

Attention, spatial representation and visual neglect:  
Simulating emergent attention and spatial memory in  
the Selective Attention for Identification Model  
(SAIM)

Dietmar Heinke, Glyn W. Humphreys  
Behavioural and Brain Sciences Centre,  
School of Psychology,  
University of Birmingham,  
Birmingham B15 2TT,  
UK,

**Psychological Review, in press**

## Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>   | <b>1</b>  |
| 1.1      | Neuropsychological impairments of visual selection . . . . .  | 2         |
| 1.1.1    | Visual neglect . . . . .  | 2         |
| 1.1.2    | Visual extinction . . . . .   | 3         |
| 1.2      | Computational models of visual selection . . . . .  | 4         |
| 1.2.1    | The dynamic routing circuit . . . . .   | 6         |
| 1.3      | Architecture of SAIM . . . . .  | 7         |
| 1.3.1    | Dynamic routing circuits in SAIM . . . . .  | 8         |
| 1.3.2    | The knowledge network . . . . .   | 10        |
| 1.3.3    | Shifting attention and the location map . . . . .   | 12        |
| 1.3.4    | Discussion . . . . .  | 14        |
| 1.3.5    | Testing the model . . . . .   | 14        |
| <b>2</b> | <b>Modelling normal visual selection</b>  | <b>16</b> |
| 2.1      | General properties of SAIM . . . . .  | 16        |
| 2.1.1    | Study 1. Bottom-up mapping . . . . .  | 16        |
| 2.1.2    | Study 2. Accessing stored knowledge and attention- switching:<br>Bottom-up biases . . . . .           | 23        |
| 2.1.3    | Study 3 Accessing stored knowledge and attention-switching: Top-<br>down biases . . . . .             | 26        |
| 2.2      | Study 4. Wholistic perception . . . . .   | 31        |
| 2.3      | Study 5: Cueing . . . . .   | 32        |
| 2.3.1    | (i) Spatial cueing . . . . .  | 32        |
| 2.3.2    | (ii) Cueing objects and cueing space . . . . .  | 36        |
| 2.3.3    | (iii) Position of cueing within an object . . . . .   | 39        |
| 2.3.4    | (iv) Inhibition of return (IOR) . . . . .   | 40        |
| 2.3.5    | Discussion of Study 5 . . . . .   | 42        |
| 2.4      | Study 6 . . . . .   | 43        |
| 2.5      | Conclusion . . . . .  | 44        |
| <b>3</b> | <b>Modelling pathological attention and selection</b>   | <b>45</b> |
| 3.1      | Neurobiological status of SAIM . . . . .  | 45        |
| 3.2      | Lesioning SAIM . . . . .  | 45        |
| 3.3      | Study 7: Effect of 'vertical lesioning': Visual extinction without neglect. . . . .                   | 47        |
| 3.4      | Study 8. Effects of combined 'vertical & horizontal lesioning': Field-<br>dependent neglect . . . . . | 50        |
| 3.5      | Study 9. Effect of 'horizontal' lesioning: Object-based neglect . . . . .                             | 52        |
| 3.5.1    | (i) Neglect of single objects . . . . .   | 52        |
| 3.5.2    | (ii) Multiple objects: Sequential neglect . . . . .   | 55        |
| 3.5.3    | (iii) Multiple objects: 'Sticky attention' and impaired spatial memory . . . . .                      | 57        |
| 3.5.4    | (iv) Multiple objects: 'Between-object neglect' . . . . .   | 58        |
| 3.6      | Study 10. Template effects: Top-down filling in and premature acceptance . . . . .                    | 59        |
| 3.6.1    | (i) Top-down filling in, in neglect . . . . .   | 59        |
| 3.6.2    | (ii) Effects of premature acceptance . . . . .  | 61        |
| 3.7      | Study 11: Neglect within- and between-objects . . . . .   | 63        |

|          |   |            |
|----------|---|------------|
| 3.8      | Study 12. Line bisection. . . . .                                     | 66         |
| 3.9      | Study 13. Cueing in neglect . . . . .                                 | 67         |
| 3.9.1    | (i) Posner et al. (1984) . . . . .                                    | 67         |
| 3.9.2    | (ii) Pavlovskaya et al. (1997) . . . . .                              | 70         |
| 3.10     | Conclusions . . . . .   | 70         |
| <b>4</b> | <b>General Discussion</b>   | <b>72</b>  |
| 4.1      | Qualitative fit and the refutation of models . . . . .                | 72         |
| 4.2      | Useful properties of the model . . . . .                              | 75         |
| 4.3      | Relations between SAIM and other models of visual selection . . . . . | 78         |
| 4.4      | Emergent properties of SAIM . . . . .                                 | 81         |
| 4.5      | Factors not covered . . . . .   | 83         |
| 4.6      | Biological plausibility . . . . .                                     | 84         |
| <b>5</b> | <b>References</b>   | <b>87</b>  |
| <b>A</b> | <b>The design of SAIM</b>   | <b>100</b> |
| A.1      | The contents network and the FOA . . . . .                            | 100        |
| A.2      | The selection network . . . . .                                       | 100        |
| A.3      | Knowledge Network . . . . .   | 102        |
| A.4      | The complete model . . . . .  | 103        |
| A.5      | Gradient Descent . . . . .  | 103        |
| A.6      | Attention Switching . . . . .   | 104        |
| A.7      | Cueing Studies . . . . .  | 105        |
| <b>B</b> | <b>Lesion</b>   | <b>105</b> |
| <b>C</b> | <b>Parameters</b>   | <b>105</b> |
| C.1      | Model . . . . .   | 105        |
| C.2      | Lesions . . . . .   | 106        |

## Abstract

We present a computational model of human visual attention, termed SAIM (Selective Attention for Identification Model), based on the idea that there is competition between objects for recognition. SAIM uses a spatial window to select visual information for object recognition, ensuring both that parts are bound correctly within objects and that translation invariant object recognition can be achieved. We show how such a competitive model can provide a qualitative account of a range of psychological phenomena on both normal and disordered attention. Simulations of normal attention demonstrate two-object costs on selection, effects of object familiarity on selection, global precedence, spatial cueing both within and between objects, and inhibition of return. When simulated lesions were conducted, SAIM demonstrated both unilateral neglect and spatial extinction, depending on the type and extent of the lesion. Different lesions also produced view-centred and object-centred neglect, enabling both forms of neglect to be simulated within a single patient. We discuss the relations between SAIM and other models in the literature, and we highlight how emergent properties of competition within the model can unify (i) object- and space-based theories of normal selection, (ii) dissociations within the syndrome of visual neglect, and (iii) 'attentional' and 'representational' accounts of neglect.

## 1 Introduction

Visual scenes typically contain many objects, yet actions may only be addressed to a few. Consequently there needs to be some form of selection so that only representations of behaviourally relevant objects are made available to response systems (Allport 1987). Selection is also likely to be a necessary part of successful visual object identification. For example, when multiple objects are present there may be problems in binding their parts together without incorrect (illusory) bindings being formed (Feldman & Ballard 1982, Hummel & Biederman 1992). A possible solution to this problem is to select only one object for binding at a time, so that the parts of irrelevant objects do not compete in this binding process (see Treisman 1998).

Psychological models traditionally hold that visual selection is spatially-based, assuming (for instance) that selection involves some form of internal spotlight or zoom lens that activated attended regions or leads to the inhibition of unattended regions (Eriksen & Yeh 1985, Posner 1980, Treisman & Gormican 1988). Strong evidence for spatial selection comes from studies of cueing which show that detection of a target is enhanced when it is preceded by a spatially valid cue (Posner 1980). Models of selection need to account for how spatial cueing may facilitate responses to target stimuli.

Other work, though, demonstrates that selection is influenced by grouping between the parts of objects, so that parts belonging to the same object tend to be selected together. To give three examples: (i) the detrimental effects of invalid spatial cues are reduced if cues and targets fall within the same perceptual object (Egley et al. 1994, Kramer et al. 1997); (ii) 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 a single object (Baylis & Driver 1993, Duncan 1984, Vecera & Farah 1994); and (iii) effects of response congruency between distractors and targets are determined by whether stimuli group, with interference being stronger between members of a group than between members of separate groups (Baylis & Driver 1992). To accommodate such results, theorists have suggested that object recognition processes can interact with spatial selection to enable grouped parts to be selected together. For example, activation in stored representations may feedback to location-coded representations of parts, forcing spatial attention to spread across whole objects (Farah 1990, Humphreys & Riddoch 1993); alternatively, recognition procedures may be able to recover parts of known objects even when some are processed sub-optimally (i.e., without attention, see Mozer (1991)). Thus in some circumstances selection appears to be 'object-based' and not simply 'space-based'.

In the present paper we aim to provide a formal model of how spatial and object information may interact to enable objects to be selected for identification: SAIM (the Selective Attention for Identification Model). The model is applied to 'classic' studies of visual selection in normal participants demonstrating both space-based and object-based constraints on performance. In SAIM selection is mediated by a form of 'attentional window' that is tuned to the location of a target; however this tuning process is influenced by higher-level properties of objects, making selection sensitive to object- as well as spatial information. In addition to this, we show that SAIM can capture various neuropsychological impairments of visual selection, that are difficult to account for in other frameworks.

## 1.1 Neuropsychological impairments of visual selection

### 1.1.1 Visual neglect

Arguments for both space-based and object-based effects on visual selection are demonstrated in a dramatic fashion in neuropsychological disorders such as visual neglect and extinction. Following damage to temporo-parietal and fronto-parietal regions, patients may fail to respond to stimuli presented on the side of space opposite to their lesion: visual neglect (see Husain & Kennard 1996, Humphreys & Heinke 1998, Robertson & Marshall 1993, Walker 1995). Performance can be affected by the position of an object relative to the patient's body (Karnath et al. 1991, Riddoch & Humphreys 1983) and also, in some cases, by the positions of parts relative to an object. For example, 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 (Young et al. 1991), suggesting that neglect is then determined by the positions of the letters in the word. There can also be neglect of the contralesional parts of objects when objects are rotated so that these parts appear in the ipsilesional visual field (Behrmann & Tipper 1994, 1999, Driver & Halligan 1991) and patients can detect stimuli on the ipsilesional side of an object better than stimuli on the contralesional side, even when the former stimuli fall in the contralesional field and the latter in the ipsilesional field (Driver et al. 1992, Humphreys & Heinke 1998). Indeed neglect tied to the positions of parts in objects can be doubly dissociated from neglect tied to the positions of stimuli in retinal and/or body-space; thus in Humphreys & Heinke (1998) a case series of patients was reported in which some showed neglect based on the positions of parts within objects irrespective of the positions of the objects in the visual field, whilst for others there was little effect of stimulus position within an object but there were large effects of position in the field. We have termed these disorders "within-object" and "between-object" neglect the latter title being appropriate because neglect is apparent only when there are multiple objects in the field and thus it does not reflect the field position per se (see Humphreys 1998, Humphreys & Heinke 1998).

Both within-object and between-object neglect can even be demonstrated in a single patient. Humphreys & Riddoch (1994, 1995), for example, reported a patient with bilateral lesions who showed left-side neglect within objects but neglect of the right side of space when stimuli were coded as separate objects (see also Costello & Warrington 1987, Cubelli et al. 1991, Riddoch et al. 1995a,b, for similar cases). This contrasting pattern occurred even with the same stimulus, depending on how the stimulus was represented for the task. There was left neglect when single words were read as perceptual wholes but right neglect of the identical visual stimulus when the patient was required to read aloud individual letters (treating each letter as a separate perceptual object). Humphreys and Riddoch proposed that the bilateral lesions in this patient resulted in deficits to different representations of visual space: there was damage to one representation in which elements were coded with respect to their position within an object (a "within object spatial representation") and damage to another representation in which elements were coded as separate objects (a "between object spatial representation").

These different forms of neglect are related to problems in visual selection. For example, cueing patients to attend to the affected side can improve their performance significantly (Posner et al. 1984, Riddoch & Humphreys 1983), and impaired responding to contralesional stimuli is most pronounced when patients are cued to the ipsilesional side (Posner et al. 1984). Cues to attend to parts within objects, or to the positions of multiple objects

in the visual field, can also be differentially effective for the contrasting problems. In the case reported by Humphreys & Riddoch (1994, 1995), cueing to attend to the left side of individual objects improved within-object neglect, though cueing to the left side of a page filled with words had no effect. In contrast, cueing to the right side of a page reduced the neglect of separate whole objects on the right; cueing to the right sides of individual objects then had little effect. Egly et al. (1994) further report that left parietal damage selectively affects the ability to shift attention from one object to another, whilst right parietal damage affects attentional shifts from one location to another (see also Buck et al. (1999), for evidence for a similar distinction in patients with varying patterns of degenerative change). There are grounds here for arguing that forms of selection operate with each type of representation, with contrasting deficits in selection arising after different brain lesions.

### **1.1.2 Visual extinction**

An apparently milder problem in visual selection occurs in patients showing visual extinction. The term extinction is used to refer to patients who can detect a single stimulus presented on the contralesional side of space, but who fail to detect the same item when it is exposed simultaneously with an item on the ipsilesional side (e.g. Karnath 1988). This deficit arises when there is competition between the ipsi- and contralesional items for selection, when consequences of a spatial bias are exposed (see Duncan et al. 1997). 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 compete as separate objects (Driver & Mattingley 1995, Gilchrist et al. 1996, Humphreys 1998, Ward et al. 1994). In such instances a putatively spatial deficit can be over-ridden by grouping between the parts of objects. The nature of the spatial code involved in such cases remains unclear however. An item at the centre of the field can be detected or extinguished as a function of whether the competitor appears on the ipsi- or contralesional side (Humphreys et al. 1996b). Absolute retinal position seems less important here than the relative locations of stimuli.

These data demonstrate both dissociable and interactive effects of object- and spatial-coding on visual selection and, as we shall argue, they provide strong constraints on functional models of selection in vision. However, whether disorders such as unilateral neglect and extinction are themselves directly due to impaired attentional processes has been the subject of heated debate. Some authors have argued for there being a specific disorder of visual attention; for example, poor disengagement of attention from ipsilesional stimuli (Posner et al. 1984), poor automatic orienting of attention to contralesional events (Riddoch & Humphreys 1983), or inappropriate automatic orienting of attention to ipsilesional stimuli (Ladavas et al. 1990). In contrast, other authors have argued that these problems are not due to impaired attention but rather they reflect damage to particular spatial representations (e.g. Bisiach & Luzzatti 1978, Caramazza & Hillis 1990). The representations involved may vary according to the reference frame on which they are coded (e.g., some being retinotopic, others object-based; Caramazza & Hillis (1990), Miceli & Capasso (2001)), and this may help explain why some deficits are influenced by the position of the stimulus relative to the patient whilst others are affected by the positions of parts with respect to the object (see above). Also access to damaged repre-

representations may be improved by valid attentional cueing and disrupted by invalid cueing, making it difficult empirically to separate 'attentional' and 'representational' accounts of neglect and extinction.

The model we will present, SAIM (for Selective Attention for Identification Model), attempts to achieve translation invariant object identification by mapping retinally-coded input through to higher-level 'object' representations which respond to stimuli irrespective of their original presentation location. Such translation-invariant identification is a pre-requisite for any successful model of object recognition (see Marr 1982). To ensure that the mapping is successful, and that binding problems are not introduced by high-level representations being activated by multiple stimuli, there is competition for selection between objects so that stored representations are activated primarily by one object in the field; selective processing of objects is an emergent property of this competition. Interestingly, following damage to representations mediating the mapping from retinal to high-level descriptions, forms of visual neglect and extinction occur. These deficits reflect biases in competition for access to high-level representations, and they are influenced by spatial cueing and by the presence of competing objects in the visual field. However, since attentional selection here is an emergent consequence of competition between particular spatial representations of objects, it makes little sense to describe neglect and extinction as either 'attentional' or 'representational' in nature; within our framework, neglect and extinction are both 'attentional' and 'representational' deficits.

## 1.2 Computational models of visual selection

SAIM is a computational model of both translation invariant object recognition and visual selection. A preliminary version of the model was presented in (Humphreys & Heinke 1998, Heinke & Humphreys 1997, Humphreys & Heinke 1997). In this paper we report the model in full and apply it to a range of experimental and neuropsychological data on visual selection.

Over the past decade, a number of explicit computational models of visual attention have been developed. Some have been used to simulate experimental results on normal visual selection (e.g., the 'Feature Gate model', Cave (1999); Guided Search, Cave & Wolfe (1990), Wolfe (1994); MORSEL, Mozer (1991, 1999); Mozer & Sitton (1998); SEarch via Recursive Rejection [SERR], Humphreys & Müller (1993); SeLective Attention Model [SLAM], Phaf et al. (1990)). Some have been applied to neuropsychological data (MORSEL, Mozer et al. (1997); Mozer (1991) Mozer (1999); SErr, Humphreys et al. (1992), Humphreys et al. (1996b); the 'spatial transformation model' of Pouget & Sejnowski (1997)). Further categories of model have been developed in order to illustrate neurobiological constraints on selection (Deco & Zihl 2001, Niebur et al. 1993, Olshausen et al. 1993, 1995, Pouget & Sejnowski 1997, Pouget & Snyder 2000, Tsotsos et al. 1995, Usher & Niebur 1996), and to solve particular computational problems in vision (e.g., view-invariant object recognition; Hinton (1981b,a)). Common to many of these models is the use of competitive interactions between processing units (representing in some instances objects or object parts, in others positions in a spatial attention network), which enforce forms of 'winner take all' behaviour so that units given some initial advantage inhibit other units. This advantage can be based on a spatially coincident precue (benefitting one position in an image over others; (e.g. Mozer & Sitton 1998), on grouping support between like elements (Humphreys & Müller 1993), or on top-down activation



from memory representations for targets (Phaf et al. 1990, Usher & Niebur 1996). This idea, of selection by forms of competition between stimulus representations, is also found in psychological theories such as the 'integrated competition' model (Duncan 1998, Duncan et al. 1997). In addition, following selection of one target by means of these competitive interactions, computational models have employed forms of self-inhibition of targets (and their associated spatial representations) to enable other items to be selected subsequently (Humphreys & Müller 1993, Koch & Ullman 1985, Olshausen et al. 1993, 1995). SAIM also incorporates both winner-take-all competition and self-inhibition following selection, to enable selective processing to emerge. Winner-take-all competition is affected by both bottom-up and top-down factors. Top-down factors include stored object representations in addition to representations in 'working memory' pre-activated according to task goals (cf. Usher & Niebur 1996).

SAIM is closest in architecture and processing operations to the 'dynamic routing circuit' model proposed by Olshausen and colleagues (Olshausen et al. 1993, 1995), and to the model of view-invariant object recognition proposed by Hinton (1981b,a). Olshausen et al. provided an 'in principle' analysis of the model, showing that it could successively select individual (known) letters when multiple letters were present in the visual field; however they did not apply the model to either psychological or neuropsychological results on visual selection in humans. Hence it is unclear whether such an approach could provide a viable psychological model. Also SAIM extends the Olshausen et al. model by incorporating top-down constraints on selection that operate in parallel with bottom-up processes of grouping, a step that is necessary to accommodate data on both normal and impaired human visual selection. For example, visual neglect and even extinction can be reduced if the letters form a known word rather than a nonword, even when attempts are made to rule-out guessing biases (e.g., see Kumada & Humphreys (2001), for evidence on extinction; see Riddoch et al. (1990); Sieroff et al. (1988), for evidence on neglect). This lessening of neglect and extinction occurs even though there is no low-level information that would favour the grouping of letters in words over those in nonwords. In this last instance, grouping seems to be contingent on top-down knowledge, by dint of letters activating a common stored representation for the word. Using top-down constraints to influence selection seems to be a useful procedure for capturing such results. Hinton's model used top-down knowledge to guide perception, but it has not been applied in detail to modelling processing when multiple items are present, nor does it address a large body of psychological data (though see Hinton & Lang (1985)). Our aim in implementing SAIM was to provide a test of this combined bottom-up and top-down approach to selection by assessing its qualitative fit to psychological and neuropsychological data. To illustrate the general properties of SAIM, we first describe properties of the dynamic routing circuit.

### 1.2.1 The dynamic routing circuit

In the dynamic routing circuit, size and translation invariant object identification is achieved by mapping retinal input through multiple stages into a smaller 'focus of attention' (FOA). When multiple objects are present, the circuit has to select one in order to avoid overlapping features from different objects being mapped in to the FOA; any such mapping can disrupt object identification by distorting the input to the recognition system: there would be a form of 'binding problem'. The routing circuit contains two components, one that performs the mapping into the FOA from the retina and one that controls this mapping according to the task demands and the nature of the stimuli present. We term these the 'contents' network (reflecting the contents of the visual field) and the 'selection' network (which selects between multiple objects, to optimise object identification).

One way to conceptualise the contents network in such models is as a connection matrix specifying all possible relationships between locations on the retina and locations in the FOA, as illustrated in the simple case for a one-dimensional input with just 4 retinal locations and 3 locations in the FOA, in Figure 1. Here a high level of activation in one unit in the contents network instantiates a particular correspondence between a given retinal location and a given location in the FOA. Now in the example given, there is Gaussian input present, centred on location 2 in the retina; the goal of the circuit is to map this input into the FOA. For example, to map the input from retinal locations 1-3 into units 1-3 in the FOA, the units shown in black in the contents network need to be active and other units in this network inactive (these would map other retinal units into the same units in the FOA). However, if the Gaussian input was centred at location 3 on the retina, and this center of activity was to be mapped into unit 2 in the FOA, then the lower units in the contents network (shown here in white) would need to be active and the others inactive. The role of the selection network in the model is to gate activity in the contents network, enabling activity to be passed on from some but not other units in the contents network. This 'selects' which mapping is implemented from the retina to the FOA. Units in the selection network are shown on the left side of Figure 1. In this example, the units in the selection network are mutually inhibitory with one another so that they compete to establish which units in the contents network have most effect on units in the FOA. Units in the FOA are activated by multiplicatively combining activity from the contents network with activity from the selection network. This interaction between the contents and the selection network enables activity to be mapped from the retina to the FOA in a purely bottom-up manner, depending upon which units in the contents and selection networks are most strongly activated by the input.

It is possible for networks of this sort to achieve size invariance (see Olshausen et al. 1993, 1995). Size invariance can operate by introducing multi-scale representations, each acting at different resolutions, and the selection network structure can be extended to operate at each resolution with a switching procedure being required to move from one resolution to the next. In addition, the number of units and connections required within the selection network can be reduced by having these units connect to several units in the contents network, with the strengths of the connections varying in a Gaussian fashion (e.g., maximum weight to the central unit and reducing weight to units either side of this in the contents network). However, these additional factors are not of central concern

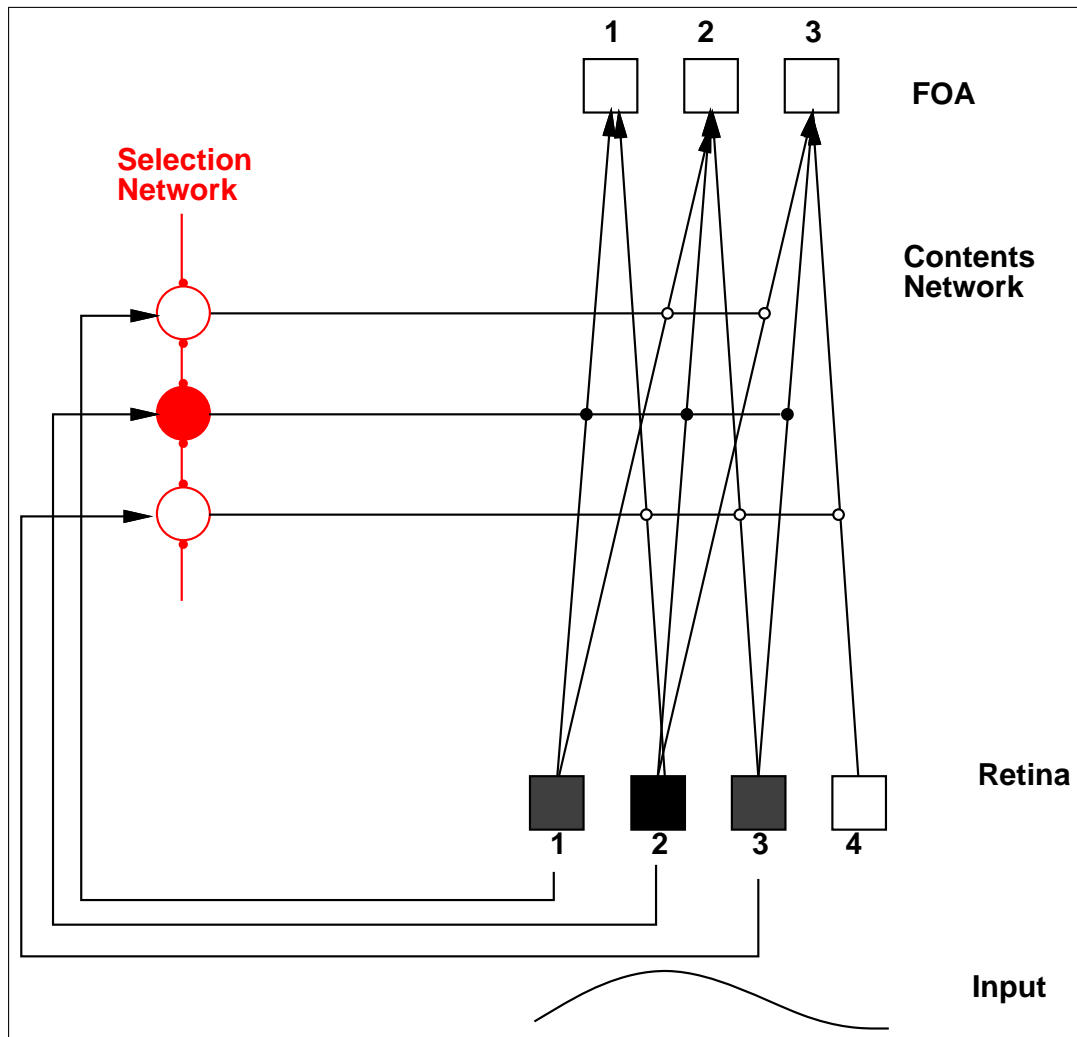


Figure 1: Illustration of a small-scale dynamic routing circuit. Each unit in the contents network represents a correspondence between the retina and the FOA; the selection network determines which correspondences are instantiated. In this example, Gaussian input activates retinal unit 2 most, and units 1 and 3 to a lesser degree. The circuit acts to map retinal input into the FOA, based on competition between units in the selection network. Not all connections are shown.

here, where we deal with the issue of whether the dynamic routing circuit approach to translational invariance can capture important aspects of visual representation and selection in humans. For this reason, and to simplify the analysis, SAIM uses one-to-one mappings from the selection to the contents network (with no Gaussian spread), and it operates at just one level of resolution. We return to discuss issues about biological plausibility in the General Discussion.

### 1.3 Architecture of SAIM

In addition to having a dynamic routing circuit to achieve translation invariant mapping into the FOA, the full version of SAIM also contains a higher-level 'knowledge network', to simulate object recognition processes, and a mechanism to enable attention to be shifted

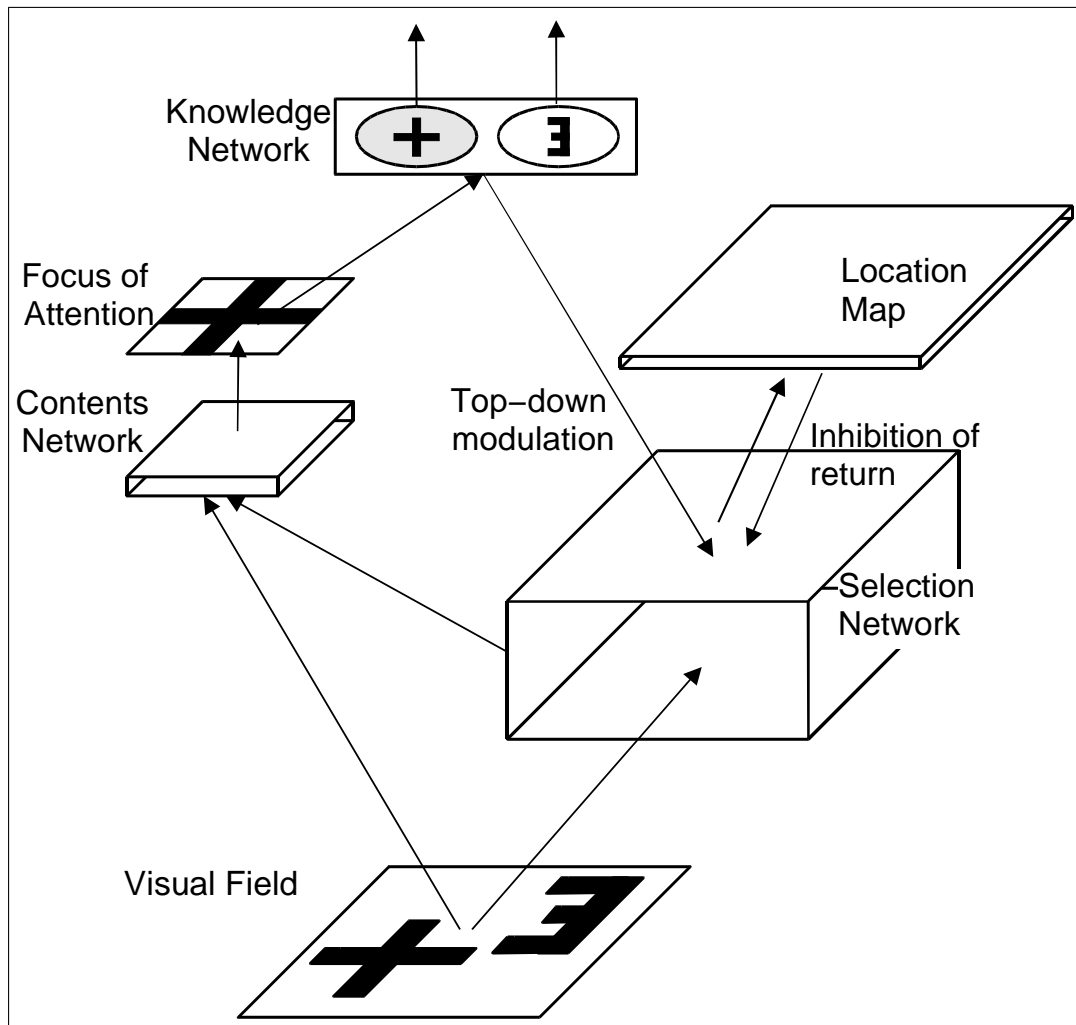


Figure 2: The architecture of SAIM.

from an object once recognized to other objects in the field (using a location map). Figure 2 shows the full architecture. The details are outlined in the following paragraphs.

### 1.3.1 Dynamic routing circuits in SAIM

SAIM uses one-to-one mapping from the retina to all locations in the FOA, mediated by connections between the contents and selection networks. In principle this allows any arbitrary mapping to be achieved from the retina to the FOA. However, such mappings should not be arbitrary. For example, in order for subsequent recognition to be successful, neighboring units on the retina should be mapped into neighboring units in the FOA (e.g., units 1 and 2 on the retina into units 1 and 2 in the FOA, in Figure 1), and the relative orientations of the retinal units needs also to be preserved. Constraints needed to preserve neighborhood relations and relative orientation information are embodied into the architecture of SAIM by means of the connection weights between units in the selection network, and they are illustrated in Figure 3, which provides an example of how a one-dimensional input might activate both the contents and the selection networks. The constraints incorporated into the selection network in SAIM are that: (i) neighborhood relations in the retinal input tend to be preserved in mapping through to the FOA; (ii) one

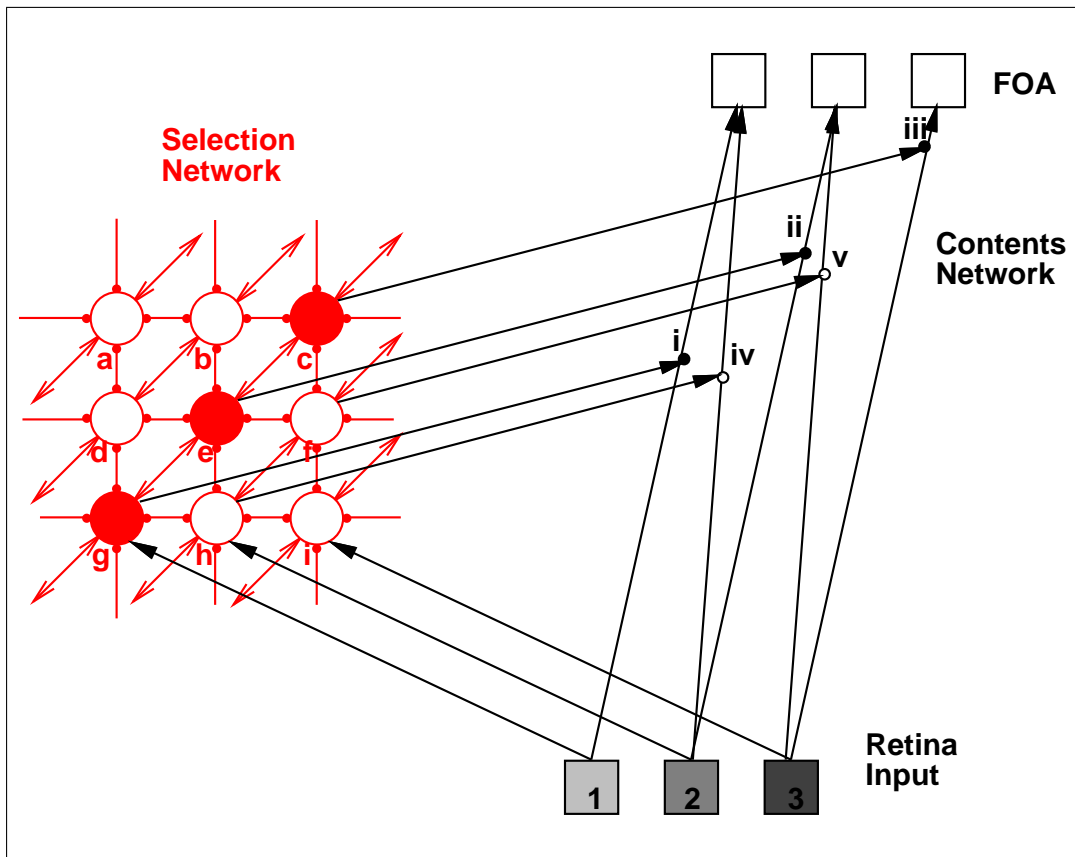


Figure 3: An illustration of how connections between units in the selection network impose constraints on the mapping from the retina to the FOA in SAIM. 'Vertical' and 'horizontal' connections in the selection network are mutually inhibitory whilst 'diagonal' connections are mutually excitatory; see the text for details. Within the Selection network columns contain units activated by one given position on the retina; rows contain units that activate one position in the FOA.

retinal unit should not be mapped more than once into the FOA; and (iii) one unit in the FOA should not receive input from more than one retinal unit (the converse of (ii)). These constraints define the activity patterns in the selection network that are permissible and others that are not. Following a procedure suggested by (Hopfield 1984, Hopfield & Tank 1985), we defined an energy function for which the minimal values were generated by just the permissible activity values. To find these minima, a gradient descent procedure was applied resulting in a differential equation system, given in the Appendix. The differential equation system defines the topology of the network (i.e, the strengths and signs of the connections between units).

The constraints are captured in the following way. First, there are excitatory connections between units in the selection network that would preserve neighborhood relations in mapping from the retina to the FOA; these are indicated by the arrow connections between units along the diagonals of the selection network shown in Figure 3. The strength of the excitatory connections has a gaussian function with respect to the distance between units; more proximal units have stronger connections than more distant units. This produces the sensitivity of the model to the spatial proximity of pixels in an image. Second, there are inhibitory connections between units in the selection network that

would map the same retinal location into different units in the FOA. In Figure 3 these are illustrated by the  $-o$  connections along the vertical axes of the selection network. Although only connections between neighboring units are shown for illustrative purposes, there are inhibitory connections (with equal weights) between all units along the vertical and horizontal axes of the selection network. For example, in our figure units e and h would map retinal location 2 into different units in the FOA (via units ii and iv in the contents network); hence units e and h are mutually inhibitory. Third, there are inhibitory connections between units that would map from different units in the retina into the same unit in the FOA. These are illustrated by the  $-o$  connections along the horizontal axes of the selection network shown in Figure 3. For example, units e and f 'compete' to gate activity into unit 2 in the FOA (via units ii and v in the contents network). The implementation of these constraints in the dynamic routing circuit is somewhat analogous to the way in which constraints are effected in the Marr-Poggio network for stereo depth (see Marr & Poggio 1976, Olshausen et al. 1993).

As we shall demonstrate, one of the emergent properties of these constraints is that the units in the selection network that receive the most support are those that lie at the location of what approximates to the Gaussian mass of a shape<sup>1</sup>. The greater the Gaussian mass (determined by, amongst other things, the number of immediately surrounding active pixels), the faster a stimulus is mapped through to the FOA (objects with higher Gaussian mass receiving most neighborhood support). Furthermore, the location of the Gaussian mass of a shape tends to be translated into the center of the FOA (since this receives most surround support).

The model as described enables input to be mapped into the FOA in a manner that is translation invariant; any region of the retina can be attended<sup>2</sup> We did not introduce constraints that could reflect retinal acuity in the model, though this could be done by having inhomogeneities in resolution across the retina<sup>2</sup>.

### 1.3.2 The knowledge network

In addition to having a dynamic routing circuit to achieve translation invariant mapping into the FOA, the full version of SAIM also contains a higher-level 'knowledge network', the activation of which achieves object recognition. The knowledge network contains 'template' units, each of which represents a known object. Template units are activated on the basis of learned weights which connect with pixels in the FOA. Note that, in general, objects are mapped so that the location of their Gaussian mass falls at the center of the FOA (see above). Consequently, stored knowledge of objects is sensitive to the spatial positions of features relative to the centroid of a shape (cf. Marr 1982). The stored units also act to bind the parts of objects to their spatial positions within the shapes, and template units are activated maximally when features fall in spatial locations that match those specified in the template. Hence templates are sensitive to viewing angle, though they are translation invariant (being activated through the FOA, irrespective of the lateral position of an object on the retina).

The introduction of the knowledge network into SAIM adds a top-down constraint on selection in the model. The input will activate most the best matching template, which,

---

<sup>1</sup>see Study 1 for further details

<sup>2</sup>We did not introduce constraints that could reflect retinal acuity in the model, though this could be done by inhomogeneities in resolution across the retina

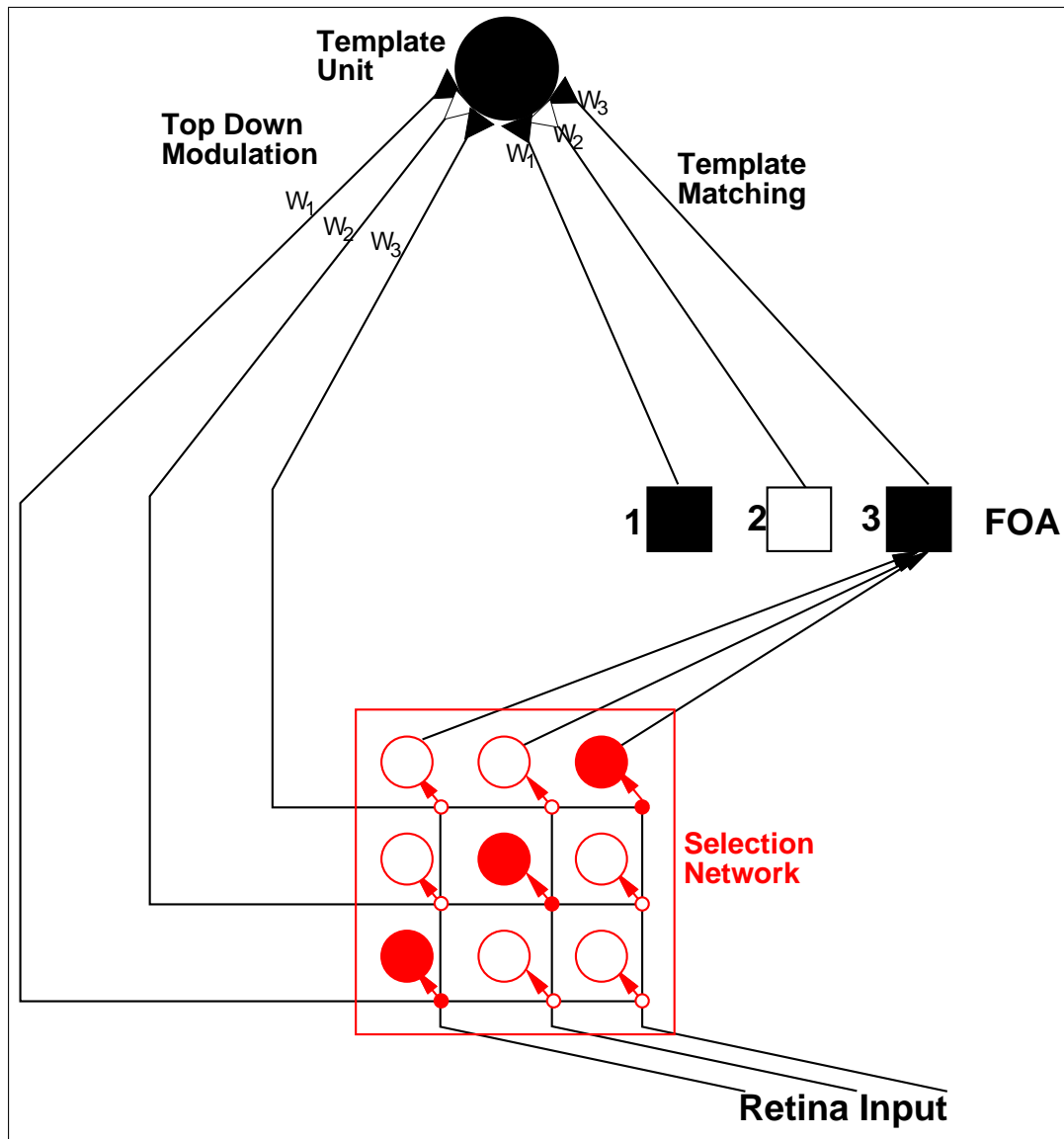


Figure 4: Top-down feedback in SAIM. Activation from template units in the knowledge network feedback to activate units in the selection network supporting activation in an appropriate attended location (in the FOA). Only relevant connections are shown. Note that top-down connections are the same strength as the bottom-up ones, from the FOA to the templates. The top-down connections from one location in the template (and the FOA) feed-back to all units in the Selection network that would activate this location in the FOA and template (i.e., they feed-back to all units within one row of the Selection network as shown in this figure).

given the 'winner take all' properties of the network, leads to inhibition of templates that match less well. To integrate this constraint into SAIM, an energy function was defined for the knowledge network and added to the energy function in the selection network. Here the gradient descent procedure results in a topology which connects the template units in a winner- take-all (WTA) fashion (see the Appendix for details).

The templates in SAIM generate top-down modulation of the selection network. The top-down modulation can be thought of in the following way. Each template modulates

the effect of input coming from the retina into the selection network via weighted top-down connections. For simplicity, the weights on these top-down connections are set to be the same as the weights on the bottom-up connections into the template unit. Activity in a template unit is then multiplied by the top-down weights (in a standard activation function) to influence the selection network. This is done by feeding back activation, via the weighted connections, to the horizontal layers within the selection network, where the units correspond to one location in the FOA (e.g., in Figure 4, feedback via W3 from the template unit would be transmitted to the top horizontal row of the selection network, where units correspond to position 3 in the FOA). Essentially, the top-down activity from a template is used as another input that influences activity in the selection network, with the top-down activity being multiplied with rather than simply being added to the bottom-up activity in the selection network (by bottom-up activity here we refer to excitation from the retina and from the inter-level connections in the selection network). This top-down activity provides positive, multiplicative support for units in the selection network that will lead to the selection of pixels in the FOA that match the template; it also decreases support multiplicatively for units that would lead to the selection of other pixels. In effect, it provides a representation of how well bottom-up activity matches the template units and it 'sharpen's' activity in the selection network to favour known objects (due to its multiplicative operation). This process is illustrated in Figure 4. Here activity in the FOA leads to bottom-up activation of a template unit, via weighted connections  $w_1$ - $w_3$ . Activity for each location in the template is then fed-back to units in the selection network corresponding to that location, with the feed-back being modulated by the same weights ( $w_1$ - $w_3$ ). The feed-back support is the highest for units in the selection network with matching retinal activity. This is shown by the dark circles in the selection network in Figure 4, whereas units where the bottom-up activity does not match are inactive (in white). Note that, because the top-down modulation from the templates is not tied to any specific retinal location, it is translation invariant; it is applied to all units in the selection network that would support the active template. Nevertheless, since top-down activity combines with bottom-up activity within SAIM, and bottom-up activity is tied to retinal location, the effect of top-down feedback is to improve selection for a particular object in a particular part of the visual field<sup>3</sup>.

### 1.3.3 Shifting attention and the location map

One further important property of SAIM is its ability to shift attention between items, once stimuli are identified. The competitive nature of the model means that, by and large, identification is limited to one item at a time. When multiple objects are present, there needs to be some mechanism to enable each to be selected in turn. In the model switching of attention is accomplished by top-down inhibition which takes place once a template unit reaches a given threshold: there is temporary suppression both of the selected template and of the selected location. This enables other items in the field to then win the competition for selection, so that they in turn become attended. Suppression of previous winners provides a form of 'inhibition of return' (IOR; cf. Posner & Cohen (1984)). To enable the location of selected items to be inhibited, activity in units corresponding to

---

<sup>3</sup>Where units in the selection network are inactive, due to lack of bottom-up support from the retina, the top-down feedback is multiplied by zero and has no effect. Feedback is effective only for units in the selection network activated in a bottom-up fashion.



each retinal location in the selection network is summed and passed on to a 'location map', which essentially represents which locations currently hold items of attentional interest. Early in the processing of a display with multiple objects, several locations in the map will be occupied. However, as the dynamics of competition within the selection network evolve, so active units in the map will decrease until only one remains; this represents the location of the attended object. Once the target is selected, the active unit in the location map inhibits input from relevant location in the visual field, so that it is no longer represented in the selection network (see Figure 2, where the inhibitory link from the location map to the selection network is specified). This leads to a discontinuity in the energy function from which the settling process must start again, but in this case there is a decreased likelihood that spatial regions previously attended are subsequently selected. Inhibited positions in the location map also provides SAIM with a form of spatial memory for previously attended regions of field.

Several previous models of visual attention have incorporated the idea of a location map to represent regions of current interest in the visual field (e.g., see Koch & Ullman (1985) and Wolfe (1994), who both use a 'saliency' map with some similar properties). Unlike these models, however, we do not use the location map as the locus of the competitive interactions that determine selection; in SAIM these interactions take place in the selection network. This enables SAIM to separate out the factors that determine selection, which include grouping factors (such as inter-element proximity) and stored knowledge, from information about the locations that attended objects occupy. In addition, the location map in SAIM can provide some implicit indication of the number of items in the field (the initially activated locations), and it modulates the switching of attention from one selected object to other to-be-selected objects. In the General Discussion we elaborate on why these are useful properties of the model and how they fit with neuropsychological and physiological results.

There are some interesting implications of the way in which the location map, and inhibition of return, is implemented in SAIM. For example, units within the selection network can be activated (by excitatory spreading of activation) even if they are unoccupied on the retina. Thus when a target is mapped into the center of the FOA, the 'empty' regions around it (in the FOA) are active within the selection network and in the location map. Following selection, any inhibition of the relevant units in the location map, and associated active units in the selection network (via IOR), suppresses units for these surrounding, 'empty' regions of field. It follows that inhibition effects should not be confined to previously attended objects but they should be found with objects presented in retinal positions adjacent to those of previously attended objects. Effects along these lines have been reported by Kim & Cave (1995), Klein (1988) and vonMühlenen & Müller (in press). In addition, there is also temporary inhibition of the template of the target first identified. This produces a bias against attending to the same object again, and a bias towards attending to new objects (cf. Gibson & Egeth (1994); Tipper et al. (1991)), a strategy with obvious advantages for organisms in the real world. Note that template-inhibition is translation invariant, since the same template is activated by objects irrespective of their position in the visual field. There is inhibition of specific, known and previously attended objects, as well as inhibition of the local spatial area surrounding a previously attended target. In the General Discussion we return to consider how well this 'object specific' aspect of processing in SAIM relates to evidence on 'object-based' IOR. Following selection and the imposition of IOR, the model begins processing afresh, with it now biased against

attending to previously occupied retinal positions and previously attended objects.

### 1.3.4 Discussion

The architecture of SAIM is of general interest because the model has no commitment to certain experimental paradigms; for example it can be used to model detection, letter recognition, object recognition, even search. This distinguishes SAIM from at least some other models of visual attention, which have been designed specifically to capture particular procedures (eg., cueing attention, in Cohen et al. (1994); search for T-elements, in Humphreys & Müller (1993); selection of letters based on their colour; Phaf et al. (1990)). This broader approach should, in the longer term, enable SAIM to function as an architectural framework for a wide range of visually-mediated tasks.

SAIM also operates in a deterministic manner, and so all effects are stable across multiple runs of the model with the same input. Given the same input, the model will always produce the same result (even if the effect is small). Performance of the model is measured in terms of network iterations (where 1 iteration involves updating all units in the model on one occasion). For simulations involving object identification, identification was judged from the activation of a template unit to a set threshold (.9). For simulations of detection performance (where the stimulus need not be identified), the duration of selection was judged to be the average time taken for units in the selection network to be activated. Since activation in the selection network biases activation in the FOA, this is equivalent to setting a threshold for target detection, based on the integration of activation within the FOA. Parameters for the network are given in the Appendix.

Our goal in this paper is to capture qualitative patterns of performance that can be observed in studies of human visual selection; for example, that there is a bias towards attending to the center of mass of a stimulus, that there are costs to selection when two relative to when one stimulus is presented (particularly for the object selected second), that there are costs when attention is cued to the wrong spatial position or object. We ask, can a model which employs the general architectural constraints of SAIM capture the required pattern of constraints on selection found in human subjects, despite the simplifying assumptions made? We return to consider the relations between qualitative and quantitative modelling in the General Discussion. In the General Discussion we also elaborate on factors the model does not accommodate and on how a qualitative comparison could lead to SAIM being falsified.

### 1.3.5 Testing the model

In order to test SAIM, we explore the behaviour of the model as a function of a number of variables known to affect both normal and disordered visual selection in humans (including factors such as (i) presenting two rather than one object, (ii) spatial and object-cueing, and (iii) the effects of stored knowledge on visual neglect, etc.). Altogether we simulate 16 different sets of results from the psychological literature using essentially a single set of parameters. In Study 1, we present the model in its 'bottom-up' form (without the knowledge network), simply to show how bottom-up factors help determine the operation of the model when it is tested in its full form. This helps us to understand the operation of the model, even though simulations of psychological data are always based on the full version of SAIM (with the knowledge network incorporated). The parameters for these

initial, bottom-up simulations were kept constant when the knowledge network was added to the model. The only changes subsequently made to the model were to alter the degree of top-down feedback to favor one target over another, either by increasing the starting level of activation for one template in the knowledge network or by increasing the weights connecting to and from one template (allowing Hebbian learning to take place). These last changes were made solely to explore the impact of object familiarity on selection; in this sense the changes do not constitute attempts to fit the model to the data but rather they provide predictions of how a particular variable (object familiarity) will impact on selection within the model.

## 2 Modelling normal visual selection

In this first section, we assess whether SAIM provides a reasonable account of visual selection in normal (non-brain damaged) humans. We explore both the general properties of the model (bottom-up procedures for selection, performance with single vs. multiple items, attention switching in the full version of the model and effects of stored knowledge on selection), and its application to paradigms used to examine visual selection in humans (e.g., the 'Posner' cueing paradigm, and the combined effects of spatial and object information in selection). SAIM provides a general framework for understanding selective processes in vision, including those found in visual search. However our interests here are focused on performance of the model when no more than two items are present in the field. As we shall show in the section on attention switching (Study 3), the mechanisms present in the model can transfer to search in a straight- forward way.

### 2.1 General properties of SAIM

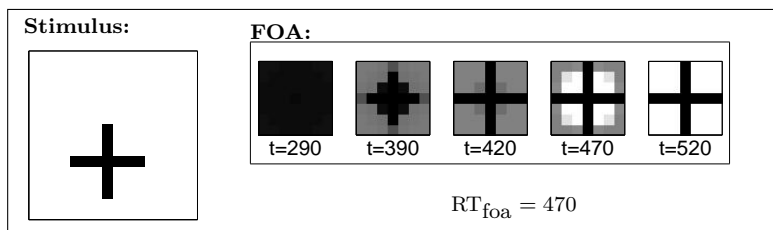


Figure 5: Illustration of the behavior of SAIM in bottom-up mode, when presented with a + stimulus.  $t$  gives the time in network iterations.  $RT_{foa}$  gives the time for the detection threshold to be reached in the FOA (again in network iterations). Activation is depicted within the FOA as processing time increases.

#### 2.1.1 Study 1. Bottom-up mapping

##### Method

As we have outlined, SAIM incorporates both bottom-up and top-down constraints on selection (e.g., inter-element proximity and stored object knowledge). In the full version of the model, these factors interact in determining selection, as we illustrate in the following studies. However, due to interactions from stored knowledge, it can be difficult to evaluate how the bottom-up constraints influence selection in their own right. To show how these constraints operate, we report three sets of simulations in which the model operated without the knowledge network. The parameters of the bottom-up part of the model (i.e., the full architecture without the knowledge network; see Figure 2) were the same as in the simulations using the full architecture. These bottom-up simulations, then, simply allow us to see how the model is biased to select some perceptual structures over other, even prior to it learning object representations for the stimuli. In these and all other simulations shown here, the FOA had a fixed size: 7 x 7 pixels. The retina was 16 x 16 pixels. The contents and the selection networks instantiated all connections between the retina and the FOA.

First, we presented a set of pixels arranged as a cross placed in the lower part of the retina

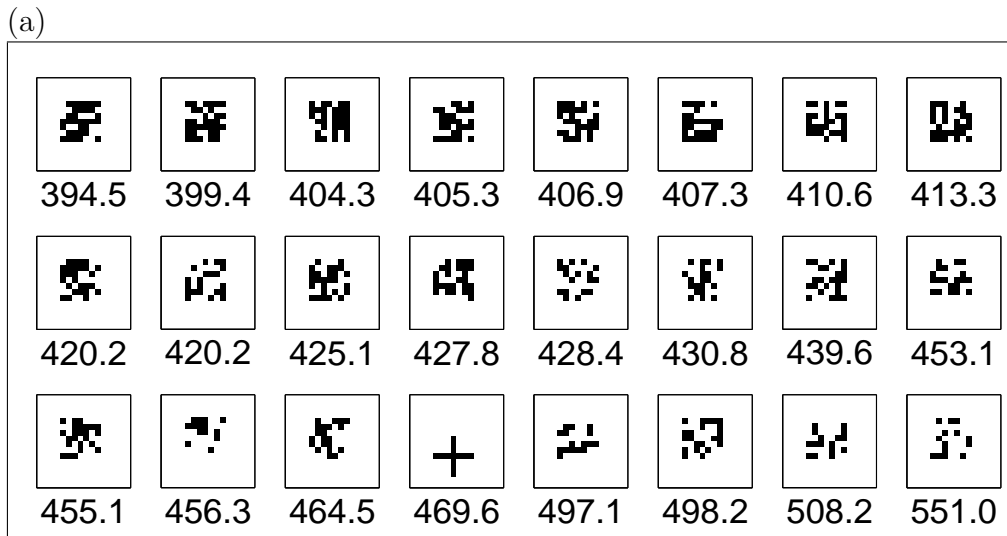


Figure 6: (a) Examples of stimuli covering a random arrangement of pixels within a central 7 x 7 area of SAIM’s retina. The numbers denote the time taken for convergence to be reached within the FOA.

(see Figure 5). Activation was then mapped across network iterations until the threshold in the FOA was surpassed. This happened to take 497 iterations. More important, the example shows that activation that is not at the center of the retina is nevertheless mapped into the center of the FOA, due to the constraints of proximity, uniqueness and so forth that characterise interactions within the selection and contents networks. The same held irrespective of where the stimulus was in the field. Attention can fix on an object irrespective of the object’s retina location. In addition, Figure 5 illustrates that activation in the FOA begins at the center of gravity of the shape, and then gradually spreads out. This ‘spread of attention’ is again brought about by proximity constraints in the model, since there is most neighborhood support for pixels at the center of gravity of shapes.

Second, we examined the rate of convergence of activation in the FOA when we presented the model with 32 randomly chosen ‘patterns’ (random pixel arrangements within a 7 x 7 area on the retina)(see Figure 6a for the full set of patterns and for the times to convergence in the FOA). RTs to attend to the patterns differed, and did so in a systematic fashion<sup>4</sup>. To understand how these RT differences arose, we analysed the patterns for: (i) the number of pixels present (the mass); (ii) the Gaussian mass (the sum of pixels weighted by a Gaussian function so that pixels at the centre of the function are weighted most strongly, with a standard deviation set to 6); (iii) a ‘textured’ Gaussian mass (the texture measure reflected the number of white-to-white pixel transitions present across both horizontal and vertical axes, normalised by the number of pixels; this value was high when the pixels are part of a homogeneous area<sup>5</sup>). In Figure 6b we present the RTs (in network iterations) as a function of each measure for each pattern (the mass, the Gaussian mass and the textured Gaussian mass). It is clear that each measure can account for a significant amount of variance, with the Gaussian mass measure accounting for more than

<sup>4</sup>This set of simulations further illustrates the stability of the model across variations in the stimuli.

<sup>5</sup>The ‘textured gmass’ measure, given in Figure 6, was based on the Gaussian mass x (7 x the normalised texture measure).

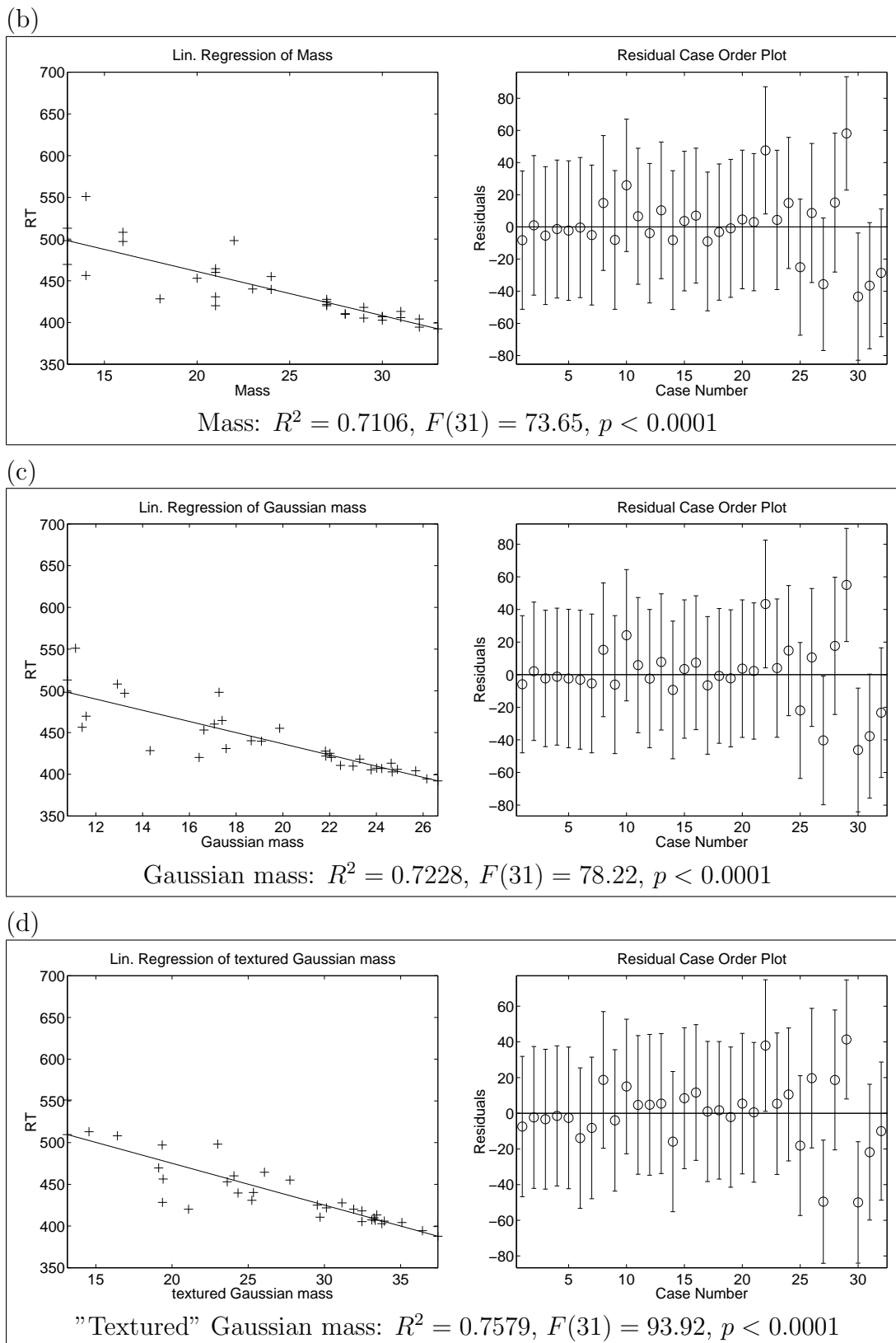


Figure 6: (b) (d) represent correlations between the time taken for convergence within the FOA and measures relating to the mass of the shape. The highest correlation is with the textured Gaussian mass measure (see the text for details).

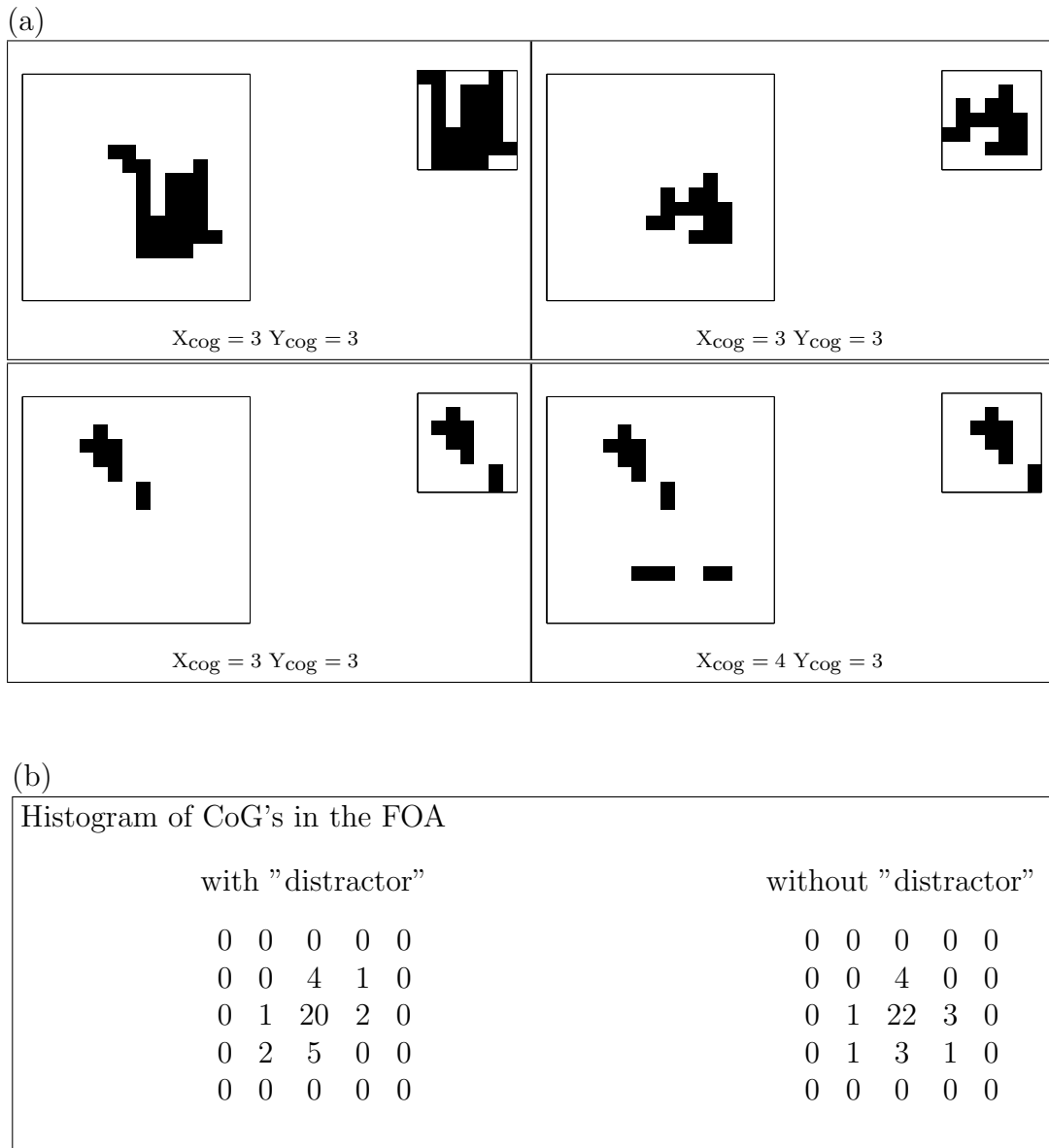


Figure 7: Examples of stimuli being placed with their center of gravity at or adjacent to the central pixel within the FOA. The top sections and bottom left section in (a) illustrate patterns positioned with their center of gravity at the center of the FOA ( $X_{cog}=3$ ,  $Y_{cog}=3$ ); the bottom right section in (a) indicates how selected patterns can shift their position within the FOA if there are other stimuli also in the field (now  $X_{cog}=4$ ,  $Y_{cog}=3$ ). (b) Provides 2 frequency accounts of where the center of gravities of shapes are placed within the FOA, according to whether a distractor is present or absent.

the measure of the number of pixels, and the textured Gaussian mass measure providing the best fit. These analyses illustrate that RTs in the model are influenced by something like the Gaussian mass of the pattern (perhaps reflecting also the textural homogeneity of pixels). The higher the textured Gaussian mass measure, the faster the RT. This bottom-up bias in SAIM is caused by the proximity constraints that affect interactions within the selection network. There is maximal support for patterns whose pixels are close and that have no gaps between them; this holds across a broad set of stimuli. The textured

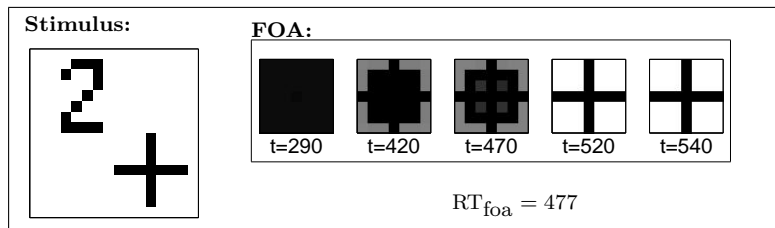


Figure 8: Competition for selection within the bottom-up version of SAIM. Here activation in the FOA is shown when two stimuli are present in the visual field. Note that the time to attend to the + (in the FOA) is increased relative to when the + is presented alone (Figure 5).

Gaussian mass (textured gmass) determines the bottom-up 'saliency' of a stimulus for the model.

In a third study of bottom-up constraints on the model, we presented SAIM with 2 random patterns where each pattern was based on a random arrangement of pixels within a 7 x 7 area on the retina. Performance was compared with when only a single random pattern was presented (defined by covering only a 7 x 7 area). Examples of the single and double patterns are shown in Figure 7. When the patterns had different 'textured gmass' values, the pattern that was mapped through to the FOA first was the one with the largest textured gmass (see also Figure 8). When a single random pattern was presented, its center of gravity tended to be mapped through to the center of the FOA (pixel 3,3 in the FOA), due to the proximity constraints within the selection network. However, this could be affected by the second pattern. For example, in Figure 7a, lower panel right, the center of gravity of the selected figure was placed at pixel 4,3 in the FOA when 2 patterns were present, but pixel 3,3 when only the selected figure occurred (Figure 7a, lower panel left). This small 'shift effect' indicates that an unselected stimulus can have some influence on how attention is fixed upon a selected item, due to spreading activation within the selection network. Figure 7b presents the numbers of occasions (across 35 different sets of single and double patterns) where the center of gravity of the pattern with the larger textured gmass was sited at particular pixels in the FOA (thus the center value in the matrix represents the center of the FOA). Note that the center of gravity of the pattern tended to be placed at the center of the FOA on slightly fewer occasions when a 'distractor' pattern was present (the center of gravity was mapped to pixel 3,3 in the FOA on 20/35 occasions), than when it was absent (when the center of gravity was mapped into pixel 3,3 on 22/35 occasions).<sup>6</sup>

Figure 7 illustrates that there is competition for selection in SAIM when two rather than one stimulus is presented; only one of the two presented stimuli is selected at one time, in the FOA. This same point is illustrated in Figure 8, where we present the results from a simulation using stimuli that can be easily labelled (a + and a 2, rather than the random patterns shown in Figures 6 and 7; we will use these figures as examples throughout the rest of the paper). The stimulus 2 (Figure 8) has a textured gmass value of 14.42. The textured gmass value for the + is 19.12. Each stimulus covers a 7 x 7 pixel area on the retina. Given these two stimuli, SAIM maps the + into the FOA. The + wins the competition for selection because it has a higher bottom-up saliency (a higher gmass value) than the 2. In this simulation, the + took 477 iterations for convergence to take

<sup>6</sup>In all cases SAIM selected an area of field 7 x 7 pixels in area, due to the fixed size of the FOA.



place within the FOA.

One thing to note from this last simulation is the contrast in the time for convergence to take place within the FOA when two rather than one stimulus is present, even when the stimulus is the first to be selected amongst multiple shapes (e.g., contrast the convergence time of 477 iterations when the 2 accompanied the + with the time of 470 iterations when the cross appeared alone; Figure 5). There is a small processing cost when there is competition for selection even for the stimulus that is preferred for selection (and so selected first). Several researchers have also shown that the mere presence of other items in the visual field can delay human identification of a target. This 'filtering cost' occurs even when the stimuli are presented at separate spatial locations and when distractor stimuli have little resemblance to targets (Eriksen & Hoffman 1972, Treisman et al. 1983), making it unlikely that effects are due to a cost in switching attention on occasions when distractors are selected prior to targets. SAIM produces a filtering cost effect due to competition within its selection network when several items appear simultaneously in the field, and even for the target selected first. This characteristic of SAIM is robust across wide variations in the stimuli used (e.g., see Figures 9, 10, Figure 11(c) compared with Figure 9(a) for further examples).

## Discussion

These simulations illustrate several points:

1. SAIM can map activity into its FOA in a purely bottom-up manner, determined by the constraints operating within the selection network. It does this in a selective manner. Stimuli that are spatially separated can be parsed into two separate objects based on proximity relations between pixels in the image, and there is competition between the objects for selection so that only one is selected at a time.
2. Representations formed within the FOA tend to be based around the center of gravity of the shapes. SAIM maintains that our attention tends to be focused at the center of gravity of objects and that the attended representations used for subsequent object recognition will use a reference frame fixed on this location. Here SAIM is supported by psychological evidence. For instance, eye movements to visual stimuli tend to fall at the center of gravity of shapes (Findlay 1982, McGowan et al. 1998, Ottes et al. 1985), and even when free vision is allowed subjects tend to bisect lines (and presumably fix attention) according to the relative center of mass of the stimulus (Shuren et al. 1997). Also studies of shape perception show that axis-based descriptions are frequently used both to identify and to match shapes across successive presentations (Humphreys (1983), Humphreys & Quinlan (1987); see also Marr (1982)). Since SAIM's attention tends to be fixed at the center of mass of an object, there is a natural basis for recognition to operate using descriptions centered on this point, which is typically coincident with where a principal axis would be defined. Furthermore, Hirsch & Mjolsness (1992) showed that displacement discrimination in random dot patterns can be best described by a model which uses a globally-computer center of mass parameter, and Morgan et al. (1990) have argued that many visual illusions (including the Müller-Lyer) are based on deviations produced by global center-of-mass computations performed by early vision.
3. For SAIM not all of an object is attended simultaneously; there is a time course

to focusing attention in which the center of a shape is attended first (the pixels at the center of the FOA gaining the initial strongest activations levels) and then attention spreads outwards (see Figure 5). In contrast to the spread of activation in the selection network, the coding of shape at earlier stages of the model operates in parallel across an object, and indeed such parallel coding is necessary in order that attention can be drawn to the center of gravity of a shape. Although we have noted above psychological evidence indicating that attention is biased to the center of gravity of shapes, we know of no data on the detailed time course of this process, though SAIM predicts that attention should spread across objects, from the center of gravity outwards. The same result occurs in SAIM when the top-down component of the model is added (e.g., see Figure 9 for an illustrative example).

It is tempting to relate this last aspect of the model, in which the center of a shape is attended before the whole, to psychological studies on wholistic effects on pattern recognition, in which identification responses to global shapes can be faster than those to their constituent parts (Navon 1977). However this would be misleading. Though attention in SAIM is initially focused at the center of a shape, other parts can be rapidly attended following on from this, and all of the parts within the FOA will activate stored templates in the full version of the model. There will be an early advantage for local parts mapped into the center of the FOA, but this will be balanced against activation of templates by a greater number of pixels within the global shape. Whichever template is activated most will depend on the relative balance between the local and global sources of information. This is illustrated by the simulation of wholistic perception, presented in Study 4, where we introduce templates into the model.

4. Bottom-up mapping of activation into the FOA acts to 'group' pixels together even though they are spatially separated (see Figures 6 and 7). Bottom-up grouping in SAIM is based solely on proximity relationships in the image. If pixels are sufficiently close, they will be mapped into the FOA as part of the same object. More sophisticated bottom-up grouping processes undoubtedly play a role in human visual attention (collinearity, closure etc.; see Donnelly et al. (1991); Elder & Zucker (1993); Rensink & Enns (1995), for evidence), but these are not simulated within this version of the model. Within the full version of the model, when top-down as well as bottom-up constraints operate, then grouping can also be imposed by top-down knowledge.
5. A good measure for the speed of convergence of activity in the FOA is not the total mass of the shapes (e.g., the summed number of pixels) but rather the textured gmass, as defined above. Since the gmass increases as the distance between pixels in a shape reduces, it is sensitive to proximity grouping in shapes. Reaction times (RTs) to stimuli should reflect the ease of grouping based on proximity. It should be noted, however, that we are not predicting that, in humans, a + would be identified faster than a number 2 because it has a higher textured gmass value. RTs in humans depend on many factors, including the exact physical properties of the stimuli, their relative familiarity, and forms of grouping not implemented in SAIM (e.g., collinearity of edges; see (4) above). What is critical is our simulation

of a qualitative difference between the times to attend to the two shapes which differ in their textured gmass, and the + and 2 stimuli simply illustrate this for the model. The simulations with a wider set of random pattern (Figures 6 and 7) illustrate that both the faster RTs and the selection bias for shapes with a higher textured gmass holds across a range of stimuli. We predict that textured gmass is an important factor in determining both the speed with which attention is deployed and its positional focus.

### 2.1.2 Study 2. Accessing stored knowledge and attention- switching: Bottom-up biases

Study 1 used a version of SAIM which had no knowledge network in order to illustrate the nature of bottom-up biases on selection in the model. A network without the ability to match stimuli to stored knowledge might be used to simulate certain tasks performed with humans (e.g., simple target detection) but it cannot be applied to studies using identification measures, for which access to stored knowledge is necessary. In addition, the bottom-up version of the model has no procedure for attention switching. From Study 2 on, then, we use a version of SAIM which incorporates stored knowledge, and uses this knowledge to modulate selection. The full version of SAIM also has the ability to switch attention from a first to a second stimulus in the field. In this version of the model, we can also test whether the processing cost in selecting two vs. one stimulus is particularly large for the stimulus selected second. This result in humans has been reported by Duncan (1980), amongst others.

#### Method

The bottom-up parameters in SAIM remained constant. The parameters for the top-down components making up the knowledge network are given in the Appendix. The knowledge network was also used to introduce into SAIM the capacity to scan attention (see also Olshausen et al. 1993). Template units were given an identification threshold (.9), after which identification was considered to have taken place. Once this threshold was passed, inhibition was applied (see above).

The stimuli presented to the model were a + and a 2, which we observed previously to show asymmetric effects in selection when presented together in the field (Figure 8, Study 1; there was a bottom-up bias favouring the +). In Study 2 we examined the identification of these stimuli both when presented in isolation and when presented together. When the stimuli were presented together we asked whether attention could be switched from the object that was 'better' in a bottom-up sense (the +) to the object that was worse (the 2). The initial activation values of the two templates, and their thresholds, were set to be equal so that any differences between the stimuli would reflect bottom-up factors.

#### Results and discussion

Figure 9 (a) and (b) illustrates the time course of activation in the template units when the 2 and the + stimuli were presented alone. There remained a bias to favour the +, with the number of iterations required for the + to reach threshold at the FOA and template levels being 438 and 610 iterations respectively. This compares with 543 and 765

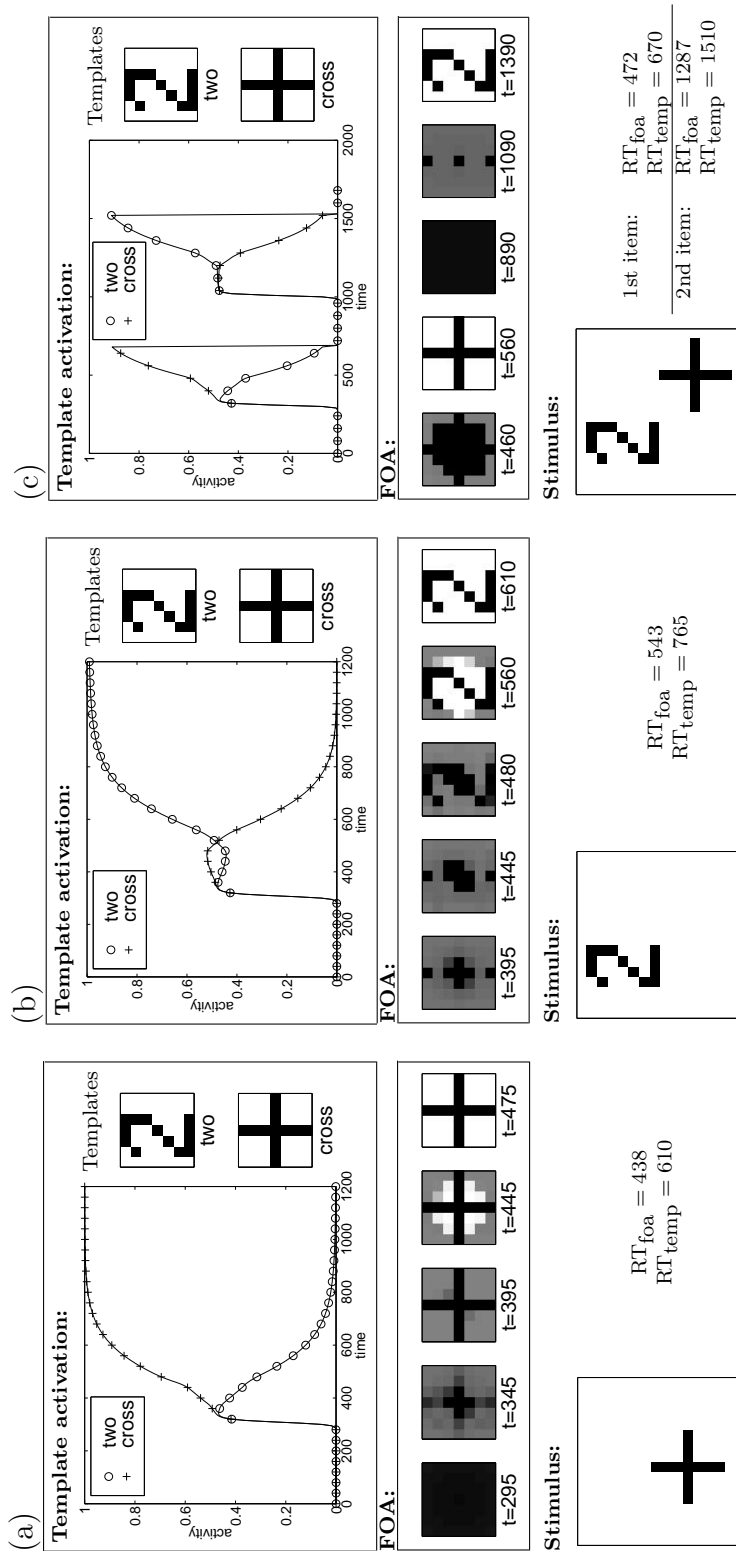


Figure 9: Performance of the full version of SAIM, with templates for the stimuli + and 2. (a) + presented alone; (b) 2 presented alone; (c) 2 and + presented together. Activation is depicted in the FOA and in the template units. RT<sub>temp</sub> is the time taken for the identification threshold to be passed, at the template level (in network iterations). When the 2 and + are presented simultaneously, SAIM selects first the + and then the 2, with the 'winner' being the stimulus with the higher textured gmass measure.

|                                    | + (salient stimulus) |          | two (less salient stimulus) |          |
|------------------------------------|----------------------|----------|-----------------------------|----------|
|                                    | FOA                  | Template | FOA                         | Template |
| Cost<br>(two items - single items) | 34                   | 60       | 744                         | 745      |

Table 1: Filtering and switching costs when 2 rather than 1 target is in the visual field, in SAIM (measured in network iterations)

iterations for the 2<sup>7</sup>.

Figure 9c demonstrates the consequences of presenting both the + and the 2 together in the field. The + is selected and identified first, but following this, selection and identification proceeds for the less good stimulus, the 2. SAIM switches attention from the first to the second item. As in the pure bottom-up version of the model, there are general consequences of having two items in the visual field: the time for attention to converge and for identification to take place is increased even for the better of the two stimuli. This effect is somewhat larger here than we observed in the purely bottom-up version of the model, due to slightly different convergence properties within the larger network. However this simply emphasises the robustness of the effect. Table 1 gives the mean increases, in network iterations. For the less good of the two stimuli (the 2), the costs of identifying two relative to one item were increased even more than was the case for the preferred item (the +). Such an increased cost for the stimulus selected second in a two-item display has been found in studies of target selection under brief presentation conditions (Duncan 1980). SAIM captures this 'switching cost' effect on the poorer item.

Study 2 demonstrates a bias in selection in SAIM: given multiple items in the field, and equal top-down biases for the stimuli (see Study 3 below), SAIM tends to select first the object that is 'better' in the sense of having higher priority for the selection network: here, the object with the largest textured gmass (the + before the 2). In the version without the knowledge network and attention switching (Study 1), this simply leads to the + being attended first. In the full version of SAIM (Study 2), the + is both attended and identified first, and then attention is switched to the 2. This bias for the 'better' object is increased when the good and less good objects are presented together relative to when each is presented alone (in the single object baseline conditions). When the good and less good stimuli are presented briefly together SAIM predicts that the less good stimulus is less likely to be identified, even when it can be identified when presented in isolation, since the time to select the second stimulus will be limited by the exposure duration. Evidence for such a bias was reported by Farah et al. (1991) (see also Gorea & Sagi (2000)). Farah et al. attempted to simulate extinction effects in patients with brain lesions by exposing normal subjects to stimuli in one visual field that were degraded. They found that the degraded stimulus could be identified as well as a brighter one when each was shown in isolation; however there was a bias to identify only the brighter one when the bright

<sup>7</sup>Note that there is a general slowing of RTs when the knowledge network is added, relative to when a smaller model is used (without the knowledge network; e.g., Study 2). This slowing is due to the additional knowledge layer added to the network, which requires that activation be transmitted through a further set of units and that there is further competition to be resolved. With multi-item displays, the benefit is that it enables stored knowledge to be used for segmentation in instances where bottom-up segmentation is difficult; Study 6 here provides an example of this.

and dim stimuli appeared simultaneously. This is analogous to spatial biases in selection found in patients showing spatial extinction, as we go on to demonstrate by simulation (see Study 6). As noted in the Introduction, competitive biases of this form, favouring the better of two stimuli, are also suggested by a number of current general accounts of selection in humans, including the integrated competition account suggested by Duncan and colleagues (Desimone & Duncan 1995, Duncan 1998, Duncan et al. 1997)<sup>8</sup>.

One other point to note here concerns the magnitude of the filtering and selection costs (for the + and the 2 respectively, in Studies 1 and 2), relative to differences in the time to attend to and identify the stimuli themselves. For example, the difference in the time to attend to the + and the 2 (judged by RT<sub>foa</sub>) was 105 iterations, whilst the filtering cost for the + was 34 iterations (the selection cost for the 2 was somewhat larger: 744 iterations; Figure 9). However it is difficult to make any conclusions from this, since few psychological studies have compared the magnitude of filtering and selection costs with the time taken to fix attention upon a whole stimuli. For now we simply note that, for SAIM, differences in the time to attend to particular stimuli can be comparable to costs in processing stimuli under attentionally demanding conditions.

### 2.1.3 Study 3 Accessing stored knowledge and attention-switching: Top-down biases

In Study 3 we showed that SAIM is sensitive to bottom-up biases in visual selection, tending to select the better of two objects, where 'better' is defined in terms of bottom-up factors such as the magnitude of the textured gmass. It is also possible to bias selection in the model in a top-down manner. The effects of top-down pre-activation were examined in two ways in Study 3. First, we set the initial value of the template units to favour the 2 over the + (that is, this value was set against the usual bottom-up bias in SAIM, which favours the + out of the + and the 2; Figure 8). For this, the template unit for the 2 was pre-set to 0.51. Subsequently, we introduced another form of top-down bias by changing the weights on the connections into the 2 template, so that these were larger than the weights into the template for the +. The first form of bias may be thought equivalent to a short-term priming effect, in which there is pre-activation for the stored representations of a particular item. The second form of bias may reflect the influence of longer-term learning, which can lead to more familiar objects having larger weights to their stored representations. For example, in a Hebbian learning process, the weights on the connections into and out of the 2 template would come to be larger than those into and out of the + template, if a 2 is presented more often to the model than a +.

Performance of the model is presented in Figure 10 (10 a shows performance with a pre-activated template; 10 b shows performance with an increased weight setting).

It is clear that, in contrast to when no top-down bias is set, the 2 now wins the competition for selection, and the template unit for the 2 is activated to a greater degree than that for the +. The same result holds for both form of top-down bias (pre-activating a template and increasing the weight connections to and from the template). This demonstrates that

---

<sup>8</sup>One result that does not appear consistent with this prediction was reported by Jonides & Yantis (1988). They presented a bright stimulus amongst dim ones, either of which could constitute a target letter. Both the bright and the dim target letters were equally affected by the number of dim stimuli in the field. If the bright stimulus was always selected first, it should show smaller effects of the number of dim items present. However it is not clear that the differences in intensity were sufficient to induce a competitive bias in this study.

bottom-up biases within the model can be modulated by top-down expectations. There is a cost to this, however, which is that it takes a relatively long time for convergence to be achieved within the network. Compare Figures 9c and 10. In 9c, data are presented from the full version of the model without a top-down bias favouring the poorer of the two stimuli (the 2). If we measure the times taken to map the item first selected (the +) into the FOA and to activate the associated template, we find increases of 33 and 60 iterations when the + & 2 appear together relative to when the + is presented alone (Figure 9(a)). This is the 2-item cost (see also Study 2). Now consider when the top-down activation is set against the bottom-up bias (Study 3). Performance in the two-item condition (Figures 10 (a) and (b)) needs to be compared with that in the one-item condition when only the 2 is presented, with the network biased towards the 2 (by pre-activation of its template or by growth of its weight connections). Data for the 2 alone, for this biased version of the model, are presented in Figures 10 (c) (with a pre-activated template) and 10 (d) (with an increase in the weightings). When the template was pre-activated, the 2-item costs were 97 and 335 iterations at the FOA and template levels respectively. When the weights were increased, the 2-item costs were 15 and 120 iterations respectively. In both cases, there remained a selection cost when 2 rather than 1 item was present.

Now to some degree, this prolongation of RTs in the model is a function of the strength of the top-down activation; if top-down activation were even stronger, RTs would be faster to select the object that is less perceptually good but more familiar. On the other hand, there is a balance to be struck in introducing top-down modulation within SAIM's WTA selection network, since small top-down increments can have large effects on performance. Stronger biases can lead to an insensitivity to bottom-up stimulation (even to 'hallucinatory' behavior, when no stimulus is present), which is not appropriate. The bias introduced here was sufficient to affect selection without dominating it (i.e., producing false misidentifications when a single stimulus was presented) For example, the + presented alone, or even with another stimulus that does not resemble the 2, was still recognized even when the template for the 2 was pre-activated. This is illustrated in Figure 11, where we show the performance of the model when it is biased towards a 2 (by pre-activating the 2's threshold) but no 2 stimulus is presented. In Figure 11(a), a + is presented along with a diagonal line, which should not strongly activate the template for the 2. The + is attended and identified. However, when the + is accompanied by a bracket stimulus, which has pixels in common with the 2, then SAIM can select the distractor mistakenly as a 2, even when another identifiable stimulus (having its own template) is present - the + (Figure 11b). Figure 11(c) demonstrates that this does not happen when the representation for the 2 is not primed (for the simulations shown in Figure 11(c), the templates for the 2 and + had the same value). Schvaneveldt & McDonald (1981) showed that humans too make false alarms to distractors when primed to identify a particular target and when presented with degraded distractors that overlap with the expected target (e.g., making a false alarm to the nonword MOHTER when primed with FATHER). We have explored the level of overlap in SAIM required to produce a false identification when the model is primed to expect a particular target, and, under the parameter conditions used for our simulations, an overlap of about 55% of the pixels could trigger a trial on which a false alarm response is made (when the stimulus itself has no corresponding template). Although this has not been explored systematically in psychological studies, we suggest that it is not too dissimilar to the results reported by Schvaneveldt & McDonald (1981); e.g., where 4/6 letters are shared in the same position

between the stimulus MOHTER and the primed representation MOTHER). There is clearly a balance between making the recognition system sensitive to momentary context and running the risk of false misidentifications occurring. The most general point we conclude is that, whilst top-down biases can overcome bottom-up stimulus preferences, there will remain some cost to selection of the primed item (relative to when it appears alone). Top-down biases that are even stronger can reduce such a cost, then the false-alarm rate will rise.

Interestingly surprisingly few studies have been conducted into the effects of effects of what might be construed as top- down biases on visual selection itself. One result that is consistent with SAIM was reported by Shiffrin & Schneider (1977), (Experiment 4d). Shiffrin & Schneider assessed the selection of a target learned through a varied mapping scheme against a distractor that had previously been learned as a target using a consistent mapping scheme. Consistent mapping should lead to representations having larger weights into and out of their stored representations, as captured by the simulations shown in Figure 10(b). Shiffrin & Schneider found that detection of the target learned via varied mapping was disrupted by the simultaneous presentation at an irrelevant location of a former target learned under consistent mapping conditions. This cost was worse than when the irrelevant location contained a former target learned through varied mapping. Selection was biased by the effects of prior learning (see also Irwin et al. (2000), for converging empirical evidence).



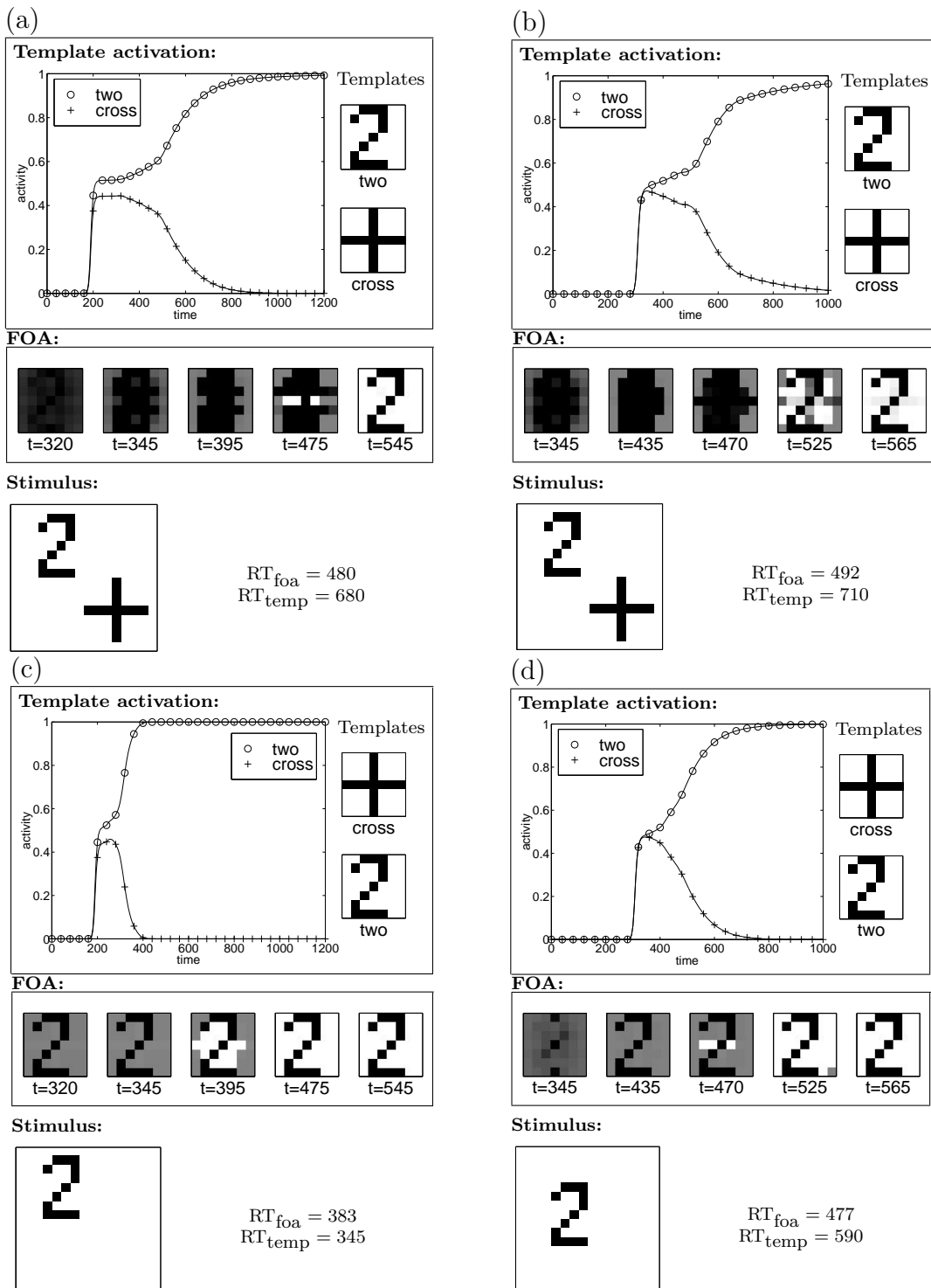


Figure 10: Effects of top-down bias on selection. In (a) and (b), a 2 and a + are presented simultaneously, with the model biased to favor the 2. In both instances, the 2 now wins the competition for selection over the +. In (c) and (d), a 2 was presented alone to the model. In (a) and (c) there was pre-activation of the template for the 2. In (b) and (d) the weights were increased on connections into and out of the 2 template.

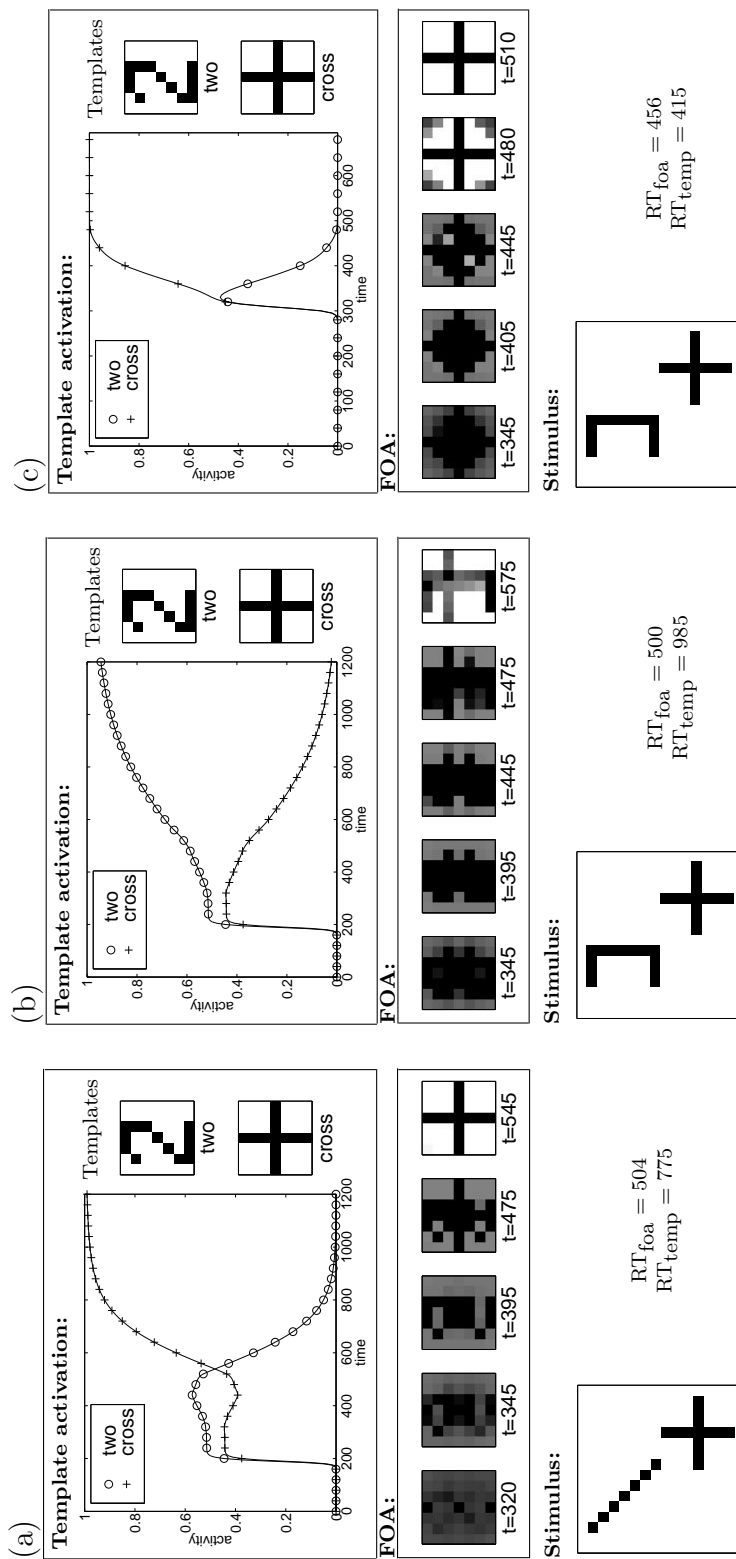


Figure 11: Performance of the top-down version of the model when biased towards a stimulus not presented (the 2). (a) gives performance with a + and another stimulus dissimilar to a 2 (a diagonal line); (b) gives performance with a + and a stimulus similar to a 2 (l) - note that the template for the 2 is activated; (c) shows performance with the same stimuli as in (b) but without a top-down bias for a 2, when the cross wins the competition for selection.

## 2.2 Study 4. Wholistic perception

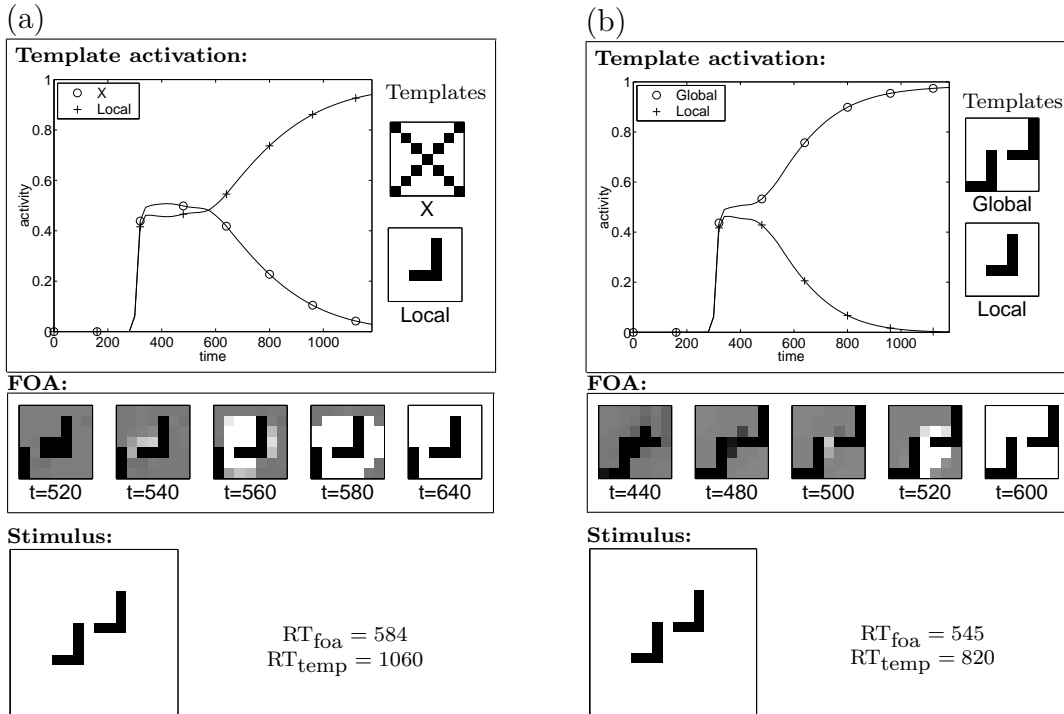


Figure 12: Global precedence in SAIM. (a) illustrates performance when a template does not exist for the whole, when a part is identified; (b) shows that a 'whole' compound stimulus is detected in preference to a component parts, when templates exist for both the whole and the part. RTs are faster for the response to the whole than to the part.

As we have already noted, there is evidence in human perceptual report for participants being able to attend to and identify more easily whole shapes than the constituent features. This was shown by Navon (1977) in some now classic studies using compound shapes (large letters made up of small letters). Navon found that RTs were overall faster to global relative to local letters, and that the identities of the global letters interfered with responses to local letters when the letters had incompatible responses. Many other studies have demonstrated similar effects, though the precise pattern of results depends on factors such as the density of the local letters (Martin 1979), the spatial uncertainty about the target (Grice et al. 1983), and the form of grouping between the elements comprising the compound letters (Han et al. 1999, Kimchi 1994). On the other hand, the time course of mapping visual information into SAIM's FOA suggests some advantage for local over global properties of stimuli (with the caveat discussed in Study 1). In Study 4 we tested whether there really was a local advantage when an identification task was used and the stimuli comprised either: visual elements that could each be identified as a distinct target at a local level (with no template for the more global whole), or the elements could activate a template for the perceptual whole as well as for each part (with a template for the global whole and for each part).

### Method


The full network was used, without any top-down priming of particular templates. Two

tasks were employed: select a global target and select a local target. To examine selection of the global target, two templates were employed: one for a whole, compound stimulus



made up of unconnected local elements (the  ) and one for a part of the same stimulus



( Local – see Figure 12). Each template was coded so that its center of gravity would fall at the center of the FOA. In this case, identification of the perceptual whole based on activation of the template for the whole stimulus was pitched against identification of a part of the same object based on activation of the template for the part. To foreshadow the results, we found that the whole was selected prior to the part. To compare identification time for the part, when SAIM was only 'set' for this element, we re-ran the simulation but introduced a 'dummy' global template (a global X). The global X template has minimal pixels in common with the stimulus presented, but it equates the networks used to simulate the local and global tasks in terms of the number of templates present. RTs in SAIM are influenced by the number of templates held in memory (see footnote 7), so all comparative simulations need to be based on the networks having the same numbers of templates.

## Results and Discussion

Figure 12a plots the activation in the FOA and the template units as convergence is achieved for the compound stimulus when the 'part' and X templates were used. Figure 12b gives the activation to the same stimulus when the X template was replaced with one for the whole, compound stimulus.

Figure 12b shows that, when a template exists for the global shape, it is this shape that wins the competition for selection and identification (even though a template for the local part also exists). There is global precedence. Under these circumstances identification of the part can only take place by inhibition of the template for the global shape following its identification. Figure 12a demonstrates that SAIM is capable of identifying a part, when the template for the part does not compete with one for the whole. However, identification times remain slow, when compared to those for an identifiable global shape (part (b)).

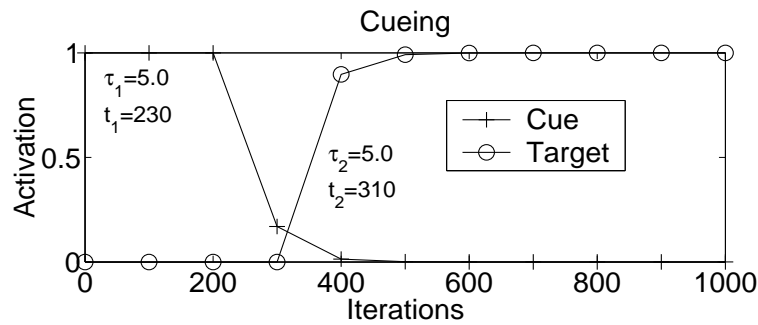
It is also of interest to note that when the part is identified (Figure 12a), other sections of the cross are mapped into the FOA. This is because of proximity effects between the pixels in the compound shape, which act in a bottom-up manner against the top-down influence to attend to just the part. Again there is an interaction between bottom-up and top-down factors in determining performance. In human observers similar interactions can be seen. For instance, we have only limited ability to attend to parts of stimuli when they form strong bottom-up groups with surrounding elements (e.g. Rensink & Enns 1995). Also the magnitude of global precedence effects in letter identification tasks increases when local letters group based on proximity (Han et al. 1999).

## 2.3 Study 5: Cueing

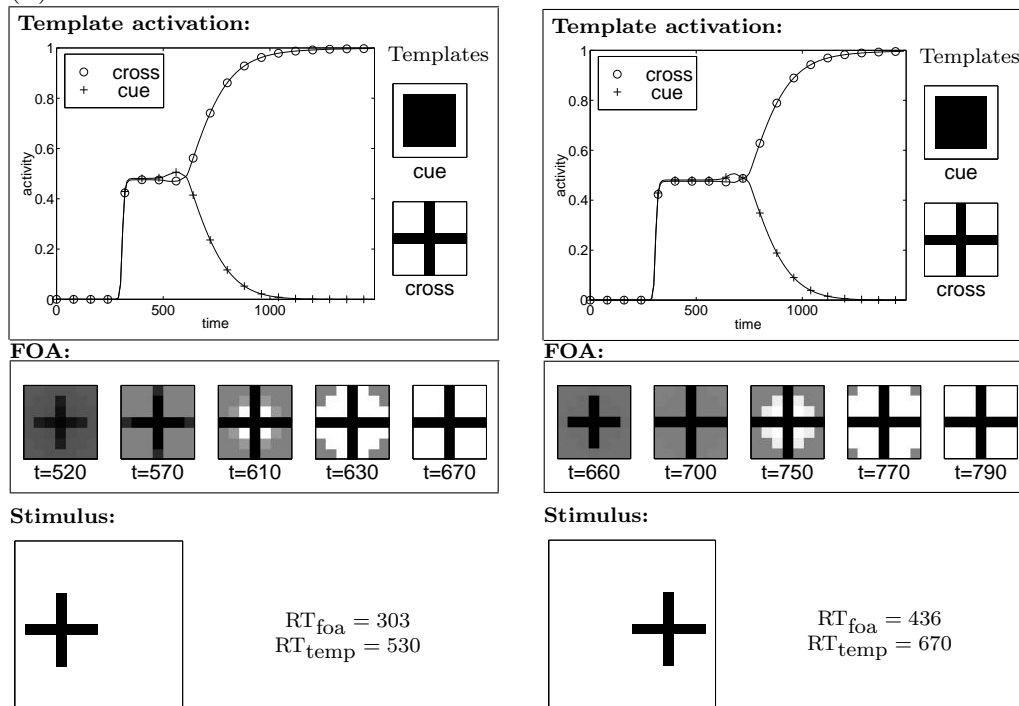
### 2.3.1 (i) Spatial cueing

As discussed in the Introduction, one of the classic pieces of evidence that has been taken to indicate that visual selection in humans is spatially mediated is the spatial cueing effect:

(a)



(b)



(c)

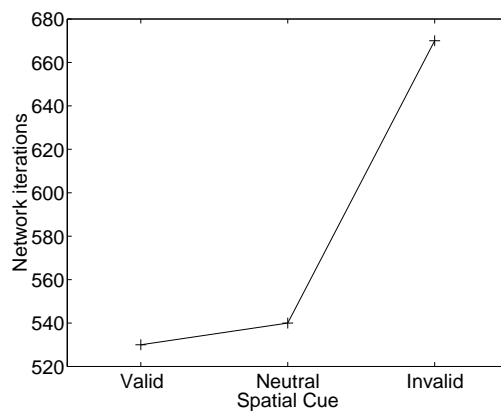


Figure 13: Spatial cueing in SAIM. (a) Illustrates the parameters governing the decay of activation from the cue and the rise in activation for the target. (b) Depicts activation within the FOA at different intervals following the onset of a valid (top) or invalid spatial cue (bottom). (c) Gives the time for convergence to be reached in the FOA as a function of whether a valid, neutral or invalid cue was presented.

RTs to detect a target are faster when it falls at the same location as a preceding spatial cue (the valid cueing condition) relative to when the cue falls at a different spatial location (the invalid cueing condition Posner (1980)). This result has typically been interpreted in terms of a spotlight analogy of attention; processing is speeded because the cue directs the spotlight of attention to the location of a target in an internal spatial representation. Other evidence suggests that different modes of cueing attention can be distinguished. Peripheral visual cues seem to affect performance in a relatively automatic way, over a short time interval ('exogenous' cueing). RTs to targets falling at the same locations as such cues are faster than RTs to targets falling at uncued locations, even when the cue is not predictive of the target's location across trials (cf. Müller & Rabbit 1989). Cues can also exert influences over longer time intervals if they are predictive of targets, and even if they only provide symbolic information about the target's location (e.g., if the cue is a central arrow pointing to the location of a target but not appearing at that actual location; 'endogenous' cueing). The effects of exogenous cues are difficult to prevent and (for example) can be unaffected by adding secondary task loads to performance (Jonides 1981). Endogenous cues are not effective if unpredictable of target locations across trials, and are affected by adding secondary loads to tasks. In these simulations we attempt to model exogenous cueing effects on target detection.

## Method

The cue was a square of 5x5 pixels and the target a cross with arm sizes of 7 pixels horizontally and 7 pixels vertically. Cueing was modelled by presenting the cue on one side of the visual field for a short duration (1 iteration). This was sufficient to allow the cue to generate activation within the FOA. Subsequently, the target cross was presented, centred either at the cued location or at a location on the opposite side of the field, beginning at iteration 400, and this remained in the field until activation converged in the FOA. There was also a neutral cue condition, in which 2 cues were presented simultaneously at each of the two possible target locations (Müller & Rabbit 1989, e.g.). Performance was tested using both the bottom-up version of SAIM (without the knowledge network) and the full version (with the knowledge network and with a template for the target stimulus). The parameters for the bottom-up components of the model were kept constant across the two test situations. In all previous simulations we examined SAIM's responses to individual stimuli presented in a single time step. Here, however, SAIM received two stimuli consecutively. To model performance under these new conditions, we assumed that representations in the FOA, once activated, would decay over time. This may correspond to something like the decay of an iconic representation of the cue (cf. Coltheart 1983)<sup>9</sup>. The decay parameter for the cue, and the activation parameter for the target, are illustrated in Figure 13a.

## Results and Discussion

In Figure 13b we illustrate the growth in activation in the FOA and the template units in the full version of the model, and Figure 13c gives the times (in network iterations) for

---

<sup>9</sup>The energy functions used for the original version of SAIM were not designed to work with dynamically changing stimuli, but assumed instead that the input was unchanging. Introducing a decay parameter after a stimulus has appeared provides a way to approximate a dynamic change within the model, ensuring that there is stable evolution of the energy function. The decay parameter has no effect on the performance of SAIM in any of the other simulations we present, where static stimuli are presented.

SAIM to identify the target in the valid, neutral and invalid trials (here performance is shown following a left side cue only, with a left and right side target). SAIM manifests a cueing effect, with RTs on valid trials being faster than those on invalid trials, and with RTs on neutral trials falling in-between. These effects are found across a range of values for the decay parameter, provided the target is presented prior to activation from the cue having delayed back to baseline. Essentially the same pattern of results occurs in a pure-bottom up version of the model, consistent with the effects being due to the exogenous properties of the cue.

SAIM manifests an effect of spatial cueing because excitatory connections within the selection network maintain their activity even after disappearance of the cue, due to the self-sustaining properties of activation in competitive-co-operative networks (Amari 1982). In the valid condition there is sustained activity at the location of the target; this boosts the target's activation and produces faster mapping of the target into the FOA. In the invalid condition, activation from the target, at one location, must compete with the sustained activity from the cue, slowing the time taken to translate the target into the FOA. This effect will arise whenever targets follow shortly after cues presented at a given location, and hence it should be an automatic consequence of cueing. Note, however, that the cueing effect is not simply due to sensory summation of activation from the target and cue in a common location. There was a benefit for valid trials relative to neutral trials, even though the cue appeared at the location of the target in both instances. In the neutral condition, though, any sensory effects of the cue were modulated to some degree by competition between the two cues presented.

The cueing effect examined in SAIM mimics a situation in which the cue is presented briefly and 'automatically' affects target processing so-called 'exogenous' cueing. As SAIM would predict, the effect of this exogenous cue can decrease if too long a delay is introduced between the cue and the onset of the target (Cheal et al. 1994, Nakayama & Mackleben 1989). In addition to this, humans can employ a form of 'endogenous' cueing, based on intentional allocation to the location of a target (Cheal et al. 1994). This is not implemented in this version of the model.

## 2.3.2 (ii) Cueing objects and cueing space

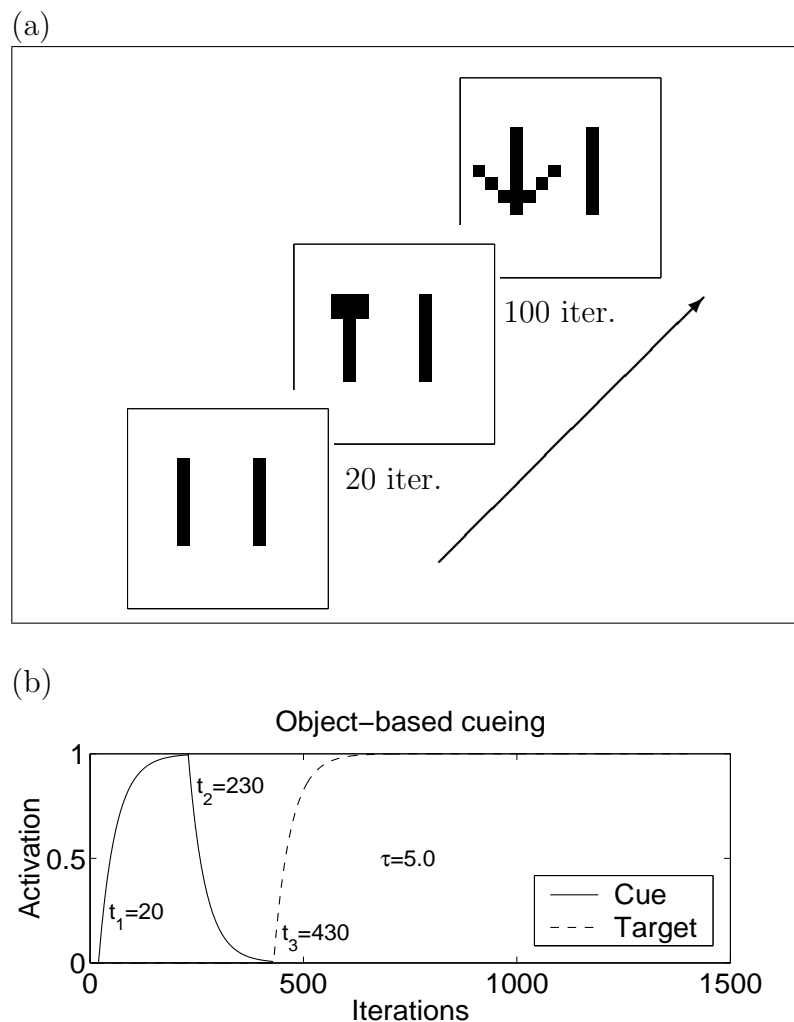


Figure 14: Simulation of space-based and object-based effects of spatial cueing in SAIM. (a) The stimuli and presentation times (the cue was a small square and the target an arrow). (b) The parameters governing the decay and rise of activation for the stimuli in the simulation.

Egley et al. (1994) first reported effects on target detection of not only cueing the spatial position of where a target would appear but also which object it would appear in. Subjects were presented with two rectangular boxes at the start of a trial. A cue was presented by highlighting one part of a box, and this was followed by a target spot which appeared at either the same location (same object, same position), at a different location but within the same object (same object, different position), or at a different location and also in the other object (different object, different position - here the distance between the target and cue was matched to that between the target and cue in the same object, different position condition). Subjects made a simple RT response to the onset of the target. Egley et al. found that RTs were fastest in the same object, same position condition, followed by the same object, different position condition and then the different object, different position condition. These results are difficult to accommodate in a pure space-based model of



(a)

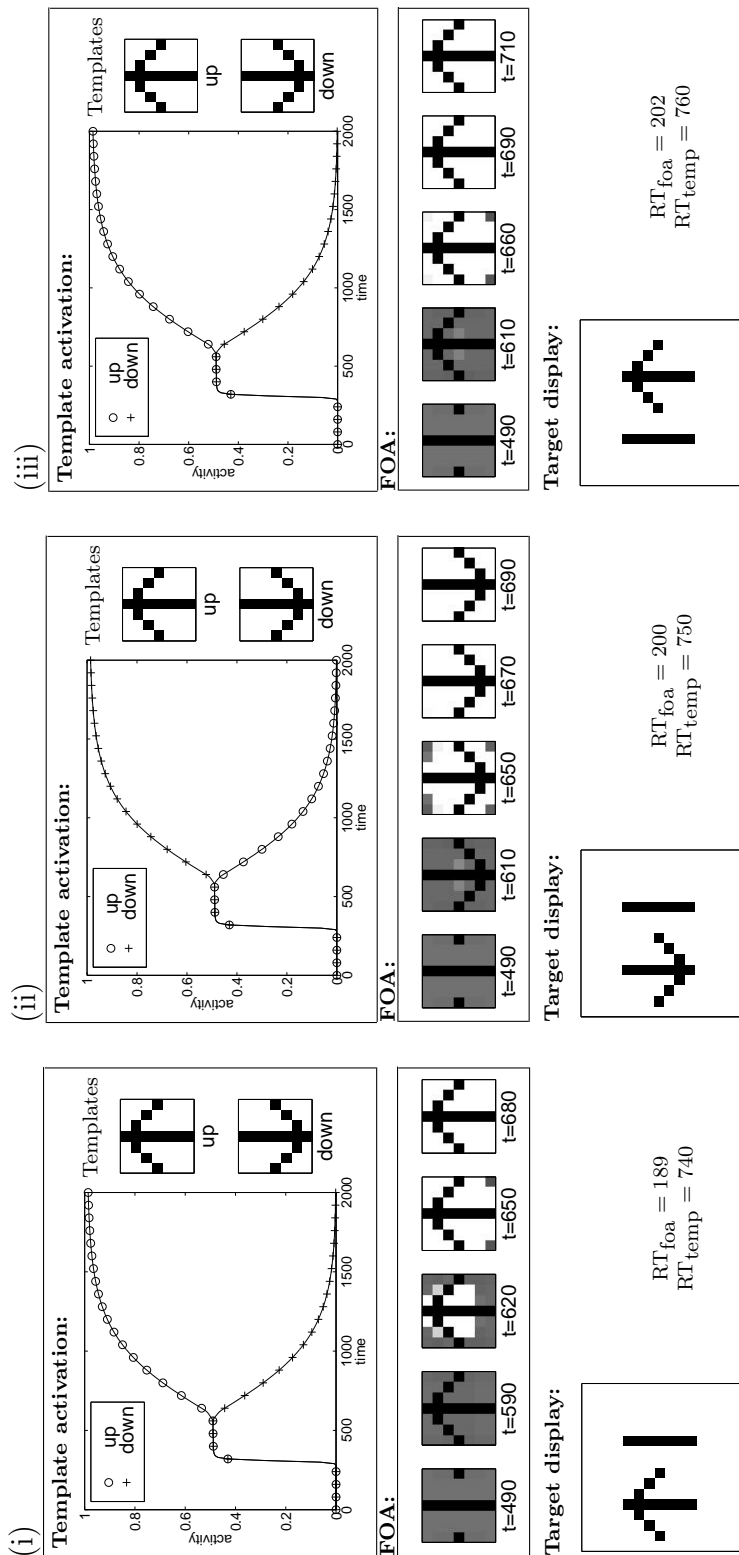


Figure 15: (a) Activation within the FOA and template units when the cue and target (i) fell within the same object at the same location, (ii) fell within the same object but at different locations, or (iii) fell in different objects and at different locations.

(b)

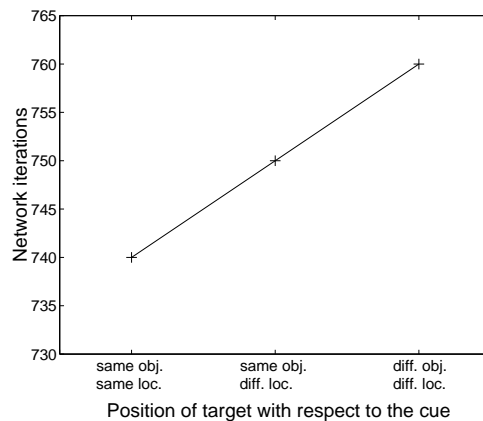


Figure 15: (b) The times taken for the target template to reach threshold as a function of whether the cue and target fell at the same or different locations, within the same or different objects.

visual attention, since the 'spotlight' of attention should be the same distance from the target in the two different position conditions, irrespective of whether the cue and target are in same and different objects. However, the effects of the spatial relations between cues and targets within the cued object also suggest that spatial proximity within objects is important. The cued part of the prime object seemed to be attended more than the uncued part. A model such as SAIM, which combines spatial and object-based effects in selection, may be able to capture such effects better than either pure space- or pure object-based models.

To simulate the results of Egly et al., we presented a 'prime' display composed of two vertical line elements (similar to the rectangles used by Egly et al). One of the two lines was then cued, by presenting a small square which fell at the far tip of the critical line. Subsequently, a vertex shape was exposed, which created an arrow target with one of the two primer lines. Crucially, the vertex could fall at the cued location (same object, same location), it could fall at the opposite end of the cued line (same object, different location), or it could fall on the other line (different object, different location). The separation between the cue and the target vertex was matched in the same object, different location and the different object, different location conditions (5 pixels, in each case; Figure 14a). We measured the time taken for SAIM to detect the vertex in the FOA (equivalent to Egly et al.'s measure of simple RT to the target dot). In the version presented here, we also included a template which consisted of an error pointing up or down, corresponding to the line plus a vertex at either its top or bottom. This is akin to human subjects coming to store a template for the rectangle plus the target dot in each of the probed location, in Egly et al.. However, essentially similar results occur in the bottom-up version of SAIM without the template. Using the template here simply ensures that the network architecture is held constant across different simulations. We assessed whether the time for the target to be mapped into the FOA was affected by whether it fell within the cued shape or not.

## Method

Initially two lines, each of 1 pixel across and 7 pixels down, were presented. Next a 3 x 3

pixel square appeared at the end of one of the primer lines. After this, the target vertex appeared. The activation and decay parameters for the model were the same as those used for the study of space-based cueing (above). The cue was presented for 210 iterations and it was followed immediately by the target. The timing and activity parameters are shown in Figure 14b<sup>10</sup>. The full version of SAIM was used, with templates for an upright and an upside down arrow.

## Results and Discussion

Figure 15a illustrates the dynamics of activation within the FOA, and the accumulation of activation in the template units, when the cue and target appeared: (i) in the same object at the same location, (ii) in the same object at a different location, and (iii) in a different object, at a different location. Figure 15b gives the times for the templates to each threshold, yielding the different identification responses.

Detection (and identification) times were fastest when the target appeared both in the same object and at the same location; they were slower when the target fell in the same object but at a different location, and they were slower again when the target fell the same distance away but in an uncued object (different object, different location). Performance of the model is affected by the distance separating a target from the original center of attention and by whether the target falls within the originally attended object; there are both space and object-based characteristics of attention. The object-based effect on selection here occurs because activation is propagated from the cued part of the object to other parts, by dint of their spatial proximity. This conveys an advantage for a target subsequently presented on the cued object relative to one presented on the uncued object (which does not benefit from spreading activation within the selection network).

### 2.3.3 (iii) Position of cueing within an object

Another study reporting effects of object properties on the allocation of visual attention was published by Pavlovskaya et al. (1997). They examined the accuracy of detecting briefly presented, masked letter-like targets, which were preceded by visual cues. The cues fell at the center of mass of the target or at a position within the shape but displaced from the center of mass. They found that cueing at the center of mass of the shapes led to better performance than cueing away from the center of mass. This result is consistent with attention normally being drawn to the location of the center mass in order to discriminate the shape.

To test for similar effects in SAIM, we presented a spatial cue to the model and followed it by a target shape. The cue fell at the center of mass of the target or it fell within the target but away from the center of mass. We ask whether performance is affected by the position of the cue relative to the center of mass.

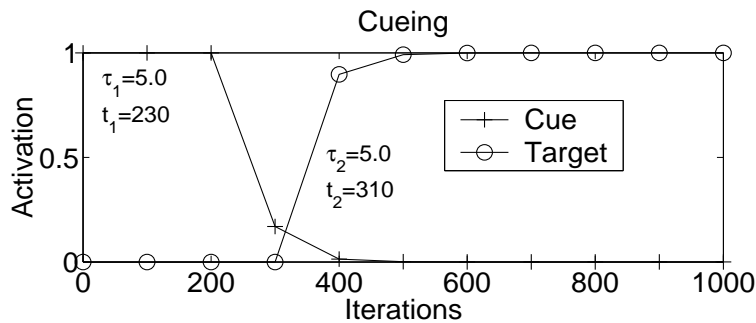
## Method

The cue was a 3 x 3 square. The target was a 7 x 7 pixel cross. Performance was measured in terms of the time taken to identify the target. The cue fell so that its center of mass fell between +/- 3 pixels away from the center of mass of the target shape along the

---

<sup>10</sup>Note that, in Figure 14b, the activation function for the cue starts at 0, whilst it started at 1 in Figure 13a. This difference arose because, for the data in Figure 15b, the cue was presented against a background already containing two primer stimuli, producing a longer rise-time in its activation.

(a)



(b)

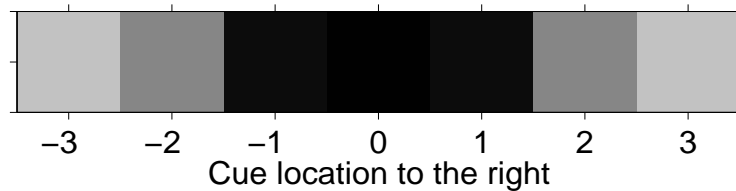


Figure 16: (a) The parameters governing decay and rises in activation for the cue and target in the simulations of Pavlovskaya et al. (1997). (b) RTs in SAIM as a function of the spatial relations between the cue and the center of textured mass in the target. The darker the shading, the faster the RT. At position 0 the location of the cue and the center of textured mass of the shape coincide.

horizontal meridian (to simulate the data of Pavlovskaya et al. (1997)). The parameters were the same as for the studies of spatial cueing using the full version of the model.

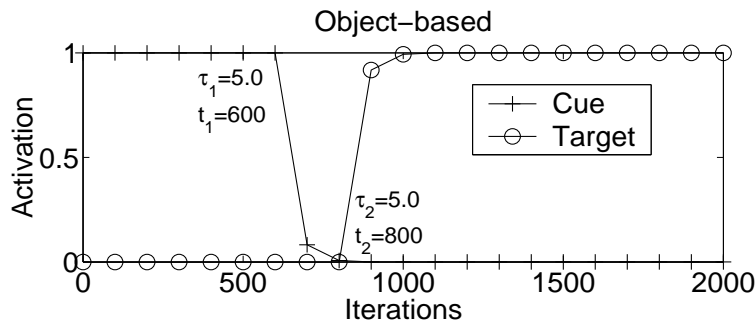
## Results and Discussion

In Figure 16 we illustrate the RT for the target to be identified, according to the spatial relations between the location of the textured mass of the cue and the location of the textured mass of the target shape along the horizontal axis. At point 0 the centers of the textured mass of each shape are perfectly aligned; at point 3 the center of the textured mass for the cue fell 3 pixels to the right of the center of the textured mass of the target. The darker the shading, the faster the RT. The figure shows that RTs varied as a function of the spacing between the centers of textured mass of the shapes, with fast RTs occurring when the locations of the textured masses were aligned. This mimics the results reported by Pavlovskaya et al., that human RTs vary according to the spatial distance between a cue and the center of mass of a shape.

### 2.3.4 (iv) Inhibition of return (IOR)

When there are multiple items in the field, SAIM needs to switch attention so that each item is selected and identified in turn. It does this by inhibiting locations in the location map and in the selection network associated with the previously attended region of field (to prevent attention returning to areas of field already sampled). In addition there is inhibition of the template for the target (an effect that is not retinally-specific). When attention has shifted from a cued object in one location, there should thus be a cost to

(a)



(b)

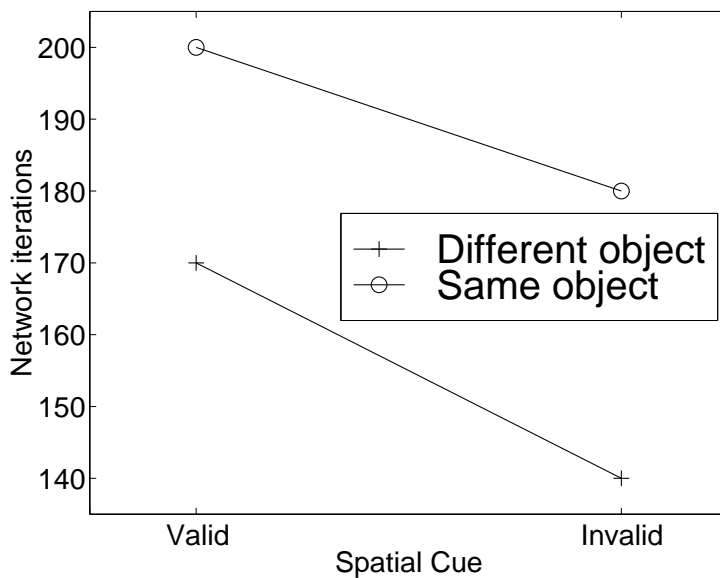


Figure 17: Inhibition of return in SAIM (with a long cue- target separation of 800 iterations). (a) Parameters governing the decay and rise in activation for the cue and target. (b) RTs for convergence to be reached in the FOA according to whether valid or invalid spatial cues have appeared, in either the same or in different objects.

identification when a target is subsequently presented at the same location or when the same object is presented but at a different location. There should be both space-based and object-based 'inhibition of return' (cf. Gibson & Egeth (1994); Maylor (1985); Posner & Cohen (1984); Tipper et al. (1991)).

We examined these properties of SAIM in a simulation using the full version of the model in which the cue and target used in our earlier study of spatial cueing were presented at a longer SOA, with the cue being exposed until it was identified (i.e., until thresholds were reached in both the FOA and the template units). The full model allowed attention-shifting to take place. Following identification of the cue, there was subsequent inhibition to enable attention to be shifted to other items in the field.

## Method

The target was a +, 7 pixels across by 7 pixels high. On trials with the 'same object' cue, the + also appeared as the cue. On 'different object' cue trials, the cue was a 5 x 5 pixel square (the target was the same, on 'same' and on 'different object' trials). The cue was focused either at the center of the target (when the target was subsequently presented) or at an equivalent position on the opposite side of the visual field (on valid and invalid trials respectively). The cue appeared for 700 iterations prior to the target being presented. This was longer than the cue time used produce the 'early' facilitation effect (400 iterations, see (i) above), and it was sufficient for the cue to be attended and identified. There were 4 conditions. Cue and target were presented in the same or in opposite positions, and they could be the same or different objects. The decay parameters for the templates are illustrated in Figure 17a.

## Results and Discussion

The times for template thresholds to be passed are given in Figure 17b. RTs were prolonged on valid (same position) relative to invalid (different position) trials, and when the cue and target were the same relative to when they were different objects (though targets were the same in both instances). These effects of 'spatial' and object-based' IOR were roughly additive. This matches the human data. For example, Tipper et al. (1991) used a cueing procedure in which one of four boxes was illuminated; the boxes then moved to new positions. RTs were slowed when targets were presented in the previously cued box even after it had moved into a new position (object-based IOR; see also Gibson & Egeth (1994); Tipper et al. (1999)). RTs were also slowed when targets appeared in a new box that had moved into the previously cued position (location-based IOR). RTs were slowed even more when the same object was cued in the same location, with the effects combining in an additive fashion.

Now, as currently implemented, SAIM has no way of sustaining activation within the selection network once a stimulus disappears from the visual field. This means that IOR can only be produced by using a long cue duration (to ensure that the cue is identified). In contrast to this, most studies of IOR in humans have used short cue durations and varied the interval between the cue and the target (Maylor 1985, Posner & Cohen 1984); effects with long cue durations have proved controversial due to possible masking effects from cues on targets (e.g., Tassinari et al. (1994, 1998); see Lupianez & Weaver (1998)). However in studies with short cue durations, the exposures are still sufficient to enable cues to be identified, perhaps because such stimuli are typically unmasked and so may benefit from iconic persistence. Provided identification takes place, and some (iconic) representation of the location of the cue remains, then SAIM generates IOR.

### 2.3.5 Discussion of Study 5

SAIM is sensitive to the effects of spatial cueing on its performance. With short cue-target intervals RTs are faster when cues and targets appear in the same spatial positions than when they appear in different spatial positions. However, when cues are presented long enough to be attended and identified, there is both spatial and object-based IOR in order to allow the model to select new items. In each respect, SAIM captures psychological data on visual cueing effects. In addition to these aspects of spatial selection, SAIM manifests effects of object properties on selection. When part of an object is cued, there

is facilitation for targets presented within the same object relative to targets presented outside the object but the same distance from the attended part. Also cueing effects on attention to a target shape are increased if the cues fall at the center of textured mass of the shape. These results match data reported by Egly et al. (1994) and Pavlovskaya et al. (1997). In addition there is object as well as spatial IOR, due to inhibition of both previously attended locations and objects. The ability of SAIM to simulate object effects on selection extends other models that have been used to simulate cueing but without incorporating an object processing module (e.g. Cohen et al. 1994).

## 2.4 Study 6

In the final study in this section, we examined whether SAIM would be able to select one of two spatially overlapping stimuli. Some of the strongest evidence for object-based selection in vision comes from studies using spatially overlapping stimuli. Using such stimuli it has been demonstrated that subjects are able to select several attributes from one object with little or no cost, while there is a cost for selecting attributes from both of the objects present (e.g. Behrmann et al. 1998, Duncan 1984). Since this cost in selection occurs even when stimuli fall in the same spatial region, it is difficult to explain in a purely spatial model (e.g., if all attributes are selected together if they fall within the same spatial window). Here we ask whether SAIM too can be biased to select one but not both of two overlapping shapes.

### Method

SAIM was given two templates, one for a horizontal 'handle' and one for a vertical line. We then presented these two stimuli together on SAIM's retina, so that they overlapped within a 7 x 7 region. Note that SAIM normally selects all the pixels falling within such an area (see Study 1). To provide a 'set' for the model to select one or other of the overlapping stimuli, we pre-activated one of the two templates (using the same parameters as those used in Study 3). We assessed whether SAIM then selected all the pixels associated with one pattern but not the other (even though the other pattern fell within the attended spatial region).

### Results and Discussion

Figure 18(a) presents the results when SAIM was biased towards the 'handle' target, and Figure 18(b) the results when the bias was towards the vertical line target. In each case, SAIM selected the object whose template was pre-activated and only pixels conforming to that objects remained active within the FOA; pixels corresponding to the other object present within the attended spatial region were not activated. Thus, even though SAIM operates using a window of spatial attention (the FOA), it is capable of 'pure' object-based selection. This occurs in the model because of object-based (top-down) feedback to the selection network, which favors pixels belonging to the pre-set target over pixels belonging to other stimuli in the same spatial region. We have obtained similar effects with a variety of stimuli.

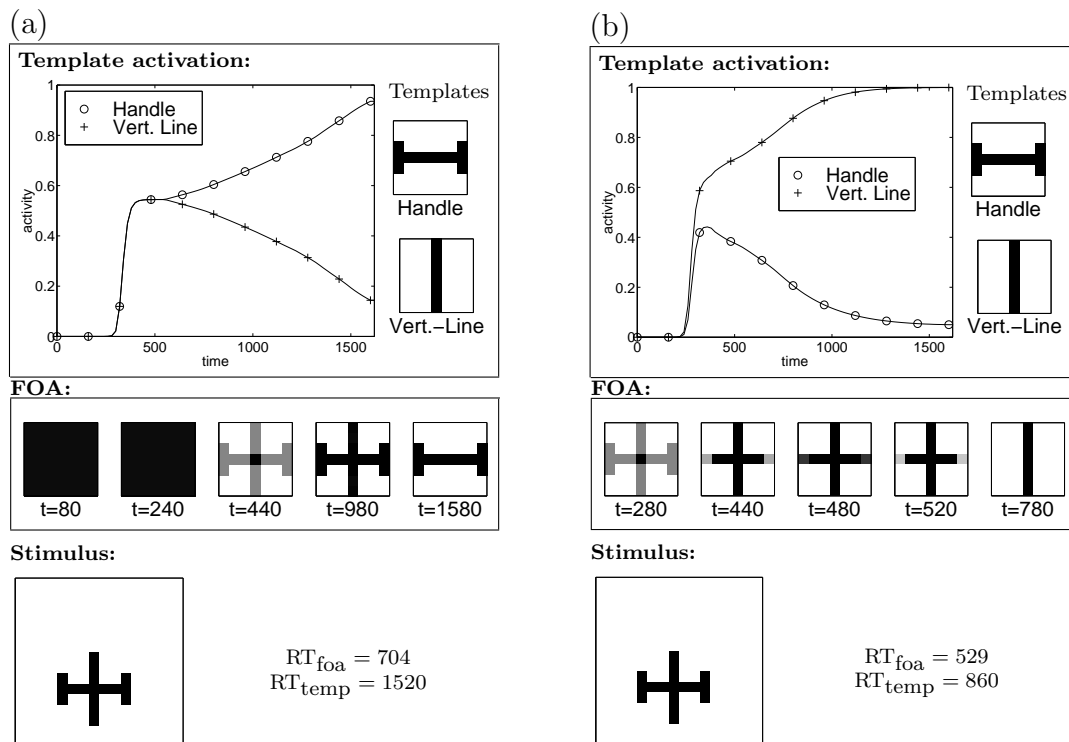


Figure 18: SAIM's ability to select one of a pair of spatially overlapping stimuli. (a) SAIM is pre-set to select the handle stimuli; (b) SAIM is pre-set to select the vertical line target.

## 2.5 Conclusion

We have simulated some of the 'classic' results supporting arguments for spatial selection in human visual attention, such as the effects of spatial cueing on target detection. SAIM captures such effects by incorporating a spatially limited FOA which determines which retinal input is mapped through to recognition units that respond to stimuli from all areas of visual field. The assignment of the FOA to a particular area of field can be biased by spatial cueing. However, SAIM also shows effects of object-coding as a natural part of the dynamics of the mapping process (e.g., from the effects of proximity within the selection network) and due to its interactive nature (in which activation within the knowledge network is fed back to the selection network). Hence the model is capable of selecting pixels belonging to a single object and rejecting pixels belonging to another object within the same spatial region. The general properties of the model, such as its tendency to align attention to the center of shapes and the consequences of biasing competition for selection (from either bottom-up or top-down factors) also make it attractive as a model of selection, linking to a broad body of psychological data (cf. Duncan 1998).



### 3 Modelling pathological attention and selection

As we pointed out in the Introduction to this paper, neuropsychological studies provide powerful constraints on models of human visual selection. In particular, studies of patients with unilateral visual neglect suggest that both spatial and object-based factors influence selection, and that forms of dissociation can occur in which patients seem to express separate forms of neglect according to the spatial positions of separate objects in the field or according to the positions of parts within objects. A full account of human visual selection needs to accommodate these examples of pathology as well as patterns of normal performance, and, as we will review, some forms of neglect remain difficult for competitor accounts of visual selection. Accordingly, we attempted to simulate effects of brain lesions within SAIM.

#### 3.1 Neurobiological status of SAIM

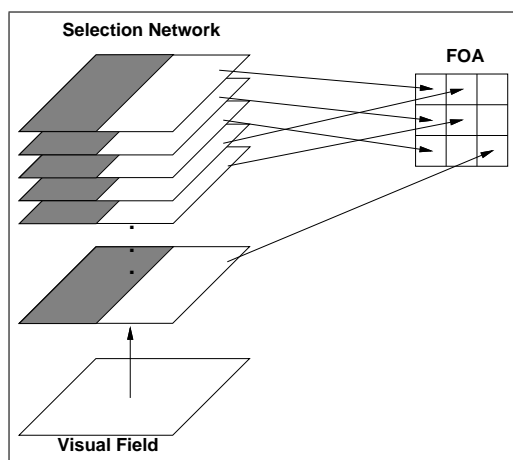
When introducing their dynamic switching circuit model, Olshausen et al. (1993) motivated their arguments by relating the structure of the model to the neural circuits mediating human selective visual attention. In particular, they argued that the contents network (using our term) relates to the architecture of the pattern or 'what system in the ventral cortex. This network is concerned with taking the contents of the retina and passing them forward, through the FOA, to high level cells concerned with the recognition of stored objects (e.g., associated with the inferotemporal cortex in primates; Tanaka (1993). The selection network, they suggested, is distributed between the pulvinar and the parietal cortex, and forms part of a 'where system, concerned with coding the spatial locations of stimuli and with using this information for action. They argued that the pulvinar was a likely candidate for such a network because of the high levels of connectivity between this structure and various regions of the cortex, as would be required in a true neural implementation of the selection network. It can also be argued that the posterior parietal cortex provides a spatial map of objects that are of potential interest for attention (Gottlieb et al. 1998) and, furthermore, it is strongly implicated in spatial transforms that map retinal information into other frames of reference for both object recognition and action (e.g. Duhamel et al. 1992, Zipser & Andersen 1988). In SAIM the location map denotes the positions of objects of attentional interest, and we implement a mechanism for spatial transformations between different reference frames in the selection network. Thus aspects of the selection network and the location map in SAIM can be linked to posterior parietal cortex.

Given these general relations between parts of the model and anatomical structures in the brain, we may look to mimic effects of particular brain lesions by simulating lesions within related parts of the model. Disorders such as unilateral neglect and simultanagnosia are associated with lesions to the posterior parietal cortex in humans. To simulate these disorders, we contrasted the effects of 'damage to the different portions of the selection network.

#### 3.2 Lesioning SAIM

There have now been numerous studies on the effects of simulated lesions on the behavior of connectionist models (see Ellis & Humphreys 1999, McLeod et al. 1998, Olson &

Vertical lesion



Horizontal lesion

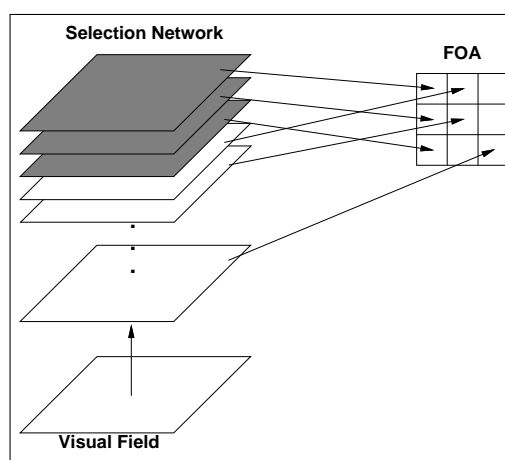


Figure 19: Illustration of 'vertical and 'horizontal lesioning in SAIM. The vertical lesion affects units in the selection network receiving input from one side of the visual field. The horizontal lesion affects units mapping into one side of the FOA.

Humphreys 1997, for reviews). Lesions have been simulated by reducing the weights on connections, eliminating connections, adding noise into the activation functions or some combination of these manipulations, and it is by no means clear exactly which procedure may be best for mimicking the effects of neural damage in humans. In the main, lesions were introduced to SAIM by damaging 'intrinsic connections to units within the selection network, which may be analogous to damaging connections from the pulvinar to the posterior parietal cortex or to damaging connections between neurons within the posterior parietal cortex; damage to these regions is associated with forms of visual neglect (Gaffan & Hornak 1997, Heilman & Valenstein 1979). There were two different main patterns of lesioning, which respectively affected (i) connections between units in the selection network that received input from one part of the visual field (we term this a 'vertical lesion), and (ii) connections between units that transmitted their output into one side of the FOA (we term this a 'horizontal lesion). Figure 20 illustrates the two patterns (here, for simplicity, lesioning is shown as a step function; however, in the simulations it was graded linearly across space; see below). The two patterns of lesion were examined because, within the architecture of the model, they are likely to generate different forms of spatial disorder. A 'pure vertical lesion, solely affecting units in the selection network receiving input from one side of the field, may produce a deficit linked to the position in the visual field where a stimulus appears. In a severe form this could mimic a field cut (though there is no effect on early coding of visual information and, indeed, activation in the contents network would be the same as in the unlesioned case; however the lesion in the selection network will prevent this activation in the contents network being translated into the FOA). With a milder lesion, it is possible that a stimulus presented in the affected part of the field can be detected under optimal circumstances but not when other items are present, competing for selection. This may lead to a form of extinction, separating the deficit from a simple field cut. We demonstrate this in Study 7. In contrast, a 'pure horizontal lesion may produce deficits solely affecting representations on one side of the FOA, and may be unaffected by the visual field where a stimulus falls; there may be a

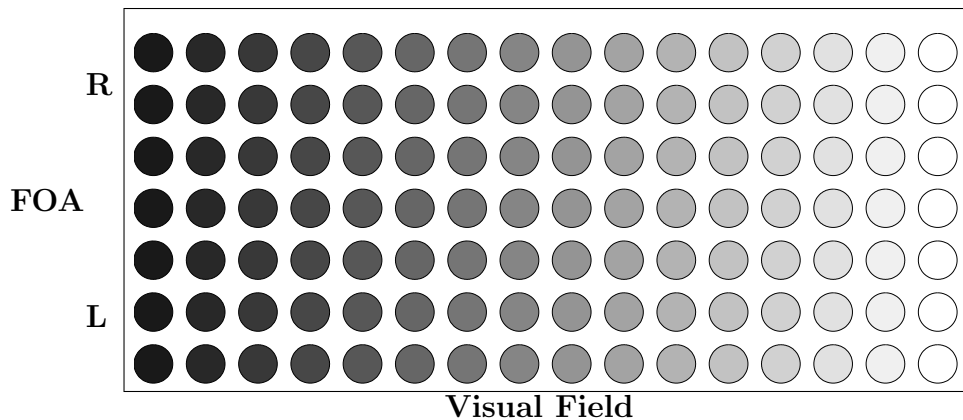
form of 'object-centred neglect. Combinations of 'vertical and 'horizontal lesions may produce variations in the relative contributions of visual field and object-based coding in neglect. As we shall demonstrate, this has interesting implications for understanding contrasting forms of visual neglect.

As with other attempts to model unilateral neglect (e.g. Mozer et al. 1997, Driver & Pouget 2000), the lesions that were given to SAIM were graded across space (not all-or-none, as in Figure 19). Such graded lesions can be conceptualized in various ways. For instance, there may be some proportion of cells within each hemisphere that have receptive fields for both rather than one side of space, with ipsilateral locations close to fixation being more likely to be represented bilaterally than ipsilateral locations far from fixation. Indeed Corbetta et al. (1995), in a PET study of visual attention shifts, showed that cells in the right parietal lobe were activated by shifts in the right as well as in the left visual field, though activation of the left parietal lobe by ipsilateral (left) field stimuli was somewhat weaker. Consequently damage to the right parietal lobe may affect not only the left visual field but also areas in the right visual field to some degree. Also left field areas around fixation may be supported by cells within the left hemisphere, as well as by cells in the right hemisphere. The net result of right hemisphere damage may be a graded impairment across space, with any deficit being worse for stimuli presented on the far left. Neuropsychological data also suggest that there is a graded deficit. For instance, in clinical tests of neglect such as star cancellation, performance can often decrease in rather a uniform manner with proportionally more items being cancelled as a function of each relative shift from the ipsi- to the contralesional side (Marshall & Halligan 1990).

### **3.3 Study 7: Effect of 'vertical lesioning': Visual extinction without neglect.**

Our first study began by examining the effect of a 'vertical' lesion on the model. As noted above, after a severe 'vertical' lesion SAIM is unable to respond to anything presented on the lesioned side of the visual field. However, a milder lesion may allow a stimulus to be responded to when presented alone, though performance may suffer when competitors are introduced into the visual field. In the present simulation, the lesion was relatively mild and graded across the visual field (see Figure 20a for illustration). The lesion also affected the mapping into all parts of the FOA (and so it extended fully along the y dimension of the matrix of selection network units shown in Figure 20a), so that there was no bias favouring attention to one side of an object. With this lesion, if an object is selected, there should not be a tendency to attend to just part of it. We tested performance using the + and 2 stimuli, for which there is a differential bottom-up bias in the model: without damage, SAIM will select the + rather than the 2 stimulus, due to the + having a higher textured gmass (see Study 1). Initially the + was presented in isolation at various locations on the retina, starting from the far left and moving to the far right. Subsequently, in the 2-item condition, the + appeared within the left (lesioned) region of visual field and the 2 appeared in the right (unlesioned) region. To demonstrate effects purely of bottom-up constraints, we used the version of the model without the knowledge network, and traced the activation of the stimuli within the FOA; subsequently the knowledge network was added and essentially similar results occurred. The parameters for the lesioning were as shown in the Appendix. The vertical lesion involved a smooth (graded) transition across the visual field.

(a) The graded vertical lesion



(b) The graded vertical and horizontal lesion

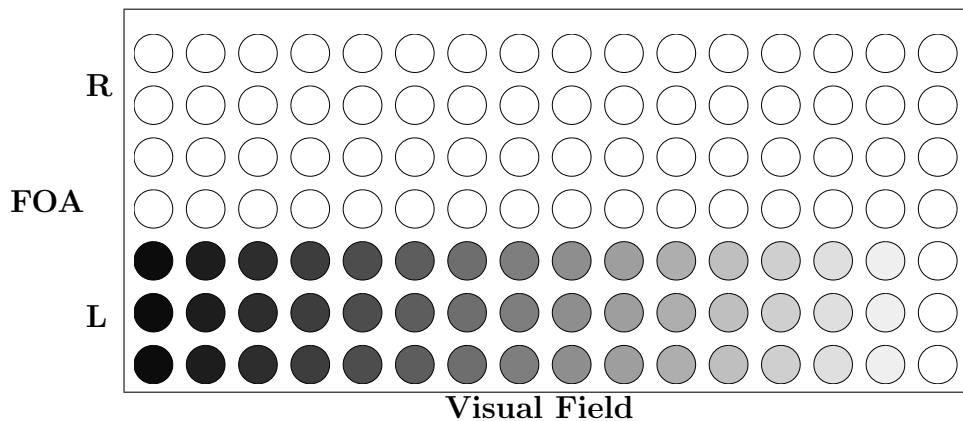
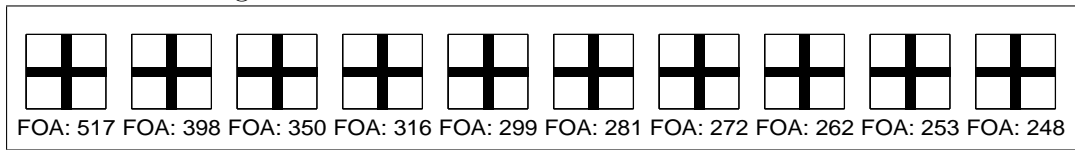


Figure 20: Example of the graded lesions used in the simulations. (a) depicts a vertical lesion affecting one side of the visual field more than the other; (b) depicts a combined vertical + horizontal lesion; here the lesion differentially affects units on one side of the field more than units on the other, and it differentially affects units on one side of the FOA (bottom rows in the selection network map into units on the left of the FOA).

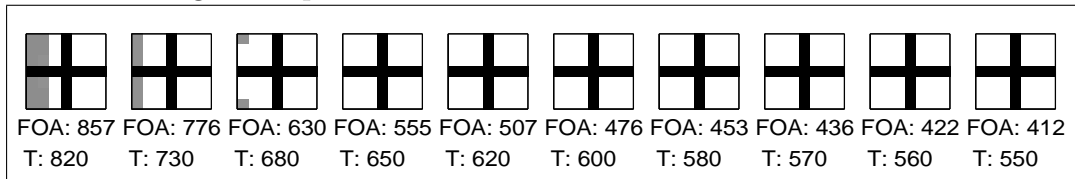
Figure 21 illustrates the mapping into the FOA after the vertical lesion for the + presented in isolation in the field. In this figure, we present the activation within the FOA for each position that the + was presented in the visual field, from left to right. With the lesion chosen, the stimulus was mapped successfully into the center of the FOA irrespective of its position in the visual field. There was however evidence of a cost in the time taken for the + to be attended, with RTs being considerably slower to left than to right-side stimuli. This result occurred with both the bottom-up only version of the model (Figure 21a) and with the full version of the network, with templates added. SAIM remained able to identify stimuli presented on the left, though identification times were slowed (Figure 21b).

Figure 22 illustrates performance when the + was presented concurrently with a 2, and the + was in a lesioned area of field (on the left). The stimuli were presented for 850 iterations

(a) Without Knowledge:



(b) With Knowledge: Templates: cross and two



Example stimuli position in the field:

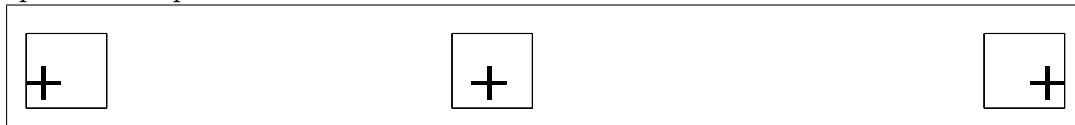


Figure 21: Effects of a vertical lesion on SAIM's ability to map visual information into the FOA, shown as a function of the position of the stimulus in the visual field (from far left to far right). (a) the bottom-up version of SAIM (without knowledge); (b) the full version (with knowledge). FOA: gives the detection threshold within the FOA; T gives the identification threshold within the template units.

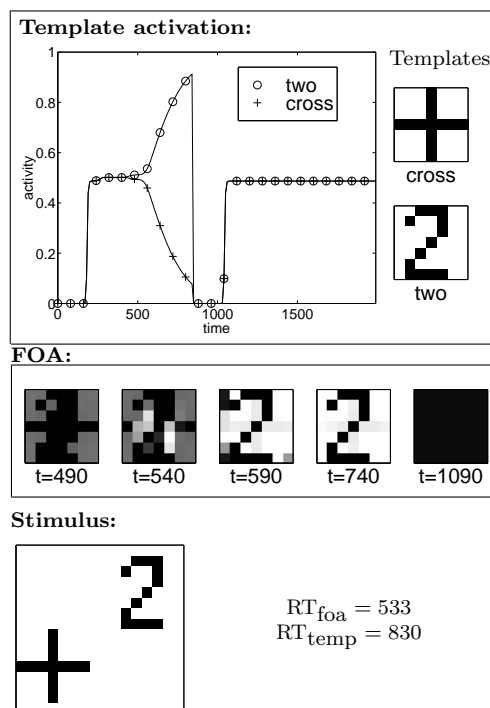


Figure 22: Effect of vertical lesioning when the normally preferred stimulus (the +) is presented in the impaired area of field (on the left).

and then turned off, to simulate the effects of short duration on human performance, when extinction effects can be observed. In its lesioned state, the model shows a bias to select the 2 in preference to the + (though in the unlesioned state the bias was in the opposite direction; Study 1). In addition, under the limited exposure duration only the item in the 'good' field is detected and the item in the impaired field is 'extinguished'. Following this, activation continued to be cycled through the model, but this was no longer supported by bottom-up input. The cycling of activation around the model moves the template units from their initialisation value (0) into a state of equilibrium in which they are activated at a 0.5 level (the equilibrium level for all units participating in the energy minimization process)<sup>11</sup>. At this level of activation there is no evidence to support the presence of any particular stimulus, so that the + is not identified. Note that the time taken to identify the + in isolation was less than the 850 network iterations used here, even for the lesioned version of the model, for nearly all stimulus locations (Figure 21b). SAIM can identify a single item in its lesioned field under durations that would produce extinction when a second item is presented on the unlesioned side.

Another interesting aspect of the network's performance is as follows. The demonstration of extinction depends on a balance between the bias in selection due to the lesion and biases in selection due to the stimuli. For instance, if the 2 were smaller, and the spatial bias from the lesion no greater, then the + may still win the competition into the FOA. However, the time for convergence to take place in the FOA would be more protracted than when the model is not lesioned, as indicated by the slowed convergence times relative to when the + is presented in isolation. It now follows that, under reduced exposure conditions the item in the impaired field might be selected whilst the item in the intact field is extinguished (if a short exposure is used, leaving insufficient time to switch attention to the second new stimulus). For SAIM, this counter-intuitive result emerges from the general slowing in the convergence of activation in the FOA after lesioning. Interestingly there are now several reports in the neuropsychological literature demonstrating that patients with right hemisphere damage can show abnormally poor report of right field items when cued to select a left field item, under conditions of brief stimulus exposure (Humphreys et al. 1996a). For SAIM this result is not an anomaly but rather an emergent property of the general slowing of attentional convergence in SAIM after lesioning.

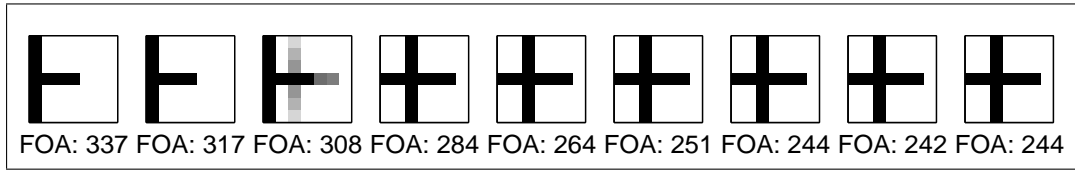
### 3.4 Study 8. Effects of combined 'vertical & horizontal lesioning': Field-dependent neglect

In Study 8, we examined what we will term a combined 'vertical and horizontal' lesion. The lesion used is presented in Figure 20b. For this lesion there was a graded 'vertical' deficit affecting units in the selection network more on one side of the visual field than the other (the x axis of the matrix in Figure 20b). However the lesion was not extended across all units feeding into the FOA, rather it only affected units feeding into one side of the FOA (so, in Figure 20b, the lesion progresses only half way up the y-axis, to affect units on the left side of the FOA more than units on the right side). This produces a form of 'horizontal' lesioning, in which the competition to map into one side of the FOA is unbalanced. The lesion values are given in the Appendix. SAIM was presented with

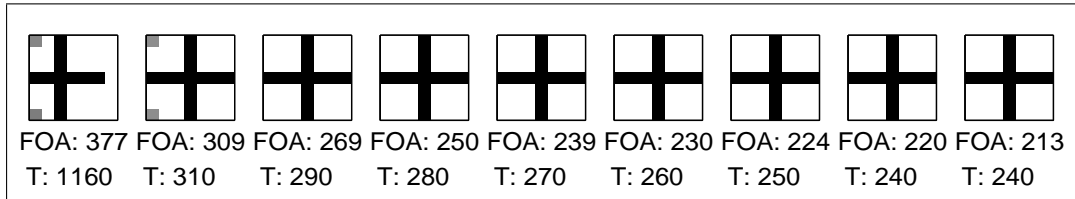
---

<sup>11</sup>This value of 0.5 would be attained here for templates corresponding to stimuli not even present in the field, since it is generated simply by the internal dynamics of the network.

(a) Without Knowledge:



(b) With Knowledge: Templates: cross and two



(c) Extinction

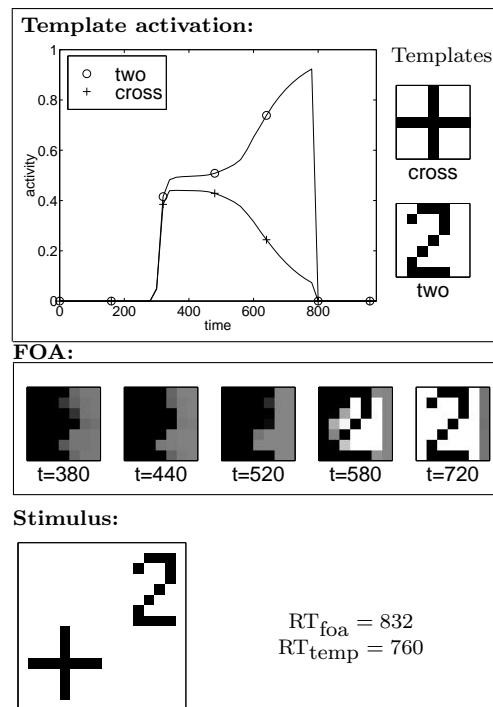


Figure 23: Effect of combined vertical + horizontal lesioning. In (a) and (b) performs in the field, from left to right. (a) the bottom-up version of the model; (b) the full version. (c) depicts the selection preference for a 2 over a + (in the lesioned area of field), which would result in extinction for the + under brief presentation conditions.

a single +, 7 x 7 pixels in size, at various locations on the retina. Subsequently it was presented with a + in the lesioned part of the field along with a 2 in the unlesioned area, to pit the biasing effects of the lesion against the biasing effects of stimulus salience (the + having a higher textured gmass than the 2).

With the combined vertical & horizontal lesion, there was some neglect of features on the affected side of the stimuli when it fell in the left visual field. This is apparent in Figure 23a, where we again show activation in the FOA as a function of the field position of the +. When the + was in the left field, none of the pixels in its left arm were transmitted

into the FOA. However, as the object was moved from the impaired into the less affected region of field, so neglect recovered. There is a form of field-dependent neglect. In addition to this, there was some tendency to neglect the left most pixel in the cross even when the stimulus was in the right visual field, when the bottom-up version of the model was used (Figure 23a). In the full version of the model (Figure 23b), there was generally less neglect (i.e., more pixels on the left of the stimulus were brought into the FOA) and, for right-field stimuli, neglect was eliminated.

These results indicate:

1. the combined vertical and horizontal lesion produced more neglect than the vertical lesion alone. For example, there was no neglect of the left pixels of the cross after the vertical lesion, even for the bottom-up version of the model when the + appeared in the right visual field (Figure 21); there was left neglect of a cross in the left field with a combined vertical and horizontal lesion (especially in the bottom-up version of the model; Figure 23a). This is interesting because, in terms of the number of units damaged, the vertical lesion alone is more severe than the combined vertical and horizontal lesion (see Figure 20), and the gradient of the lesion across the visual field was the same in both instances. This illustrates that what is important in the model is not the magnitude of the lesion, per se, but whether the lesion unbalances competition for mapping information into the FOA. The combined vertical and horizontal lesion unbalances this mapping, to favor one side of the FOA.
2. as illustrated by the example in (i) above, a horizontal lesion has some tendency to introduce 'object-based' neglect, affecting features on one side of the object even when the object is in the less affected field (Figure 23a).
3. in the full version of the model, there can be recovery of affected information due to top-down activation from templates (Figure 23b vs. Figure 23a).

We examine these last two factors further in Studies 9 (object-based neglect) and 10 (top-down filling in).

In addition to this, we again tested the lesioned model when the + appeared in its lesioned (left) field and the 2 in its intact field. As in Study 7, there was a bias in the lesioned model to select the 2 rather than the +. Under conditions of reduced exposure duration (under 800 iterations), there would again be a tendency for the + to be fully extinguished, even though parts of the + can be detected when it is presented in isolation (cf. Figure 23c with Figure 23b). In further examples discussed below (and shown, e.g., in Figure 27), we demonstrate that SAIM can simulate a form of extinction even with prolonged viewing conditions when a combined vertical plus horizontal lesion has been imposed (when the model becomes 'stuck' on the object first selected). Thus extinction-type effects are not necessarily contingent on the use of reduced exposures.

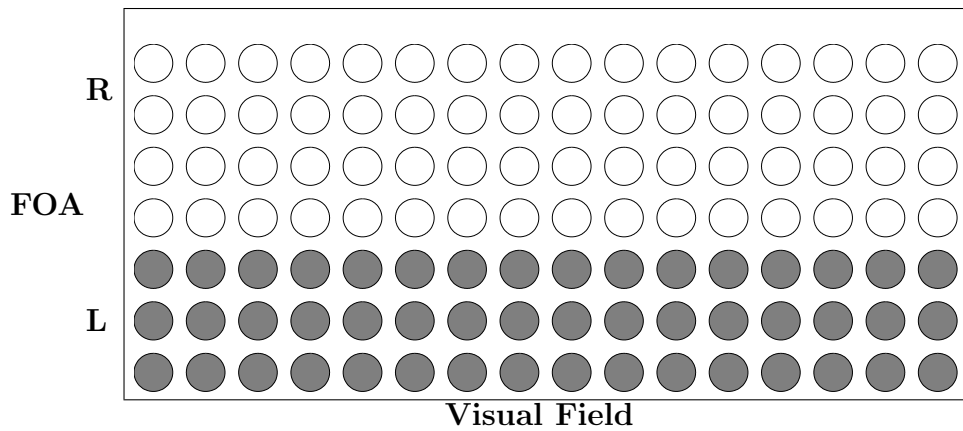
### 3.5 Study 9. Effect of 'horizontal' lesioning: Object-based neglect

#### 3.5.1 (i) Neglect of single objects

Having shown that different combinations of 'vertical' and 'horizontal' lesioning produce both 'pure' extinction (Study 7) and field-dependent neglect (Study 8), in Study 9 we

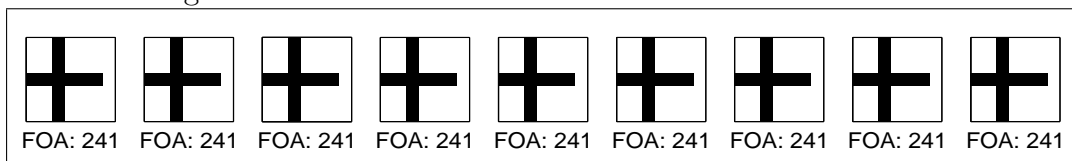


(a) Lesion



(b) Object-based neglect

Without Knowledge



With Knowledge: Templates: cross1 and two

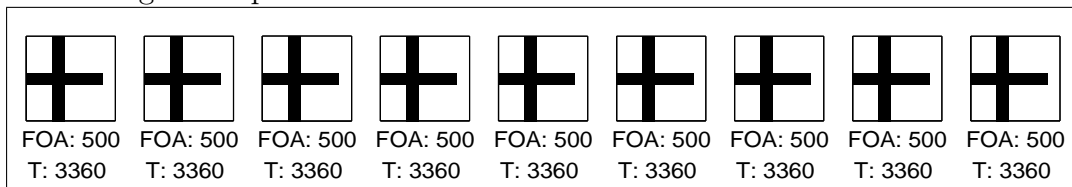


Figure 24: Effects of a horizontal lesion. (a) indicates the lesion used; (b) illustrates object-based neglect within the FOA as the stimulus is moved from left-to-right across the visual field.

examined the effects of a horizontal lesion alone. The lesion is illustrated in Figure 24a. There was damage to units in the selection network that transmit activation through to the left side of the FOA (along the y-axis of the matrix in Figure 24a). Unlike the cases where a vertical lesion was introduced, there was now no gradient to the damage across the visual field (x-axis, Figure 24a).

SAIM was presented with a 7 x 7 pixel +, which could appear at any of six positions in the visual field (moving from left to right). The model was run both with and without the knowledge network. The patterns of convergence achieved in the FOA, along with the times to attend to and identify the +, are illustrated in Figure 24b.

Figure 24b shows that, across all the field positions, there was neglect of the left most pixel in the cross, and the cross was in all cases mapped too far to the left of center of the FOA. This also occurred when identification rather than detection was measured (in the full network, Figure 24b bottom row). Here SAIM manifests a form of 'object-based'

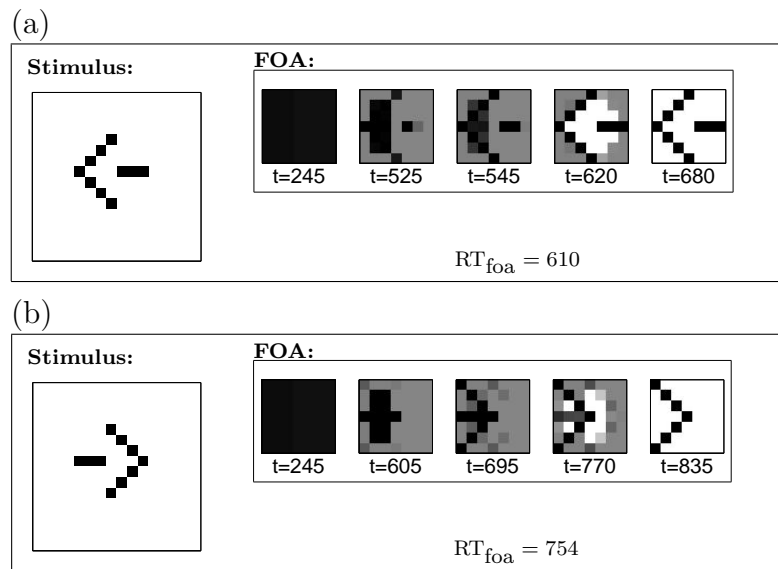


Figure 25: The effect of object structure on neglect following a horizontal lesion: the  $->$  is neglected more than the  $<-$ .

neglect, in which the left features of the object are neglected even when the stimulus falls in the right visual field (cf. Young et al., 1991). Interestingly, the model also showed no effect of field position on RT. It suggests that patients with 'pure' object-based neglect may respond with equal facility to stimuli in both fields.

We also evaluated SAIM's performance with 2 other objects, using the bottom-up version of the model: a left-facing arrow ( $<-$ ) and a right-facing arrow ( $->$ ). These stimuli were chosen not because they have been used to any great degree in studies of neglect patients but because, for the model, there is a contrast in the positions of elements relative to the center of the textured gmass of the shapes. The stimulus has more pixels to the left of the center of its textured gmass than the . With a lesion affecting processing on the left of the shapes, the will have more support for pixels falling on the neglected side relative to the . There should be more neglect of the left-most pixels in the . By contrasting SAIM's performance with these shapes, we test whether neglect, as expressed by the model, is sensitive to the position of elements relative to the center of the textured gmass of the shapes. Effects on neglect of the positions of elements relative to the center of mass of stimuli have been reported by Grabowecky et al. (1993) and by Pavlovskaya et al. (1997). Right hemisphere lesioned patients show less neglect when there are larger numbers of elements on the left of the shape.

The arrows each had 11 pixels. The same horizontal lesion was used again (Figure 24a), and the stimulus was presented with its center of textured gmass at the center of the visual field. The results, showing the pixels activated in the FOA, are depicted in Figure 25. SAIM manifested different degrees of neglect with the different stimuli. The right-facing arrow ( ) was subject to most neglect; all of its bar section was omitted from the FOA. In contrast the left facing arrow ( ) was not neglected at all; its center of textured gmass was mapped into the center of the FOA.

These contrasting effects with different shapes reveal an important aspect of SAIM's performance following lesioning; namely, the effect of lesioning is more complex than simply producing a consistent bias against stimuli in one part of the field, or the parts

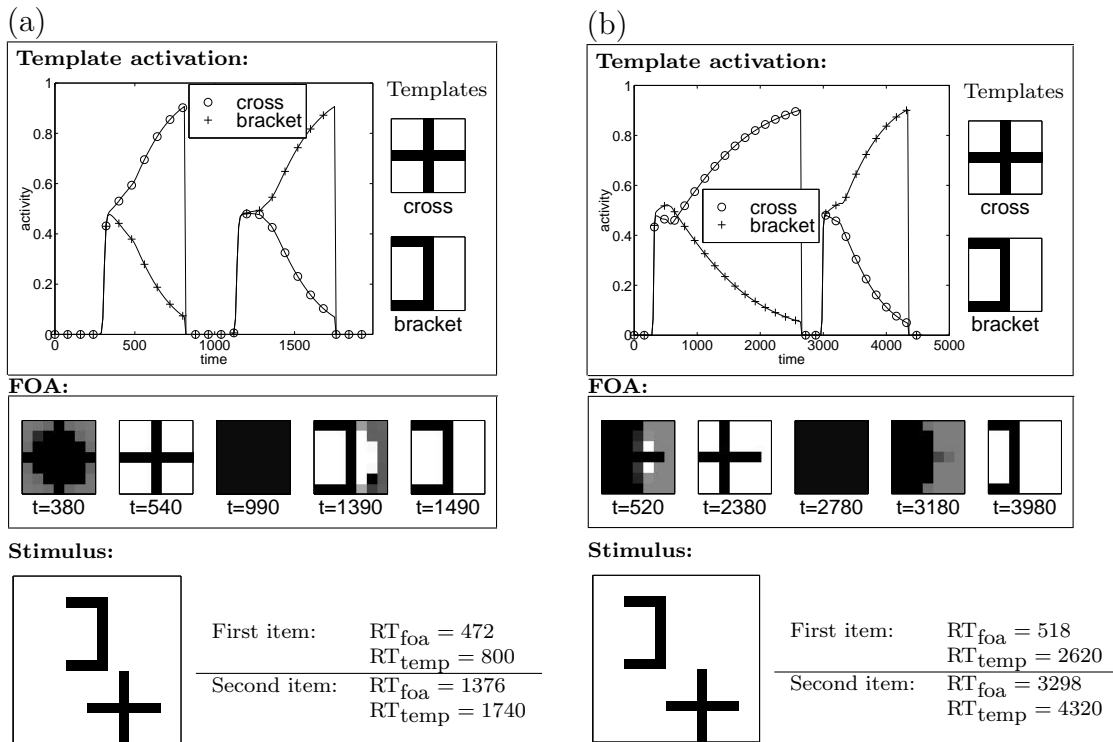


Figure 26: Selection and switching attention in an unlesioned version of the model. First the + is selected followed by the ]. (a) The unlesioned model. (b) The model after a left, horizontal lesion; note that there is now left, object-based neglect for each stimulus.

on one side of an object. This is because computation of the center of the textured gmass of a stimulus, and alignment within the FOA, depend on interactions within the selection network which are determined by: the degree of spatial bias, the magnitude of the textured gmass, the spatial spread of the shape, the proportions of the shape that fall either side of the center of the textured gmass, and the location of the center of the textured gmass relative to the spatial lesions. For example, here the has 5 pixels to the left and 3 to the right of center; the has the opposite proportions. There is less neglect for the after lesioning because there are increased numbers of pixels on the left relative to the right of the center of the textured gmass of the shape, and this counter-acts the reduced activation due to left-side lesioning. The is neglected more because the bottom-up spatial bias from the stimulus combines with the spatial bias due to lesioning, which favors pixels to the right of the center of the textured gmass of the shape. These results mimic the effects found in neglect patients (see Grabowecky et al. 1993, Pavlovskaya et al. 1997).

### 3.5.2 (ii) Multiple objects: Sequential neglect

In a second assessment of the effects of horizontal lesions on SAIM's behaviour, we examined performance when multiple stimuli were presented, when the model should normally switch attention from one object to another. Studies of neglect patients using simple drawings tasks (Gainotti et al. 1986, Marshall & Halligan 1994), and tasks requiring the discrimination of features within shapes (Humphreys & Heinke 1998), have demonstrated that forms of 'sequential neglect' can occur - there is neglect of each object selected in

a sequence. For instance, in drawing a patient may fail to include some details on the left of one item, move onto the next item and then fail to include its left-side details too. This can produce neglect of features on the left of objects that fall further into the right (non-neglected) field than non-neglected features on the right side of an object. It is not clear how well this result can be explained in other models of visual attention. For instance, in the MORSEL model of attention and object recognition, attention operates within a two-dimensional retinotopic map (Mozer 1991, Mozer & Sitton 1998). Neglect can be modelled in terms of a graded lesion, affecting attentional activation differentially from left-to-right across the map (e.g., with the lesion being more severe on the left). Now within such a model, effects of within-object neglect can be mimicked if there are graded lesions across the attentional network, so that there is always a bias for one side of an object over another irrespective of the object's position on the retina (Mozer et al. 1997, e.g. see). Competition between units in the attentional network will then ensure that one part of the object gets privileged access to stored knowledge over the other. This can simulate forms of object-based neglect in which there is neglect of left-side features in objects presented within the right field. However, the same graded lesion that produces this object-based effect would also lead to reduced activation of an object placed in the left field, and indeed the activation of the right-side features of such an object may be less than the activation of the left-side features of an object in the right field. Perception of the left field object should be poor. This is because neglect of parts of an object are intrinsically intertwined with neglect of parts of the visual field<sup>12</sup>. In contrast to this, in SAIM neglect of parts of objects can be treated separately from neglect of parts of the field. Consequently, as indicated below, SAIM demonstrates the same form of sequential object-based neglect as human patients. This proves to be crucial in the simulations of neglect within- and between-objects, presented in Study 11.

The lesioned version of SAIM was presented with 2 stimuli: a + and a bracket (]). The bracket had 13 pixels like the +, but it had a lower textured gmass (15.80). When the model was unlesioned, it first selected the + and then the ]; also each stimulus was fully mapped into the FOA (Figure 26a). After the horizontal lesion had been imposed (as in Figure 26a), SAIM attended sequentially to the + and then the ] as before, but for each stimulus there was a shift in its position within the FOA, with the left most pixel for each being neglected. This is shown in Figure 26b. SAIM shows sequential left-neglect of each object selected. Left-neglect occurred for each object because the model has difficulty in mapping input into the left side of the FOA, so that the right-side features play a more dominant role in the mapping process. The object is consequently displaced in the FOA and neglect occurs. This result matches data from human patients (Driver et al. 1992, Gainotti et al. 1986, Humphreys & Heinke 1998). The contrast with MORSEL at least in part reflects the fact that MORSEL uses a single view-dependent representation of attended space. There is only a single locus for deficits of attention, which must be graded across space to produce within-object neglect (irrespective of the positions of individual objects in the field). The penalty is that the gradient will also affect multiple objects according to their locations in the field. SAIM, however, creates a translation-invariant representation in its FOA. Damage to the processes involved in mapping into one side of the FOA produces a form of neglect that is specific to the features on one side of the object irrespective of the object's lateral position in the field and even when there is no

---

<sup>12</sup>Additional procedures that can be used to simulate object-based neglect in MORSEL are elaborated in the General Discussion.

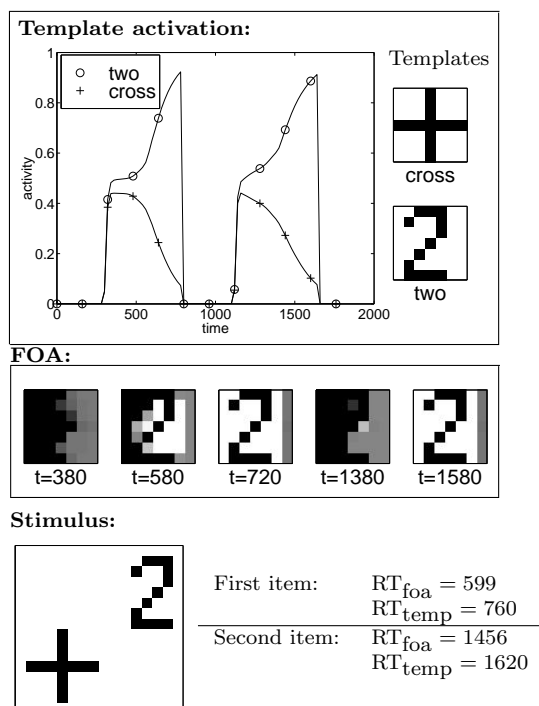


Figure 27: Example of 'sticky' attention and between-object neglect following a combined vertical + horizontal lesion (affecting both the left visual field and the left side of the FOA). The 2 in the less-lesioned field is selected in preference to the + in the more lesioned field (though the bottom-up bias is the opposite). In addition, in this example SAIM was unable to reject the stimulus first selected due to misalignment of the item in the FOA with its template. There is 'sticky attention' and neglect of a second object falling on the affected side.

bias across the field.

### 3.5.3 (iii) Multiple objects: 'Sticky attention' and impaired spatial memory

In the simulations using the + and the ], SAIM was able to inhibit the first stimulus after selecting it (the +), and consequently the second item could then be selected (the ]; see Figure 26b). However in other simulations with the lesioned version of the model we have found that rejection of first attended stimuli can sometimes be difficult. This difficulty arises because stimuli can be spatially distorted when they are mapped into the FOA (as here, where stimuli are shifted to the left of the center of the FOA). Even if a template is activated to threshold, due to a partial match with the attended figure, suppression based on active units in the selection network and on active pixels in the target's template may be inaccurate (note that active pixels in the template may not coincide with activation mapped through to the FOA, due to the spatial distortion in the FOA). On some occasions, this can lead to the model having difficulty in rejecting an object first selected, so that it continues to be selected even after inhibition of the selection network and the template. This is shown in Figure 27, where we examined the effects of combined vertical and horizontal lesions on the full version of the model (note that there is view-dependent neglect after this form of lesioning; see Study 8). In Figure 27 we illustrate performance when the 2 and the + from Study 2 were presented to the

model. The 2 was on the right of the retina and the + on the left. Due to the spatial bias from lesioning, the model selected the 2 prior to the +, overruling the standard bottom-up bias (Study 1). However, due to the spatial shift of the 2 within the FOA, inhibition applied after the 2 had been identified failed to suppress units associated with the 2. The net result was that, when the display was re-sampled, the model remained vulnerable to the spatial bias in selection, and the 2 was again selected. In other simulations using pure 'horizontal' lesions to simulate object-dependent neglect we have found similar effects. On such occasions, SAIM can appear to have severe problems in disengaging its attention from the object first identified. Problems in attentional disengagement are associated with damage to the parietal lobe (Posner et al. 1984), and they will be formally examined in SAIM using the Posner cueing procedure (Study 13).

Interestingly, even under conditions of 'sticky attention', SAIM does reject the item first selected at a microgenetic level (the 2, in Figure 28), but it then re-selects it<sup>13</sup>. Clinically this may be demonstrated when, on some occasions, patients select one item in search and then re-sample it shortly afterwards without necessarily knowing that they have previously selected the same location. This might be attributed to a problem in visual working memory on the part of patients, who seem to have 'forgotten' that they had already selected a particular item or location (see Husain et al. (2001), for data showing this and for an interpretation in terms of impaired spatial working memory in patients). In SAIM, though, the same problem emerges not due to a spatial memory problem but to the spatial distortion in selection which allows previously selected items still to compete for attention. SAIM's map of previously inhibited locations, which provides a working memory for the model, is disrupted by the lesion.

#### 3.5.4 (iv) Multiple objects: 'Between-object neglect'

The problem in re-engaging attention on one object also gives rise to one other interesting result. This is that, in SAIM, there can be neglect of an object that falls on the contralesional side of other objects present, a feature termed 'between-object' neglect by Humphreys & Riddoch (1994, 1995). In Figure 27, there is neglect of the + because it falls to the contralesional side of the 2. Note that this is not simply neglect of one part of the field since the + (or at least parts of it) can be selected from the same location when presented alone (Figure 24). What is neglected is determined by parsing of the field into objects and the selection of one object before another. The relations between neglect within- and between-objects was examined further in Study 12, where we show the independence of these phenomena by pitting the critical lesions against one another to simulate patterns of performance in human patients, who, following bilateral brain damage, have deficits affecting opposite sides of within- object and between-object spatial representations (see Costello & Warrington 1987, Humphreys & Riddoch 1994, 1995, Riddoch et al. 1995a).

---

<sup>13</sup>We term this a microgenetic level because it may be difficult to observe in a patient, as tests would need to be administered in the small time window following the inhibition of the first template.

### 3.6 Study 10. Template effects: Top-down filling in and premature acceptance

In Study 9, the templates the objects were activated to threshold even though parts of objects were neglected within the FOA. From this it might be tempting to assume that neglect may be less pronounced on tasks requiring stimulus identification (a response derived through template activation) than on tasks sensitive to attention to the full spatial extent of the stimulus. This also seems to be the case for other simulations of neglect, for example in MORSEL. MORSEL uses a pull-out network to generate better identification of words than nonwords, even when there is reduced activation of the left letters in the stimuli due to impaired attention to that spatial region (Mozer & Behrmann 1990). However, the recovery of the left letters for report does not lead to a spread of attention to the affected region, since there are no top-down connections from the recognition model to influence the attention system. In contrast to this, there is neuropsychological evidence indicating that there is greater spread of attention in neglect patients when a stimulus activates stored knowledge. For example, (Brunn & Farah 1991) found that neglect patients were better able to report the colors of contralesional letters in words than in nonwords, suggesting that stored knowledge modulates attention to the affected side. Also simple detection tasks, not requiring stimulus identification, can be performed more accurately when the parts to be detected make a known stimulus (e.g., Kumada & Humphreys (2001), Ward et al. (1994); see the Introduction). These results are consistent with stored knowledge having a direct, modulatory effect on attention. The architecture of SAIM allows for such modulatory effects to occur, since activation from templates is fed-back to bias the selection process. In Study 11 we test whether stored knowledge can influence neglect and extinction in the model, not only in identification tasks but also in terms of simple stimulus detection. We also assess whether template activation can be shown to be sensitive to neglect. In these simulations, the full version of the model was always used.

#### 3.6.1 (i) Top-down filling in, in neglect

The effects of top-down knowledge on neglect were tested in two ways: (i) by varying top-down modulation of the selection network by template units (is the degree of neglect a function of the magnitude of any feedback?); and (ii) by varying whether stimuli jointly activated a common template (so that stimuli could share top-down support). Performance was assessed using a 'horizontal' lesion only (see Figure 24a). First we compared performance with the usual degree of top-down modulation (top-down activation set to 3.0), with performance when a higher top-down parameter value was employed (5.0). In Figure 28a we present the data for the standard version of the model when presented with a + stimulus. The model selects the +, but manifests neglect on the left most pixel (which is not mapped into the FOA). The threshold for the + template shows a very prolonged time to reach threshold. The slow activation of the template here is partially due to the neglect of the + in the FOA and partially due to competition from the template for the 2; the 2 and the + share pixels and so generate competition at the template level. In Figure 28b we present the results when top-down activation is increased. Two effects are clearly apparent. One is that activation of the + template to threshold takes place much more quickly. The second is that top-down feedback from the template helps to

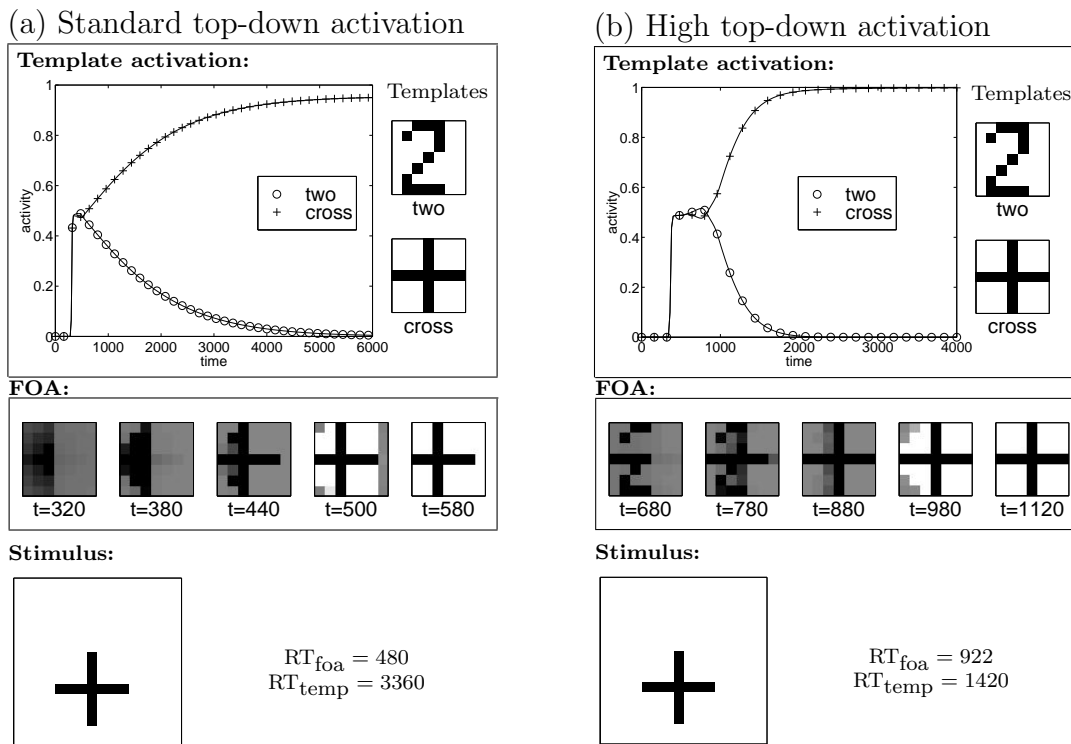


Figure 28: Effects of top-down 'filling-in' after a horizontal lesion. (a) standard top-down activation; (b) high top-down activation. Only a + is present in the field. There is neglect of the left-most pixels of the cross with standard but not with high top-down activation.

correct for the spatial distortion in mapping activation from the + into the FOA, and it 'fills in' the neglected pixel on the far left of the shape. Strong top-down feedback can help compensate for the effects of neglect and leads to a spread of attention to the affected region.

Another way to show the same effect in SAIM is to contrast the processing of stimuli which can be grouped by activating a common stored memory (template) and stimuli which activate different templates. Items which share the same template will both receive feedback as they are processed and they will be mutually supportive for selection. Items which activate different templates, however, will compete for selection. If there is a spatial bias in the selection process (due to lesioning) then, in the latter case, there can be neglect for the losing item. The same spatial bias need not produce neglect though, if the items activate a common template, since top-down activation from the template aids the recovery of the 'part' on the affected side. This is illustrated in Figure 29. In this simulation, SAIM was presented with an I and a T in its field, along with templates either for both individual letters alone (Figure 29a) or for the individual letters plus also the letters pair (i.e., a template for the word IT; Figure 29b). The model was run with a 'left horizontal' lesion. Quite different effects emerged according to whether the stimuli activated a common template. When the stimuli activated a common template, both items were mapped into the FOA. However, when the stimuli activated competing templates, the item contralesional to the lesion (the I) was neglected. SAIM first selected the T, but there was some distortion in its mapping into the FOA. As a consequence of this spatial distortion, the locations corresponding to the T were not inhibited, allowing it to be selected again.



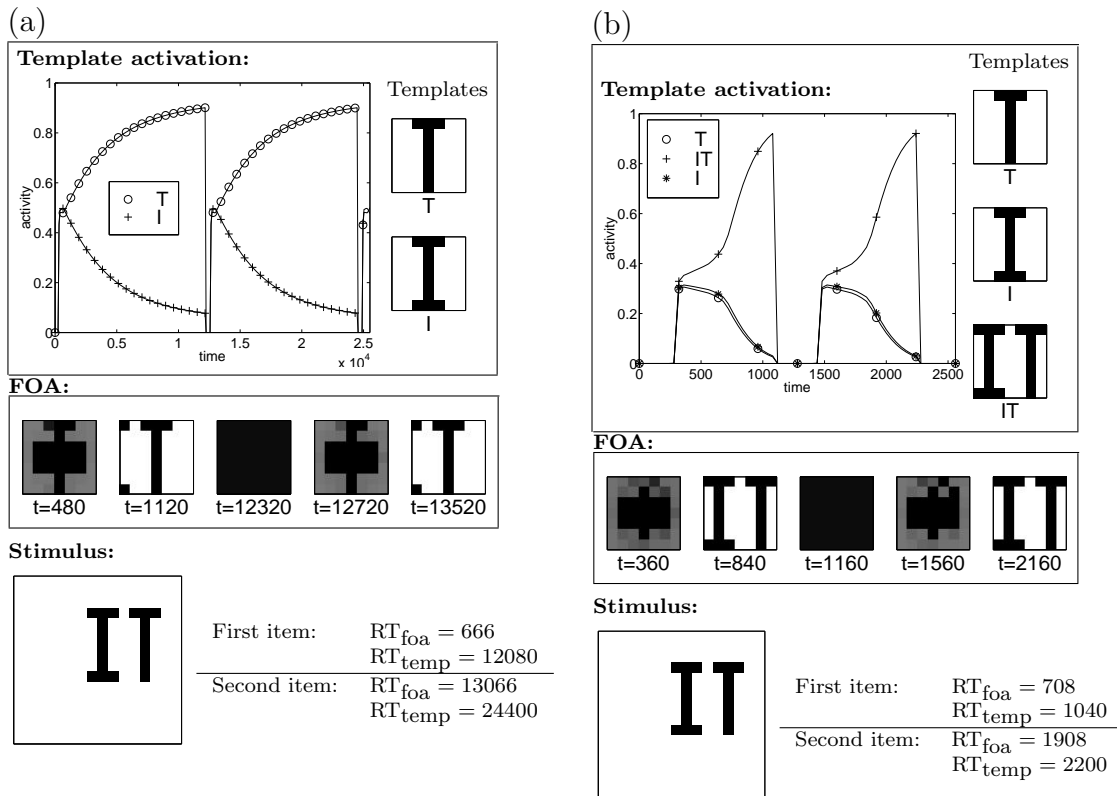


Figure 29: Effects of grouping by stored knowledge on extinction. The stimulus comprises 2 letters: I and T. In (a) there are separate templates for each letter and neglect of the left-most letter (the I). In (b) there is no neglect when a template for the word IT is added, though lesioning is the same in both instances.

Our purpose in this study is not to evaluate which version of SAIM gives the best fit to the human data, but rather to demonstrate the general case that top-down feedback can modulate neglect in the model. Note that the biases on attention in SAIM can be purely top-down, in the sense that operate differentially on stimuli that group equally in a bottom-up manner (the I and the T in IT). In humans, the differential report of letter colors with words and nonwords (Brunn & Farah 1991) may similarly reflect top-down effects of stored knowledge, given that bottom-up cues should group letters equally well in these two classes of stimuli.

### 3.6.2 (ii) Effects of premature acceptance

The prior study indicates that stored knowledge can help to ameliorate neglect and extinction. In this respect the model captures important effects present in human patients. For patients, however, contact of parts of a stimulus with stored knowledge may occasionally be detrimental. This occurs when partial information from the 'good' side of a stimulus matches a stored representation and causes what we will term 'premature acceptance'. An example might be where a patient is given a relatively unfamiliar compound word such as 'dingbat'. Activation of the template for 'bat' may lead to premature acceptance and identification of the stimulus as this word 'bat' (see Sieroff et al., 1988). We examined this issue by presenting SAIM with a + stimulus under conditions in which it had a template

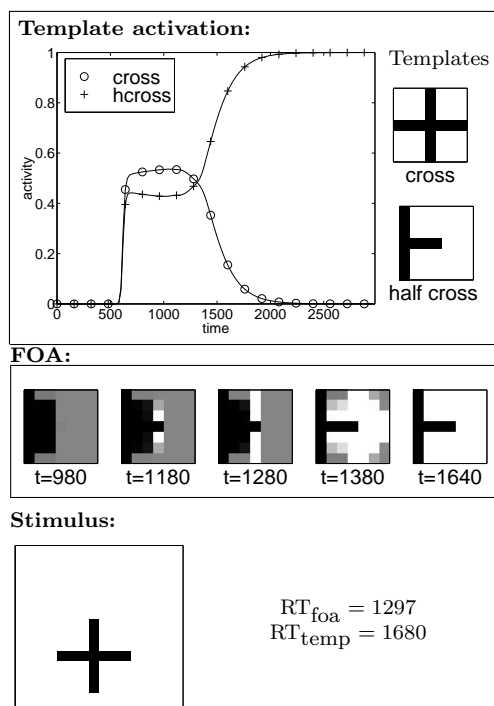


Figure 30: Effects of 'premature acceptance'. A cross is presented to a lesioned network having templates for the part as well as the whole. The template for the part in the good field wins the competition for selection.

corresponding to just the right part of the + (as well as having a template for the whole stimulus). A horizontal lesion was introduced to produce neglect. Remember that there is relatively little neglect of the + stimulus when there is a template for this stimulus and no template corresponding to the ipsilesional part (Figure 23b). However, with a lesion of the same severity there was quite severe neglect when there was a template for the ipsilesional part as well as one for the whole stimulus; this is illustrated in Figure 30. In this instance, stored knowledge of the part led to neglect of the whole.

The simulations shown in Study 10 demonstrate that activation of stored knowledge can either help or hinder report of a whole stimulus. Whether there is top-down filling in or premature acceptance of a part as the whole object takes will depend on a number of factors. To begin with, it will depend on the existence and relative strength of any stored representations for both the part and the object - as the simulations with the IT stimulus illustrated, a representation for the whole object can override biases to select the part. Performance also varies as a function of the bottom-up information in the stimulus (see Study 9, with the stimulus; Figure 25). Though this variance in SAIM's performance might be worrying at first sight, the fact of the matter is that human patients can show the same degree of variance or more when presented with subtly different stimuli. What SAIM does is to provide an abstract metric that we can use to try and classify human performance, based on factors such as: the familiarity of the overall stimulus, the strength of bottom-up grouping between the parts of the stimulus, whether the ipsilesional parts can be identified as an object in their own right, and the number of parts falling on the ipsi- or contralesional side relative to the center of textured gmass (Figure 25). These factors can be manipulated systematically in studies of neglect patients, to test the model.

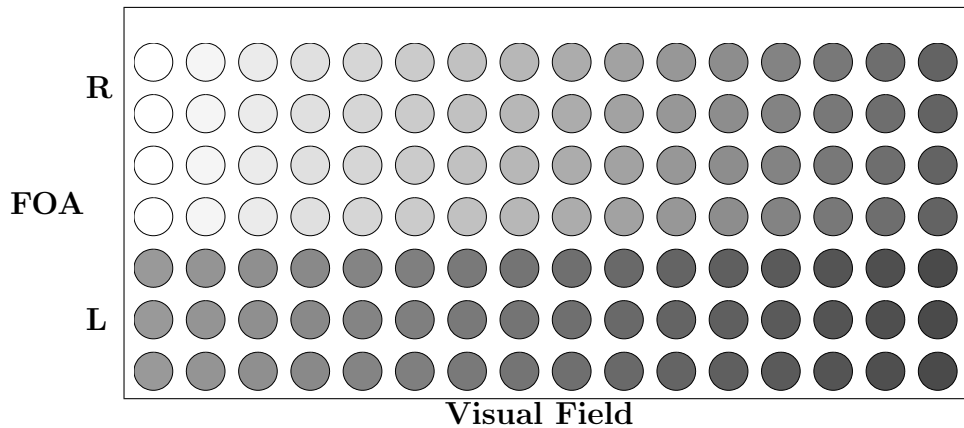


Figure 31: The 'bilateral' lesions that result in within- object and between-object neglect on opposite sides.

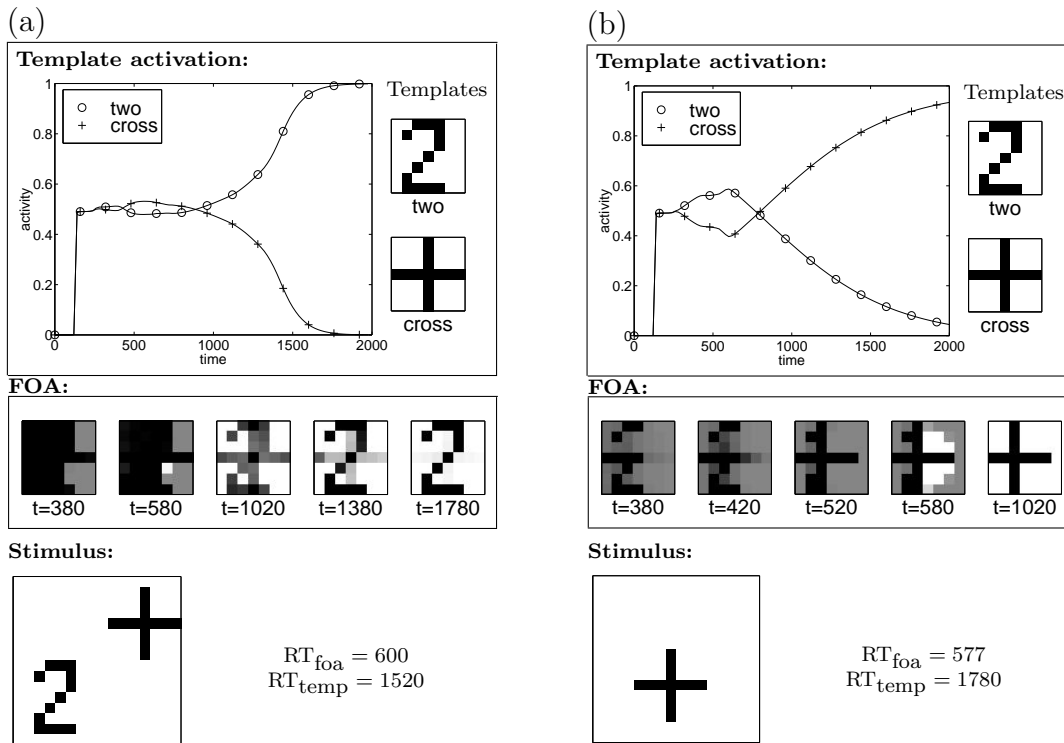


Figure 32: Simulation of within-object and between-object neglect on opposite sides, within a single network. (a) between-object neglect; there is a preference for selecting the 2 over the + when a lesion affects units receiving input from the right part of the visual field. (b) within-object neglect; when the cross is selected, there is neglect of the left-most pixels due to a lesion affecting the mapping into the left side of the FOA.

### 3.7 Study 11: Neglect within- and between-objects

In the Introduction we described results suggesting that contrasting patterns of neglect can be found in different patients; for example, some patients seem to show neglect primarily within the representation of a single object (Humphreys & Heinke 1998, Tipper & Behrmann 1996, Young et al. 1992), whilst others show neglect between separate ob-

jects without necessarily neglecting the parts within objects (Humphreys & Heinke 1998). These two patterns can be demonstrated using the same tests across different patients, indicating that they arise from dissociable impairments (Humphreys & Heinke 1998). The two types of deficit have been described as respectively within-object and between-object neglect. Whilst the term within-object neglect is relatively self-explanatory, the term between-object neglect needs a brief explanation. Patients showing between-object neglect will omit objects presented to the left of another object. It is not necessarily visual-field dependent however, since a single object in the left field may be discriminated (Humphreys & Heinke 1998). Neglect is contingent on the presence of multiple objects, and neglect is demonstrated in terms of a bias away from the contralesional objects in such sets.

Within many accounts of neglect, such a pattern of dissociable impairments is difficult to account for. Take again the MORSEL model (e.g. Mozer 1991, Mozer et al. 1997), in which object-based neglect can be simulated by introducing a graded lesion across its retinally-coded 'attentional field' (see above). To avoid neglect between objects, there would need to be a relatively shallow gradient to any lesion. However we might then expect all the parts of any identified object to be recovered too. Similarly a lesion that produces neglect of one of two separate objects is also likely to produce a gradient within an attended object, leading to neglect within- as well as between-objects. Particularly problematic is the finding that, following bilateral lesions, there can be both within- and between-object neglect, but for opposite sides of space (Costello & Warrington 1987, Cubelli et al. 1991, Humphreys & Riddoch 1994, 1995). How could opposite gradient effects be manifest within a single 'attentional field'?

These two results, of (i) between-object neglect with minimal neglect within objects, and (ii) position within object overriding effects of position within field, in within-object neglect, are produced by SAIM. For example, in the example shown in Figure 27 (Study 9), a spatial shift in the position of the object within the FOA leads to the object not being adequately inhibited once selection has taken place: attention remains 'stuck' on the object selected first. Nevertheless, all the parts of the selected (right field) object are attended. Thus the model selects the object in the unlesioned field and then has difficulty in selecting the second, contralesional object; there is neglect between objects. In contrast, in the example shown in Figure 26b, there is consecutive neglect of the features in both the left and the right-side objects, with the features on the right of the left-side objects being detected whilst there is neglect of the features on the left of the right-side object. There is within-object neglect, where position-in-object is more important than position-in-field. These different patterns of neglect can arise in SAIM because the selection network not only plays the role of selecting one from the multiple objects present, but it also acts to transform stimuli from a retinal to an object-dependent coordinate system. Lesions to the selection network can affect retinally-dependent mapping (the 'vertical' lesions) or object-dependent mapping (the 'horizontal' lesions), in which case different patterns of disturbance emerge. A mild horizontal lesion, combined with a vertical lesion, can produce neglect between-objects; a more severe horizontal lesion produces neglect within-objects. SAIM can account for double dissociations between patients within within-object and between-object neglect (Humphreys & Heinke 1998).

SAIM can also accommodate patterns of dissociation within single patients, since there can be two opposite lesion gradients, at different loci, to produce the two forms of neglect. This provides a particularly strong verification of SAIM's combined approach both

to visual selection and to translation invariant recognition. To illustrate this, we administered two lesions to the model: a vertical lesion modulated the effects of inputs into the selection network, more severely on the right than the left; a horizontal lesion modulated outputs from the selection network into the FOA, more severely on the left than the right (see Figure 31). These two lesions were chosen to capture the pattern of brain damage in a patient JR, reported by Humphreys & Riddoch (1994, 1995). JR had a left hemisphere lesion leading to neglect of the right side of multi-object displays, and a right hemisphere lesion leading to neglect of the left side of single objects.

Figure 32 shows performance of the model when it was presented with a + in its right field and a 2 in its left field. When lesioned the normal selection bias for the + was reversed: the 2 was selected prior to the +. If exposure times are limited, if the patient uses a too lax search criterion (failing to scan all the objects present), or if there are problems in inhibiting objects first selected (Figure 27) The likelihood of attention becoming 'stuck' will depend on the properties of the object, just as some objects showing more neglect than others (Study 9 (i)). The more an object is displaced within the FOA, the greater the likelihood of poor disengagement. The 2 used in this simulation is relatively robust to such displacements, but allows us to examine processing of the + when it is subsequently selected (Figure 27)<sup>14</sup>, this bias will lead to neglect of objects falling to the right of the object first selected. In contrast, when only the + was present, there was neglect of the left-most pixels (Figure 32b). Left neglect arose within this object because of the damage to the process mapping from the selection network to one side of the FOA. There is a selection bias favoring the left-most of two separate objects, along with a separate bias favoring the right parts within individual objects. Under appropriate circumstances, there will be both between-object and within-object neglect, but on opposite sides of space, within a single network subject to two lesions.

We might ask, however, whether it is neurally plausible to have 2 different forms of lesioning of SAIM's selection network. It is tempting to suggest that units in SAIM's selection network receiving input from the left visual field are in the right hemisphere and those receiving input from the right field are in the left hemisphere. Neglect of the left visual field would then be associated with right hemisphere damage, and neglect of the right field with left hemisphere damage. However, this would mean that units in each hemisphere feed into both sides of the FOA (e.g., via the rows of the matrices shown in Figure 20). It then becomes difficult to attribute patterns of within- and between-object neglect on opposite sides of space to separate lesions in the two hemispheres (e.g., right hemisphere damage producing left neglect within-objects and left hemisphere damage producing right neglect between objects; see Humphreys & Riddoch (1994, 1995)). On this view, within-object neglect should not be associated with a lesion to a particular hemisphere. Consistent with this, Caramazza & Hillis (1990) have reported apparent object-based neglect following left hemisphere damage, whilst others have reported effects after right hemisphere lesions (Humphreys & Heinke 1998, Young et al. 1992). Another way to conceptualise the lesion effects in SAIM might be in terms of damage to connections into or out of the selection network. For instance, all connections into the left side of the FOA may reside in the right hemisphere and may be selectively damaged by a right hemisphere lesion. This

---

<sup>14</sup>The likelihood of attention becoming 'stuck' will depend on the properties of the object, just as some objects showing more neglect than others (Study 9 (i)). The more an object is displaced within the FOA, the greater the likelihood of poor disengagement. The 2 used in this simulation is relatively robust to such displacements, but allows us to examine processing of the + when it is subsequently selected.

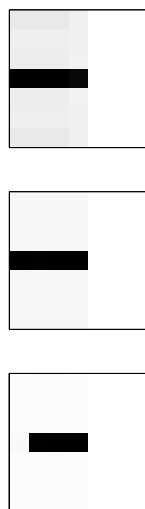


Figure 33: Line bisection in the lesioned version of SAIM. The stimuli were lines that were either 7 (top), 5 (middle) or 3 pixels long (bottom).

would disrupt output from the selection network to the left side of the FOA, producing left neglect within objects. Right neglect between objects would be due to damage to units in the left hemisphere that receive input from the right visual field. Alternatively it might be that similar selection networks exist in both hemispheres, perhaps with the networks being mutually inhibitory. Strong activation of the selection network in one hemisphere may then lead to inhibition of the equivalent network in the other hemisphere, with the relative magnitude of activation in each network being determined by the task (e.g., language tasks generating stronger left hemisphere activation; spatial tasks stronger right hemisphere activation). This notion of task and hemisphere-specific activation can be used to account for at least some patients who are reported as having a unilateral lesion but still show neglect on opposite sides of space (Cubelli et al. 1991, Riddoch et al. 1995a). In such cases, strong task-based activation of the damaged network could lead to it suppressing the network in the opposite (undamaged) hemisphere. It can also accommodate evidence that split brain patients can direct serial search simultaneously on both sides of space (Luck et al. 1994), which would follow if the networks in the two hemispheres were disconnected. We might also presume that connections from the left and right fields are stronger into units in the opposite hemisphere, and output into the left and right sides of the FOA are stronger in the opposite hemispheres. In this case, a vertical lesion to the selection network in the left hemisphere could generate between-object neglect of items in the right field, whilst a horizontal lesion to the selection network in the right hemisphere could generate left-side neglect within objects.

### 3.8 Study 12. Line bisection.

In Studies 7-11, we have demonstrated general properties of SAIM after lesioning the selection network. In Studies 12 and 13, we apply the model to two specific effects reported in the literature on patients with extinction and/or unilateral neglect: (i) the effects of length on line bisection, and (ii) performance in the Posner cueing task.

Study 12 examined the effects of length on line bisection. Line bisection is one of the

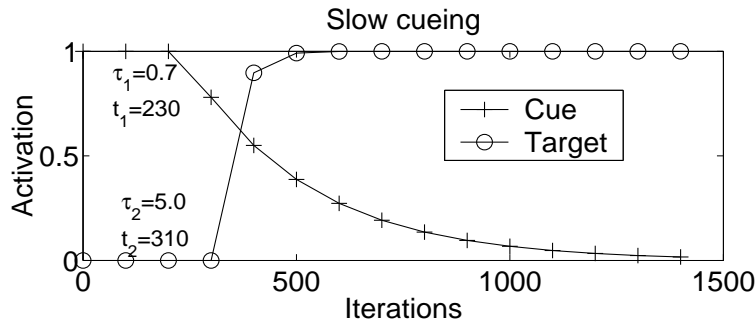


Figure 34: Example of the parameters used to determine decay and rise times in activation for cues and targets in the simulation of the effects of spatial cueing with a lesioned model.

traditional clinical tasks used to assess unilateral neglect. Patients are asked to mark where they think the center of a line falls. Patients with neglect after right hemisphere damage typically bisect to the right of the true centre (Bisiach et al. 1976, Halligan & Marshall 1988, Marshall & Halligan 1990, Riddoch & Humphreys 1983). In addition, the size of the rightwards error is typically related to the length of the lines: errors are shifted further rightwards for longer lines (Marshall & Halligan 1990, Riddoch & Humphreys 1983). We tested whether SAIM, when lesioned to produce left neglect, shows a rightwards bias in line bisection and whether it too is affected by line length.

The model was first tested after horizontal lesioning, using a single template in each case, always matched to the length of the stimulus presented. SAIM was presented with horizontal lines either 3 or 7 pixels long. Limitations on the size of SAIM's retina and on the size of the FOA constrained prevented more extensive manipulations of line length. The results are presented in Figure 33. In each case, the horizontal line was displaced towards the left of the FOA, with the right-most pixel set at the center of the FOA. With the longer line, the right-most two pixels failed to be attended at all (falling outside the FOA). If the perceived center of the line is judged to be where the center of the FOA falls, then the true centre was displaced by 1, 2 and 3 pixels for the line lengths of 3, 5 and 7 pixels. The magnitude of the displacement increased as a function of line length.

### 3.9 Study 13. Cueing in neglect

#### 3.9.1 (i) Posner et al. (1984)

Numerous papers have demonstrated that the magnitude of neglect shown by a patient can be reduced by cueing the patient to attend to the affected side (e.g. Riddoch & Humphreys 1983, Robertson et al. 1992). Indeed this effect of cueing has been taken as one of the prime pieces of evidence to suggest that neglect reflects a disorder of visual attention (see Riddoch & Humphreys 1987, for one example). Perhaps the best known example of the effect of cueing is that reported by (Posner et al. 1984). They tested the detection of ipsi- and contralesional targets in patients with unilateral parietal lesions. Targets were preceded by brief peripheral cues that fell either at the location of targets or on the opposite side of fixation to where targets appeared. Posner et al. found that there was relatively little difference in RTs to detect ipsi- and contralesional targets when cues were valid; however there was a marked difference on invalid trials: there was then particularly slow detection of contralesional targets. Posner et al. interpreted their findings as

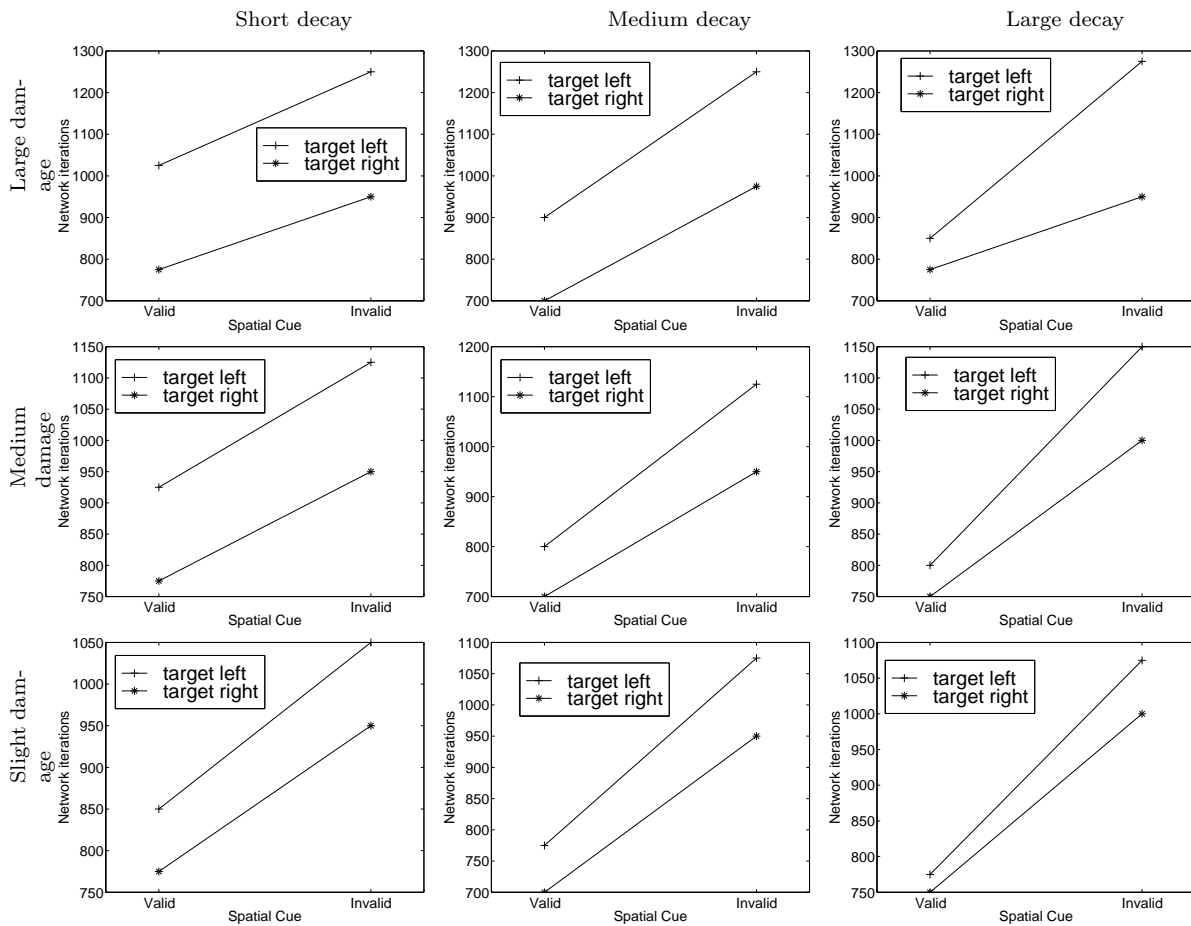


Figure 35: RTs for a lesioned version of SAIM to identify targets preceded by valid or invalid spatial cues. Performance is shown as a function of the magnitude of the lesion (rows) and the decay time for the cue (columns)(see the text for details).

indicating that the patients were selectively impaired at disengaging attention from ipsilesional cues to detect contralesional targets. We assessed whether such a disengagement problem would occur in SAIM following lesioning to one side of the selection network.

The patients tested by Posner et al. did not have gross forms of neglect, though some were reported as having extinction. To generate extinction we used a 'vertical' lesion (as in Study 7), but varied the strength of the lesion to provide a test of the generality of the findings (see the Appendix for information on the magnitude of lesioning). With one exception, noted below, the conditions otherwise matched those used to examine spatial cueing in the unlesioned version of the model when detection alone was measured (Study 6). The one change made to the lesioned version of the model was to vary the parameter for the decay of activation in the model; this was set to either 5.0 ('normal'), 1.5 or 0.7 at the lower parameter values there was slowed decay of activation from the cue, enabling it to have an increased effect on responses to the target (see Figure 34). This turned out to be important for the results.

The times taken for activation in the FOA to reach threshold are given in Figure 35. When cues were valid, the time for convergence to be achieved in the FOA was less for targets presented to the ipsilesional field than for targets presented to the contralesional field. However, when the cues were invalid, this advantage for ipsilesional targets in-



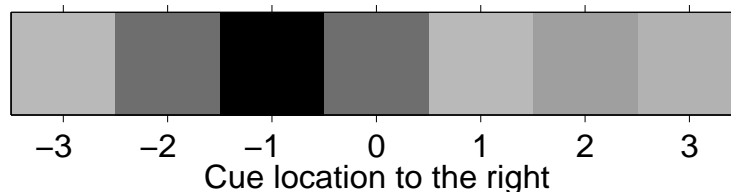


Figure 36: Effects of cueing the center of mass of a shape, following a left-side lesion. At position -1 the cue falls to the left of the center of mass of the target shape.

creased. This general pattern occurred across all the parameters varied here, but it was most pronounced when the temporal decay of the cue was increased and as the magnitude of the lesion increased (from the bottom to the top graphs in Figure 35). With a faster decay of activation from the cue, there was a relatively stronger effect of the lesion than the cue, so that RTs were slow to contralesional stimuli even with a valid cue RTs. The overall pattern of data matches that reported by Posner et al. (1984). The deficit on contralesional targets is increased under conditions of invalid cueing.

In SAIM these effects of invalid cueing reflect competition in the mapping of stimuli through to the FOA. A valid cue can facilitate this mapping by pre-activating units in the selection network that represent a correspondence between the position of the target in the visual field and the FOA. An invalid cue disrupts this mapping, since it generates initial competition for a different mapping between the retina and the FOA. When input into the FOA from the uncued area of field is weakened by lesioning, the effects of competition are increased. In this respect SAIM behaves similarly to other recent connectionist simulations (Cohen et al. 1994, Humphreys et al. 1996b); like other models it shows an 'attentional disengagement deficit' through exaggerated competition for selection within a winner-take-all network. It again demonstrates that disengagement deficits are not necessarily contingent on having a module to perform this processing operation (cf. Posner et al. 1984), but rather such deficits can reflect unbalanced competition due to a spatially asymmetric lesion.

Our finding that the parameter governing the temporal decay of activation is important for this result is also of interest. In the main part, studies of deficits in spatial attention have concentrated on what we may term 'single factor' accounts of any impairments. One exception to this rule was reported by (Duncan et al. 1999). These investigators attempted to characterise patients with unilateral right hemisphere lesions and signs of neglect in terms of the parameters of the 'TVA' model of Bundesen (1990). They found that, in addition to any deficits in spatial selection, the patients also tended to have slowed visual processing (e.g., their letter identification performance improved at a slower rate than normal as stimulus exposure duration increased, even for stimuli presented in the ipsilesional field). This non-lateralised deficit also predicted the degree of neglect shown by patients in standard clinical tests. A similar argument has been made by Husain et al. (1997), in a study using rapid serial visual presentation of letters at fixation. These studies indicate that non-lateralised processing deficits can contribute to the clinical picture of neglect in many patients. A slowed decay of activation, as implemented in SAIM, would lead to reductions in the speed of processing rapidly presented stimuli.

### 3.9.2 (ii) Pavlovskaya et al. (1997)

In (i) above we used a lesion that generates extinction rather than neglect in SAIM. Effects of cueing have also been reported on patients with frank neglect (e.g. Pavlovskaya et al. 1997, Riddoch & Humphreys 1983, Robertson et al. 1992). We now ask, is the result in (i) generalisable to lesions that generate neglect in the model? In addition, Pavlovskaya et al. demonstrated that cueing effects in neglect were affected by the position of the cue relative to figural properties of subsequent targets. In Study 6 (iii) we simulated their data for normal subjects. Here we assessed whether SAIM, when lesioned, could capture their neuropsychological results.

With neglect patients, Pavlovskaya et al. found that cues to the left of the true center of mass of a shape facilitated shape discrimination (unlike control subjects, who show best discrimination when cued to the center of mass). They proposed that this displaced cue shifted the location where the patient habitually attended (which fell to the right of the center of mass), so that attention now fell at the center of gravity of the target, facilitating its identification. In our simulation of this in SAIM we used a combined vertical and horizontal lesion, to generate neglect as well as extinction (as in Study 8). To measure identification the full version of the model was used, with a single template for the target. The model was presented with a cross that fell on the left of the retina, where some degree of neglect occurs when no cue is given (Figure 21). Prior to the target occurring, a cue was presented either at the center of mass of the target shape or 1-3 pixels away on the left or right arm of the shape. In an unlesioned state, the effect of the cue is larger when it falls at the center of mass of the target (Figure 16).

Apart from the lesion, the procedure was the same as in Study 6 (iii). Figure 36 presents the results, with darker shading representing faster RTs. At point 0,0 the cue falls at the center of the textured gmass of the target. At point -1, 0 it falls 1 pixel to the left but at the same vertical height as the location of the center of the textured gmass of the target. + and - values on the vertical dimension have been averaged together for the figure. RTs to identify the target were fastest when the cue fell to the left of the object's center of textured gmass then when it fell at the center itself, and they were slowest when the cue fell to the right of the center of the textured gmass. Again this result is not dependent on the use of a cross as the target, and similar effects of cueing in relation to the center of gmass of the shape occur with other stimuli.

This result demonstrates two main points. First, that cueing can affect patients who show neglect as well as extinction. Second, that the magnitude of the cueing effect is related to figural properties such as the center of the textured gmass of the shapes, matching again the human data (Pavlovskaya et al. 1997). SAIM generates these results because (i) neglect involves a shift in where the center of attention is located within a shape, and (ii) cueing the location where features appear gives these features a 'head start' for mapping into the center of the FOA. This can be sufficient to overcome a spatial bias in the mapping process, when attention is cued to the features that would otherwise suffer the bias.

## 3.10 Conclusions

In Section 2 we have shown how simulated lesions affect the performance of SAIM. SAIM can be lesioned in different ways. We have examined lesioning that affects units that

receive input from one part of the visual field ('vertical lesions) and lesioning that affects units that transmit output to one side of the FOA ('horizontal lesions). Lesions that include some 'vertical component are sensitive to the positions of items in the visual field, and produce: spatial extinction, 'disengagement deficits after spatial cueing, and effects of line position on bisection. These lesions can also produce problems in which the model continues to select an object even after it has been attended before, due to poor template-based inhibition ('sticky attention). In such circumstances there can be neglect between two objects along with only a mild effect on recovering all the parts within an object (i.e., there is between-object neglect; Humphreys & Heinke (1998), Humphreys & Riddoch (1994, 1995). 'Horizontal lesions, in contrast, produce object-dependent neglect that is affected by the positions of parts relative to the center of mass of the shapes, and neglect is then uninfluenced by the positions of the shapes in the field. When subject to such lesions, the model may select objects successively from different field locations but then neglects the parts on the affected side of each one (i.e., there is within-object neglect). A 'vertical lesion on one side, when coupled with a 'horizontal lesion to the other, can even generate between- object and within-object neglect on opposite sides of space within the same network. This double dissociation within a single patient is difficult to account for in other models, but it matches the neuropsychological data (Costello & Warrington 1987, Humphreys & Riddoch 1994, 1995). Also, neglect can be mediated by properties of objects. 'Better objects, whose parts group or that fit with a stored stimulus description, will show less neglect than objects that are less good, depending on other factors such as field position and the magnitude of any lesion. Finally, data on the effects of cueing on patients with parietal lesions can be simulated, especially when there are changes in the decay parameter within the model, so that slowed processing of stimulus information results. These cueing effects are sensitive to the relative positions of the cue and the centre of textured gmass of the target shapes.

## 4 General Discussion

SAIM is a model of translation invariant object recognition, in which invariance is achieved by mapping activation through an attentional window (the FOA) en route to the activation of stored templates for objects. Multiple objects, when present, compete to win the mapping through the FOA. Hence SAIM provides a computationally-motivated account of visual selection for object identification. Our simulations have shown that the model may not only have useful computational properties (see also Olshausen et al. 1993, 1995), but also that it can be applied to a broad body of psychological and neuropsychological data. SAIM provides a qualitative account of such data. Within the boundaries of how SAIM represents stimuli, the model predicts qualitative patterns of performance in a systematic fashion, according to how stimuli are coded within the model. This approach, however, needs some justification, since in many simulations modellers attempt to relate network performance (e.g., timed in network iterations) directly to actual reaction time or accuracy data. We first consider the utility of qualitative modelling, and whether such a modelling approach could be refuted. We subsequently review three aspects of the model that we believe make it attractive as a general approach to modelling human visual selection: (1) the way in which SAIM links objects to space; (2) the way in which SAIM combines selective processing with its coding of objects in particular representational schema; and (3) its ability to integrate some seemingly disparate phenomena from the syndrome of unilateral neglect within a single framework. We finally discuss (i) the relations between the model and others in the literature, (ii) some emergent properties of the model (in terms of both visual selection and spatial working memory), (iii) some results on visual selection that the model does not address and (iv) the biological plausibility of the framework in SAIM and how it may be extended.

### 4.1 Qualitative fit and the refutation of models

SAIM is by no means the first connectionist model that has been used to provide a qualitative fit to data; for example, simulations of word recognition using the interactive activation and competition framework also took a similar approach (McClelland & Rumelhart 1981), and an analogous approach has been used in nearly all other models of visual attention (e.g. Humphreys & Müller 1993, Mozer 1991, Phaf et al. 1990). The aim of a qualitative approach is to demonstrate patterns of behavior in the model that match those found in human subjects. In the normal (unlesioned) version of SAIM, the patterns of behavior include:

- (i). the effects of the center of textured gmass on both object coding (the center of textured gmass is the frame around which other parts are linked, within the FOA) and attention (which is drawn first to the center of textured gmass; Study 1);
- (ii). the costs of selection when there are multiple objects in the field (Study 2);
- (iii). the tendency of the model to select the better' of two objects first (Study 2);
- (iv). the tendency to make false alarms to stimuli similar to expected targets (Study 3);
- (v). the faster identification of the whole rather than the parts of objects (Study 4);

- (vi). the effect of spatial cueing on visual attention, including effects of the center of mass of the object and inhibition of return (Study 5);
- (vii). the effects of object-based properties on both attentional cueing and on inhibition of return (Study 5); and
- (viii). the ability to select one of two overlapping shapes, presented in the same spatial region (Study 6).

In the lesioned version of the model the patterns of behavior simulated include:

- (i). visual extinction (Study 7);
- (ii). field-dependent and object-based neglect (Studies 8 and 9);
- (iii). the effects of top-down knowledge on helping recovery from neglect (Studies 8 and 10) and extinction (Study 10);
- (iv). the effect of the center of mass of an object on neglect (Study 9);
- (v). neglect errors due to premature acceptance of a known part of an object (Study 10);
- (vi). neglect between' as well as within objects', even within the same model when subject to two lesions (Study 11);
- (vii). the effects of line length and cueing on neglect (Study 12); and
- (viii). the effect of cueing on neglect, including the interaction between the position of the cue and the center of mass of the object (Study 13).

In all of the above aspects, the performance of the model matches the patterns of effects found with human subjects. We suggest that this attests to the power of the model as a framework for understanding human visual selection. There was only one significant change to the parameters needed for a particular simulation (apart from the variations in top-down activation, but these variations were the model's means of reflecting the familiarity of the objects) this was the change in the decay parameter, which influenced the effect of a cue on responses to a subsequent target when the model was lesioned. However, even this change can be related to evidence on non-spatial deficits in processing associated with visual neglect and extinction (e.g. Duncan et al. 1999).

Now the way that SAIM produces some effects does not in every case match precisely the factors that are likely to be important in human perception. An example is the effect of the goodness' of the object on selection. As noted above, given good' and less good' objects in the field, SAIM tends to select the good object prior to the less good one. For SAIM the goodness' of an object can be defined bottom-up, in terms of the relative proximity of the pixels in the object and the magnitude of its textured gmass (Study 2). Goodness' can also be defined in a top-down manner, in terms of whether an object has a pre-activated template (Study 3). In humans comparable studies have evaluated selection with degraded and undegraded objects (Farah et al. (1991); or high and low contrast stimuli, Gorea & Sagi (2000)), and selection with stimuli subject to learning via varied or consistent mapping (Shiffrin & Schneider 1977). In addition, neuropsychological studies show that better' objects win the competition for extinction, when object goodness has

been varied in terms of whether line elements are collinear and whether a closed or non-closed pattern is presented (Humphreys et al. (1994); see also Ward & Goodrich (1996)). In its present form, SAIM cannot simulate some of these effects. For example, SAIM is not sensitive to factors such as edge collinearity or closure, and so would show no bias for a closed over a non-closed pattern, if the patterns were matched for the magnitude of their textured gmass. We do not believe this is crucial, however. The qualitative pattern of performance shown by SAIM indicates that, within a competitive model of this sort, an advantage gained for one pattern over another (whether due to bottom-up grouping or top-down activation) will pan out in terms of unbalanced competition for selection. By building a more sophisticated front-end to the model, we would be likely to simulate effects of closure and collinearity on selection precisely because these factors would also influence competition within the model, the general principles of which have been explored here.

An alternative to the qualitative approach would be to attempt to match the number of iterations in the model to a measure of human performance, such as reaction time (see Seidenberg & McClelland 1989, for one example), so that a quantitative fit is provided. The difficulty with this is that our simulations have been applied to a substantial range of studies, which have used widely different stimuli and tasks. The simulations with SAIM show that the time taken to attend and to identify objects depends on factors such as the number of objects in the set, the center of mass of objects, the spacing between the objects and so forth. The procedure of providing a quantitative fit would involve making several arbitrary assumptions and run the risk of providing a spurious validity that is essentially dependent on the setting of particular parameters. Instead of this, we have shown that a single set of parameters in SAIM generate reasonable results across substantial variations in stimuli (e.g., see Figures 6 and 7). We suggest that a qualitative approach, in which a model matches large data sets from psychological studies, can provide useful constraints on theory (see also Mozer & Sitton 1998, for a similar argument).

One legitimate concern with a qualitative approach to simulation, though, is whether such models could ever be refuted. We suggest that SAIM can be refuted, but what is important is to distinguish attributes that are critical to the operation of the model and those that take the form of implementational detail (see Ellis & Humphreys 1999, for a more extended discussion of this argument). We point out three factors that are critical to the current operation of SAIM: (i) that there is translation invariance by means of mapping visual information through an attentional window; and (ii) that mapping into the window operates over time, controlled by bottom-up factors such as element proximity, and (iii) and that object-based effects in the model are contingent on the use of stored templates. On each of these points SAIM is open to refutation.

Take factor (i). The means by which translation invariance is achieved in SAIM leads to outputs from the attentional window being coded in terms of the relations of parts to the object, irrespective of the lateral position of the objects in the field. This predicts that lesioning can have the same selective effect on object identification irrespective of lateral object position, as is found in object-based neglect. However, if forms of object-based neglect did not exist (e.g., if all such disorders were bound to the retinal positions of stimuli, say), then the model would again be incorrect in a fundamental point. Given SAIM's architecture, it has to break down in this particular way. In addition, our simulations of object-based neglect indicated that RTs to a target need not vary dramatically as a function of whether the target fell on the contra- or on the ipsilesional sides of space

(Figure 23). RTs in patients with object-based neglect need not be strongly affected by visual field. This is clearly open to empirical test.

Now take factor (ii). In SAIM the mapping into the attentional window is guided by factors such as inter-element proximity, which provides perhaps the simplest bottom-up constraint on selection without even this constraint, SAIM would fail to map parts of objects together in any sensible way. Again, the model would be incorrect if it turned out that human perception was not influenced by proximity or by factors such as the center of mass of an object. It turns out, however, that SAIM's sensitivity to the center of the textured gmass of a shape not only accounts for existing data (e.g., using attentional cueing paradigms) but also leads to new predictions about the time course of attention to objects. SAIM holds that objects are not attended in a single step but rather attention spreads over time, from the center of mass of the object to its outer fringes. The model is open to refutation on this point.

Finally, consider factor (iii). SAIM implements object-based effects on selection in a quite specific manner: by means of activation from templates for known objects. Now many psychological experiments on this topic use very basic stimuli (e.g., simple boxes), and it is not clear whether any object-effects then reflect stored representations for these items, or whether they reflect novel object files' set up on the fly, on each trial of the experiment (cf. Kahneman & Treisman 1983). SAIM, however, predicts that effects specific to known objects should be apparent in these studies. Tipper (2001) has recently explored this in studies of inhibition of return. His data suggest that our attention is biased away from a particular face, having attended to that face previously, as would be expected if there were inhibition of the stored representation for that face. SAIM also predicts that selection should be influenced by stored knowledge even when bottom-up factors do not favor one object over another. As we have noted, neuropsychological data on visual extinction are consistent with this. For example, there is less extinction for letters in words than for letters in nonwords, though bottom-up information is much the same for these stimuli (Kumada & Humphreys 2001). For SAIM the effects of stored knowledge should not only occur in tasks requiring read out' from high level representations (e.g., word identification) but also in low level' tasks not requiring identification of the whole stimulus as a single unit' (e.g., in detection or in tasks such as identifying the colors of individual letters; cf. Brunn & Farah (1991). If top-down effects are found only on high-level tasks, this would contradict the model.

## 4.2 Useful properties of the model

There are several properties of SAIM that make it attractive as a framework for understanding visual selection. We consider three.

1. Linking objects to space.

Both psychological and neuropsychological evidence suggests that visual selection is affected by both the spatial relations between stimuli (Posner 1980, Posner et al. 1984) and by grouping relationships between image features (Baylis & Driver 1993, Donnelly et al. 1991, Duncan 1984, Egly et al. 1994, Humphreys & Heinke 1998, Vecera & Farah 1994, Ward et al. 1994). General frameworks have been proposed to account for such effects (e.g., suggesting that top-down influences from object coding and grouping activate areas of field to which a spatial attention system spreads; see

Farah (1990), Humphreys & Müller (1993), but none have hitherto been worked out in detail. SAIM provides detailed mechanisms by which object properties may modulate spatial selection. In the simulations presented here, two mechanisms are explored: bottom-up grouping by proximity relations within the selection network (see Study 1), and top-down grouping by template activation and feed-back to the selection network (Study 4). We have shown that, armed even with just these mechanisms, SAIM is able to accommodate data indicating effects of both object coding and spatial proximity on selection.

SAIM also simulates such effects in a way that differs from some other computational models, though other models do maintain that grouping influences our ability to select parts of stimuli together. For example, in the MORSEL model of object coding and selection (Mozer 1991), grouped elements may be selected together because they jointly activate a stored representation; however grouping between elements does not directly determine the locations where spatial attention falls (indeed in MORSEL high-level groups are abstracted from location information, and so cannot easily provide useful information concerning the positions of their features). SAIM, in contrast, directs spatial attention to cover the area where parts of the same object fall, by top-down feed-back into the selection network. There is evidence in favour of this account. Kim & Cave (1995), for example, had subjects identify a target letter defined by its position. The target appeared along with a distractor of the same color (the grouped distractor) and a distractor of a different color (the non-grouped distractor). Immediately following the display, a probe could be presented. RTs to the onset of the probe were faster when it appeared in the location of the grouped distractor than when it appeared in the location of the non-grouped distractor. Thus, when a group is selected, attention seems to be directed to the locations of its members. This is in accord with the general principles of SAIM, though the model as presented here does not have the capacity to group by color and so cannot simulate these results precisely. SAIM also predicts that object-coding should not only influence spatial attention when stimuli are grouped by bottom-up cues but also when grouping and selection are determined by stored knowledge (even by being the more familiar of two objects). We have examined this in preliminary neuropsychological studies in which we briefly presented two stimuli that differed in goodness' at variable locations. Patients with parietal lesions can select the better of two stimuli, extinguishing the other, irrespective of where the stimuli appear in the field (see Humphreys et al. 1994, Ward & Goodrich 1996, for evidence). We followed the stimuli by two letters and asked patients to identify the letters. There was improved report of letters that fell at the location of the better figure. This is consistent with object-based selection leading to enhanced processing of the selected locations, consistent with SAIM's approach. It is not clear how other models would capture such results.

The contrary case is when objects are spatially overlapping (or adjacent), when spatial selection by means of some undifferentiated spotlight or zoom lens may not be possible (cf. Duncan 1984). Within the framework of SAIM, selection of overlapping figures is possible if the template for one object can be activated selectively, enabling top-down biases to guide selection (Study 7). This process may also be helped further by appropriate scaling of the size of the attentional window, so that parts of each figure are selected over time. The issue of scaling will be examined in future



work. The over-arching point, though is that the distinction between spatial and object-based selection is artificial for SAIM: both types of information interact to determine selection.

2. Spatial representation and attention. In addition to blurring the distinctions between space-based and object-based theories of selection, SAIM also blurs traditional accounts in neuropsychology that maintain that visual neglect is either a disorder of spatial representation or a disorder of attention. We believe this is an important point. For SAIM, spatial selection is a necessary process for translation invariant object recognition, and translation invariant object recognition involves coding the properties of stimuli into different forms of spatial representation (from one coded within a retinal co-ordinate system to one that is dependent on the center of mass in an object). Neglect in the model is indeed a disorder of attention, since lesioning can prevent all of the parts of a stimulus from being attended. However for SAIM neglect is also a disorder of spatial representation, since it disrupts the formation of certain forms of representations of stimuli (e.g., the formation of an object-dependent representation, within the FOA). For SAIM it does not make sense to discuss neglect as either a disorder of spatial representation or attention; it is a disorder of both.
3. Integrating different aspects of the neglect/extinction syndrome. To a disinterested party, syndromes such as unilateral neglect may seem to consist of a multitude of deficits, with the symptoms of different patients appearing unrelated or even frankly contradictory with one another. Examples relevant to the simulations presented here are the dissociations between view-dependent and object-based neglect, the relations between neglect and extinction. These differences in symptoms within the neglect/extinction syndrome may detract from attempts to use neuropsychological data to influence models of normal object recognition and attention. SAIM, however, provides a framework that enables the contrasting symptoms to be understood in relation to one another and in relation to functional components of normal cognition. The model shows that complex neuropsychological disorders can be understood in a systematic manner.

A first example of how SAIM can accommodate different forms of neglect is in our simulations of view-dependent neglect, on the one hand (after a vertical' lesion of the selection network), and object-based neglect on the other (after a horizontal' lesion). For SAIM, the two forms of neglect are a natural consequence of how translation invariance is achieved with activation mapped into the selection network from a retinal co-ordinate system, and mapped out of the selection network (into the FOA) into an object-based co-ordinate system (coding the positions of left and right features relative to the center of mass of an object). These forms of neglect are not contradictory but rather reflect breakdowns at different stages of object selection.

A second example is the relation between visual neglect and extinction. Very often extinction is thought to be a moderate form of neglect. On the other hand, there is evidence that neglect and extinction can doubly dissociate (e.g. Cocchini et al. 1999). In addition to some patients having extinction without neglect (as expected if extinction can occur with a less severe lesion), other patients can manifest neglect of single stimuli but do not have particular difficulties when presented with

multiple items. This dissociation is difficult to explain if there is only a single locus within a model where a spatially sensitive lesion can be effected. SAIM, however, can be lesioned at different loci to generate view- and object-based deficits. The object-based deficit does not necessarily generate a problem in selecting multiple stimuli (Figure 26); there can be neglect without extinction. In contrast, a mild field-dependent deficit will generate extinction without neglect, due to differential competition for selection between stimuli in the contra- and ipsilesional fields. The model provides a natural account of such dissociations.

In addition to the dissociations we have noted, there are in fact several other contrasts between patients that SAIM does not capture. For example, there are patients whose neglect appears to be confined either to near or far space (Cowey et al. 1994, Halligan & Marshall 1991), and there are patients whose neglect is pronounced when motor actions have to be addressed to objects but not when perceptual judgements are made (e.g. Milner & Harvey 1995). Our suspicion is that these additional dissociations will be captured once models begin to link perception to action, allowing the disorders to arise from lesions that translate visual information into representations within various forms of action-coded co-ordinate systems. In future extensions we aim to use representations computed within the location map in SAIM to guide action to stimulus locations, and we will explore whether additional dissociations emerge. Even without these extensions, though, SAIM provides a framework for accounting for dissociations within perceptual forms of neglect.

The existing neuropsychological evidence does not only indicate dissociations between patients, but also associations between symptoms in particular cases. An example of this is the relation between visual neglect and apparent impairments in spatial working memory, where neglect patients may unknowingly select previously attended locations and stimuli in search tasks (Husain et al. 2001). Based on the neuropsychological data alone it is difficult to judge whether such associations are simply due to anatomical accident (a large lesion affecting two independent processes), or whether they are indeed functionally related (see Humphreys & Price 2001, for one recent discussion). Computational models can provide a reasoned account of why such associations arise. SAIM carries an implicit spatial working memory of previously attended locations, based on positions subject to inhibition within its map of locations. Following lesioning, however, the ability to inhibit previously attended locations can be disrupted – a deficit caused by shifts in the locations of stimuli within the FOA, and so associated with neglect. Under these circumstances, SAIM can be drawn back to re-select an object (and spatial locations) it has previously attended. The model operates as if there is a disturbance to its spatial working memory. For SAIM, this behavior is not due to a lesion to a spatial working memory system that is independent of the system concerned with spatial selection; rather it is an emergent consequence of the disorder in spatial selection.

### 4.3 Relations between SAIM and other models of visual selection

Psychological models of visual selection have traditionally been based on a two-stage process in which early pre-attentive processes, which are spatially parallel, are followed by later attention-dependent processes, which are spatially serial (Neisser 1967, Treisman

1988, Treisman & Gormican 1988). The serial, attentional processes have been conceptualised in terms of 'spotlight' and 'zoom lens' analogies. In many models (e.g., Treisman's Feature Integration Theory (FIT), Treisman (1988, 1998), attention is required in order to bind the parts of objects together, and this process must take place before objects contact stored memories. SAIM has a general architecture that is consistent with this distinction, between spatially parallel and then spatially serial processes. Initial activation within both the selection and the contents networks takes place in parallel across an image and irrespective of the number of objects present. This is followed by a serial stage, in which only a single object is selected. One computational reason for the serial selection process in SAIM is to facilitate binding of features within objects, since selection prevents competition from features in other objects from being available for binding. In addition to this, SAIM, like FIT, binds features by dint of their activating stored memories (for SAIM templates that specify the positions of parts relative to the center of gravity of the shape). However SAIM differs from traditional models in that the pre-attentive and attentional stages for the model are not serial and discrete, rather both pre-attentive and attentive states emerge together, in parallel over time. Also there is partial activation of stored knowledge from multiple objects as the model selects stimuli, and this knowledge is used interactively to facilitate selection. For SAIM, parts of an object that are initially attended help to determine how subsequent pre-attentive processing operates (biasing selection towards other parts of the same object). SAIM is an interactive, two-stage model, and does not mimic either strict early- or strict late-selection.

SAIM models visual selection in terms of competitive interactions between stimuli, with these interactions being biased by several factors: the size and perceptual goodness of objects (at least as defined in terms of the model; e.g., how well parts group by proximity), how much objects activate stored templates, the familiarity of the templates, spatial biases (e.g., after lesioning). These over-arching principles of selection via competition are common across several models of visual attention (e.g. Bundesen 1990, 1998, Humphreys & Müller 1993, Mozer 1991), and they are embodied in general accounts of selection such as the integrated competition approach (Desimone 1998, Duncan 1998, Duncan et al. 1997). The particular forms of this competition, however, are of course specific to the model and the way in which the selection network operates.

As we pointed out in the Introduction, there have been several other attempts to develop detailed computational models of visual selection, some of which have been designed around specific experimental tasks (Cohen et al. 1994, Humphreys & Müller 1993), some of which have been applied to a range of experimental procedures (e.g., the MORSEL model, Mozer (1991), Mozer & Sitton (1998), and some of which have been applied to neurophysiological data (Olshausen et al. 1993, Usher & Niebur 1996, Koch & Ullman 1985). The most wide-ranging model, as far as psychological applications are concerned, is MORSEL. MORSEL has two main parts. The first is an object recognition system, which operates in a hierarchically organized fashion. Initial units in the system are activated according to the locations where stimuli fall on the retina. These units also encode specific features of objects (e.g., edges of a given orientation), and so extend the simple pixel-coding procedure used by SAIM. Units at higher levels in the object recognition system pool activation from units at lower levels responding to the same features at different locations, so that gradually activation becomes viewpoint-independent. At the highest level of the model, stored representations are based on distributed coding of viewpoint-independent features. A 'pull-out net' serves to clean-up activation, so that degraded

inputs can be recovered; this will favour known over unknown objects. However, when there are multiple objects in the field, the model encounters problems of binding, since activated features of different objects may be linked together via the pull-out network. The second part of the system is an attentional network. Units in this network gate activation entering into the object recognition network, so that it is biased to favour attended objects. This reduces the binding problem. Units in the attentional network are retinotopically organized, and they operate in a winner-take-all fashion so that attention tends to become focussed on one location (though the size of this focus can vary, since self-supporting neighbourhoods are used). Attention units are activated either in a bottom-up fashion (exogenously), according to the strength of activation of features units at each retinal location, and in a top-down fashion (endogenously), by pre-setting some units to be active at an attended region of field.

MORSEL, like SAIM, has been used to simulate a wide range of psychological phenomena including spatial cueing and forms of neuropsychological disorder such as visual neglect (e.g. Mozer 1991, Mozer et al. 1997). It undoubtedly provides a useful framework for conceptualizing interactions between object recognition and visual attention, and, more than SAIM, it has been extended to enable patterns of data on human visual search to be simulated (Mozer & Sitton 1998). However, we suggest that SAIM's linkage between translation-invariant coding and attention provides a more natural account of dissociations within neuropsychological disorders such as neglect. As discussed in Study 11, MORSEL has some problems in explaining dissociations between neglect within- and neglect between-objects. Within MORSEL spatially graded lesions, across the attentional network, can generate deficits that are more severe on one side of an object than the other irrespective of where the object appears in the field, capturing a form of object-based neglect. However, it may be problematic for this account to explain how, with a graded lesion, a complete left side object would be neglected whilst there is no neglect of the parts within a right side object (in between-object neglect; see Humphreys & Heinke (1998). Finally, cases in which there is apparently neglect between objects on one side of space accompanied by within- object neglect on the other side seem especially puzzling (cf. Humphreys & Riddoch 1994, 1995). For similar reasons, there can be problems in explaining double dissociations between neglect and extinction. Extinction could arise under conditions in which the lesion is not so severe as to produce neglect, but, when deficits arise from lesioning a single site (the attentional module, in MORSEL), a lesion producing neglect should also lead to problems when multiple stimuli are present. The neuropsychological data indicate that this is not always the case (Cocchini et al. 1999). Nevertheless there are means by which MORSEL can simulate object-based effects in neglect. Take as one example a study by Behrmann and Tipper (Behrmann & Tipper (1994, 1999), Tipper & Behrmann (1996) see also Driver & Halligan (1991). Behrmann and Tipper presented patients with a barbell stimulus that rotated in the plane so that the original left-side part fell in the right field (and vice versa for the original right- side part). They found that right hemisphere lesioned patients were slowed to targets appearing in the right-side (ipsilesional) part, after the barbell had been rotated. The same result did not occur when the parts were not joined together to make a barbell; performance then remained worse for left-side targets even after rotation. To mimic this situation, Mozer (1999) presented stimuli dynamically (at one location for a set number of iterations, followed by the next location and so on). When lesioned, MORSEL initially allocated more attention to the right relative to the left side of the bar bell. Interestingly, due to the

dynamics of the attentional network, activation was maintained on the same (right) side of the bar bell as the stimulus was rotated so that its right side gradually fell into the left field (likewise, the left side of the bar bell remained relatively inactive when rotated to fall into the right [good] field). This demonstrates that a view-specific attentional system is capable of generating apparent object-based effects, based on a form of dynamic tracking of attention to an object, over time.

However, even without using tracking procedures (and even with static stimuli) SAIM can simulate many of the results that support the existence of both view- and object-based neglect. SAIM captures such effects because the selection network is dependent on both the positions of objects in the visual field and the positions of their parts within an object-dependent FOA. These dependencies are embodied in different sets of connections within the selection network, making it vulnerable to contrasting forms of lesioning. If the mappings from one side of the visual field into the selection network, and from the selection network into one side of the FOA, are affected by damage in one hemisphere, then a unilateral lesion will tend to produce a deficit on one side of view-dependent or object-dependent space; bilateral lesions may produce deficits on opposite sides of space if they affect different representations.

#### 4.4 Emergent properties of SAIM

There are several properties of SAIM that emerge not by being specifically programmed but because of the way in which processing operates within the model. These emergent properties capture aspects of psychological performance, and also make predictions for future empirical tests of the model. We highlight five examples.

First, as discussed in the last section, SAIM tends to place its center of attention at the location of the center of mass of a shape. This characteristic arises out of the excitatory interactions that operate within the selection network, which give most support to pixels at the location of the center of mass. Though not an explicit aim of the model, this property has useful psychological application.

Second, SAIM shows a time course of selection in which some parts of an object are initially given more weight than others: locations at the center of the textured mass of a shape are activated first within the FOA, followed by those further from the center of this location. This predicts that attention should not be captured by all features within an object simultaneously, a prediction that we are currently exploring empirically. However, SAIM also groups the parts of an object together so that all the parts tend to fall within the FOA (albeit over a certain time course). It can thus be applied to evidence showing that multiple features of objects can be selected together (see Duncan 1984). Indeed, once selected, parts within the same object may not be ignored, even if they are detrimental to task performance; this meshes with evidence that grouping can sometimes be detrimental to performance, when the task requires selection of a member of a group (e.g. Pomerantz 1981, Rensink & Enns 1995). SAIM predicts that there should be selection of all object properties together, but over a systematic time course.

Third, SAIM manifests general costs to selection when there are multiple items in the field. These costs occur even when a distractor is successfully rejected in the competition for selection, and even when it does not contain features that should strongly activate a target template. Psychological evidence for generic costs on visual selection from the presence of 'non-competitive' stimuli in the field comes from studies of 'filtering costs'

(Eriksen & Hoffman 1972, Treisman et al. 1983). SAIM shows how such costs can arise within a competitive selection system.

Fourth, like other competitive models of visual attention (Cohen et al. 1994, Humphreys et al. 1996b), SAIM demonstrates the emergent property of 'impaired disengagement' of attention after it is lesioned and when it is given an invalid spatial cue. It is clear that models do not need a specific module for attentional disengagement for this process to take place, and for it to break down in a way that mimics human performance. In competitive models, impaired disengagement of attention is a natural consequence of unbalanced competition for selection, when a salient precue increases the unbalance that already exists due to a spatially selective brain lesion. In fact, SAIM shows two forms of attentional disengagement in its behaviour. The first, which is similar to previous demonstrations, involves a delay of shifting attention to a target on impaired side of space after a cue has been presented on the ipsilesional side. Under short exposure conditions, this delay will result in spatial extinction. A second disengagement problem, however, arises when SAIM fails to inhibit representations activated by a target first selected. This arises because there is spatial distortion in the lesioned version of the model even when stimuli are attended (see Figure 27). This second form of disengagement problem was not a planned property of the model, but, as we have noted above, it can be linked to apparent deficits in working memory in search tasks. In addition, it is through this property that SAIM can neglect of whole-objects that fall to the contralesional side of other objects in the field, as is found in between-object neglect.

Fifth, SAIM was designed to produce translation-invariant object recognition. However, the final stages of object recognition in the model are not view-invariant, in the manner that theories such as those of Biederman (1987) and Marr (1982) are. Those theories involve the extraction of invariant properties of objects (non-accidental features, features relations relative to the main axis of an object), that can then be matched to stored representations to enable objects to be identified from any viewpoint. The representation itself may be view-independent. This contrasts with the templates for objects in SAIM, which are coded for view; objects whose parts fall in positions within the FOA that do not match the positions specified in the template will be identified less well. The representation mapped into the FOA may be thought to have some similarities with Marr's suggestion of an axis-based description for invariant recognition, in that the representation is focused on the center of mass in a shape. However, SAIM has no easy means of dealing with rotation, either in depth or in the plane. The co-ordinates of its attended representation are determined by how parts fall on the retina in relation to the center of mass. Rotating an object will change the relations of the parts to the center of mass of the object on the retina, disrupting recognition. In this sense, the object recognition part of SAIM is closer to template-matching models of object recognition than to view-invariant models. The topic of template vs. view-invariant procedures in human pattern recognition has been hotly debated over the past few years (see Biederman & Kalocsai 1997, Edelman & Duvdevani-Bar 1997) and runs tangential to SAIM's main concern with visual selection. Nevertheless, some properties of SAIM's performance are a consequence of its view-specific approach. One, that we have discussed above, is its problem in inhibiting attended objects whose representations are distorted within the FOA. A second relates to the form of within-object neglect observed after the model is lesioned. This form of neglect means that parts that fall on the contralesional side of objects will be neglected irrespective of the left-right positions of objects in the field

(see Study 9). However, if the object is inverted, the former contralesional parts will then fall on the ipsilesional side of the center of the textured mass of the object and the former ipsilesional parts will fall on the contralesional side. SAIM predicts that the parts neglected across these occasions will change, in that the contralesional features as the object falls on the retina will always be the ones that are neglected. This matches several reports in the neuropsychological literature of patients who can show neglect of left-side features of objects in the right field, but who continue to show neglect of the left parts of objects as they fall in the field when the objects are inverted (Humphreys & Riddoch 1995, Young et al. 1992). SAIM uses a form of object-based representation, but one in which parts remain coded for their retinal positions with respect to the center of mass of the object. With this representation the model is able to capture data on both translation-invariant neglect and on the effects of object center of mass on neglect (e.g. Pavlovskaya et al. 1997, see Study 13). We need to be cautious when using these last results to argue for the involvement of full object-centred representations in object recognition and neglect.

#### 4.5 Factors not covered

Not surprisingly, given the wealth of data on human visual selection, SAIM fails to capture many findings in the field. For example, the model does not simulate the observation that patients with left neglect can bisect very short lines slightly to the left (rather than the normal 'neglect' pattern in which bisections are shifted to the right, see Marshall & Halligan (1990)). At present, SAIM predicts that there should either be no neglect for short lines or, as with longer lines, bisection responses should be shifted to the ipsilesional side. A similar pattern to that produced by SAIM was generated by MORSEL when lesioned (see Mozer et al. 1997). Mozer et al. proposed that the contralateral (left) shifts with very small lines might reflect a constant bias, found even with normal subjects, towards the left side. Consequently, when the model shows minimal neglect (with short lines), bisection tends to veer to the left. It is possible for a similar assumption to be built into SAIM, although it would stand outside the current framework.

The model also does not deal with factors such as the hemispace of stimulus presentation, that can affect the performance of neglect patients (Karnath et al. 1991, Riddoch & Humphreys 1983). Such effects suggest that the positions of stimuli with respect to a patient's body can be more important than the position of the line on the retina. It may be that the form of coding within the selection network might need to be elaborated, to allow body or head position to influence mapping, as well as the positions of objects on the retina. Similarly, there may be input from sensory input other than vision, and this may explain why forms of cross-modal extinction and neglect can arise (see Mattingley et al. 1997). This is beyond the scope of the present model.

Other pieces of neuropsychological evidence not simulated here include the apparently 'true' object-centred forms of neglect found when patients neglect parts according to their positions relative to an arbitrary part of the object, and so neglect these parts even when objects are rotated or inverted. For example, Caramazza & Hillis (1990) reported a patient who showed neglect of the right end letters in words when reading. When the words were rotated so that these right end letters now fell at the left end of the words on the retina, neglect of the same letters (now on the left) was found. As noted above, this is not the result found in other cases, and it is not the result that SAIM would generate (ne-

glect would occur on the new features now falling at the right end of the inverted word). Similarly, SAIM does not easily explain the findings reported by Behrmann & Tipper (1994, 1999), Tipper & Behrmann (1996) (see above), when stimuli were rotated from one position into another. For SAIM some form of correction for small rotations could occur within the FOA if top-down knowledge is allowed to play a large part in mapping retinal input to stored knowledge (e.g., top-down activation would support the features of objects in the positions they would occupy if the objects were upright), but this may also lead to too strong top-down modulation of processing in other circumstances. Additional possibilities are to (i) incorporate some form of dynamic tracking of input, as done in MORSEL, at least when objects move so that formerly left parts shift into the right field (cf. Behrmann & Tipper 1994, 1999), or (ii) add in some form of mental rotation process, so that objects are first transformed into an upright orientation and then identified. There is some suggestion that this strategy may be adopted by patients whose neglect shifts with the positions of the features in the objects (Buxbaum et al. 1996). This remains to be examined in full in the context of the model.

Other factors not incorporated into the present version of SAIM are: the use of more sophisticated grouping procedures (e.g., using collinearity, closure, connectedness), the processing of features other than intensity and shape (e.g., color, texture, depth, motion), and the instantiation of endogenous as well as exogenous attention.. There is abundant psychological and neuropsychological evidence for more complex shape features playing a role in early grouping processes (Donnelly et al. (1991), Elder & Zucker (1993), Gilchrist et al. (1996, 1997), Kovacs & Julesz (1993); similarly grouping may also operate on feature values such as color, texture, depth and motion (e.g. Nakayama & Silverman 1986, McLeod & Driver 1993). By incorporating processes sensitive to these additional factors, SAIM should be able to provide an even broader account of human visual selection. In addition, a full account of human visual selection will need to incorporate procedures for endogenous as well as exogenous orienting of attention. Experimental, and neuropsychological data converge in suggesting that there are distinct procedures involved in the voluntary as opposed to the automatic control of attention (e.g. Jonides 1981, Jonides & Yantis 1988, Müller & Rabbit 1989, Rafal & Robertson 1995). Sustained activation to attended locations within the saliency map may have the effect of biasing selection in a top-down fashion, so that endogenous as well exogenous effects can be captured. Work is underway to implement these changes in an updated version of the model.

## 4.6 Biological plausibility

As we have noted, SAIM uses processes similar to the dynamic routing circuit proposed by Olshausen et al. (1993, 1995) to achieve translation invariant pattern recognition, whilst extending that model by incorporating features such as top-down modulation of selection and a saliency map. Olshausen et al. argue for the biological plausibility of their model. They suggest that aspects of the pattern recognition system (in SAIM, the knowledge network) mimic cells in inferotemporal cortex, concerned with representating the properties of known visual shapes. Such cells have large receptive fields (e.g. Gross et al. 1972), and so respond irrespective of where critical stimuli appear on the retina. A similar lack of tuning to retinal position exists for template units here. Also, though the templates in SAIM are translation- invariant, they are sensitive to the spatial positions of parts from a particular vantage point. Hence SAIM is sensitive to view angle. Again this is consistent



with many cells within the inferotemporal cortex (Tanaka 1993).

Olshausen et al. argue that the selection network corresponds to the pulvinar, a sub-cortical structure that is richly connected to areas of both ventral and dorsal cortex and which may play a critical role in modulating cortical activation. Consistent with this, studies of the effects of lesioning the pulvinar in both monkey (Desimone & Ungerleider 1989) and in humans (Rafal & Posner 1987), are associated with problems in visual selection. The location map in the model may be associated with the posterior parietal cortex, which holds a spatial map of putative objects in the visual field (see Gottlieb et al. 1998). In SAIM this map will not specify the identities of stimuli, only their locations. This is consistent with the general distinction between a ventral visual system concerned with object recognition (in SAIM the contents and knowledge networks) and a dorsal system concerned with spatial vision (the selection network and the location map) (Ungerleider & Mishkin 1982). There is also neuropsychological evidence that the parietal lobe is involved in coding a small number of objects but without specifying their identities. For instance, patients with ventral brain lesions that spare the parietal lobe can judge rapidly (in parallel) small numbers of objects present in the visual field but they are impaired at coding stimulus identities in the same manner; in contrast, parietal lesions leave intact the parallel coding of stimulus identity but not number (Dehaene & Cohen 1994, Humphreys & Heinke 1998).

In their analysis of patients with neglect between objects and neglect within objects, Humphreys & Heinke (1998) used data from lesion overlap to suggest that neglect between objects was caused by posterior parietal lesions whilst neglect within objects was linked to more anterior dorsal lesions. These arguments were based on small numbers of patients, and so caution must be effected in any judgements; nevertheless such a proposal is consistent with the idea that input from separate objects generates the initial competition in the selection network, and with outputs from the network being passed to more anterior dorsal areas (which perhaps represent explicit spatial representations of currently attended visual information; the FOA). Deficits in the posterior parietal cortex could also lead to selection becoming 'stuck' on previously attended objects, creating neglect between objects as we have shown (Figure 27). The neuroanatomical distinctions made by Humphreys and Heinke fit with the general idea that contrasting lesions are linked to the different forms of neglect in SAIM.

Attentional operations in SAIM can also be linked to physiological data on visual selection. For example, the effect of mapping one area of visual field into the FOA is to 'tune' the receptive field of template units so that they no longer respond to stimuli across the retina but only to objects falling within the attended region of space. This is akin to the findings of Moran & Desimone (1985), who first reported that cells in area V4 of monkey cortex responded differentially according to whether the animal attended to the location of a target object. It was as if the receptive field of the cell shrank to fit only the attended area. We have also shown that 'priming' a template unit biases selection towards a stimulus consistent with the unit (Study 2), so that this stimulus can now win the competition for selection over another item that would otherwise have done so. This mimics data on attentional priming in area IT. Chelazzi et al. (1993) found that a precue priming activity in cells in IT, so that these cells then fired more rapidly when a target was subsequently presented along with a distractor. These authors suggested that the precue provided a competitive bias favouring cells tuned to the target over those tuned to the distractor. Template pre-activation in SAIM produces just this effect.

Our aim in implementing SAIM was not necessarily to provide a model that is accurate in all respects at a neural level, and there are several aspects of SAIM that are neurally implausible. For example, an enlarged version of the model, still using only a single scale for object representation, may require far more connections than is feasible in the brain. Also, in the version used in the present simulations the FOA was of a fixed (and limited) retinal size. This is not realistic since, within some limits, people can attend to objects of varying sizes. Our aim, however, was to show that, even with these constraints, some of the basic assumptions of SAIM are useful for capturing selective processing in human vision (particularly competition between units which seek to transform retinal input to a representation that is translation invariant). The constraints due to the number of connections needed, and the fixed size of the FOA, may be overcome by future work in which multi-scale processing is incorporated and in which there are consequently means of shifting attention in scale as well as in space. The consequences of this require further, detailed experimentation. For now, the simulations suggest that processes involved in selection for translation invariant recognition capture many aspects of human attention in both normality and pathology.

## **Acknowledgements**

This work was supported by grants from the European Union to both authors and from the Medical Research Council (UK) to the second author. We thank two anonymous reviewers and Mike Mozer for their detailed comments.

## References

- Allport, D. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In Heuer, H. & Sanders, A., editors, *Perspectives on perception and action*, pages 395–419. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Amari, S. (1982). Competitive and Cooperative Aspects In Dynamics of Neural Excitation and Self-Organization. In Amari, S. & Arbib, M. A., editors, *Competition and Cooperation in Neural Nets*, pages 10–30.
- Baylis, G. & Driver, J. (1993). Visual Attention and Objects: Evidence for Hierarchical Coding of Location. *Journal of Experimental Psychology: Human Perception and Performance*, 19(3):451–470.
- Baylis, G. C. & Driver, J. (1992). Visual parsing and response competition: The effects of grouping. *Perception & Psychophysics*, 51:145–162.
- Behrmann, M. & Tipper, S. P. (1994). Object-based attentional mechanisms: Evidence from patients with unilateral neglect. In C., U. & M., M., editors, *Attention and performance XV*. Cambridge, Mass.: MIT Press.
- Behrmann, M. & Tipper, S. P. (1999). Attention Accesses Multiple Reference Frames: Evidence from Visual Neglect. *Journal of Experimental Psychology: Human Perception and Performance*, 25:83–101.
- Behrmann, M., Zemel, R. S., & Mozer, M. C. (1998). Object-based attention and occlusion: Evidence from normal participants and a computational model. *Journal of Experimental Psychology: Human Perception and Performance*, 24:1011–1036.
- Biederman, I. (1987). Recognition-by-components: A Theory of Human Image Understanding. *Psychological Review*, 94(2):115–147.
- Biederman, I. & Kalocsai, P. (1997). Neurocomputational bases of Object and Face Recognition. *Philosophical Transactions of the Royal Society: Series B*, 352:1203–1220.
- Bisiach, E., Capitani, E., Colombo, A., & Spinnler, H. (1976). Halving a horizontal segment: A study on hemisphere-damaged patients with focal cerebral lesions. *Archives Suisses de Neurologie et de Psychiatrie*, 118:199–206.
- Bisiach, E. & Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex*, 14:129–133.
- Brunn, J. L. & Farah, M. J. (1991). The relation between spatial attention and reading: Evidence from the neglect syndrome. *Cognitive Neuropsychology*, 8:59–75.
- Buck, B. H., Black, S. E., Behrmann, M., Caldwell, C., & Bronskill, M. J. (1999). Spatial- and object-based attentional deficits in Alzheimer's disease: Relationship to SPECT measures of parietal performance. *Brain*, 120:1229–1244.
- Bundesen, C. (1990). A Theory of Visual Attention. *Psychological Review*, 97(4):523–547.

- Bundesen, C. (1998). A computational theory of visual attention. *Philosophical Transactions of the Royal Society*, 353:1271–1282.
- Buxbaum, L. J., Coslett, H. B., Montgomery, M. W., & Farah, M. J. (1996). Mental rotation may underlie apparent object-based neglect. *Neuropsychologia*, 34:113–126.
- 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.
- Cave, K. R. & Wolfe, J. (1990). Modeling the role of Parallel Processing in Visual Search. *Cognitive Psychology*, 22:225–257.
- Cheal, M., Lyon, D. R., & Gottlob, L. R. (1994). A Framework for Understanding the Allocation of Attention in Location-precued Discrimination. *The Quarterly Journal of experimental psychology*, 47A(3):699–739.
- Chelazzi, L., Miller, E., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, 363:345–347.
- Cocchini, G., Cubelli, R., DellaSala, S., & Beschin, N. (1999). Neglect without extinction. *Cortex*, 35:285–313.
- 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.
- Coltheart, M. (1983). Iconic memory. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 302:283–294.
- Corbetta, M., Shulman, G., Miezin, F., & Petersen, S. (1995). Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science*, 270:802–805.
- Costello, A. D. & Warrington, E. K. (1987). The dissociation of visual neglect and neglect dyslexia. *Journal of Neurology, Neurosurgery and Psychiatry*, 50:1110–1116.
- Cowey, A., Small, M., & Ellis, S. (1994). Left visuo-spatial neglect can be worse in far than in near space. *Neuropsychologia*, 32:1059–1066.
- Cubelli, R., Nichelli, P., Bonito, V., De Tanti, A., & Inzaghi, M. G. (1991). Different patterns of dissociation in unilateral neglect. *Brain and cognition*, 15:139–159.
- Deco, G. & Zihl, J. (2001). Top-down selective visual attention: A neurodynamical approach. *Visual Cognition*, 8(1):119–140.
- Dehaene, S. & Cohen, L. (1994). Dissociable mechanisms of subitizing and counting: Neuropsychological evidence from simultanagnosia patients. *Journal of Experimental Psychology: Human Perception and Performance*, 20:958–975.

- Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society*, pages 1245–1256.
- Desimone, R. & Duncan, J. (1995). Neural mechanisms of selective attention. *Annual Review of Neuroscience*, 18:193–222.
- Desimone, R. & Ungerleider, L. (1989). Neural mechanisms of visual processing in mokeys. In *Handbook of Neurophysiology*, volume 2, chapter 14, pages 267–299. Elsevier.
- Donnelly, N., Humphreys, G. W., & Riddoch, M. J. (1991). Parallel computation of primitive shape descriptions. *Journal of Experimental Psychology: Human Perception and Performance*, 17:561–570.
- Driver, J., Baylis, G. C., & Rafal, R. D. (1992). Preserved Figure-Ground Segmentation in Visual Matching. *Nature*, 360:73–75.
- Driver, J. & Halligan, P. W. (1991). Can visual neglect operate in object-centred coordinates? An affirmative single case study. *Cognitive Neuropsychology*, 8:475–496.
- Driver, J. & Mattingley, J. B. (1995). Normal and pathological selective attention in humans. *Current Opinion in Neurobiology*, 5:191–197.
- Driver, J. & Pouget, A. (2000). Object-centred visual neglect, or relative egocentric neglect? *Journal of Cognitive Neuroscience*, 12:542–545.
- Duhamel, J.-R., Carol, L. C., & Goldberg, M. E. (1992). The Updating of the Representation of Visual Space in Parietal Cortex by Intended Eye Movements. *Science*, 255:90–92.
- 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. (1998). Converging levels of analysis in the cognitive neuroscience of visual attention. *Philosophical Transactions of the Royal Society*, 353:1307–1318.
- Duncan, J., Bundesen, C., Olson, A., Humphreys, G. W., Chavda, S., & Shibuya, H. (1999). Systematic analysis of deficits in visual attention. *Journal of Experimental Psychology: General*, 128.
- Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology*, 7:255–261.
- Edelman, S. & Duvdevani-Bar, S. (1997). A Model of Visual Recognition and Categorization. *Philosophical Transaction of the Royal Society: Series B*, 352:1191–1202.
- 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.

- Elder, J. & Zucker, S. (1993). The effect of contour closure on the rapid discrimination of two-dimensional shapes. *Vision Research*, 33:981–991.
- Ellis, R. & Humphreys, G. W. (1999). *Connectionist psychology: A text with readings*. London: Psychology Press.
- Eriksen, C. W. & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics*, 12:201–204.
- 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.
- Farah, M. J., Monheit, M. A., & Wallace, M. A. (1991). Unconscious perception of "extinguished" visual stimuli: Re-assessing the evidence. *Neuropsychologia*, 29:949–958.
- Feldman, J. A. & Ballard, D. H. (1982). Connectionist Models and Their Properties. *Cognitive Science*, 6:205–254.
- Findlay, J. (1982). Global visual processing for saccadic eye movements. *Vision Research*, 22:1033–1045.
- Gaffan, D. & Hornak, J. (1997). Visual neglect in the monkey: Representation and disconnection. *Brain*, 120:1647–1657.
- Gainotti, G., d'Erme, P., Monteleone, D., & Silveri, M. C. (1986). Mechanisms of unilateral spatial neglect in relation to laterality of cerebral lesions. *Brain*, 109:599–612.
- Gibson, B. & Egeth, H. (1994). Inhibition of Return to Object-Based and Environment-Based Locations. *Perception and Psychophysics*, 55:323–339.
- Gilchrist, I. D. and Humphreys, G. W., Riddoch, J., & Neumann, H. (1997). Luminance and Edge Information in Grouping: A Study Using Visual Search. *Journal of Experimental Psychology: Human Perception and Performance*, 23(2):464–480.
- Gilchrist, I., Humphreys, G. W., & Riddoch, M. J. (1996). Grouping and Extinction: Evidence for Low-Level Modulation of Selection. *Cognitive Neuropsychology*, 13:1223–1256.
- Glendinning, P. (1995). *Stability, Instability and Chaos: an introduction to the theory of nonlinear differential equations*. Cambridge University Press.
- Gorea, A. & Sagi, D. (2000). Failure to handle more than one internal representation in visual detection tasks. *Proceedings of the National Academy of Sciences*, 97:12380–12384.
- Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, 391:481–484.

- Grabowecky, M., Robertson, L. C., & Treisman, A. (1993). Preattentive Process Guide Visual Search: Evidence from Patients with Unilateral Visual Neglect. *Journal of Cognitive Neuroscience*, 5(3):288–302.
- Grice, G. R., Canham, L., & Boroughs, J. M. (1983). Forest before trees? It depends on where you look. *Perception & Psychophysics*, 33:121–128.
- Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex. *Journal of Neurophysiology*, 35:96–111.
- Halligan, P. W. & Marshall, J. C. (1988). How long is a piece of string? A study of line bisection in a case of visual neglect. *Cortex*, 24:321–328.
- Halligan, P. W. & Marshall, J. C. (1991). Spatial Compression in Visual Neglect: A Case Study. *Cortex*, 27:623–629.
- Han, S., Humphreys, G. W., & L., C. (1999). Uniform connectedness and classical Gestalt principles of perceptual grouping. *Perception & Psychophysics*, 61(4):661–674.
- Heilman, K. M. & Valenstein, E. (1979). Mechanisms underlying hemispatial neglect. *Annals of Neurology*, 5:166–170.
- 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. (in press). Computational Models of Visual Selective Attention: A Review. In Houghton, G., editor, *Connectionist Models in Psychology*. Psychology Press.
- Hinton, G. E. (1981a). A parallel computation that assigns canonical object-based frames of reference. In *Proceedings of the seventh Internal Joint Conference on Artificial Intelligence*, pages 683–685.
- Hinton, G. E. (1981b). Implementing semantic networks in parallel hardware. In G. E., H. & J. A., A., editors, *Parallel models of associative memory*, pages 161–188. Hillsdale, NJ: Erlbaum.
- Hinton, G. E. & Lang, K. J. (1985). Shape Recognition and Illusory Conjunctions. In *Proc. of 9th IJCAI*.
- Hirsch, J. & Mjolsness, E. (1992). A Center-of-Mass Computation Describes the Precision of Random Dot Displacement Discrimination. *Vision Research*, 32(2):335–346.
- 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.
- Hopfield, J. J. & Tank, D. (1985). "Neural" Computation of Decisions in Optimization Problems. *Biological Cybernetics*, 52:141–152.

- Hummel, J. E. & Biederman, I. (1992). Dynamic Binding in a Neural Network for Shape Recognition. *Psychological Review*, 99(3):480–517.
- Humphreys, G. W. (1983). Frames of reference and shape perception. *Cognitive Psychology*, 15:151–196.
- Humphreys, G. W. (1998). Neural representation of objects in space: A dual coding account. *Philosophical Transactions of the Royal Society*, 353.
- Humphreys, G. W., Boucart, M., Datar, V., & Riddoch, M. J. (1996a). Processing of fragmented forms and strategic control of orienting in visual neglect. *Cognitive Neuropsychology*, 13:177–203.
- Humphreys, G. W., Freeman, T. A. C., & Müller, H. M. (1992). Lesioning a Connectionist Model of Visual Search: Selective Effects on Distractor Grouping. *Canadian Journal of Psychology*, 46:417–460.
- Humphreys, G. W. & Heinke, D. (1997). Representing and attending to visual space: A computational perspective on neuropsychological problems. In *4th Neural Computation and Psychology Workshop Connectionist Representations: Theory and Practice*, pages 99–112, University of London, England.
- 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. (1996b). Competitive mechanisms of selection by space and object: A neuropsychological approach. In A. F., K., M. G. H., C., & G. D., L., editors, *Converging operations in the study of visual attention*, page ?? Washington, DC: American Psychological Association.
- Humphreys, G. W. & Price, C. J. (2001). Cognitive neuropsychology and functional brain imaging: Implications for functional and anatomical models of cognition. *Acta Psychologica*, 107:119–153.
- Humphreys, G. W. & Quinlan, P. T. (1987). Visual object processing: A cognitive neuropsychological approach. London: Lawrence Erlbaum Associates.
- 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. (1995). Separate coding of space within and between perceptual objects: Evidence from unilateral visual neglect. *Cognitive Neuropsychology*, 12:283–312.



- Humphreys, G. W., Romani, C., Olson, A., Riddoch, M. J., & Duncan, J. (1994). Non-spatial extinction following lesions of the parietal lobe in humans. *Nature*, 372:357–359.
- Husain, M. & Kennard, C. (1996). Visual neglect associated with frontal lobe infarction. *Journal of Neurology*, 243:652–657.
- Husain, M., Mannan, S., Hodgson, T., Wojciulik, E., Driver, J., & Kennard, C. (2001). Impaired spatial working memory contributes to abnormal search in parietal neglect. *Brain*, 124.
- Husain, M., Shapiro, K., Martin, J., & Kennard, C. (1997). Abnormal temporal dynamics of visual attention in spatial attention in humans. *Nature*, 385:154–156.
- Irwin, D. E., Colcombe, A. M., Kramer, A. F., & Hahn, S. (2000). Capture of the eyes by color and onset singletons. *Vision Research*, 40:1443–1458.
- Jonides, J. (1981). Voluntary vs. automatic control over the mind's eye's movement. In Long, J. B. & Baddeley, A. D., editors, *Attention and performance IX*, pages 187–204. Hillsdale, NJ: Erlbaum.
- Jonides, J. & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, 43:346–354.
- Kahneman, D. & Treisman, A. (1983). Changing views of attention and automaticity. In R., P., R., D., & J., B., editors, *Varieties of attention*. New York: Academic Press.
- 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.
- Kim, M.-S. & Cave, K. R. (1995). Spatial attention in visual search for features and feature conjunctions. *Psychological Science*, 6:376–380.
- Kimchi, R. (1994). The role of wholistic/configural properties versus global properties in visual form perception. *Perception*, 23:489–504.
- Klein, R. (1988). Inhibitory tagging system facilitates visual search. *Nature*, 334:430–431.
- Koch, C. & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4:219–227.
- Kovacs, I. & Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Science*, 90:7495–7497.
- Kramer, A. F., Weber, T. A., & Watson, S. E. (1997). Object-based attentional selection-grouped arrays or spatially invariant representations?: Comment on vecera and farah (1994). *Journal of Experimental Psychology: General*, 126:3–13.

- Kumada, T. & Humphreys, G. W. (2001). Lexical recovery on extinction: Interactions between visual form and stored knowledge modulate visual selection. *Cognitive Neuropsychology*, 18(5):465–478.
- Ladavas, E., Petronio, A., & Umiltà, C. (1990). The deployment of visual attention in the intact field of hemineglect patients. *Cortex*, 26:307–317.
- Luck, S. J., Hillyard, S. E., Mangun, G. R., & Gazzaniga, M. S. (1994). Independent attentional scanning in the separated hemispheres of split-brain patients. *Journal of Cognitive Neuroscience*, 6:84–91.
- Lupianez, J. & Weaver, B. (1998). On the time course of exogenous cueing effects: A commentary on Tassinari et al. *Vision Research*, 38.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.
- Marr, D. & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, 194:283–287.
- Marshall, J. C. & Halligan, P. W. (1990). Line bisection in a case of visual neglect: Psychophysical studies with implications for theory. *Cognitive Neuropsychology*, 7:107–130.
- Marshall, J. C. & Halligan, P. W. (1994). The yin and yang of visuo-spatial neglect: A case study. *Neuropsychologia*, 32:1037–1057.
- Martin, M. (1979). Local and global processing: The role of sparsity. *Memory and Cognition*, 7:479–484.
- Mattingley, J., Driver, J., Beschwin, N., & Robertson, I. H. (1997). Attentional competition between modalities: Extinction between touch and vision after right hemisphere damage. *Neuropsychologia*, 35:867–880.
- Maylor, E. (1985). Facilitatory and Inhibitory Components of Orienting in Visual Space. In *Attention and Performance XI*. Hillsdale, N.J.:Lawrence Erlbaum Association.
- McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88:375–407.
- McGowan, J. W., Kowler, E., Sharma, A., & Chubb, C. (1998). Saccadic Localization of Random Dot Targets. *Vision Research*, 38(6):895–909.
- McLeod, P. & Driver, J. (1993). Filtering and physiology in visual search: A convergence of behavioural and neurophysiological measures. In D., B. A. & Weiskrantz, L., editors, *Attention: Awareness, selection and control*. Oxford: Oxford University Press.
- McLeod, P., Plunkett, K., & Rolls, E. T. (1998). *Introduction to connectionist modelling of cognitive processes*. Oxford: Oxford University Press.

- Miceli, G. & Capasso, R. (2001). Word-centred neglect dyslexia: evidence from a new case. *Neurocase*, 7:221–237.
- Milner, A. D. & Harvey, M. (1995). Distortion of size perception in visuospatial neglect. *Current Biology*, 5(1):85–89.
- Mjolsness, E. & Garrett, C. (1990). Algebraic Transformations of Objective Functions. *Neural Networks*, 3:651–669.
- Moran, J. & Desimone, R. (1985). Selective Attention Gates Visual Processing in the Extrastriate Cortex. *Science*, 229:782–784.
- Morgan, M. J., Hole, G. J., & Glennerster, A. (1990). Biases and sensitivities in geometrical illusions. *Vision Research*, 30:1793–1810.
- Mozer, M. (1991). *The perception of multiple objects: a connectionist approach*. The MIT Press.
- Mozer, M. C. (1999). Do attention and perception require multiple reference frames? Evidence from a computational model of unilateral neglect. In *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society*, pages 456–461. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mozer, M. C. & Behrmann, M. (1990). On the interaction of selective attention and lexical knowledge: A connectionist account of neglect dyslexia. *Journal of Cognitive Neuroscience*, 2:96–123.
- 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. & Rabbit, P. M. (1989). Spatial cueing and the relation between the accuracy of "where" and "what" decisions in visual search. *Quarterly Journal of Experimental Psychology*, 41A(4):747–773.
- Nakayama, K. & Mackleben, M. (1989). Sustained and transient components of focal attention. *Vision Research*, 29(11):1631–1647.
- Nakayama, K. & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, 320:264–265.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9.
- 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.
- Ottes, F. P., Van Gisbergen, J. A. M., & Eggermont, J. J. (1985). Latency dependence of colour-based target vs nontarget discrimination by the saccadic system. *Vision Research*, 25:849–862.
- 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.
- Pomerantz, J. R. (1981). Perceptual organization in information processing. In Kubovy, D. & Pomerantz, J. R., editors, *Perceptual organization*. Hillsdale, N.J.: Lawrence Erlbaum Assoc.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32:3–25.
- Posner, M. I. & Cohen, Y. (1984). Components of Visual Orienting. *Attention and Performance*, pages 531–556.
- Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal injury on covert orienting of attention. *J. of Neuroscience*, 4(7):1863–1874.
- Pouget, A. & Sejnowski, T. J. (1997). A new view of hemineglect based on the response properties of parietal neurones. *Philosophical Transactions of the Royal Society of London: Series B*, 352:1449–1459.
- Pouget, A. & Snyder, L. H. (2000). Computational approaches to sensorimotor transformations. *Nature Neuroscience*, 3:1192–1198.
- Press, W. H. (1992). *Numerical Recipes in C: the art of scientific computing*. Cambridge University Press.
- Rafal, R. D. & Posner, M. I. (1987). Deficits in human visual spatial attention following thalamic lesions. *Proceedings of the National Academy of Science*, 84:7349–7353.
- Rafal, R. D. & Robertson, I. (1995). The neurology of visual attention. In M., G., editor, *The cognitive neurosciences*. Cambridge, Mass.: MIT Press.

- 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.
- Riddoch, M. J., Humphreys, G. W., Burroughs, E., Luckhurst, L., Bateman, A., & Hill, S. (1995b). Cueing in a case of neglect - modality and automaticity effects. *Cognitive Neuropsychology*, 12:605–621.
- Riddoch, M. J., Humphreys, G. W., Burroughs, E., Luckhurst, L., Bateman, A., & Hill, S. (1995a). "Paradoxical neglect": Spatial representations, hemisphere-specific activation and spatial cueing. *Cognitive Neuropsychology*, 12:569–604.
- Riddoch, M. J., Humphreys, G. W., Cleton, P., & Fery, P. (1990). Interaction of attentional and lexical processes in neglect. *Cognitive Neuropsychology*, 7:479–518.
- Robertson, I. & Marshall, J. C. (1993). *Unilateral neglect: Clinical and experimental studies*. London; Psychology Press.
- Robertson, I., North, N., & Geggie, C. (1992). Spatio-motor cueing in unilateral neglect: Three single case studies of its therapeutic effectiveness. *Journal of Neurology, Neurosurgery and Psychiatry*, 55:799–805.
- Rumelhart, D. E. & McClelland (1988). *Parallel Distributed Processing; Explorations in the Microstructure of Cognition; Volume 1: Foundations*. A Bradford Book, The MIT 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. & McClelland, J. L. (1989). A distributed, developmental model of word recognition. *Psychological Review*, 96:523–568.
- 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.
- Shuren, J. E., Jacobs, D. H., & Heilman, K. (1997). The Influence of Center of Mass Effect on the Distribution of Spatial Attention in the Vertical and Horizontal Dimensions. *Brain and Cognition*, 34:293–300.
- 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.
- Tanaka, K. (1993). Neural mechanisms of object recognition. *Science*, 262:685–688.

- Tassinari, G., Aglioti, S., Chelazzi, L., Peru, A., & Berlucchi, G. (1994). Do peripheral non-informative cues induce early facilitation of target detection? *Vision Research*, 34:179–189.
- Tassinari, G., Aglioti, S., Chelazzi, L., Peru, A., & Berlucchi, G. (1998). On the time course of exogenous cueing effects: A response to Lupianez and Weaver. *Vision Research*, 38:1625–1628.
- Tipper, S. P. (2001). Frames of references in attention. In *South Carolina Symposium on Attention*. Columbia, South Carolina.
- 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.
- 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. & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95:15–48.
- Treisman, A., Kahneman, D., & Burkell, J. (1983). Perceptual objects and the cost of filtering. *Perception & Psychophysics*, 33:527–532.
- 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.
- Ungerleider, L. G. & Mishkin, M. (1982). Two cortical visual systems. In Ingle, D. I., Goodale, M. A., & Manfield, R. J. W., editors, *Analysis of visual behavior*. Cambridge, Mass.: MIT Press.
- Usher, M. & Niebur, E. (1996). Modeling the Temporal Dynamics of IT Neurons in Visual Search: A Mechanism for Top-Down Selective Attention. *Journal of Cognitive Neuroscience*, 4(8):311–327.
- Vecera, S. P. & Farah, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General*, 123:146–160.

- vonMühlenen, A. & Müller, H. M. (in press). Probing distractor inhibition in visual search: Inhibition of return. *Journal of Experimental Psychology: Human Perception and Performance*.
- Walker, R. (1995). Spatial and Object-based Neglect. *Neurocase*, 1:371–383.
- Ward, R. & Goodrich, S. (1996). Differences between objects and non-objects in visual extinction: A competition for attention. *Psychological Science*, 7:177–180.
- 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.
- Young, A. W., Hallowell, D. J., & Welch, J. (1992). Neglect and visual recognition. *Brain*, 112:51–71.
- Young, A. W., Newcombe, F., & Ellis, A. W. (1991). Different impairments contribute to neglect dyslexia. *Cognitive Neuropsychology*, 8:177–192.
- Zipser, D. & Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of posterior parietal neurons. *Nature*, 331:679–684.

## A The design of SAIM

### A.1 The contents network and the FOA

As explained in the main text, SAIM uses one-to-one mapping from the retina to all locations in the FOA, mediated by the connections between the contents and selection networks. This allows any arbitrary mapping to be achieved from the retina to the FOA to generate translation invariance. The following equation describes the mapping produced through the units in the contents network:

$$y_{ij}^{FOA} = \sum_{k=1}^N \sum_{l=1}^N y_{kl}^{VF} \cdot y_{ikjl}^{SN} \quad (1)$$

Here,  $y_{ij}^{FOA}$  is the activation of units in the FOA,  $y_{kl}^{VF}$  the activation of units in the visual field and  $y_{ikjl}^{SN}$  the activation of units in the selection network.  $N$  is the size of the visual field. The indices  $k$  and  $l$  refer to retinal locations and the indices  $i$  and  $j$  refer to locations in the FOA. Note that this notation is kept through all of the equations used in SAIM. Equation 1 shows that units in the contents network were of the "sigma-pi" type (Rumelhart & McClelland 1988).

### A.2 The selection network

The mapping from the retina to the FOA is mediated by the selection network. In order to achieve successful object identification, the selection network has to fulfill certain constraints when it modulates the mapping process. These constraints are that: (i) units in the FOA should receive the activity from only one retinal unit; (ii) activity of retinal units should be mapped only once into the FOA; (iii) neighborhood relations in the retinal input should be preserved in mapping through to the FOA.

In order to satisfy these constraints we used an energy function approach suggested by Hopfield (1984), where minima in an energy function are introduced as a network state in which the constraints are satisfied. In the following derivation of the energy function parts of the whole function are introduced and each part relates to a particular constraint. At the end (Sec. A.4) the sum of all parts lead to the complete energy function satisfying all constraints.

Most parts of the energy function used here are based on the function taken from Mjolsness & Garrett (1990):

$$E_{WTA}(y_i) = a \cdot \left( \sum_i y_i - 1 \right)^2 - \sum_i y_i \cdot I_i \quad (2)$$

This energy function defines a winner take all (WTA) behavior, where  $I_i$  are the inputs and  $y_i$  the outputs of the units. This energy function is minimal when all  $y_i$ 's are zero except one  $y_i$ , and the corresponding  $I_i$  has the maximal value of all  $I_i$ s. The advantage of this WTA approach over other approaches is the fact that the number of connections between units increase only linearly, whereas the number of connections, for instance, in the original formulation of Hopfield (1984) increases quadratically (see Mjolsness & Garrett 1990, for detailed discussion). For the first two constraints of the selection network this function



was extended to "two state winner take all" (WTA2 behavior):

$$E_{WTA2}(y_i) = a \cdot \left( \sum_i y_i - 1 \right)^2 \cdot \left( \sum_i y_i \right)^2 - \sum_i y_i \cdot I_i \quad (3)$$

Here we introduce a second minimum into the energy function with the term  $(\sum_i y_i)^2$ , where all units can stay zero. This equation adds a threshold function to the WTA behaviour, whereby the output activity of the units stays zero as long as the input activity is below a certain level. With the parameters used in the simulations this threshold property of the two state winner-take-all equation leads to here having to be at least three pixels in the visual field in order to activate the network. Furthermore, strong "vertical" lesioning (see the main text for details) can lead to a visual field cut. The threshold property is explored in Study ... Apart from this additional feature, the resulting network shows the same competitive behavior as the single state WTA network.

Now, to incorporate the first constraint, that units in the FOA should receive the activity of only one retinal unit, the equation of the "two state winner-take-all" equation turns into:

$$E_{WTA2}^{(1)}(y_{ikjl}^{SN}) = \sum_{ij} \left( \sum_{k,l} y_{ikjl}^{SN} - 1 \right)^2 \cdot \left( \sum_{k,l} y_{ikjl}^{SN} \right)^2 \quad (4)$$

Where the term  $(\sum_{k,l} y_{ikjl}^{SN} - 1)^2$  ensures that the activity of only one retinal unit is put through to FOA units (compare with Equation 2).

To incorporate the second constraint, the "two state winner-take-all" equation turns into:

$$E_{WTA2}^{(2)}(y_{ikjl}^{SN}) = \sum_{kl} \left( \sum_{i,j} y_{ikjl}^{SN} - 1 \right)^2 \cdot \left( \sum_{i,j} y_{ikjl}^{SN} \right)^2 \quad (5)$$

Where the term  $(\sum_{i,j} y_{ikjl}^{SN} - 1)^2$  makes sure that the activity of retinal units is mapped only once into the FOA (again compare with Equation 2).

So far, the last term of Eqn. 3 ("input term") has been taken out of Equation 4 and 5. It is now reintroduced by a common input from the visual field:

$$E_{input}(y_{ikjl}^{SN}) = - \sum_{kl} y_{ikjl}^{SN} \cdot y_{kl}^{VF} \quad (6)$$

For the neighborhood constraint the energy function was based on the Hopfield associative memory approach:

$$E(y_i) = - \sum_{\substack{ij \\ i \neq j}} T_{ij} \cdot y_i \cdot y_j \quad (7)$$

The minimum of the function is determined by the matrix  $T_{ij}$ . For  $T_{ij}$ s greater than zero the corresponding  $y_i$ s and  $y_j$ s should either stay zero or become active in order to minimize the energy function. In the associative memory approach  $T_{ij}$  is determined by a learning rule. Here, we chose the  $T_{ij}$  so that the selection network fulfills the neighborhood constraint. The neighborhood constraint is fulfilled when units in the selection network which receive input from the adjacent units in the visual field, and control adjacent units in the FOA are active at the same time. Hence, the  $T_{ij}$  for these units in Equation 7

should be greater than zero and for all other units  $T_{ij}$  should be less than or equal zero. This leads to the following equation:

$$E_{neighbor}(y_{ijkl}^{SN}) = - \sum_{i,j,k,l} \sum_{\substack{s=-L \\ s \neq 0}}^L \sum_{\substack{r=-L \\ r \neq 0}}^L g_{sr} \cdot y_{ijkl}^{SN} \cdot y_{i+r,k+s,j+r,l+s}^{SN} \quad (8)$$

with  $g_{sr}$  being defined by a Gaussian function:

$$g_{sr} = \frac{1}{A} \cdot e^{-\frac{s^2+r^2}{\sigma^2}} \quad (9)$$

where  $A$  was set, so that the sum over all  $g_{sr}$  is 1.

The effect of  $g_{sr}$  is equivalent to the effect of  $T_{ij}$  in Eq. 7. However, due to the properties of  $g_{sr}$ , energy is minimized not only by the activation of immediate neighbors ( $s = -1, r = -1, s = 1, r = -1$ ) but also by the activation of units in the selection network controlling units in FOA which receive input from more widely space pixels in the visual field (all other values of  $r$  and  $s$ ). Note that activation of these units in the selection network does not contribute as much as activation from immediate neighbors to minimizing the value of the energy function, since  $g_{sr}$  decreases with greater distance in a Gaussian fashion. The slope of decrease is determined by  $\sigma$ .

### A.3 Knowledge Network

In order to introduce effects of stored knowledge into SAIM, a simple template matching approach was used. Here, the match between templates and contents of the FOA was determined using a scalar product as a similarity measure:

$$I_m^{temp} = \sum_{I=1}^M \sum_{j=1}^M y_{ij}^{FOA*} \cdot w_{ij}^m \quad (10)$$

Where  $w_{ij}^m$ 's are the templates and  $M$  is the size of the FOA. In the knowledge network the templates are formed from the weights that connect units in the FOA to the template units ( $y_m^{KN}$ ).

Together with the introduction of the knowledge network the visual field was recoded:

$$y_{kl}^{VF*} = 2 \cdot y_{kl}^{VF} - 1 \quad (11)$$

This equation transforms the input from the visual field from zeros and ones ( $y_{kl}^{VF}$ ) to ones and minus ones ( $y_{kl}^{VF*}$ ). The contents network extracts the contents of the FOA out of this recoded visual field:

$$y_{ij}^{FOA*} = \sum_{k=1}^N \sum_{l=1}^N y_{kl}^{VF*} \cdot y_{ijkl}^{SN} \quad (12)$$

This new FOA code has the advantage that the knowledge network can distinguish between FOA pixels which correspond to contents of the visual field (ones or minus ones) and pixels where the selection network has not fully converged or has stayed zero (values between minus one and one). In the FOA code of the bottom-up version of SAIM, it is

not possible to distinguish between these two states since a zero FOA pixel can be either caused by a zero pixel in the visual field or by units in the selection network which stay zero during network iterations.

A WTA is used to detect the best matching template. The same energy function as in Eqn. 2 was used with  $I_m^{temp}$  as input:

$$E_{knowledge}(y_m^{KN}, y_{ikjl}) = \frac{a^{KN}}{2} \cdot \left( \sum_m y_m^{KN} - 1 \right)^2 - \sum_m y_m^{KN} \cdot I_m^{temp} \quad (13)$$

Here  $K$  is the number of template units.

## A.4 The complete model

The complete energy function of SAIM, which satisfies all constraints, is simply the sum of the different energy functions:

$$E_{total}(y_m^{KN}, y_{ikjl}^{SN}) = a_1 \cdot E_{WTA2}^{(1)}(y_{ikjl}^{SN}) + a_2 \cdot E_{WTA2}^{(2)}(y_{ikjl}^{SN}) + b_2 \cdot E_{input}(y_{ikjl}^{SN}) \quad (14) \\ + b_1 \cdot E_{neighbour}(y_{ikjl}^{SN}) + b_3 \cdot E_{knowledge}(y_m^{KN}, y_{ikjl}^{SN})$$

The coefficients of the different energy functions weight the different constraints against each other.

## A.5 Gradient Descent

The energy function defines minima at certain values of  $y_i$ . To find these values a gradient descent procedure can be used:

$$\tau \dot{x}_i = - \frac{\partial E(y_i)}{\partial y_i} \quad (15)$$

The factor  $\tau$  is antiproportional to the speed of descent.

In the Hopfield approach  $x_i$  and  $y_i$  are linked together by the sigmoid function:

$$y_i = \frac{1}{1 + e^{-m \cdot (x_i - s)}}$$

and the energy function includes a leaky integrator, so that the descent turns into:

$$\tau \dot{x}_i = -x_i - \frac{\partial E(y_i)}{\partial y_i} \quad (16)$$

using these two assertion the gradient descent is performed in a dynamic, neural-like network, where  $y_i$  can be related to the output activity of neurons,  $x_i$  the internal activity and  $\frac{\partial E(y_i)}{\partial y_i}$  gives the input to the neurons.

Applied to the energy function of SAIM, it leads to two sets of dynamic units (neurons), one in the knowledge network and one in the selection network:

$$\tau^{SN} \cdot \dot{x}_{ikjl}^{SN} = -x_{ikjl}^{SN} - \frac{\partial E_{total}(y_m^{KN}, y_{ikjl}^{SN})}{\partial y_i^{SN}} \quad (17)$$

$$\tau^{KN} \cdot \dot{x}_i^{KN} = -x_i^{KN} - \frac{\partial E_{total}(y_m^{KN}, y_{ikjl}^{SN})}{\partial y_i^{KN}} \quad (18)$$

For the selection network the gradient descent procedure leads to the following set of equations:

$$\dot{x}_{ikjl}^{SN} = -x_{ikjl}^{SN} - I_{ikjl}^- + b_1 \cdot I_{ikjl}^+ + b_3 \cdot I_{ikjl}^{td} + b_2 \cdot y_{kl}^{VF} \quad (19)$$

with

$$I_{ikjl}^- = a_1 \cdot \frac{\partial E_{WTA2}^{(1)}(y_{ikjl}^{SN})}{\partial y_{ikjl}^{SN}} + a_2 \cdot \frac{\partial E_{WTA2}^{(2)}(y_{ikjl}^{SN})}{\partial y_{ikjl}^{SN}} \quad ; \quad I_{ikjl}^+ = \frac{\partial E_{neighbor}(y_{ikjl}^{SN})}{\partial y_{ikjl}^{SN}} \quad ;$$

$$I_{ikjl}^{td} = \frac{\partial E_{knowledge}(y_{ikjl}^{SN}, y_m^{KN})}{\partial y_{ikjl}^{SN}} = y_{kl}^{VF*} \cdot \sum_m^K w_{ij}^m \cdot y_m^{KN}$$

The input into the units in the selection network consists of inhibitory input  $I_{ikjl}^-$ , excitatory input  $I_{ikjl}^+$  and top-down modulation  $I_{ikjl}^{td}$ .

For the knowledge network the gradient descent procedure results in the following equation:

$$\dot{x}_m^{KN} = -x_m^{KN} - a^{KN} \cdot \left( \sum_m^K y_m^{KN} - 1 \right) + I_m^{temp} \quad (20)$$

In order to perform the gradient descent and to simulate the behaviour of SAIM on a computer, a temporally discrete version of the descent procedure was implemented on a computer. This was done by using a trapezium approximation (Press 1992), where for each unit in the model the following equation was iterated:

$$x(t) = e^{-\frac{T_0}{\tau}} \cdot x(t-1) + \frac{T_0}{2\tau} \cdot (e^{-\frac{T_0}{\tau}} \cdot I^*(t-1) + I^*(t-2))$$

$T_0$  is the step size,  $I^*$  the input into the unit and  $t$  the iteration step.

## A.6 Attention Switching

In order to implement attentional switching, a "location map" was computed based on activity in the selection network:

$$y_{kl}^{LM} = \sum_{i=1}^M \sum_{j=1}^M y_{ikjl}^{SN} \quad (21)$$

When a template unit in the knowledge network passes a threshold  $\theta$ , the location map is used to reduce the activity in the visual field:

$$y_{kl}^{VF}(new) = y_{kl}^{VF}(old) \cdot (1 - a^{IR} \cdot y_{kl}^{LM}) \quad (22)$$

$a^{IR}$  controls the amount of inhibition. As explained in the main text, this leads to a spatial inhibition of return. All units in the selection network and the knowledge network are set to the initial state they had at the beginning of the simulation. In addition, a second form of inhibition reduces the influence from the original template of the winning unit. The term  $I_m^{temp}$  in equation 20 is multiplied by a factor  $(1 - a^{IT})$ , where  $a^{IT}$  determines the amount of template inhibition.

## A.7 Cueing Studies

SAIM was originally not designed to operate with temporally changing input. In particular, this dynamic input would violate the basic assumption of the underlying theory of energy functions, where inputs have to be constant (e.g. Glendinning 1995). In order to maintain the proper operation of SAIM, cueing was implemented by introducing a smooth transition to a stimulus offset and onset. The smooth transition for an offset was implemented by multiplying the activity of the offsetting stimulus with  $e^{-\tau_1 \cdot t^*}$ , where  $t^*$  is the number of iterations after offset. For the onset the stimulus was multiplied by  $(1 - e^{-\tau_2 \cdot t^*})$ .  $t^*$  is number of iterations from the onset of the stimulus. In most studies the speed of the offset ( $\tau_1$ ) and the speed of the onset ( $\tau_2$ ) were set to 0.02. In Study 13 the speed of the cue offset ( $\tau_1$ ) was set to 0.7 (large decay), 1.5 (medium decay) and 5.0 (short decay).

## B Lesion

SAIM was lesioned in the following way:

$$\dot{x}_{ikjl}^{SN} = -x_{ikjl}^{SN} + d_{ikjl} \cdot (-I_{ikjl}^- + b_1 \cdot I_{ikjl}^+) + b_3 \cdot I_{ikjl}^{td} + b_2 \cdot y_{kl}^{VF}$$

$d_{ikjl}$  is the lesioning factor and its values ranged from 0 to 1, where 0 indicates a strong lesion and 1 no lesion. As explained in the main text, the lesion was only applied to connections within the selection network.

Different spatial patterns of lesions were used in the study. All lesions can be described with the following equation:

$$d_{ikjl} = \begin{cases} d_1 + m_1 \cdot \frac{k}{N-1} & 0 < i \leq M - h - 1 \\ d_2 + m_2 \cdot \frac{k}{N-1} & M - h - 1 < i < M - 1 \end{cases} \quad (23)$$

$N$  is the size of the visual field.  $M$  is the size of the FOA.  $m_1, m_2$  are the gradients of the lesion from left to right in the visual field.  $d_1, d_2$  are the lesions of the left most pixels in the visual field. The indices 1 and 2 indicate the application of the lesions to different parts of the selection network. Index 1 stands for the part of the selection network which controls the left part of the FOA and index 2 relates to the part of the selection network which controls the right section of the FOA. The parameter  $h$  determines the selection layer effected in terms of the corresponding FOA pixels where the switch from index 1 to index 2 occurs. Note that  $d_{ikjl}$  is independent of the indices  $j$  and  $l$  so that lesioning is applied homogeneously from the top to the bottom of the visual field and the FOA. For the different types of lesion, the lesioning parameters were set to different values. The parameters are given in the following section.

## C Parameters

### C.1 Model

In most studies the following parameters were used:  $s = 1$ ,  $m = 40$ ,  $s^{KN} = 1.3$ ,  $m^{KN} = 60$ ,  $a^{KN} = 10$ ,  $\tau^{KN} = 35.0$ ,  $a_1 = 2.0$ ,  $a_2 = 1.5$ ,  $b_1 = 15$ ,  $b_2 = 1.25$ ,  $b_3 = 3.0$ ,  $\tau = 1.0$ ,  $\sigma = 7.0$ ,  $L = 3$ ,  $N = 16$ ,  $M = 7$  and  $T_0 = 0.005$ . For the attentional switching the parameters

were set to:  $a^{IR} = 0.05$ ,  $\theta = 0.9$  and  $a^{IT} = 0.2$ .

In Study 6 the following parameters were modified:  $s^{KN} = 50$ ,  $m^{KN} = 0.7$ ,  $\tau^{KN} = 200.0$ ,  $a_1 = 10.0$ ,  $b_2 = 0.7$ ,  $b_3 = 600.0$ . The main objective of Study 6 was to show that SAIM is capable of suppressing overlapping parts of ignored objects. In order to achieve this suppression, the top-down influence ( $b_3$ ) had to be increased, so that the knowledge network can suppress parts in the FOA that do not match the winning template. This led to an increase of excitatory activation in the selection which, subsequently, had to be compensated by an increase of inhibition in the selection network ( $a_1$ ) and by a reduction of dynamics in the knowledge network ( $\tau^{KN}$ ,  $m^{KN}$  and  $s^{KN}$ ).

## C.2 Lesions

The studies used two different spatial pattern of lesioning: vertical lesions and horizontal lesions. For vertical lesions the damage was homogeneous with respect to the FOA, so that the lesion did not unbalance processing across different FOA pixels. Hence,  $h$  was set to zero. However, since the vertical lesion yields unbalanced processing with respect to the visual field,  $m_1$  and  $d_1$  have not to be zero. Study 7 and 10 used the parameters  $d_1 = 0.1$ ,  $m_1 = 0.9$ . In Study 13 the parameters were set to  $d_1 = 0.1$  and  $m_1 = 0.9$  for large damage;  $d_1 = 0.2$  and  $m_1 = 0.8$  for medium damage;  $d_1 = 0.3$  and  $m_1 = 0.7$  for slight damage. The horizontal lesion yielded unbalanced processing in the selection network with respect to the FOA, but left processing balanced with respect to the visual field. Hence,  $m_1$  and  $m_2$  were set to zero and  $h$  became none zero. For the horizontal lesion  $h$  was always set to 4 and  $d_2$  to 1. For Study 9, 10, 13  $d_1$  was set to 0.8. For the combination of the two types of lesioning,  $h$  was set to 4 in order to maintain the characteristics of the horizontal lesion. In Study 8  $d_2$  was set to 1 and  $m_2$  to 0. The lesioning parameters of Study 11 were  $d_1 = 1.0$ ,  $m_1 = 0.6$ ,  $d_2 = 0.6$ ,  $m_2 = 0.3$ . For Study 12  $h$  was increased from 4 to 5 and  $d_1$  was set to 0.3.