Weight-linkage in Visual Selection. *Visual Cognition*, 1(1):101–129.

- Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2):202–238.
- Wolfe, J. M. (1998). Visual Search. In Pashler, H., editor, *Attention*, pages 13–74. Psychology Press.
- Wolfe, J. M. & Cave, K. R. (1990). Deploying Visual Attention: The Guided Search Model. In Blake, A. & Troscianko, T., editors, *AI and the Eye*, pages 79–103. John Wiley and Sons.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). An alternative to the feature integration model for visual search. *J. of Experimental Psychology: Human Perception & Performance*, 15:419–433.
- Young, A. W., Newcombe, F., & Ellis, A. W. (1991). Different impairments contribute to neglect dyslexia. *Cognitive Neuropsychology*, 8:177–192.