Also in this series:

Maria Marinaro and Roberto Tagliaferri (Eds)
Neural Nets - WIRN VIETRI-96
3-540-76099-7

Adrian Shepherd
Second -Order Methods for Neural Networks
3-540-76100-4

Dimitris C. Dracopoulos
Evolutionary Learning Algorithms for Neural Adaptive Control
3-540-76161-6

John A. Bullinaria, David W. Glasspool and George Houghton (Eds)
4th Neural Computation and Psychology Workshop, London,
9-11 April 1997: Connectionist Representations
3-540-76208-6

Maria Marinaro and Roberto Tagliaferri (Eds)
Neural Nets - WIRN VIETRI-97
3-540-76157-8

Gustavo Deco and Dragan Obradovic
An Information-Theoretic Approach to Neural Computing
0-387-94666-7

Thomas Lindblad and Jason M. Kinser
Image Processing using Pulse-Coupled Neural Networks
3-540-76264-7

L. Niklasson, M. Bodén and T. Ziemke (Eds)
ICANN 98
3-540-76263-9

Maria Marinaro and Roberto Tagliaferri (Eds)
Neural Nets - WIRN VIETRI-98
1-85233-051-1

Amanda J.C. Sharkey (Ed.)
Combining Artificial Neural Nets
1-85233-004-X

Dirk Husmeier
Neural Networks for Conditional Probability Estimation
1-85233-095-3

Dietmar Heinke, Glyn W. Humphreys
and Andrew Olson (Eds)

Connectionist Models in Cognitive Neuroscience

The 5th Neural Computation and Psychology Workshop, Birmingham, 8-10 September 1998

Springer

Modelling Emergent Attentional Properties

Dietmar Heinke, Glyn W. Humphreys
Cognitive Science Centre
School of Psychology, University of Birmingham
Birmingham B15 2TT
United Kingdom

Abstract

We recently introduced a computational model, SAIM (Selective Attention Identification Model), which is capable of simulating visual disorders in brain lesioned patients, including visual neglect and extinction [12]. Here, we report that the same model can both simulate known attentional effects in normal subjects and make novel verifiable predictions. SAIM aims to achieve a translation-invariant object recognition by mapping inputs from their location on the retina to a translation-invariant "focus of attention". Inputs are competitively identified by matching to stored templates. When there are multiple items in the field, there is also competition between the items to win the mapping process. With these mechanisms, SAIM can reproduce qualitatively the results of (1) the Eriksen "flanker" experiment, where RTs increase when a target is flankered by distractors of the opposite response category; and (2) the Posner spatial cueing paradigm, where RTs increase, when the locations of cues do not match the locations of targets. In the cueing paradigm SAIM also predicts that on invalid trails the target is perceived as being shifted more into the periphery (overshoot effect). We have confirmed this prediction experimentally. In SAIM, attentional effects are emergent properties of the competition for limited resources which is needed to achieve a translation invariant object recognition. In humans, there may be no need to posit an explicit attentional system to account for emergent "attentional" effects on behaviour.

1 Introduction

Visual scenes typically contain many objects and require "attention" to be analysed. There is a large literature of experiments looking at the effects of attention on visual scene analyses (see [18] for a recent summary) and, linked to this, several attempts have been made to model the data within a connectionist framework. Here, we outline an approach that has, as its aim, the development of a model for translation-invariant object recognition. To achieve this, the model operates selecting so that only one object at a time is mapped through to recognition procedures. "Attentional" behaviour emerges out of the computational constraints of object recognition.
Previous formal models of visual attention have been developed within rather limited contexts, so that simulations are specific to a particular experimental

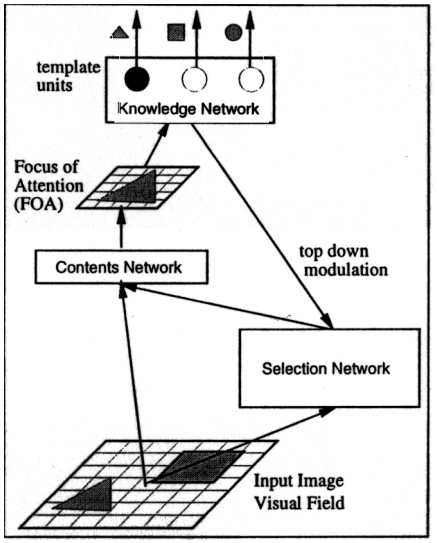


Fig. 1: Overview of SAIM. In order to achieve a translation-invariant object identification, SAIM maps the visual field through to a smaller FOA. This mapping is performed by two networks: The contents network contains "sigma-pi" units, that determine the activation values assigned to units in the FOA by combining multiplicatively activation in retinal units with that in units in the selection network. The selection network determines which retinal units have their activation values mapped through to the FOA (via the contents network). Which retinal units come to be mapped through to the FOA is determined by process of mutual constraint satisfaction between units in the selection network. The knowledge network introduces knowledge about objects into SAIM and modulates the behaviour of the selection network in a top down way.

paradigm. One example of this is the model proposed by Cohen et al. [3]. This model mimics spatial cueing effects whereby the detection of a target is enhanced when it is preceded by a spatially valid pre-cue [19] (see Sec. 3.2 for detailed description). The model has simple detection units, fed both by input units coding visual intensity and attention units, one for each visual field. Attention units compete, so that damage to one (for one visual field) leads to particularly poor responses when a target in the "impaired" field is preceded by a cue in the "intact" field, i.e., there is a "disengagement" deficit. However, the model is capable of doing no more than detecting simple stimuli and it does not encompass a broader range of phenomena such as those requiring pattern recognition.
Other models of attention, such as Guided Search by Wolfe [20] and SEarch via Recursive Rejection (SERR) by Humphreys and Müller [11] can accommodate

a broader range of findings dealing with human visual search, but remain limited as general accounts of selective processing in vision. Guided Search uses a saliency map, based on summed activation from maps that detect simple visual features. Activation in the maps depend on lateral inhibitory interactions, so that items different from their neighbours are strongly activated. Locations in the saliency map are interrogated serially by a second attentional process, with those with most activation interrogated first. Given noise in the interrogation process, patterns of human search can be captured. However, the model fails to cope with evidence showing response competition from items in the field even when a target can be attended in advance – as in the classical Eriksen flanker task [7] (see Sec. 3.1 for detailed description). The model is also constrained to visual search.

SERR [11] provides a connectionist implementation of aspects of Duncan and Humphreys' attentional engagement theory [6]. This theory holds that search is affected by competitive grouping between nontargets and between targets and nontargets, with target detection based on a match between the items processed in parallel and a memory template for the target. Humphreys and Müller [11] showed that SERR can accommodate existing data on visual search and that it can even make verifiable predictions. Also, the architecture of the model is not limited to visual search tasks, since it incorporates a mechanism for object identification: template matching. However, template matching in SERR was not modelled in detail, and it relied in memory representations for targets and distractors coded for every location in the visual field. There are severe problems in scaling up this model to deal with recognition of broad classes of object, presented in a variety of locations.

One other model that links together object recognition and attentional processing to some degree is MORSEL, introduced by Dozer [15]. MORSEL has a parallel pattern recognition system with translation invariance being achieved by mapping visual features to increasingly complex detectors which sample from increasingly wider areas of visual field. Recognition when two or more objects are presented is improved by the operation of a second, attentional network, which enhances activation in the pattern recognition system at attended locations. The attentional mechanism has some flexibility in the area of space that can be activated, and it does not influence performance in an all-or-none fashion. MORSEL has primarily been applied to word and letter recognition tasks. Whether its approach to translation-invariant recognition would be successful in a broader sphere is questionable, however, especially as it demands that every object be presented at every location in the input during the training procedure. Also the model has some difficulty accommodating the full range of neuropsychological disorders of attention that have been reported. In the syndrome of visual neglect, for example, patients may have a relatively pure "object-based" deficit in which they neglect one side of an object whatever its position in space. This can mean that (e.g.) the left side of an object can be neglected even when that object is in the right visual field, whilst the right side of an object in the left field is reported correctly [12]. It is difficult for a model using a retinally-coded attentional system to account for such effects.

Humphreys and Heinke [12] presented a model for translation-invariant object recognition that incorporated spatially selective processing of visual stimuli. The model,SAIM, is similar to the "dynamic routing circuit"-type model proposed by Olshausen [16]. Multiple objects, when present, compete for a dominant mapping to be achieved between their retinal location and a template-based recognition system, mediated by a window or "focus of attention" (FOA). Templates respond in a translation invariant manner, since input is transposed (in the mapping process) from a retinal, co-ordinate system to a system based in the centre of gravity of an object. The model is not limited to particular paradigms or classes of objects, providing templates are learned for the stimuli in question. Humphreys and Heinke examined the performance of SAIM when "lesions" were implemented at different levels of the model. They showed that forms of object-based neglect could occur, as well as, retinally-based neglect, depending on where a lesion was located (spatially selective lesions affecting the mapping into one side of the FOA produced forms of object-based neglect). This prior work shows that SAIM has the potential to model a number of paradigms and to capture a range of neuropsychological data, both of which have proved difficult for other accounts. Here, we examine whether SAIM can accommodate classic results on visual attention in normal observers and whether it is capable of generating novel, testable predictions.

2 SAIM

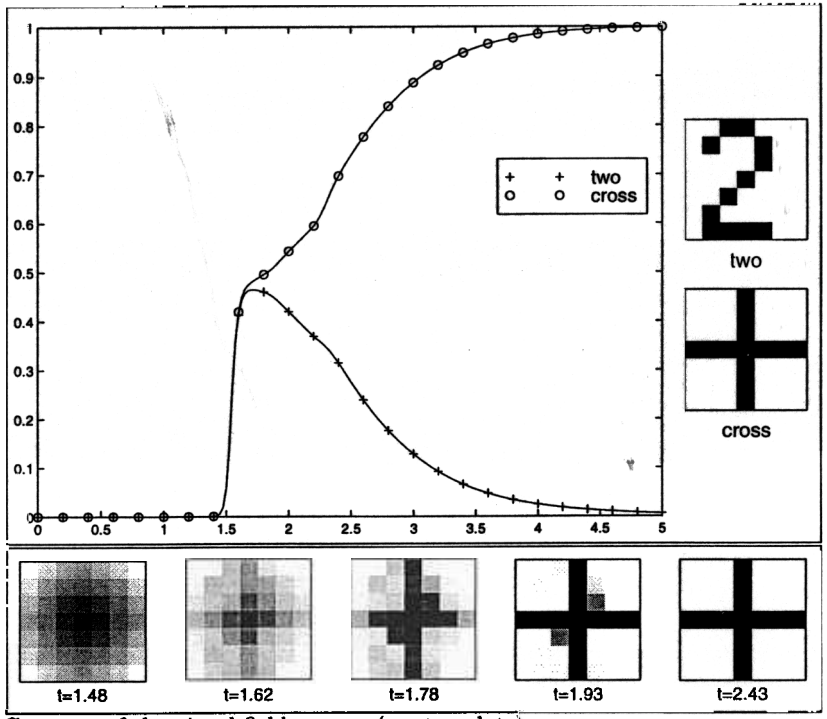
In SAIM, translation invariant object identification is achieved by mapping a retinal input through multiscale stages into a smaller FOA. This mapping is achieved by a dynamic routing circuit, that has a modular structure containing two subnetworks: one performs the mapping from the retina into the FOA (the "contents network") and a second controls this mapping (the "selection network"). Mapping from the retina to the FOA is achieved in SAIM via single stage. This single stage fails to achieve size-invariant recognition, and a multi-stage process may be necessary for this [17]. However, we assume that the general explanatory power of the model is not lost, particularly, since we maintain the key idea of interest here, which concerns spatial selection. Figure 1 gives an overview of the resulting one stage architecture of SAIM. The model used here extends previous dynamic routing circuits models, including the version of SAIM presented in [12], by adding a knowledge network, for object recognition. This networks involves recognition templates activated according to learned weights connecting them to locations in the FOA.

In order to design the topology of SAIM, we formulated constraints support the fulfilling of the computational objectives, e.g., the correct template unit should become active and the wrong ones should be suppressed, and the contents of the FOA should represent the contents of the visual field optimally. These constraints defined the minima in an energy function. A gradient descent in this energy function leads to a set of nonlinear differential equations, which define the topology between the units in SAIM. This approach follows,

in general, the work of [10], where it was applied to the travelling salesman problem. In Heinke and Humphreys [9] a detailed, mathematical description can be found. The resulting topology consists of cooperative and competitive interaction within the selection network and competitive interactions (Winner Take All) in the knowledge network.

3 Results

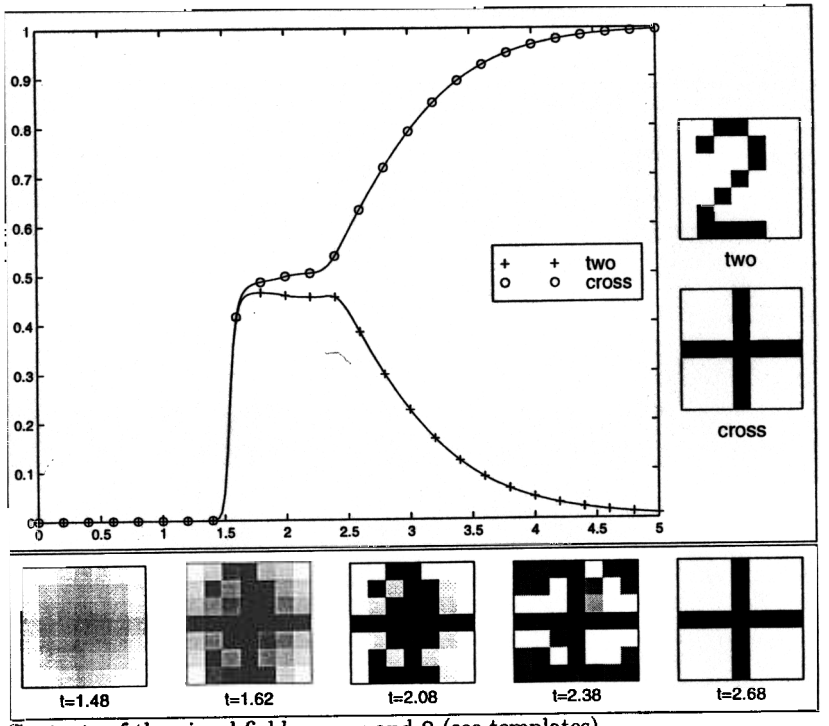
3.1 Basic Behaviour and Eriksen Flanker Task



Contents of the visual field: cross (see template)

Fig. 2: This figure illustrates the behaviour of SAIM. The pictures on the top show the time course of the template units (left) and the corresponding templates (right). The pictures on the bottom show the time course of the FOA activity.

Figures 2 and 3 show the basic behaviour of SAIM for different numbers of objects in the visual input (a cross and a cross along with the number 2). Comparison of the two results shows that the activity of the matching template



Contents of the visual field: cross and 2 (see templates)

Fig. 3: This figure illustrates the behaviour of SAIM. The pictures on the top show the time course of the template units (left) and the corresponding templates (right). The pictures on the bottom show the time course of the FOA activity.

passes a certain threshold, e.g. 0.9, at different points in time. If one assumes that passing the threshold, corresponds "object identification", then the reaction time (RT) of the model increases with the number of objects present. This result and other simulations, show that RT in SAIM depends on competitive interactions in the selection network and the knowledge network. In this framework effects of the number of items in the field, as found in visual search tasks, result from constraints involved in translation invariant object recognition; it is not necessary to assume a saliency map in order to model such results (as in the Guided Search).

The competitive interactions in SAIM enable it to be applied to the Eriksen flanker experiment [7]. Typically here the task is to decide if a letter at a known location belongs to one of the known categories and to press the corresponding lever. Two letters may be presented, e.g. H and K, belonging to one response category and two other letters, e.g. S and C, belonging to a second response

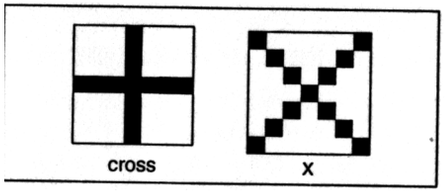
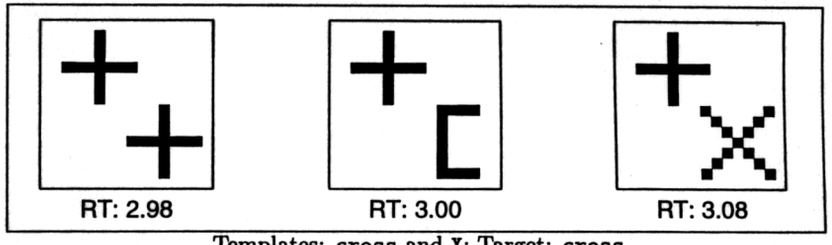


Fig. 4: These templates are used for reproducing the results of the Eriksen task.

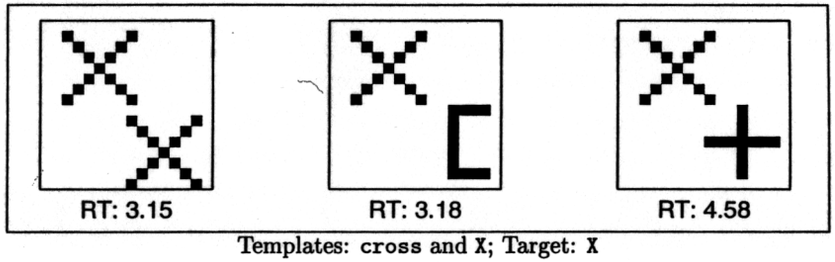
category. Displays can be compatible or incompatible. In the compatible condition the target letter may be flanked by letters that are either identical to it or by that other letter from the same response category. The incompatible displays the target flanked by letters of the opposite response category. RTs are increased in the incompatible relative to the compatible condition.
To simulate this paradigm, we assigned one template (limited to one letter) to one response category, and another to the other letter. Fig. 4 shows the templates used. Because the space in SAIM's visual field is not sufficient to have for more than two letters, only two letters in each display were used. Preknowledge of the target's location was implemented by biasing the selection network at the location of the centre of the target. A bias in SAIM is realised by giving biased units a higher activity for their initial value relative other to units.
Figure 5 gives SAIM's RTs under three conditions: (i) When the stimuli in the field activate a common template (the compatible condition, e.g. `cross` and `cross`), (ii) when only the target has a template (the neutral condition, e.g., `cross` and `C`, which has no template), and (iii) when the stimuli in the field activates competing templates (the incompatible condition, e.g., `cross` and `X`, which both have templates). RTs were slower in the incompatible condition. Note also that, relative to the neutral condition, interference is greater (with incompatible stimuli) than facilitation (with compatible stimuli). This pattern of greater interference than facilitation has frequently been observed in response competition effects in humans [4]. Here it emerges as a natural consequence of processing dynamics in the network. SAIM could easily be extended to accommodate the standard Eriksen result with more than 1 letter assigned to each response category, for example by adding a further set of category templates (employing similar dynamics), above the level of templates for individual letters.

3.2 Spatial Cueing Paradigm

In a second set of simulations, we have examined whether SAIM can capture another classic effect on visual attention: the effect of spatial cueing. Many experiments have shown that humans RT to detect a simple target are enhanced if the target is preceded by a valid spatial precue and disrupted if preceded by



Templates: `cross` and `X`; Target: `cross`



Templates: `cross` and `X`; Target: `X`

Fig. 5: These images show the simulation of the Eriksen task. The images in the columns show the RT for object the compatible, neutral and incompatible condition.

an invalid cue. For our simulations, the cue was a simple `square` and the target a small `cross` (see Fig. 6). The cues was presented for 1.3 and followed by the target `cross` which remained on until the network responded. The network response was based on a complete appearance of the target in FOA (as shown). On valid trails the cue fell at the same location as the `cross`; on invalid trials it fell at a different location.
Figure 6 shows the results, which fit those found with human subjects. The reason for the cueing effect in SAIM is due to its cooperative connections which maintain activity in selection network following the cue. This maintenance of activity due to cooperative connections is a well-known property of networks with dynamic cooperative and competitive interactions [1]. In the valid condition, activity from the target is boosted by the sustained activity in the selection network, leading to a cueing benefit. In the invalid condition, target activation has to compete against the sustained activation of the cue, causing a cueing cost. Note that the cueing effects emerged here even though the network has no explicit mechanism for engaging or disengaging attention. This reiterates the point made by other connectionist models in this field (e.g. [3]), that attentional engagement and disengement can reflect network states rather than processing mechanisms devoted to there operations.
These simulations with SAIM also showed an additional effect. The images on the bottom row of Fig. 6 depict the result when the target is in the right field and preceded by an invalid cue in the left field. The figure reveals that

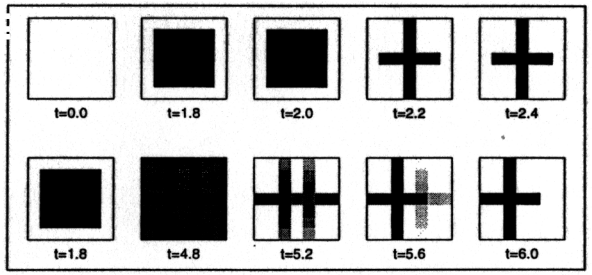


Fig. 6: These images show the results of simulating the spatial cueing effect (see text). The upper images show the simulation result in the valid condition and the lower image show the results in the invalid condition. The reaction time is clearly longer in the invalid condition ($t = 6.0$) than in the valid condition ($t = 2.2$). This result matches with the findings on the spatial cueing. However, the model shows an additional effect. In the invalid condition the location of the `cross` is misplaced. If the target appears on left side, the `cross` appears on the right of the FOA. If the target is on the right side, the `cross` appears on the left side of the FOA. This predicts an overshoot effect for the localisation of the target in the invalid condition (see text for details).

the target `cross` is misplaced within the FOA. It is shifted fro the centre of the FOA in a direction towards to the location where the invalid cue appeared. This shift can be conceptualised in the following way. Normally, in mapping from the retinal represention of an object to the FOA, the centre of gravity of the object is placed in the central part of the selection network (ensuring that this centre of gravity is then mapped to the centre of the FOA). It is as if the network is trying to align a marker for the centre of gravity (the central part of the selection network) with the centre of the object on the retina. Now, in the case of invalid spatial cueing, this alignment process is repulsed away from the cue, due to competition, so that the marker for the centre of gravity gets shifted to align with the right side of the `cross` (when the cue is left and the target right). The result of this incorrect alignment is that the target is mapped into the left side of the FOA, because the centre of gravity indicated in the selection network falls on the right side of the `cross`. From this, a prediction can be made concerning human location judgements. We might assume that human location judgements are based on where the spatial marker for the centre of gravity (activity in the central part of the selection network) aligns with the retina. On invalid cue trails, this alignment is shifted away from cue. On invalid cue trials, this alignment is shifted away from cue. Under these circumstances, judgement may be made that the target is further into the invalid field than it is really the case. In other words, there an "overshoot" of the perceived centre of gravity. This possibility was examined empirically.

4 An experimental test

To test the prediction of an overshoot under invalid cueing conditions, we ran a spatial cueing study in which subjects had to make a localisation judgement to targets. The cue was an open circle and the target a simple star (*). The centre of the cue and the centre of the target could appear at 14 different locations, 7 locations left and 7 locations right of the fixation `cross`. Following the presentation of the target, subjects saw a horizontal scale in the field where the target appeared, numbered 1 to 7. The locations of the numbers on the scale exactly corresponded to the seven possible locations of the target. The numbers appeared in increasing order from left to right. The experimental procedure was as follows: First, a fixation `cross` appeared on the screen for 1000 ms. After that, the cue appeared at one of the 14 locations for 100ms. After that the target appeared either at the same location or at the corresponding location on the opposite side of the fixation `cross`, again for 100 ms. Finally, the line of numbers appeared for an unlimited presentation time. The subjects were asked to press the number, corresponding to where they saw the target. If they did not see the target at all, they pressed the space bar. There were 196 trails and all conditions were equal likely.
The data of 15 subjects were analysed by averaging the difference between the target location and the perceived location for the valid and the invalid condition, separately. The average for the invalid condition showed that subjects were more likely to select more peripheral locations than in the valid condition. The displacement was 0.65° ($F(1,14) = 75.15, p < 0.01$). This experimental finding confirms the prediction of SAIM.

5 Discussion

In this paper we presented simulation results from SAIM, a translation-invariant object identification network, which maps inputs into a FOA by cooperative and competitive interactions. The model can account for classic findings in the literature on human visual attention, such as the Eriksen flanker and spatial cueing effects. For spatial cueing effects SAIM also made a novel prediction, that we confirmed by a simple experiment. All the presented results arise out of simple competition for spatially limited resources in mapping retinal input to an object-centred output representation in the FOA. It follows that attentional effects in the human literature may be considered emergent properties of competition for spatial mapping and not the influence of a system devoted specifically to shifting attention in space.
Of course, SAIM has many shortcomings. For instance, it does not have a feature extraction stage, it uses only simple proximity-based grouping and, in particular, the selection network has too many units and connections for it to be scaled-up easily. However, we believe that solutions to these difficulties will not change the basic results from the model.

In addition to accounting for behavioural data, the simulation results can

also be related to single cell studies. Several single cell studies have found cells in the inferior temporal cortex (IT, part of the "What-system") which code complex features, such as mixtures of colour, shape or texture in ways that are relatively independent of stimulus location (e.g., [8]). Here, the template units of the knowledge network represent a simple model of such complex feature detectors with the current emphasis on the coding of shape. Studies have also shown that cells in IT that are modulated by the attentional behaviour of animals [2; 14]. For instance responses of IT cells to stimuli are suppressed if animals have to ignore the stimuli [14]. Again, this would correspond to the behaviour of the template units in the knowledge network, where the responses to an object in the visual field are suppressed when the object is ignored by SAIM. In SAIM, this attentional suppression can occur both by precueing on target location or by priming one template so that it dominates the competition with other templates. These different forms of suppression may also be evident in physiological studies of attentional modulation [2; 14].

In contrast to the knowledge network, the selection network may be considered part of a "where-system". Units in the network respond to the location of target elements. Preactivating such units leads to spatial cueing effects. This last result can be related to single cell studies in the parietal cortex, where enhancement of activity takes place in neurons responding to the location of objects, where attention is directed [5].

The distinction between the knowledge and selection network may be mapped onto the distinction between "what" and "where" processing in the brain [13]. SAIM, however, suggests that these two pathways may interact, to produce coherent selective processing for object recognition.

References

[1] Shun-ichi Amari. Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields. *Biological Cybernetics*, 27:77–87, 1977.

[2] L. Chelazzi, E.K. Miller, J. Duncan, and R. Desimone. A neural basis for visual search in inferior temporal cortex. *Nature*, 363:345–347, 1993.

[3] J. Cohen, M. Farah, and D. Servan-Schreiber. Mechanisms of spatial attention: The relation of macrostructure to microstructure in parietal neglect. *J. of cognitive Neuroscience*, 6(4):377–387, 1994.

[4] J. Cohen and J. L McCelland. On the control of automatic process: A parallel distributed porcessing model of the Stroop effect. *Psychological Review*, 97:332–361, 1990.

[5] R. Desimone, M. Wessinger, L. Thomas, and W. Schneider. Attentional Control of Visual Perception: Cortical and Subcortical Mechanisms. In *Cold Spring Habor Symposia on Quantitative Biology*, pages 963–971. Cold Spring Habor Laboratory Press, 1990.

[6] J. Duncan and G. W. Humphreys. Visual Search and Stimulus Similarity. *Psychological Review*, 96(3):433–458, 1989.

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

[8] I. Fujita, K. Tanaka, M. Ito, and K. Cheng. Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, 360:343–346, 1992.

[9] D. Heinke and G. W. Humphreys. SAIM: A Model of Visual Attention and Neglect. In *Proc. of the ICANN'97, Lausanne, Switzerland*, pages 913–918. Springer Verlag, 1997.

[10] J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA*, 81:3088–3092, 1984.

[11] Glyn W. Humphreys and Hermann J. Müller. SEarch via Recursive Rejection (SERR): A Connectionist Model of Visual Search. *Cognitive Psychology*, 25:43–110, 1993.

[12] G.W. Humphreys and D. Heinke. Spatial representation and selection in the brain: Neuropsychological and computational constraints. *Visual Cognition*, 5(1/2), 1998.

[13] M. Mishkin, L.G. Ungerleider, and K.A. Macko. Object vision and sptial vision: two cortical pathways. *Trends Neuroscience*, 6:414–417, 1983.

[14] J. Moran and R. Desimone. Slective Attention Gates Visual Processing in the Extrastriate Cortex. *Science*, 229:782–784, 1985.

[15] M.C. Mozer. *The perception of multiple objects: a connectionist approach.* The MIT Press, 1991.

[16] B. A. Olshausen, C. H. Anderson, and D. C. v. Essen. A Neurobiological Model of Visual Attention and Invariant Pattern Recognition Based on Dynamic Routing of Information. *J. of Neuroscience*, 13(11):4700–4719, 1993.

[17] B. A. Olshausen, C. H. Anderson, and D. C. v. Essen. A Neurobiological Model of Visual Attention and Invariant Pattern Recognition Based on Dynamic Routing of Information. *J. of Computational Neuroscience*, 2:45–62, 1995.

[18] H. Pashler, editor. *Attention.* Psychology Press, 1998.

[19] M.I. Posner, C.R.R. Snyder, and B.J. Davidson. Attention and the detection of signals. *J. Exp. Psychol.*, 109:160–174, 1980.

[20] J. M. Wolfe. Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2):202–238, 1994.