

Modeling the integration of expectations in visual search with centre-surround neural fields[☆]

Joshua P. Salmon^a, Thomas P. Trappenberg^{b,*}

^a Department of Psychology, Dalhousie University, Halifax, NS, Canada

^b Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada

ARTICLE INFO

Article history:

Received 14 February 2008

Received in revised form

7 August 2008

Accepted 22 September 2008

Keywords:

Dynamic neural field model

Eye saccades

Attention

Likelihood map

Visual search

ABSTRACT

Centre-Surround Neural Field (CSNF) models were used to explain a possible mechanism by which information from different sources may be integrated into target likelihood maps that are then used to direct eye saccades. The CSNF model is a dynamic model in which each region in network excites nearby location and inhibits distant locations, thereby modeling competition for eye movements (saccades). The CSNF model was tested in a number of conditions analogous to a naturalistic search task in which the target was either (1) present in the expected location, (2) present in the unexpected location, or (3) absent. Simulations showed that the model predicted a pattern of accuracy results similar to those obtained by [Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science*, 17(11), 973–980] from human participants. However, the model predicts different saccadic latencies between conditions where Eckstein, Drescher, and Shimozaki (2006) found no significant differences. These discrepancies between model predictions and behavioural results are discussed. Additional simulations indicated that these models can also capture the qualitative flavor of eye movements in conditions with multiple targets as compared to [Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, 37(5), 617–631].

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

When an individual, or primate, moves their eyes they tend to do so in jerky stop-and-start movements referred to as “eye saccades”. Numerous neuropsychological and psychological studies have studied these eye saccades in a various experimental paradigms (for a recent review of eye movement research see Stigchel, Meeter, and Theeuwes (2006)). A debate still exists about how information from different sources is integrated to direct these eye saccades. One model is the Centre-Surround Neural Field (CSNF) model whose primary feature are centre-surround nodes that excite near-by locations and inhibit distant ones (Amari, 1977; Grossberg, 1973). These CSNF models are dynamic models in which each node represents a possible choice or location to which a person could move their eyes. Each node competes to reach threshold with the time it takes to reach threshold considered to be the reaction time or saccadic latency, and the winning

location represents where in space the saccade would be directed. The current paper compares these CSNF models to models used by Eckstein, Drescher, and Shimozaki (2006) and considers the integration of multiple targets in comparison to Findlay (1997), as well as Chen and Zelinsky (2006).

1.1. Model input: Maps

Many models use feature extraction at their core to determine which regions would be interesting/relevant. For example, Navalpakkam and Itti (2005) recently proposed a model using a combination of feature detectors tuned to different resolutions. This information was further modified/enhanced through top-down biasing to simulate previous knowledge that a person would have. Navalpakkam and Itti's (2005) model further incorporated hierarchies, and different types of what they called *saliency maps*, each of which had to be integrated to make some decision in regards to which locations the model would direct its eyes.

Maps are topographically arranged such that two equidistant locations on the map represent two equidistant locations that fall on the retina. High activity at a location on a map indicates a corresponding region of interest (or saliency) in the real-world or input (Li, 2005). Navalpakkam and Itti (2005) hypothesized the existence of at least three different kinds of maps. These three

[☆] Contributed article.

* Corresponding address: Faculty of Computer Science, Dalhousie University, 6050 University Ave., Halifax, NS, Canada, B3H 1W5.

E-mail addresses: Joshua.salmon@dal.ca (J.P. Salmon), tt@cs.dal.ca (T.P. Trappenberg).

maps included (1) an exogenous or bottom-up visual saliency, hypothesized to be in posterior parietal cortex, (2) an endogenous or top-down task-relevant map, hypothesized to be located in the pre-frontal cortex, and (3) an eye-movement guidance map in the superior colliculus that may guide the focus of attention. This final map (3) can also be called a “target likelihood map”, and is likely the site at which information from both the other two maps are integrated.

Previous research has built target likelihood maps using various methods of integrating exogenous and endogenous inputs. Tsotsos et al. (1995) implemented attentional selection using a combination of feedforward bottom-up feature extraction hierarchy and a feedback selective tuning of these feature extraction mechanisms. That location is then propagated back through the feature extraction hierarchy, through the activation of a cascade of winner-take-all networks embedded within the bottom-up processing pyramid. Wolfe (1994) built on Neisser’s pre-attentive/attentive distinction (Neisser, 1967), integrating both exogenous and endogenous saliency criteria in his Guided Search model. Deco, Pollatos, and Zihl (2002) used inhibition to mediate between exogenous and endogenous influences in a biased competition model (Duncan & Humphreys, 2002), simulating saliency in posterior parietal cortex with a model similar to that discussed here. For a review of these and other models see Shipp (2004) and Itti and Koch (2001).

Itti and Koch (2000) proposed a model with centre-surround receptive field mechanisms used to derive luminance, color, orientation contrast difference signals at multiple spatial scales within a pyramid. They then combined these feature contrast signals to build saliency maps. Their model was successful at explaining how feature singleton targets can be detected very quickly independent of set size (cf. Treisman and Gelade (1980)) but did not incorporate the influence of top-down guidance. Such top-down guidance is a key element to fixation patterns as illustrated by the classic study by Yarbus (1967) in which participants were shown to fixate on different areas in the same painting depending on their instructions. In a more recent example, Chen and Zelinsky (2006) found that top-down guidance to be more influential on eye saccades than bottom-up in a design that pitted top-down information (available in the form of a target preview) against bottom-up information (color saturation). Thus, it seems important to consider models that act on higher level information than just bottom-up salience.

This current paper focused on maps that incorporate information from multiple sources. In the context of most of the experiments in this paper, these multiple sources can represent bottom-up and top-down information, and the maps we discuss here are “target likelihood maps”, such as the map in the intermediate layer of the superior colliculus. A common problem in the literature has been deciding how exactly to combine top-down task-relevant maps with bottom-up saliency maps. Our model does not distinguish between the nature of the sources, but implements an universal mechanism for integrating them.

The CSNF model used in this study consist of a single competitive layer plus one input layer. This CSNF layer was treated as a layer in which information from potentially different sources was integrated. We did not model feature detection in this paper, but rather concentrate on the integration dynamics.

1.2. Centre-Surround Neural Field (CSNF) models

Centre-Surround Neural Field (CSNF) models are named as such because their mechanisms work similar to on-centre off-surround ganglion cells. That is, certain ganglion cells in retina increase their firing rate when stimulated in the centre of their receptive field, but

decrease their firing rate when stimulated in the periphery. Centre-Surround Neural Field (CSNF) models work in a similar way with activation at one region exciting near-by locations, but suppressing distant locations. The CSNF models we will describe in this paper are also called Continuous Attractor Neural Networks (Standage, Trappenberg, & Klein, 2005; Stringer, Trappenberg, Rolls, & Araujo, 2002) and Competitive Integration Models (Godijn & Theeuwes, 2002). The centre-surround mechanisms of CSNF models have been under investigation for a number of years (e.g. Standage et al. (2005); Trappenberg, Dorris, Munoz, and Klein (2001); Zhang (1996)), and are similar to those used by Itti and Koch (2000) and Usher and McClelland (2001).

These models have several important features including (1) They are arranged retinotopically or topographically. (2) They make predictions regardless of the source of information (e.g. exogenous or endogenous). (3) They work competitively and dynamically through time with lateral connections exciting other near-by locations and inhibiting distant ones. (4) They can be thought of as a type of Winner-Take-All (WTA) model with the winning “node” (a unit of the model) representing the most likely location for the target, and therefore the location to which an eye saccade is most likely to go. (5) These models have a certain degree of neurophysiological plausibility, for example, recent research in macaque monkeys found centre-surround mechanisms in the frontal eye-fields (Schall, Sato, Thompson, Vaughn, & Juan, 2004). (6) Finally, and perhaps most importantly, these models have great predictive power. Versions of these models have been shown to explain many aspects of eye movement data including endpoints, latencies, pro-saccades, anti-saccades, trajectories of eye saccades, and responses to distractors and target expectancies, among other things (cf. Godijn and Theeuwes (2002) and Trappenberg et al. (2001)).

CSNF models are dynamic models in which the retinotopic “nodes” act over-time on information/input from the environment. Input from the environment (external input) is processed by the model and results in increased activation at the node representing that region in space. Activity at this node then, in turn, begins to excite neighboring locations, and inhibiting distant locations. If there is only one strong input to the system a winning region/packet or bubble of activity will develop over time (see Fig. 1A). With two or more strong inputs to the system, activity at the strongest region will eventually inhibit activity at the other regions (Fig. 1B). However, if two inputs to the model are close enough to each other excitatory activity at both regions can cause the CSNF model representation to merge together to form a winning bubble of activity centered somewhere between the two inputs (Fig. 1C). This last feature of CSNF models can explain “averaging saccades” in behavioural data (the phenomenon in which eye saccades are made to a location between two close targets instead of directly to one of the targets).

Generally, even with multiple inputs to the network, a location will eventually reach threshold (“win”), however the location of the “winner” and the time taken to reach threshold will depend on the number of inputs, their relative strength, and their relative distance from each other. The winning location, or the first node to reach threshold, is a prediction about to which location an eye saccade go. The time taken for the winning location to reach threshold is a prediction about how long it takes to actually initiate the saccade (known as the *saccadic latency*).

In CSNF models, each node has its own accumulator, which interacts with accumulators at other nodes. This is opposed to diffusion models, as described recently by Smith and Ratcliff (2004), in which there is only a single accumulator. In basic diffusion models, this accumulator is initialized to start between some upper and lower threshold that represent different decisions. The model then accumulates evidence and moves either up or down until it crosses one of the two thresholds and a decision is

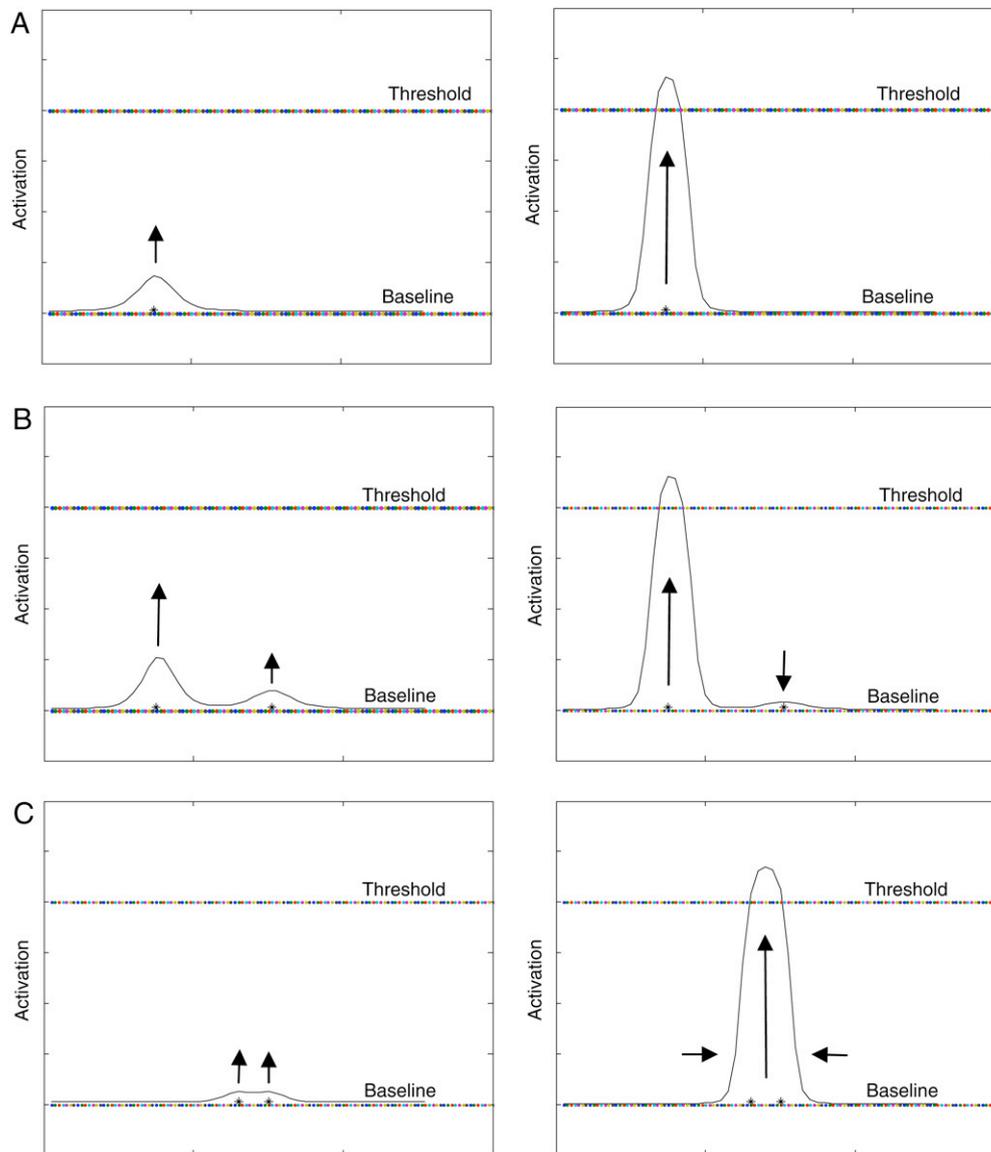


Fig. 1. Build-up of CSNF model from different starting conditions. Left columns show early states of network, and right columns show the network as it reaches threshold. Panel A show build-up of a single input (*) to reach threshold. Panel B shows how with two inputs, the larger one will inhibit the smaller one through lateral connections as it increases in activity. Panel C shows how two close inputs will merge via lateral excitatory connections to form a single winning bubble. Similar figures and explanation of model dynamics can be found in Godijn and Theeuwes (2002).

made. Usher and McClelland (2001) described the classical *random walk model* where the accumulator variable can be thought of the difference between the number green and red balls drawn from an urn. Before processing, the counter is initialized and then at each step a count is added if the ball is green or subtracted if the ball is red. CSNF models, on the other hand, are a type of leaky competing accumulator model where each node has its own accumulator initialized to zero that rises toward an upward threshold (moderated by lateral connections to other nodes). The first to reach threshold represents the decision, or in this case, the location to which the eyes are moved. Of course, as shown in Fig. 1B, an important dynamic in this architecture, in contrast to non-competing accumulator models, is that CSNF accumulators not only rise, they can also fall back to zero through competition and lateral connections between the nodes.

1.3. Current simulations

The first part of this paper focuses on comparing CSNF models to recent work by Eckstein et al. (2006). Eckstein et al.

(2006) compared two different approaches to constructing target likelihood maps for a simulated naturalistic search task with at least two highly-likely target locations. They called one approach (1) The Bayesian approach and the other the (2) Limited Attentional Resources (LAR) approach. In the Bayesian approach, two maps were built separately, a likelihood ratio map (containing visual evidence for the target) and a prior probability map (differential weighting based on cues). These two maps were then combined in a multiplicative way. The resulting posterior probability map or target likelihood map would then indicate the decision map or the location to which participants would most likely make eye saccades. In the LAR model the likelihood ratio map was built in the same way, but the prior probability map was instead used to process cued locations more accurately. The different models make different predictions in a case with only a cue and no target. Specifically, the Bayesian model predicted eye saccades to the cued location, while the LAR Model predicted better target rejection at the cued location. In other words, the Bayesian model predicted more eye saccades to the cued location, while the LAR Model did not, or no more so than chance. The Bayesian model more

appropriately captured the pattern of results from the behavioural data.

Although we see some of advantages in the modeling approach of multiplying two different maps together to choose a winner (as done by Eckstein et al. (2006) in their Bayesian model), we believe leaky competing accumulator models offer an important extension to their work. Most notably, these models offer a temporal component to the saccadic predictions absent from the models of Eckstein et al. (2006) in addition to offering a more biologically plausible implementation (cf. Schall et al. (2004)). Thus, the first part of this paper focused on applying our Centre-Surround Neural Field (CSNF) models to conditions outlined by Eckstein et al. (2006).

The Eckstein et al. (2006) model considered cases with only two locations of interest, but several studies have investigated multiple targets and distractors. For example, Chen and Zelinsky (2006) considered a design in which a target was searched for in a circular array with nine distractors. The objects in the array were all black & white photos of real objects but one of the distractors was shown in color to create visual salience and distract direction of gaze away from the target. Findlay (1997) also considered a design in which a target was one of eight objects somewhere around a ring. In addition to single-target trials, initial eye saccade endpoints were considered to conditions with no targets, two targets, or targets of different distances. In the latter case, 16 objects were presented, an inner ring with eight, and an outer ring of eight. The saccade pattern difference was considered when the target was at a near location versus a far location. Since CSNF models are easily extendable to cases with multiple targets and/or distractors, the last part of this paper incorporated a few of these simulations. The results of these simulations were compared to the behavioural data collected by Chen and Zelinsky (2006) and Findlay (1997).

2. Methods

2.1. Centre-Surround Neural Field (CSNF) model architecture

The current simulations use a Centre-Surround Neural Field model similar to those employed by in previous research (e.g. Standage et al. (2005), Trappenberg (in press, 2008b) and Trappenberg et al. (2001)). Neural field models of this sort were first introduced by Amari (1977). The dynamics of an internal state variable $u(\vec{x})$ at 'location' \vec{x} in 2-dimensions (Owen, Laing, & Coombes, 2007; Taylor, 1999) is given by:

$$\tau \frac{\partial u(\vec{x}; t)}{\partial t} = -u(\vec{x}; t) + \int_{\vec{z}} w(|\vec{x} - \vec{z}|) r(\vec{z}, t) \partial \vec{z} + I^{\text{ext}}(\vec{x}; t). \quad (1)$$

In this formula, $I^{\text{ext}}(\vec{x}, t)$ represents an external stimulus as received by the neural field, and $r(\vec{z}, t)$ is a state variable of the neural field. This rate is related to the internal state variable $u(\vec{x}, t)$ through a gain function $r = g(u)$. A sigmoid gain function of $g(u) = 1/(1 + \exp(-\beta u))$ was used. Other gain functions are sometimes biologically more appropriate but would not change the principle findings in the simulations presented here. The weight kernel $w(|\vec{x} - \vec{z}|)$ describes the interaction in the neural field as specified further below.

Period boundary conditions were used to minimize boundary effects. The result in 2-dimensions is a torus, where we used feature values contained between values of $(0-2\pi]$. Eq. (1) defines the dynamic evolution of the neural field, which is continuous in time and space (Schöner, 2007).

The interpretation of each node is not that of a simple neuron, rather, it stands for a collection (population) of neurons that

receive similar input in a certain task (Brunel & Wang, 2001; Gerstner, 2000; Wilson & Cowan, 1972), or that of a cortical minicolumn (Johansson & Lansner, 2007). The lateral interaction kernel $w(|\vec{x} - \vec{z}|)$, is chosen to be a shifted Gaussian

$$w(|\vec{x} - \vec{z}|) = A_w \left(\frac{1}{\sqrt{4\pi\sigma_r}} e^{-\Delta^2/4\sigma_r^2} - C \right), \quad (2)$$

with $A_w = 50$, $\sigma_r = 2\pi/24$, and the inhibition level (C) = 0.3, and

$$\Delta = \min(|\vec{x} - \vec{z}|, 2\pi - |\vec{x} - \vec{z}|). \quad (3)$$

Other popular choices are weight profiles in form of a Mexican-hat, but these functions result in similar behaviour when the extent of inhibition and excitation are matched in the periodic feature maps discussed here. Many studies assume this interaction structure from the start, and it is possible that such structures are genetically coded, in particular if the feature space is topographic. However, such weights can also be learned from activity-dependent Hebbian learning on Gaussian patterns with width σ_r (Stringer et al., 2002; Trappenberg, 2002; Wu & Trappenberg, 2008). Such learning might be, for example, important in the formation of place fields in the hippocampus (Stringer et al., 2002).

All CSNF simulations were done using MATLAB (Mathworks, Natick, MA). Sample code for one of these programs is provided in the Appendices A and B. The MATLAB programs and files themselves are available, by request, from the authors.

2.2. Simulations: Using CSNF models to predict eye saccades in different conditions

We first considered the performance and predictions of eye saccade data under the same conditions investigated by Eckstein et al. (2006). Eckstein and colleagues considered the predictions of their models under three different conditions: (1) target present at expected location, (2) target present at unexpected location, and (3) target absent, or cue only. Eckstein et al. (2006) described their models as simulating perception in naturalistic search. They compared performance in their models to behavioural data in which participants were searching for items in a scene in which the item would more probabilistically occur in certain locations. For example, participants would be asked to look for/at a chimney and then be shown a picture of a house. Not surprisingly, participants tended to look toward the roof of the house (that's where chimneys tend to be). Thus, previous knowledge that houses and chimneys occur together could act as a real-time "cue" towards the roof (prior probability map). The knowledge about the features of a chimney would be a separate and distinct form of knowledge than the knowledge of the "cue".¹ Specifically, the map that contains information about the chimney would be a map that contains more low-level feature-extraction information or visual evidence for target information (likelihood ratio map). These can be considered two different maps or two different sources of information.

In the CSNF model of such a naturalistic task (searching for a chimney on a picture of a house) the focus was on how information from these two different sources were integrated. Input to the model such as location of the target (chimney) and the cue (roof of house) were feed to the model at the same time. The CSNF main leaky accumulator layer then simulated the integration of these pieces of information. How information was imputed was changed based on source (as explained in the next section), but the

¹ Note, that although the "cue" in this context is not a traditional cue (it does not precede presentation of the target), however, the location of the roof/house will be referred to as a "cue" for the remainder of this paper.

mechanism by which the CSNF model integrated information was consistent at all nodes regardless of the information source. Thus, whether information was coming from a target map (likelihood ratio map) or cue map (prior probability map), it was treated in the same way by a CSNF model.

Thus, a CSNF model can simulate target absent (cue only) trials with a single input (input from cue map but not from target map). Target present at expected location trials are simulated with two inputs that are close or very close together (both maps send input to the same spatial location). Target present at unexpected location trials would be two inputs at different locations (input from two maps sent to different spatial locations in the target likelihood map).

Our final simulations considered examples with more than one target and more than one distractor. The focus of the [Eckstein et al. \(2006\)](#) was predicting, on average, the error rate of initial saccades from the target on various conditions. However, when there are multiple targets and distractors error rates in terms of numbers become less meaningful and the real interest is where the participants looked. That is, did they go to target A, target B, or somewhere in between (averaging saccade)? This has been best represented in the past by plots showing the input to the model with an overlay of all the initial saccade endpoints (draw as dots or squares). Thus, our models were modified to create figures of this type. We then used these models to simulate a number of conditions (1) eight targets around a ring with equal salience, (2) two targets with two different degrees of separation in a ring of distractors, and (3) targets of different distances from the initial fixation position. The results of these simulations were then compared to behavioural data collected by [Chen and Zelinsky \(2006\)](#) and [Findlay \(1997\)](#).

2.3. Additional details: CSNF model simulations

There are a number of details worth keeping in mind when considering the results of the current simulations. (1) In the current simulations target map and cue map were used as inputs to the models at the same time. (2) The current simulations differentiated between target and cue inputs on the basis of size, with target inputs having larger amplitudes and smaller diameters (occupying a more precise region in space). (3) The current models did not employ active inhibition of previously attended locations, as only the first eye saccade was being modeled. (4) The models included a “cost” function as explained by [Eckstein et al. \(2006\)](#) that biases attention to making shorter eye saccades. (5) The simulations used noise/noisy inputs, even though this noise is not depicted in the figures drawn in the results section. These features of the simulations will now be elaborated on.

(1) The simulations in this research differed from those used by [Trappenberg et al. \(2001\)](#) in that target maps and cue maps were fed into the current model at the same time in the current simulations. Research has suggested that target information is generally available to the visual system more quickly than cue information (of this type) because cue information requires extra processing time and, in this case, memory for the contiguity of objects. In a sense, the target information in this example is considered a bottom-up source, and the cue a top-down, although these distinctions are debatable. [Trappenberg et al. \(2001\)](#) simulated this with models in which input from bottom-up (exogeneous) were imputed to the models with top-down (endogeneous) added only after a delay. In the context of the current simulations, these timing differences did not appear to significantly alter the results. In addition, it's difficult to say to what extent the power of the “cue” has or true timing differences. For example, participants were told the target “chimney” before shown the picture, thus, there might be advanced top-down

priming of “chimneys on houses – look for house” before the task even begins. Also, the target is not a true exogeneous or bottom-up source in the traditional sense. Usually exogeneous info or salient info is that which is expected to automatically capture the visual system by standing out from the background. In these cases, no attempt was made to make the target (“chimney”) any more salient, in fact, care was taken to insure there was no strong pop-out effect ([Eckstein et al., 2006](#)). Thus, in the current simulations, there was no clear reason why information from one source should arrive sooner than the other, and target info and cue info information were always imputed into the models at the same time.

(2) In a naturalistic search task such as the ones used by [Eckstein et al. \(2006\)](#) it would be quite likely that when searching for a chimney in a picture of a house with a yard, that that chimney itself would create a strong but very localized region of activity (or saliency), after all the chimney does occupy much space, and it should have all relevant features of a chimney by definition. The roof, and upper region of the house, would also provide a sort of information (or saliency) cue, but this cue would be more diffuse (roofs occupy more space) and less potent/salient (a roof is less guaranteed to contain a chimney than chimney is likely to be a chimney). Thus, our simulations differentiated between target (e.g. chimney) and cue (e.g. roof) inputs by creating all target inputs with a large amplitude (strength) and small diameter (size), and all cue inputs with a small amplitude (strength) and large diameter (size). To put it another way, our targets were designed as strong inputs to a small spatial region, while cues were more diffuse (weaker influence but to a wider area).

Specifically, target inputs had an amplitude (strength) ≈ 15 , and a size $= 1\sigma_r$. Cue inputs, on the other hand, had an amplitude (strength) ≈ 12 , and a size $= 2\sigma_r$. Target and cue strength were modified for some examples (as stated in the results section) but their relative sizes were maintained for all simulations. Further, for these simulations we used $51 * 51$ nodes ($nm = 51$), with the resolution in degrees (dx) taken as $dx = 2\pi/(nm)$. However, the nature of these models does not depend on the number of nodes as long as there are enough nodes to sufficiently approximate the spatial continuum. With fewer or more nodes the principal behaviour of these models would be identical.

(3) The Competitive Integration Model of [Godijn and Theeuwes' \(2002\)](#) described how a fixation point is salient at the beginning of trial (when nothing else is present) but then needs to be actively inhibited when stimuli appear so that a participant can effectively disengage from the fixation to attend/move their eyes to other regions of space. This active inhibition of previously attended locations is an important feature of any model of eye saccades and eye trajectories (cf. literature on inhibition of return for example [Klein \(2000\)](#) and [Lupiañez, Klein, and Bartolomeo \(2006\)](#)). However, the current simulations were only interested in the initial saccade to a visual scene, and thus this sort inhibition was unnecessary in the current model. Thus, the current simulations did not include active inhibition of previously attended /uninteresting locations, and the only inhibition present in this simulations were those inherent in the lateral connections between the nodes (all nodes inhibit distant locations). This feature, however, could easily be added to future models, and would be important for any simulations considering subsequent saccades. It is worth reminding the reader, however, that the current model still includes inhibition, as each location in the network still inherently inhibits all distant locations through lateral connections.

(4) Although we did not include active inhibition at fixation, we did include what [Eckstein et al. \(2006\)](#) called a “cost function”. The cost function relates to the behavioural observation that during eye saccadic tasks to targets participants will favor short

saccades over long saccades. One explanation for this has been that longer saccades require more energy and therefore, when possible, shorter saccades are favored as a sort of “cost” saving mechanism. CSNF models provide an alternative theory for short saccades that perhaps shorter saccades are result of inefficient inhibition of fixation, or the fact that previous activity at fixation inherently cues near-by locations. In any event, a “cost function” was added to the simulations through the simple addition of a very diffuse (large diameter and small amplitude) input centered at fixation. Specifically, the “cost function” was an additional input to the model centered in the middle of the visual array with a amplitude (strength) = 10 and a diameter (size) = $12\sigma_r$.

(5) The visual system is very noisy and, in a true naturalistic search task, other visual stimuli can share visual properties with the target. Thus, any biologically plausible model is incomplete without some degree of noise. Noise was added to the current simulations by adding a normally distributed random variable to all locations in the model. Specifically, noise was added by multiplying each location in the external input ($I^{\text{ext}}(\vec{x}, t)$) by $1 + N_x$, where N_x for each location was chosen randomly from a normal distribution with $\mu = 0$, $\sigma = 0.5$. For single trial simulations, this noise was excluded from figures to more clearly illustrate the principle behaviour of the model. However, for all simulations over a large number of trials noise was present for each trial.

3. Results

We simulated experimental conditions of search in natural scenes as described and studied by Eckstein et al. (2006). These experiments consisted of trials where targets were either absent, present at the expected location, or present at an unexpected location. Simulations of these trials are discussed in turn.

3.1. Simulating target absent trials

In naturalistic search a target absent trial would be when a participant is asked to make an eye saccade to a chimney and then they are shown a picture of a house (cue) without a chimney (no target). Since the only salient input to the CSNF model in the “target absent” case is a cue, then this entire condition can be modeled with a CSNF that has a single input (Fig. 2A and B). Since the CSNF model is built retinotopically, activity in the network will begin increase the fastest at the location in which the input is presented (Fig. 2C). As the network is allowed to progress through time, activity around the input excites the neighbors while inhibiting distant locations and a winning bubble will emerge (Fig. 2D). This winning bubble or activity packet is where we would predict the eyes to move. The examples shown below (Fig. 2A–D) are all shown in a noiseless case. In simulations of average reaction times (RTs) we included stochastic processes by random input that represents noise in the system and unrelated objects in the environment.

Fig. 2 shows snapshots of a 2-dimensional CSNF model responding to a single input at different times. However, it is also possible to graph the complete activity of nodes in the network over-time. Of particular interest is graphing the growth in activity of the node at the winning (in this case “cued”) location compared to other locations. The time evolution of the internal states, u , of the neural fields are shown in Fig. 3A while the external states, r , are shown in Fig. 3B. The internal state of the winning node (region) begins rising, slowly at first, but eventually recruits enough neighbors that the winning region (cue) takes over the whole network rising quickly to its maximum state. The internal states of other nodes are suppressed (Fig. 3A). The external state (firing rate) of the winning node (cue) follows a similar trend of rising slowly at first and then taking off all at once (Fig. 3B). In this example, this sudden take-over appears to happen at around

$t = 20\tau$, where τ is an arbitrary time scale. In fact, if we set an arbitrary rate threshold at 0.8, with all firing rates are between 0 and 1, then we label the point at which any node/region passes threshold as the reaction time (RT) or the time it would take to initiate an eye saccade (saccade latency). The RT or saccade latency is shown to be about $t = 20.5\tau$ in this case. This simulated RT can be related to a RT of human subjects by choosing a specific time constant and by taking further sensory and response latencies into account. Here, however, we are mainly interested the relative pattern of RTs across the different conditions.

The time to reach threshold will depend on a number of factors including, but not limited to: (1) the strength (amplitude) of the input to the network, (2) the size (radius) of the input (3) the number of inputs to the network, (4) the amount of noise added to the system, (5) the level of excitation and inhibition in the model and (6) where the threshold is chosen.

3.2. Target present at expected location (simulated) trials

In the naturalistic search example, a trial with a target present at the expected location would be one in which participants are asked to search for a chimney, and then shown a picture of a house with a chimney on it. In a CSNF model this is simulated by two inputs either very close to each other, or right on top of each other. Perhaps, not surprisingly, when two inputs in this model are beside each other or very close, they add together to make very strong input at that location for the model (Fig. 4A). The result is a strong response from the network right from the start (Fig. 4B). Just like in the target absent case, an activity bubble will develop (Fig. 4C), although this time much faster. This speed is evident when the firing rate of the winning node is graphed over-time (Fig. 4D). In this case, threshold was reached for the winning node (cue and target together) at $t = 3.5\tau$, more than five times faster than in the target absent case.

If the target and cue are close together, but not exactly in the same location the network response is similar, but the extra distance costs the network in terms of speed. In this example the “target” was differentiated from the “cue” by being defined with a larger amplitude/strength but smaller radius/size. Thus, when the target and cue are shown to the system, the target appears as a brighter smaller spot than the cue (Fig. 5A). At first, the network representation of the input looks similar to the input itself (Fig. 5B). However, as locations in the network excite near-by locations and inhibit distant locations the characteristic winning activity bubble develops (Fig. 5C). This activity bubble will develop centered at the “target” location, the “cue” location, or somewhere in between (which would constitute an averaging saccade). In this particular example, the winning bubble appears to be centered above the target (Fig. 5C–D). The firing rate of the winning node or target rises, as usual, although at a slower rate than when the target and cue are at exactly the same location (Fig. 5D). Further, Fig. 5D also shows that node representing the cued location increases firing at a slower rate than the target location.

3.3. Target present at unexpected locations (simulated) trials

In the naturalistic search example, a target present at an unexpected location would be one in which a participant was asked to look for a chimney and then shown a picture in which the chimney was not on the house but in a different location (such as in a tree, or on the ground). The CSNF model simulates these trials with two inputs (a cue and a target) in different locations. When the cue and target are at different locations only one can “win”, that is, only one location can attract the eye movement. The winner in these models will depend upon the distance between the cue and the target, as well as the relative strength of the cue

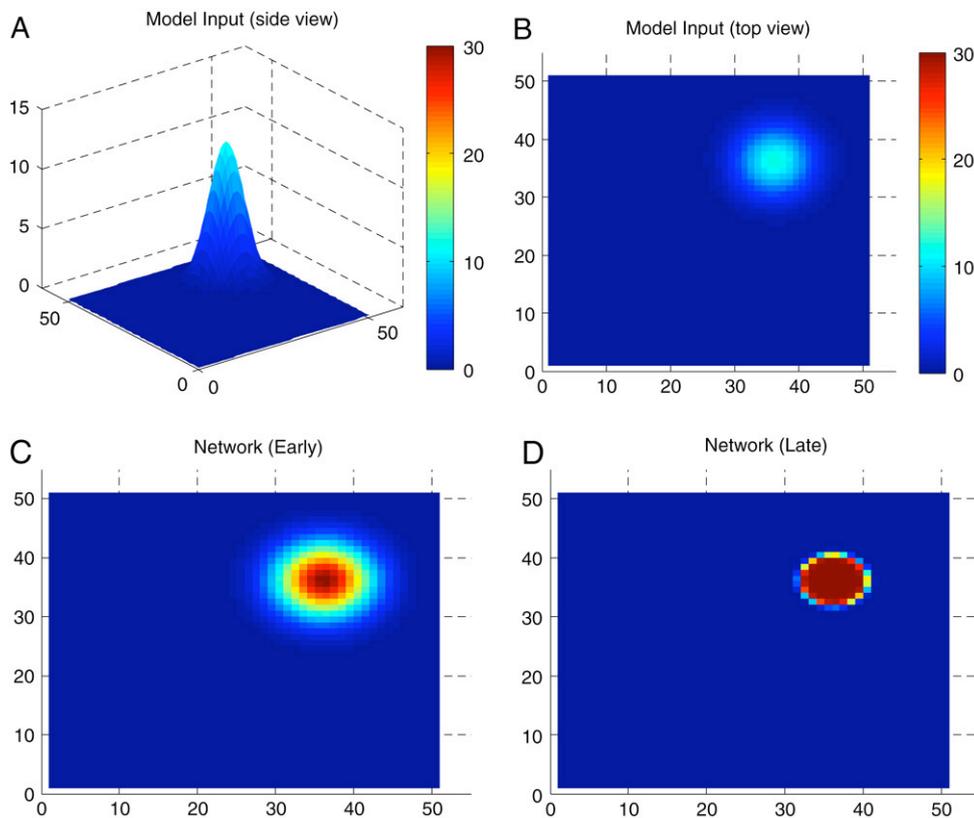


Fig. 2. A CSNF model with a single input to the model shown from the side (A) and from the top (B). The model responds with increased activity at the location of input (C). As time progresses they activity grows into a winning bubble inhibiting activity at all other locations (D). These figures show a noiseless case.

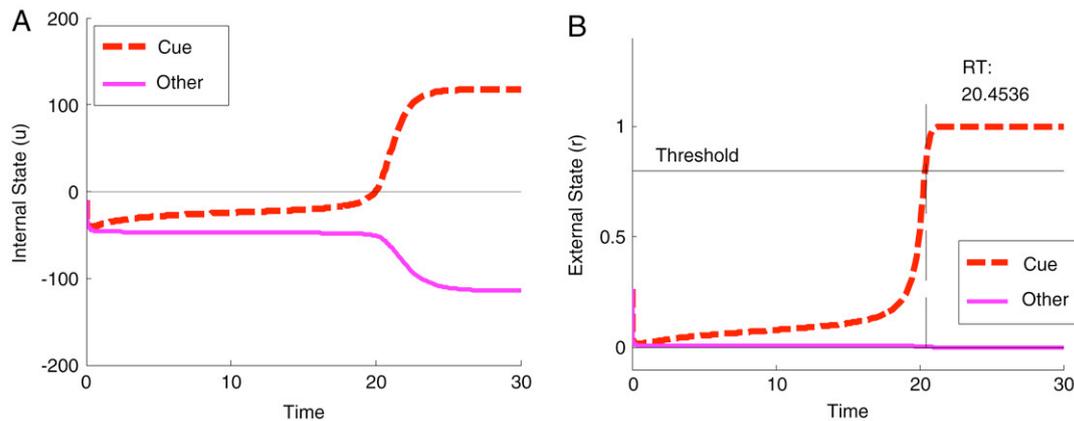


Fig. 3. Activity of select nodes in the single input case (time on the x-axis). (A) Represents the membrane potential of the cue or winning node (dashed) and another random node (solid). (B) Represents the firing rate of the cue or winning node (dashed) and another random node (solid). With a firing rate threshold set = 0.8, relative predictions about reaction times can be made.

and target. As stated earlier, the cue was set to be larger than the target (occupy more of the visual field) but the target was set to be stronger in amplitude than the cue. The relative amplitude of the target and cue can then be influenced by noise, so that sometimes the target will win, and sometimes the cue will win. Fig. 6A–B show an example of a case where the cue (top-right) is twice as large as the target (bottom-left). In addition, the amplitude of the target and cue are close enough in strength that the system resolves itself for the cue to win (Fig. 6C–D). Notice, however, that this system takes much longer to resolve itself than in the target at expected location condition.

Any parameter sufficiently changed, such as amplitude or size, will change the dynamic of the system. For example, if the size or amplitude of the cue was increased, the system would resolve

more quickly (faster RT). If the size or amplitude of the target was increased, then the system would resolve less quickly. Eventually, if the strength of the target is increased enough the target will win instead of the cue. Fig. 7 is an example of targets and cues in the same location shown in Fig. 6, only now the strength (amplitude) of the target has been increased so that it is the winning node. Again, this system does not resolve as quickly as the other non-competitive conditions ($RT = 12\tau$, Fig. 7D). The more unbalanced the balance between cue and target, the faster the system resolves. When the cue and target are almost perfectly balanced the system takes a long time to resolve, and is very sensitive to noise. When the two are almost balanced then the addition of noise to the system will tilt the balance so that sometimes the cue will win, and sometimes the target will win. A consequence of this is that

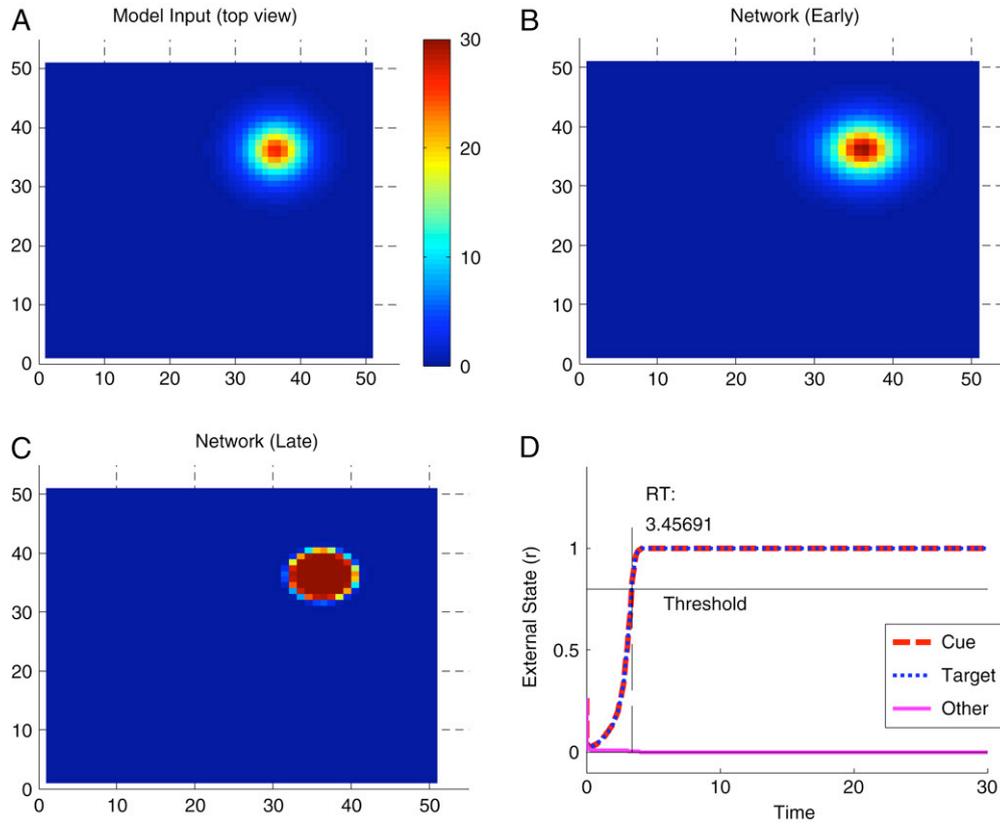


Fig. 4. Target present at expected location. (A) The input to the system. (B) The initial response of the system. (C) The system after it has reached at equilibrium state ($t = 30$). (D) The firing rate of the winning location over time.

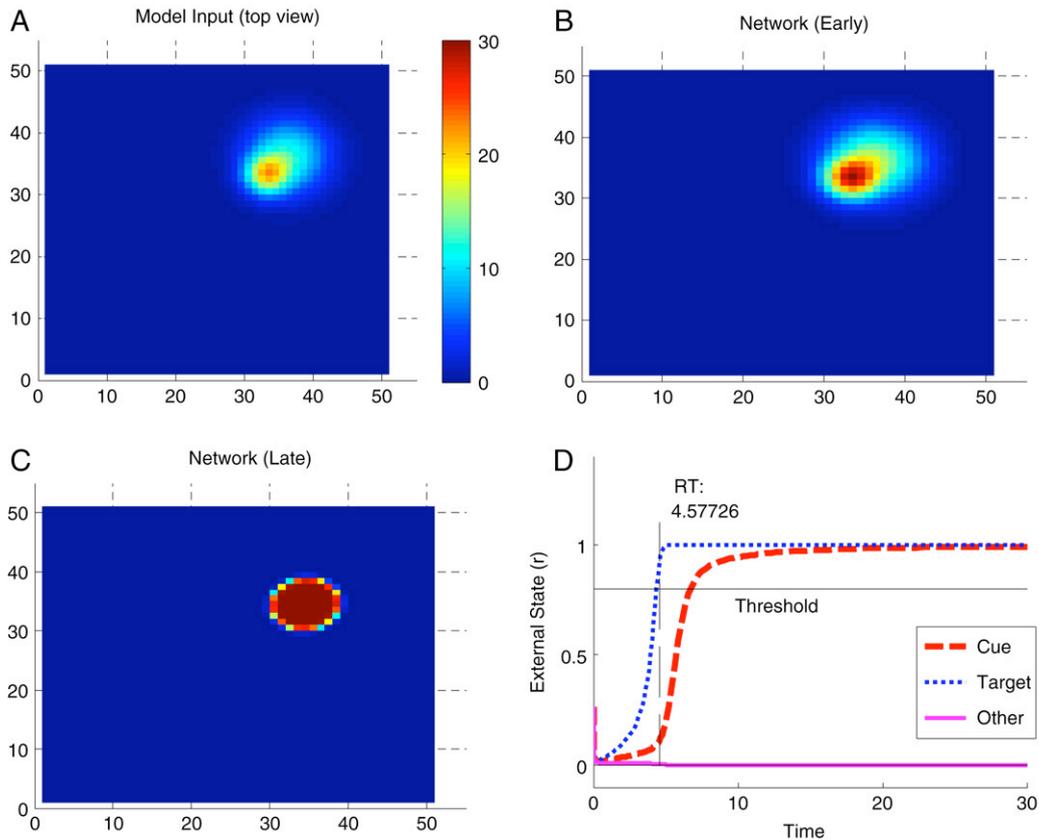


Fig. 5. Target present NEAR the expected location. (A) The input to the system. (B) The initial response of the system. (C) The system after it has reached at equilibrium state ($t = 30$). (D) The firing rate of the target (dotted) reaches threshold faster than the firing rate of the cue (dashed) meaning that an eye saccade would be generated to the target. Other locations (solid) are inhibited by all this activity through lateral connections.

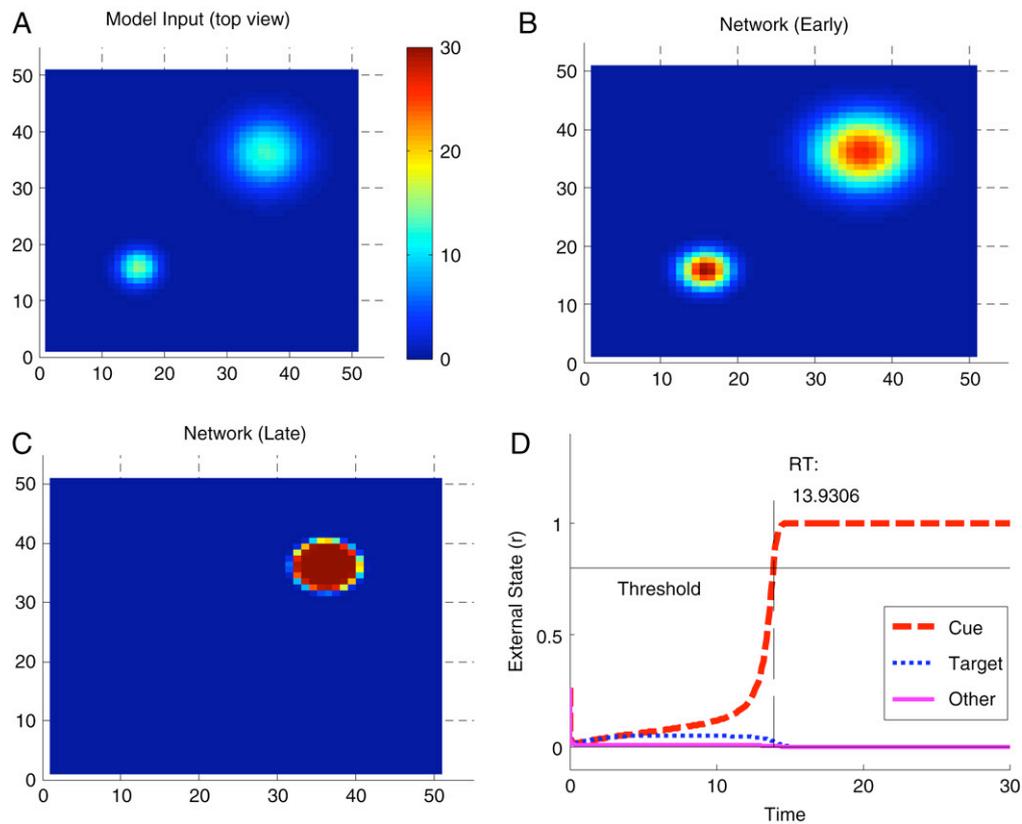


Fig. 6. Target present at the unexpected location (Cue Amp = 13, Target Amp = 15). (A) The input to the system. (B) The initial response of the system. (C) The system after it has reached at equilibrium state ($t = 30\tau$), the cue has inhibited activity at the target location. (D) The firing rate of the cue which also the winner (dashed), the firing rate of the node at the centre of the target (dotted), and another distant location (solid).

the average error (the distance between the first saccade end-point and the target) will be larger in target present at unexpected location condition compared to the other conditions.

3.4. Accuracy over many trials

When Eckstein et al. (2006) considered behavioural data under these different experimental conditions, their focus was on accuracy / error data, not speed. They measured the distance from the end-point of the first saccade to the target (only the first/initial saccade) was considered. Shorter average distance (over many trials) implied a smaller error, and therefore more accuracy. In the target absent case distance measures were taken from the end-point of the initial saccade to the location of where the target would have been if it were present in the expected location. Behavioural data collected by Eckstein et al. (2006) showed that participants were the most accurate in the target present at expected location condition and the least accurate in the target present at an unexpected location condition. This was also the pattern of results predicted by their Bayesian model.

The CSNF model can also be used to predict accuracy over many trials. This was done by considering the average Euclidean distance for many simulations under the different conditions. In the noiseless case, the network performs the same way given the same starting conditions. That is, without noise, the model is either always right or always wrong. This is, of course, a poor analogy to the real visual system that is full of noise and competing stimuli. Thus, to get more valid predictions, noise was added to each location in the model. The result of these noisy CSNF models was that target present at expected location trials were the most accurate as the combined strength of the target and cue occurring together creates a strong signal to the network that

is not easily disrupted by noise. Target present at unexpected location conditions, on the other hand, were the least accurate as sometimes the cue, which is in a different location, will win instead of the target. The percentage of the time that either cue or target will win depends on the relative strength of the two. In our implementations, that cue and target were set to have approximately equal strength such that the noise would sometimes enhance the target and sometimes enhance the cue.

Fig. 8 shows the predicted accuracy for the different trial conditions. From left to right these conditions were target present at the expected location (PE), target present at the unexpected location (PU), target absent at the expected location (AE), and target absent at the unexpected location (AU). The last “condition” (AU) comes from representations by Eckstein et al. (2006) and deserves some explanation. Eckstein et al. (2006) took distance measures for target absent trials – a distance measure from the eye saccade to where the target would be if it were at the expected location (AE) and from where the target would be at the unexpected location (AU). However, since one would never expect participants to look at unexpected locations especially when there is nothing there to attract attention, you would never expect much accuracy in target absent at the unexpected location condition (AU). Thus, the distance measures in these cases were essentially identical to taking the average distance measure between any two random points. The result is that the average distance measure for these cases indicate the average distance between any two points in the visual array which works out to be roughly half the size of the visual array.²

² Actually, it's a little less than half the size of the original visual array because targets are always chosen to be somewhat off-set from the border (never at the extreme edge).

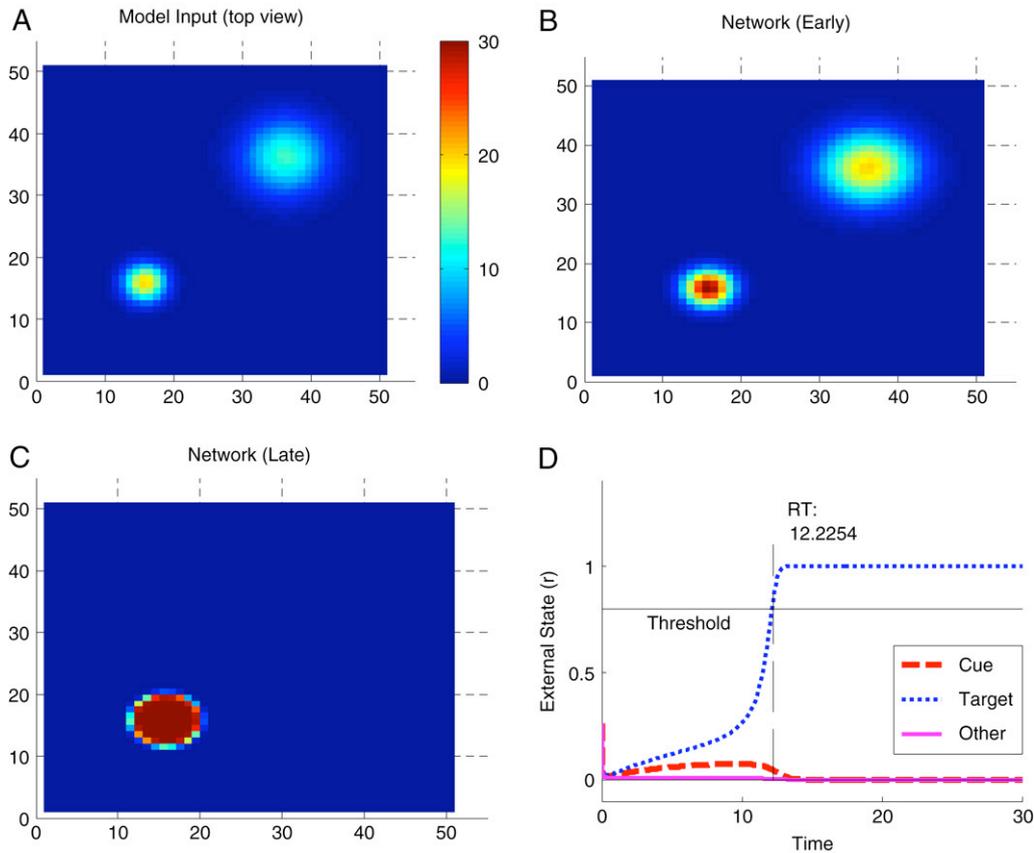


Fig. 7. Target present at the unexpected location (Cue Amp = 13, Target Amp = 20). (A) The input to the system. (B) The initial response of the system. (C) The system after it has reached at equilibrium state ($t = 30\tau$), the target has inhibited activity at the cued location. (D) The firing rate of the target which also the winner (dotted, blue), the firing rate of the node at the centre of the cue (dashed, red), and the firing rate of another distant node (solid, pink). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

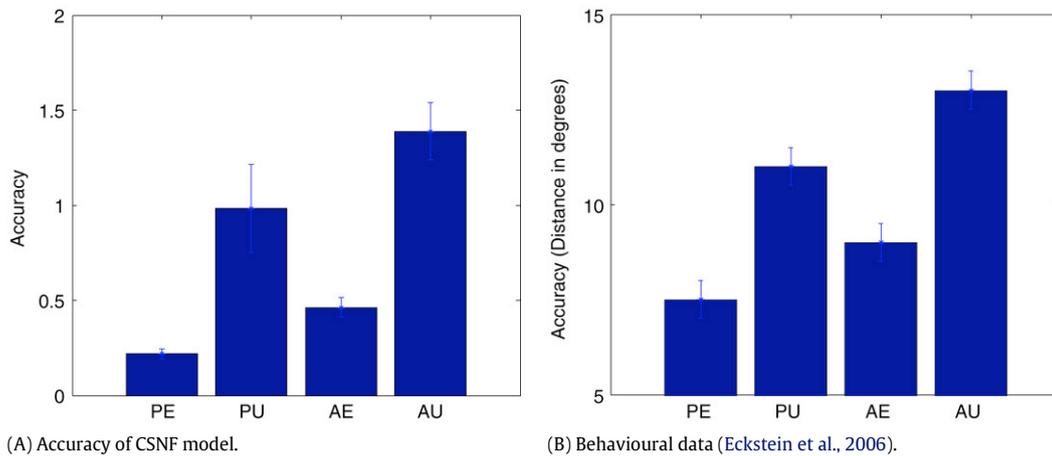


Fig. 8. CSNF model accuracy data averaged over many trials ($N = 80$, 20 per condition). Using a CSNF model with noise to reproduce the relative pattern of accuracy data observed in real participants in Eckstein et al. (2006). Note, this exact pattern depends on the presence and balance between the cost function, noise, and relative strength of the cues and target.

In some ways the errors for the target absent at an unexpected location are meaningless, but they offer a good comparative scale in which to compare the other conditions. For example in Fig. 8, the errors in target unexpected location (PU) are lower than the errors in the target absent at unexpected (AU). This means that participants were finding the target at better than chance in the target at unexpected location case (PU). Individual trials for condition PU saccades do seem to be generally directed either at the cue or at the target. Thus, the saccade distributions for these PU

trials was approximately bimodal, resulting in low accuracy (large errors from the target).

In the target absent at expected location case (AE), errors were low but not as low as in the target present at expected location (PE) trials (see Fig. 8). That reduced accuracy in target absent trials occurs simply because these trials are less robust to noise as there is only one input to the model. That is, an eye saccade to the correct location (in AE) is more likely to get disrupted by noise (an incorrect signal) in the system. Thus, target absent or AE trials

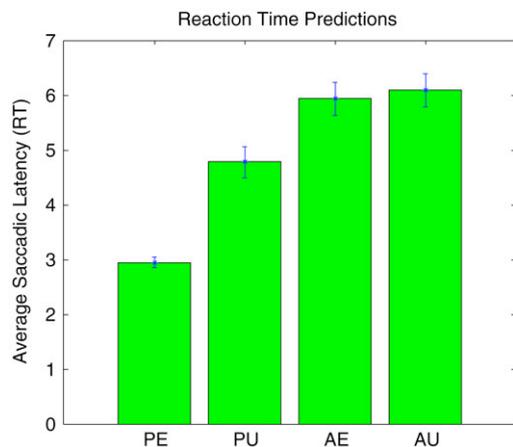


Fig. 9. CSNF reaction time (RT) or saccadic latency predictions when averaged over many trials ($N = 80$, 20 per condition). In this figure the last bar, target absent at unexpected locations (PU), has no meaning and is completely redundant with bar 3 (AE). This last bar was maintained in the figure for reasons of symmetry with Fig. 8.

(with only a single input) are more sensitive to noise in the system than target present at expected or PE trials.

3.5. Speed over many trials

Generally the CSNF model, like the visual system, is accurate where it is fast. However, we can see from the predictions of the model that there are some exceptions to this rule – in particular the target present at the unexpected location (PU) condition. As we saw in Fig. 8, this condition has the lowest accuracy compared to the two other conditions, however the CSNF model predicts reaction times (RTs) in this condition that would be faster than RTs in target absent trials (Fig. 9). Recall the difference between these two conditions is that target absent (AE and AU) trials have only one input to the model, where target present at unexpected location (PU) trials have two competing inputs. Thus, on average, the network predicts that resolution between two competing inputs will occur faster than the build-up of a single input (Fig. 9).

The CSNF model predicts accuracy and speed differences across these conditions. In the case of accuracy, the CSNF predictions match those of Eckstein et al.'s (2006) Bayesian model and the pattern of results they observed in human participants. In the case of speed, however, the CSNF model predictions did not match Eckstein et al.'s (2006) behavioural data (their Bayesian model did not make speed predictions). Specifically, Eckstein et al.'s (2006) results indicated no significant difference in saccade latencies between the conditions, with average latencies between 225 ms and 236 ms ($p = .997$). Further, these saccadic latencies are comparable, if not a little longer/slower, than saccadic latencies from other research. For example, Godijn and Theeuwes (2002) reported shorter latencies to distractors ($M = 161$ ms) than to targets ($M = 223$ ms) during a simple onset distractor task with six possible target locations. However, the lack of match between behavioural data (no latency differences) and CSNF model predictions (predicted differences) was concerning. Thus, one more simulation was considered.

3.6. Time constrained CSNF models

It is possible that the lack of difference in saccade latencies for the three conditions was related to search strategies employed by the participants. For example, searching in a naturalistic task (color photographs) may have been a sufficiently difficult task that participants began moving their eyes as soon as the picture appeared. However, the accuracy data suggests that participants

must have allowed enough time for at least some information about the “cue” and “target” to reach their target likelihood map before making a saccade (otherwise accuracy would be equivalent in all conditions). Thus, the data suggests the participants must be waiting until they have some information about the visual scene before making an eye saccade, but perhaps their strategy was to acquire a bare minimum amount information before making a saccade. We simulated this search strategy by programming a time-constrained CSNF model that forced the system to make pre-threshold saccade/decision regardless of the trial type (Fig. 10).

The CSNF time-constrained model was told to make a saccade (choose a winning location) after only $t = 2.5\tau$ (recall that the fastest RTs previously in a no-noise condition were $t = 3.5\tau$). Thus, the RTs for each condition were identical (RT constrained to 2.5τ), and no simulation ever reached threshold (firing rate of 0.8 or above). Interestingly, in all trial conditions the firing rate for the winning node was already beginning to differentiate itself from the other nodes as early as 2.5τ . In fact, it appeared, from the gradual increasing slopes of the firing rates, that as early as 1.0τ may have been enough to accurately predict the winning node (Fig. 10A–C). In addition, when many trials of this time-constrained CSNF model were averaged together the accuracy pattern appeared exactly as it did before (Fig. 10D). This suggests that a CSNF model can be forced to make a very early decision and the result is not noticeably different than when the model is given all the time it needs. If the CSNF model is a good analogy for how the biological visual system behaves in this specific experiment, then this suggests that the participant did not wait until certainty is reached but acted as soon as one area of the target-likelihood map rose to an activity level slightly higher than the others. That is, the system may act on what little information is available, and does so with a generally high degree of success.

3.7. Multiple targets

Our last simulations considered the extension to cases with more than one target. First, we considered the condition with eight targets spread equidistant around the fixation (or centre). The simulation figures show the input to the model (without noise) with initial saccade endpoints for each run/trial superimposed as black dots on top. A small degree of jitter was added to the plotting to partially simulate motor error but mostly to decrease the chances of multiple dots appearing in one location and distorting the apparent results (the sense of where the model “looked”). In the trials, noise was always added to input, otherwise all saccades would have been to the same location.

At first past, all targets inputs used the same diameter as before, but a lower strength / amplitude (Target Amp = 2). The result of the initial simulation (Fig. 11A) showed significantly fewer averaging saccades compared to the behavioural results obtained Chen and Zelinsky (2006). However, by simply increasing the width of the interaction kernel (from $\text{sig} = 2 * \pi / 12 * 0.5$ to $\text{sig} = 2 * \pi / 12 * 0.7$, Fig. 11B), the diameter/size of the targets (diameter from $1 * \text{sig}$ to $2 * \text{sig}$, Fig. 11C), or increasing both at the kernel and diameter of targets (Fig. 11D) the model predicted many more averaging saccades.

Adjusting these parameters to produce Fig. 1B–D was akin to either (1) lower visual resolution of the image, in the case of increasing the target diameter or (2) increasing the radius of the centre-surround mechanisms, in the case of increasing the lateral interaction kernel. Both seemed like justifiable adjustments to the parameters as research suggests (1) fast decisions such as eye saccades are probably done on lower resolution images. For example, recent modeling research by Rao, Zelinsky, Hayhoe, and Ballard (2002) suggests that processing of visual stimuli by humans likely is done on a coarse-to-fine scale. In addition, (2) stimuli

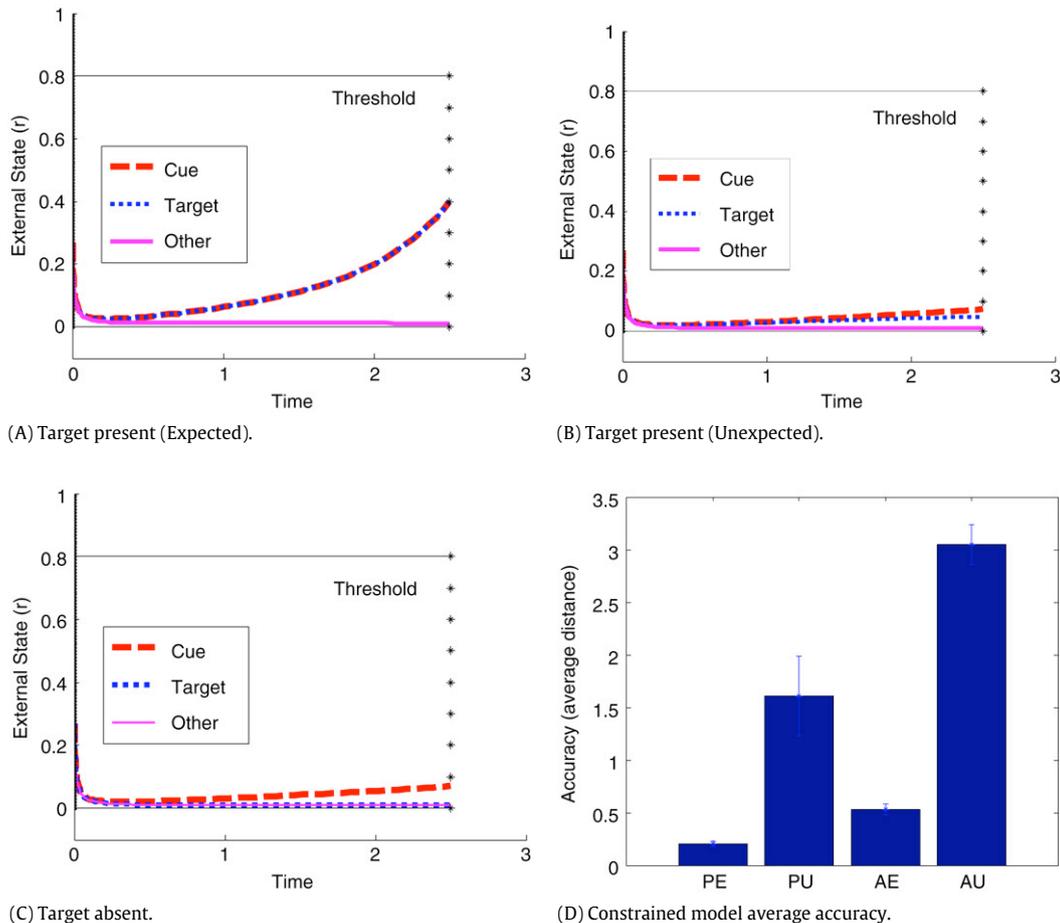


Fig. 10. A CSNF model in which the system is forced to make an early saccade. (A) In the target present at expected location the winning node (cue and target) and the target are quite clearly distinguishing themselves from the cue. (B) In the target present at unexpected location we can see the cue (dashed) is winning over the target (dotted) in this case. (C) In the target absent case, the cue (dashed) is already starting to rise above the others even after only $t = 2.5\tau$. (D) Averaging this constrained CSNF model over 80 trials. The accuracy results are virtually identical to the non-constrained CSNF model from before.

further in the periphery fall on regions with larger receptive fields. Thus, both parameter adjustments seemed reasonable. The choice of interaction kernel size will largely depend on the eccentricity of targets, and perhaps should be adjusted in future models to vary as a function of eccentricity (instead of being the same value at all nodes – as in the current simulations).

Next, we considered the case where two of the items in the ring were targets and the rest were distractors. We looked at both the case where the two targets were at adjacent locations (Fig. 12A), and the case where targets were separated by a distractor (Fig. 12B). Both simulations were done with 100 trials, and both the interaction kernels and size/diameters of the targets were kept large as before (see Fig. 11D). The result was a pattern of averaging saccades that appeared to be almost identical to the pattern seen in human data (cf. Findlay (1997)). That is, averaging saccades predicted by the CSNF model appeared to be equally frequent. With a smaller interaction kernel and/or targets of a smaller size, fewer averaging saccades were observed.

Finally, simulations with 16 objects in which the target was located either near (Fig. 13A) or far (Fig. 13B) from fixation were considered. The primary difference between the near/far model predictions was more accurate saccades in the near condition (Fig. 13A, in particular). In the far condition, inaccurate saccades went to near not distant distractors. This pattern was consistent with Findlay's (1997) behavioural data. In these particular simulations, targets with a smaller strength and a lower interaction kernel appeared to better capture the behavioural pattern.

In summary, the CSNF models, with the proper parameters, were able to replicate the pattern of initial saccade endpoints for multiple targets and distractors in numerous conditions.

4. Discussion

4.1. Summary

This research shows that Centre-Surround Neural Field (CSNF) models can account for the pattern of results predicted from behavioural experiments by Eckstein et al.'s (2006). That is, a CSNF model predicts better accuracy (in terms of eye saccades) when a target is at an expected location. Further eye saccades are more likely to go to a contextually cued location than the target at an unexpected location as indicated by lower error rates (higher accuracy) in the target absent (but cue still present) condition compared to the target present at an unexpected location condition. The CSNF model is consistent with the Bayesian model by Eckstein et al. (2006), and can be viewed as a biological plausible implementation of this model. The CSNF model has an additional advantage over models described by Eckstein et al. (2006) in that the CSNF model is a dynamic model that explains the time component allowing for predictions of saccadic latency to be made. The predictions made by a CSNF model in these conditions are also generalizable to more traditional experimental designs, as shown in work by Godijn and Theeuwes (2002) as well as Trappenberg et al. (2001).

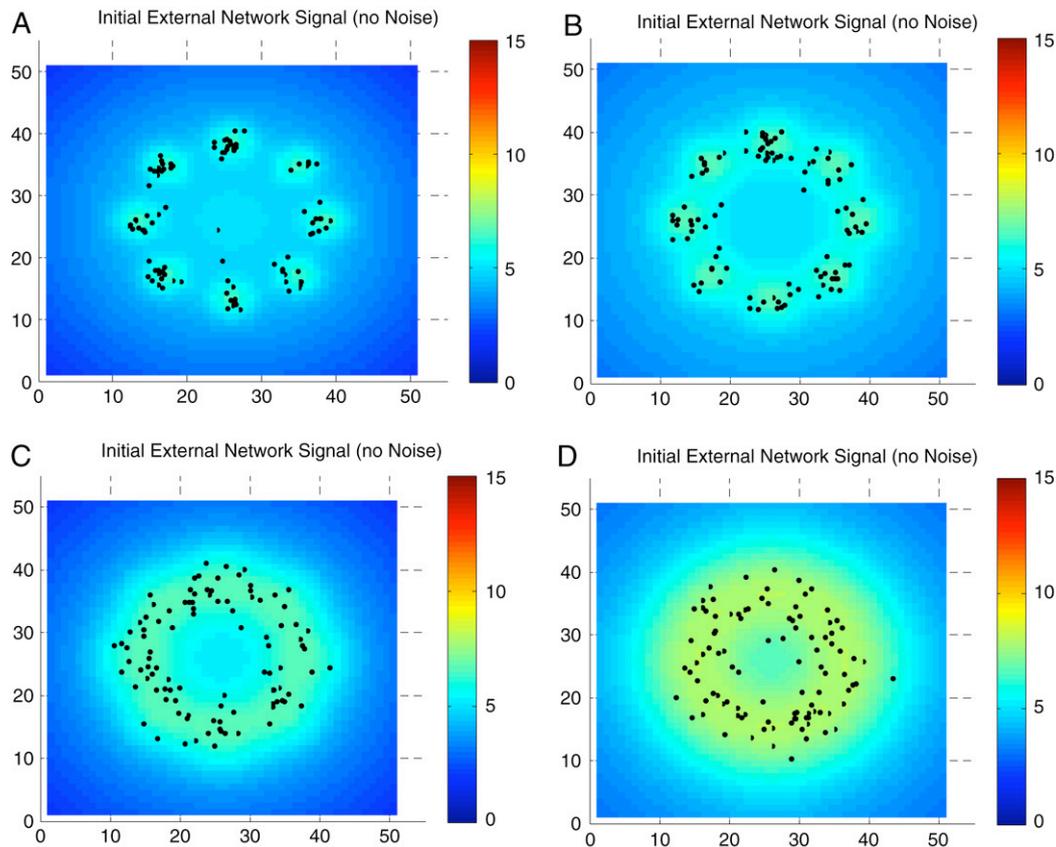


Fig. 11. Initial saccade endpoints (dots) overlaid on model input consisting of eight equal strength targets ($N = 100$). Shown here is the default model (A), the same thing with an increased width of the interaction kernel (B), increased size of targets (C), or both increased width of the interaction kernel and size of targets (D).

While our model predicted the same pattern of accuracies in the different search tasks as the Bayesian model and human data in Eckstein et al. (2006), the results on reaction times of human subjects found by Eckstein et al. (2006) offer an interesting suggestion on the processing strategy in humans. While the CSNF model shows that resolving discrepant information takes longer than processing matching input, Eckstein et al. (2006) found that reaction times were similar in all conditions. However, as discussed in the paper, it is possible to resolve this issue by assuming that the participants used a search mode in which they acted in a fixed time on minimal information. Interestingly, we could further reproduce this accuracy pattern in a modified (time-constrained) model based on the fact that competitive interactions already have a measurable effect in a short time. If the hypothesis of the different search mode is correct, then it should be possible to design an experiment in which participants are forced to adopt a different search strategy resulting in different saccade latencies for different conditions. For example, perhaps a reward for correct initial saccades would encourage participants to adopt a more conservative search strategy and wait longer to make eye saccades (leading to different saccadic latencies for different conditions). The different strategies of fixed-duration (FD) and free-response (FR) task are also discussed and Zhang and Bogacz (2008) and were used by Trappenberg (in press) to interpret the data of Roitman and Shadlen's (2002).

Our CSNF models further captured the qualitative flavor of initial eye saccades in conditions with multiple targets and distractors. First, with eight targets equidistant from fixation, the first simulation predicted most saccades to targets. However, by using a slightly higher interaction kernel or larger diameter targets more averaging saccades were predicted. Second, the CSNF model appropriately captured the pattern of initial saccade endpoints for

two targets (cf. Findlay (1997)). Third, our CSNF model was able to show the expected pattern of initial eye saccades to a manipulation of target distance. Tuning of parameters such as target diameter size or size of the interaction kernel seemed particularly important to the number of averaging saccades that were produced. This is perhaps not surprising because eccentricity (distance from fixation) is particularly important for human participants. That is, receptive fields get logarithmically larger as one moves further away from the fovea (Van Gisbergen, Van Opstal, & Tax, 1987). Since larger interaction kernels (related to size of receptive fields) result in more averaging saccades, one would predict an eccentricity effect on averaging saccades.

Future CSNF models may want to consider implementing a lateral interaction kernel that changes based on its distance from the fovea/fixation. On the other hand, larger receptive fields result in lower resolution, and in fact research by Rao et al. (2002) has suggested that processing of visual stimuli is done on a coarse to fine scale. Thus, larger diameter targets may be a legitimate way of conceiving how target information is perceived at early processing stages.

4.2. Outlook

The CSNF model used in this work was intentionally kept to a minimum and was not drastically tuned to match the experimental data. Rather, it was important to show that even this minimal version was able to fully capture an abstracted version of a naturalistic search task described by Eckstein et al. (2006) as well the pattern of results obtained by Findlay (1997) and Chen and Zelinsky (2006). These CSNF models also have a certain degree of biological plausibility. Specifically, the CSNF model's primary mechanism is centre-surround interaction which has been

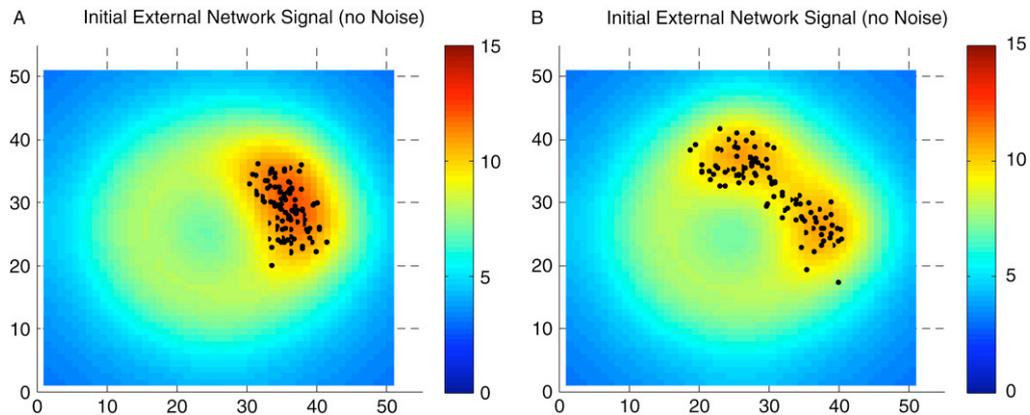


Fig. 12. Initial saccade endpoints (dots) overlaid on model input of eight objects, two of which are targets ($N = 100$). In panel (A) the targets are right beside each other, one at 3 o'clock and the other half-way between 1 and 2 o'clock. In panel (B) the targets are separated by one distractor, with one target still at 3 o'clock and the other at 12 o'clock.

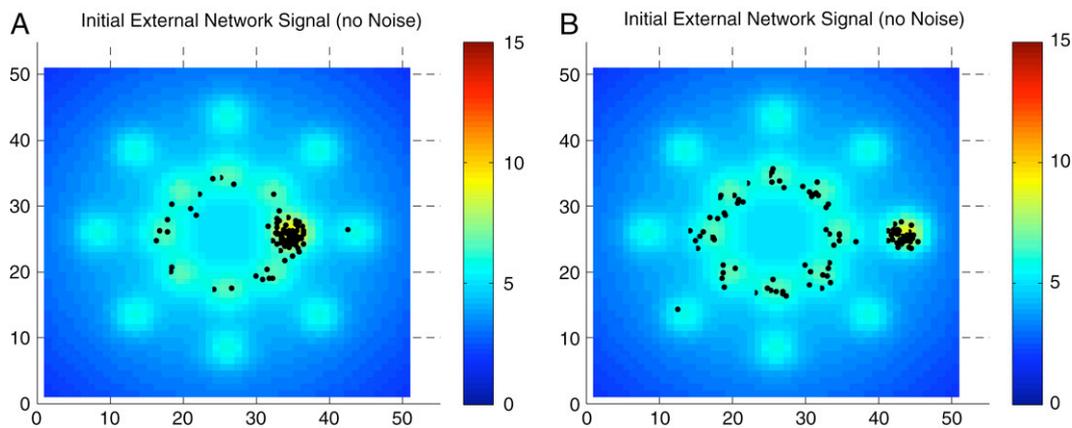


Fig. 13. Initial saccade endpoints (dots) overlaid on model input of sixteen objects, one of which was a target ($N = 100$). Simulations were done with the target at a near 3 o'clock location (A), or with the target at a far 3 o'clock location (B).

observed in the ganglion cells, as well as higher areas including the superior colliculus and frontal eye fields (Schall et al., 2004).

An emerging question in the literature has been about how a brain integrates information from different spatial maps? For example, to use the present example (searching for a chimney), information about the features of a chimney, and information about the object contiguous to chimneys (roofs and houses) are two very different sorts of information. Therefore, it's not unreasonable to expect one region in the brain to contain one map for target likelihood based on features (what Eckstein et al. (2006) called the likelihood ratio map), and another map for the location of the roof, or more generally, locations of high probability (Eckstein et al.'s prior probability map). The current simulations focused on the integration of these two types of information or maps regardless of the source. The only difference in how they were treated was at the input level: target information had a higher strength and was more localized (smaller diameter), roof/cue information had a lower strength and was more diffuse (larger diameter). After that the model integrated this information along with any noise in the input to make a decision. In principle, the sources of information were irrelevant, CSNF models could be applied to information from other/multiple sources.

There remains the concern about the relative timing of inputs. Is it fair to assume the information from two different sources would arrive at the same time? Trappenberg et al. (2001) explored this issue with their model of intermediate layers of the superior colliculus (SC). The SC receives both convergent afferents from many cortical and subcortical visual and cognitive centers related

to eye movement control and is therefore thought to be a candidate location where exogenous and endogenous visual information is integrated. In the SC exogenous information is suspected to arrive earlier through direct retinotectal projections, compared to the endogenous signals that would arise through higher cortical processes. Trappenberg et al. (2001) implemented this timing difference in the model by imputing exogenous information first, and then endogenous after a delay. Trappenberg et al.'s (2001) model was, at its core, a Centre-Surround Neural Field (CSNF) Model that effectively explained a variety of saccadic effects. Their model accounted for saccadic reaction time effects due to removal of fixation, the presence of distractors, execution of pro- versus antisaccades, and variation in target probability. Further, their model suggested a possible mechanism for the generation of express saccades.

In contrast, the current model assumed synchronous processing of inputs. However, in the naturalistic search case inputs, such as chimney location and roof location, neither are clearly exogenous or endogenous, although information about their location probably arrives from different sources in the brain. But, there are many different types of information that need to be integrated for a successful search. It is therefore more reasonable to conjecture that information integration through competitive networks is done on different hierarchical levels in cortical and subcortical areas. Thus, the SC is not the only structure where signals from different sources are combined, and the neural field model can also describe the effective interactions in a hierarchical cortical model. While predictive hierarchical

models, of which Bayesian models are a simple example, are an important description of cortical processes (Friston, 2005), our model captures the dynamics of the integration process not addressed in the model by Eckstein et al. (2006).

4.3. CSNF models and cognitive processes

Schöner has argued and shown for many years how such neural field models of the type discussed here can accurately describe the dynamics of many cognitive processes. For example, Schöner and Dineva (2007) argued that such models can explain the emergence of developmental competencies without the need to pre-suppose innate/dormant modules of knowledge and cognition. They use a dynamic field model to explain how over-coming object perseverance in infants can be explained by an increase in the strength or recurrent connections over development. The development of these connections, although continuous, may result in behaviour that appears categorical, as the system develops competency to perform the task (Schöner & Dineva, 2007). Spencer, Simmering, Schutte, and Schöner (2007) outline a similar spatial precision hypothesis (SPH) that explains how a qualitative shift in geometric category biases can be accomplished through a quantitative shift in model parameters. Thus, dynamic field models can show that the appearance of stage-based learning is really just an emergent property of gradual and continuous refinement of connections between & within neural fields.

The dynamic field theory (DFT) model outlined by Spencer et al. (2007) has seven layers, with one layer each representing: (A) a perceptual field in an egocentric reference frame, (B) a system that transforms locations from egocentric to an object-centered frame, (C) a perceptual field in object-centered reference frame, (D) a long-term memory field associated with (C), (E) a shared layer of inhibitory neurons, (F) a spatial working memory field in object-centered reference frame, and (G) a long-term memory field associated with the spatial working memory field (p. 331). Thus, each layer in this seven layer DFT serve conceptually different functions – most of the time, with some functional flexibility built into the system, especially during training. Connections between layers in this DFT model feed-forward and backward, and connections within a layer act in a centre-surround fashion, with near-by locations exciting each other, and distant location inhibiting each other. Spencer et al. (2007) show that such DFT model may explain aspects of memory such as metric memory (the location/size of objects in space), working memory, and how long-term and working-spatial memories are integrated.

Cognitive processes are continuous in space and time and it is an outstanding question how the constancy of the world arises through specific brain processes. One possibility is that information from different sources (such as exogeneous or endogeneous) are combined in a ‘cognitive map’ of which the model discussed here is a simple example. Certainly growing research suggests dynamic fields may be able to explain how representations of sensorimotor as well as metric and categorical information are maintained over time (cf. Spencer & Schöner, 2003). The appearance of perceptual decisions and the formation of perceptual categories may also be explainable through these models (cf. Schöner & Thelen, 2006). In short, dynamic field models, properly implemented, appear to be very tractable over many domains.

4.4. Conclusion

In conclusion, dynamic field models, particularly those with a centre-surround (local excitation and distant inhibition), are proving to be an effective neurophysiologically plausible, model

```
% 2-d Continuous Attractor Neural Network with Hebbian learning
% Last modified by Joshua Salmon 06-Nov-2007
clear; clf; hold on; close all;

##### Parameters #####
nn = 51; dx=2*pi/(nn); % number of nodes and resolution in deg
inhibit_level = 0.3; % Level of network inhibition
total_time = 30; % number of (total) iterations to use

% The next to parameters are used to choose the simulation condition
present = 1; % target present (=1), target absent (=0)
expected = 0; % target expected or cue and target together (=1), target
unexpected (=0)

% Cost function biases system to short saccades
cost = 1; % cost off (=0), or cost on (=1)
% Noise function adds a bit of noise to all the nodes.
noise = 0; % noise off (=0), or noise on (=1)

tar_strength = 15; % target STRENGTH or amplitude
cue_strength = 10; % cue STRENGTH or amplitude

% TARGET and CUE location between 0 & 2pi (set manually, or chosen
randomly).
x_target= 0.6*pi; %pi*(0.5+rand); % between 0.5 and 1.5 (of 0 to 2)
y_target= 0.6*pi; %pi*(0.5+rand);
x_cue = 1.4*pi; %pi*(0.5+rand);
y_cue = 1.4*pi; %pi*(0.5+rand);

%this extra location is only really used for the figures.
x_extra = 1*pi; y_extra = 1*pi;

##### Setting up the Weights #####
sig = 2*pi/12*0.5; % sigma (=0.6 mm in old paper)
w=w2dgauss(nn,sig,dx); % Symmetric and idiotetic weights

w=50*(w-inhibit_level); %increase or decrease weights to create
discrete point networks that stabilize, go to 0 or go to inf
w=reshape(w,nn*nn,nn*nn);
% note: w has been changed from a 4-D to 2-D matrix with 2601x2601

##### Experiment / Simulation #####
%%% external input to initiate bubble
u0 = zeros(nn)-10;

I_ext=zeros(nn);
ampl=10; sig2=4*pi/20;

##### Here we provide inputs to the system #####
% in_signal_pbc2d(locx, locy, ampl, sig, nn,dx)
I_ext=I_ext+in_signal_pbc2d(x_cue,y_cue,cue_strength,2*sig,nn,dx);

% Set strong signal at target (but only if target is present).
if present == 1;

I_ext=I_ext+in_signal_pbc2d(x_target,y_target,tar_strength,1*sig,nn,dx)
;
end;

% Cost function added if cost = 1
if cost == 1; I_ext=I_ext+in_signal_pbc2d(pi,pi,10,12*sig,nn,dx); end;

% Noise added if noise = 1
if noise == 1; I_ext=(1+0.5*randn(nn)).*I_ext; end;

figure(1) % This figure shows the external input pattern
surf(I_ext,'linestyle','none'); view(0,90);
```

Fig. A.1. An example of the MATLAB code used for a single simulation of a condition. In this example the simulation was for the target present at an unexpected location case.

for aspects of visual cognition as well as other general cognitive processing. Further, as elaborated in Spencer et al.'s (2007) dynamic field theory model, multi-layered versions of these models can be effective for explaining aspects of memory, and even how gradual quantitative shifts in model parameters can result in qualitative changes in behaviour – or the appearance of stage-based learning. The minimal Centre-Surround Neural Field (CSNF) model used in this paper is another example of the effectiveness of these models. Specifically, our CSNF model was able to explain the behavioural accuracy data obtained by Eckstein et al. (2006), as well data by Findlay (1997) and Chen and Zelinsky (2006). Although further behaviour research will be needed to fully resolve the discrepancy between our predictions and the saccadic latency data of Eckstein et al. (2006), we feel, overall, our model captured the essential flavour of initial eye saccades during a variety of tasks.

```

title('External Network Input')
cmin = 0; cmax = 30; colorbar_range = [cmin cmax];
caxis(colorbar_range); colorbar

% Here, we use the gain function "f1" and ode function
I_ext=reshape(I_ext,nn*nn,1);

param=0; tspan=[0,0.0001];
tic; [t,u]=ode45('rnn_ode_u',tspan,u0,[],nn,dx,w,I_ext); toc;

r=reshape(f1(u(size(u,1),:)),nn,nn);
figure(2) % This shows immediate network response.
surf(r,'linestyle','none'); view(0,90); % display the network
activity almost immediately
title('Network immediately after signal presented')

##### Capturing the time dynamic. #####
tspan=[0,total_time];
u0=u(size(u,1),:);
[t,u]=ode45('rnn_ode_u',tspan,u0,[],nn,dx,w,I_ext);
r=reshape(f1(u(size(u,1),:)),nn,nn);
r2 = f1(u); % this captures rate for all loc at each time.

% First, calculate location to which the eye moves...
% Note, this series of equations is doing the Max & Loc
[max_resp,y_loc] = max(max(r)); % this figures out y_location
[col,ind] = max(r); % this flags corresponding x
x_loc = ind(y_loc); % this grabs the x_location

% Thus we can plug the x,y coordinates into eye movement (EM)
EM = [x_loc, y_loc]; % this is the coordinates for the first saccade.

% therefore winning node is at location:
nnum_win = y_loc + nn*x_loc; %node number of winner

x_matrix_t = round(x_target/dx); y_matrix_t = round(y_target/dx);
x_matrix_c = round(x_cue/dx); y_matrix_c = round(y_cue/dx);
x_matrix_e = round(x_extra/dx); y_matrix_e = round(y_extra/dx);

% within u, the location of target & cue should be:
nnum_target = y_matrix_t + nn*x_matrix_t;
nnum_cue = y_matrix_c + nn*x_matrix_c;
nnum_extra = y_matrix_e + nn*x_matrix_e; %same for extra location.

figure(3) % plots the membrane potential over time.
plot(t,u(:,nnum_win),'black','LineWidth',9); % winner
hold on; plot(t,u(:,nnum_cue),'red--','LineWidth',6); % cue
hold on; plot(t,u(:,nnum_target),'blue','LineWidth',3); % target
hold on; plot(t,u(:,nnum_extra),'g--','LineWidth',5); % reference
title('Membrane potential, TARGET (blue), Winner (black), CUE (red),
Reference (green)')

figure(4) % plots the firing rate over time.
plot(t,r2(:,nnum_win),'black','LineWidth',9); % winner
hold on; plot(t,r2(:,nnum_cue),'red--','LineWidth',6); % cue
hold on; plot(t,r2(:,nnum_target),'blue','LineWidth',3); % target
hold on; plot(t,r2(:,nnum_extra),'g--','LineWidth',5); % reference
title('Firing Rate, TARGET (blue), Winner (black), CUE (red), Reference
(green)')

### Calculate Reaction Time (Threshold for the winning node) ###
threshold = 0.8;
[value,thresPoint] = min((r2(:,nnum_win)>=threshold).*r2(:,nnum_win) +
(r2(:,nnum_win)<threshold));

RT = t(thresPoint);
hold on; plot(t,threshold,'k-'); % threshold line
hold on; plot(RT,0:0.01:1.1,'k'); % plot the reaction time
text(RT + 5, 1.1,{'RT: '; RT});

figure(5); % Plots the network final state
surf(r);

```

Fig. A.1. (continued)

The power of dynamic field models, such as the one used in the current design, are becoming increasingly evident.

Acknowledgements

Special thanks to Dr. Klein for his advice and counseling during this research and Dr. McMullen for her continuous encouragement. We would also like to thank Jason Satel and Dominic Standage for their support throughout this project. In addition, we would like to thank the anonymous reviewers for important suggestions in the interpretation of the model and for suggesting experiments with multiple targets.

w2dgauss.m

```

function w = hebb2dGauss(nn,sig,a)
% preset Gaussian weights
%amp=50, c_inh=0.7*amp;
c=1/(sqrt(2*pi)*sig);%*amp;
w=zeros(nn,nn,nn,nn); %slows down without this statement
for x1=1:nn
for x2=1:nn
dx=min(abs(x1-x2)*a, 2*pi-abs(x1-x2)*a);
for y1=1:nn
for y2=1:nn
dy=min(abs(y1-y2)*a, 2*pi-abs(y1-y2)*a);
w(x1,y1,x2,y2)=c*exp(-(dx^2+dy^2)/(2*sig^2));%-
end
end
end
end
return

```

in_signal_pbc2d.m

```

function y = in_signal_pbc2d(lox, locy, ampl, sig, nn, dx)
y=zeros(nn);
for i=1:nn;
dx2=min((i*dx-locx)^2, (2*pi-abs(i*dx-locx))^2);
for j=1:nn;
dy2=min((j*dx-locy)^2, (2*pi-abs(j*dx-locy))^2);
%y(i,j)=ampl/(sqrt(2*pi)*sig)*exp(-(dx2+dy2)/(2*sig^2));
y(i,j)=ampl*exp(-(dx2+dy2)/(2*sig^2));
end
end
return

```

f1.m

```

function f1=rnn(u)
% gain function: logistic
beta = .1;
% alpha=.0-100*(u>0);
% f1=1./(1+exp(-beta.*(u-alpha)));
f1=1./(1+exp(-beta.*u));
return

```

rnn_ode_u.m

```

function udot=rnn(t,u,flag,nn,dx,w,I_ext)
% odefile for recurrent network
r=f1(u);
sum=w*r;
udot=-u+sum*dx+I_ext;
return

```

Fig. B.1. MATLAB code for necessary functions called by the code in Fig. A.1.

Appendix A

This is MATLAB code for a 2-dimensional CSNF model. The main code “cann2d_js2007” requires four functions “w2dgauss.m”, “in_signal_pbc2d.m”, “f1.m” and “rnn_ode_u.m” which are listed in Appendix B.

Main code:

The following is the main code for a single trial simulation. Using the current parameters, this will run a single trial simulation for the target present at unexpected location condition. However, you'll notice that in the parameters sections variables can easily be changed to run the other simulations. For example, to run a target absent expected location trial set variable present = 0, and expected = 1. The simulation can be run with or without noise and the cost function. The target & cue strength and size can also be modified, although the defaults work pretty well.

To get an average over many trials a loop would have to be added to this code. For a copy of the code with an added loop, please contact the authors. See Fig. A.1.

Appendix B

See Fig. B.1.

References

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27, 77–87.

- Brunel, N., & Wang, X. (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *Journal of Computational Neuroscience*, *11*, 63–85.
- Chen, X., & Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research*, *46*, 4118–4133.
- Deco, G., Pollatos, O., & Zihl, J. (2002). The time course of selective visual attention: Theory and experiments. *Vision Research*, *42*, 2925–2945.
- Duncan, J., & Humphreys, G. (2002). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.
- Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science*, *17*(11), 973–980.
- Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, *37*(5), 617–631.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B*, *360*, 815–836.
- Gerstner, W. (2000). Population dynamics of spiking neurons: Fast transients, asynchronous states, and locking. *Neural Computation*, *12*(1), 43–89.
- Godijn, R., & Theeuwes, J. (2002). Programming of endogenous and exogenous saccades: Evidence for a competitive integration model. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1039–1054.
- Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, *52*, 217–257.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews: Neuroscience*, *2*, 194–203.
- Johansson, C., & Lansner, A. (2007). Towards cortex sized artificial neural systems. *Neural Networks*, *20*, 48–61.
- Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*(4), 138–147.
- Li, Z. (2005). The primary visual cortex creates a bottom-up saliency map. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 570–575). Elsevier.
- Lupiáñez, J., Klein, R. M., & Bartolomeo, P. (2006). Inhibition of return: Twenty years after. *Cognitive Neuropsychology*, *23*(7), 1003–1014.
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*, 205–231.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton.
- Owen, M. R., Laing, C. R., & Coombes, S. (2007). Bumps and rings in a two-dimensional neural field: splitting and rotational instabilities. *New Journal of Physics*, *9*, 378.
- Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, *42*, 1447–1463.
- Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *The Journal of Neuroscience*, *22*(21), 9475–9489.
- Schall, J. D., Sato, T. R., Thompson, K. G., Vaughn, A. A., & Juan, C. (2004). Effects of search efficiency on surround suppression during visual selection in frontal eye field. *Journal of Neurophysiology*, *91*, 2765–2769.
- Schöner, G. (2007). Dynamical systems approaches to cognition. In R. Sun (Ed.), *Cambridge handbook of computational cognitive modeling*. Cambridge University Press.
- Schöner, G., & Dineva, E. (2007). Dynamic instabilities as mechanisms for emergence. *Developmental Science*, *10*(1), 69–74.
- Schöner, G., & Thelen, E. (2006). Using dynamic field theory to rethink infant habituation. *Psychological Review*, *113*(2), 273–299.
- Shipp, S. (2004). The brain circuitry of attention. *Trends in Cognitive Sciences*, *8*(5), 223–230.
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*(3), 161–168.
- Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamical systems approach to development. *Developmental Science*, *6*(4), 392–412.
- Spencer, J.P., Simmering, V.R., Schutte, A.R., & Schöner, G. (2007). What does theoretical neuroscience have to offer the study of behavioral development? Insights from a dynamic field theory of spatial cognition. In J. M., Plumert & J. P., Spencer (Eds.), *The emerging spatial mind* (pp. 320–361). Oxford, UP.
- Standage, D. I., Trappenberg, T. P., & Klein, R. M. (2005). Modelling divided visual attention with a winner-take-all network. *Neural Networks*, *18*, 620–627.
- Stigchel, S., Meeter, M., & Theeuwes, J. (2006). Eye movement trajectories and what they tell us. *Neuroscience and Biobehavioral Reviews*, *30*, 666–679.
- Stringer, S., Trappenberg, T., Rolls, E., & Araujo, I. (2002). Self-organising continuous attractor networks and path integration: One-dimensional models of head direction cells. *Network: Computation in Neural Systems*, *13*, 217–242.
- Taylor, J. G. (1999). Neural 'bubble' dynamics in two dimensions: Foundations. *Biological Cybernetics*, *80*, 393–409.
- Trappenberg, T. (2002). *Fundamentals of computational neuroscience*. Oxford University Press.
- Trappenberg, T.P. (2008a). Decision making and population decoding with strongly inhibitory neural field models. In D. Heinke & E. Mavritsaki (Eds.), *Computational modelling in behavioural neuroscience: Closing the gap between neurophysiology and behaviour*. London: Psychology Press (in press).
- Trappenberg, T. P. (2008b). Tracking population densities using dynamic neural fields with moderately strong inhibition. *Cognitive Neurodynamics*, *2*, 171–177.
- Trappenberg, T. P., Dorris, M. C., Munoz, D. P., & Klein, R. M. (2001). A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience*, *13*(2), 256–271.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.
- Tsotsos, J. K., Culhane, S., Wai, W., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, *78*(1–2), 507–547.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.
- Van Gisbergen, J. A. M., Van Opstal, A. J., & Tax, A. A. M. (1987). Collicular ensemble coding of saccades based on vector summation. *Neuroscience*, *21*, 641–655.
- Wilson, H., & Cowan, J. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysics Journal*, *12*, 1–24.
- Wolfe, J. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, *1*(2), 202–238.
- Wu, S., & Trappenberg, T. (2008). Learning in sparse attractor networks with inhibition. In R. Wang, F. Gu, & E. Shen (Eds.), *Advances in cognitive neurodynamics*. Berlin: Springer.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.
- Zhang, J., & Bogacz, R. (2008). Superior colliculus and basal ganglia control the saccadic response in motion discrimination tasks. In R. Wang, F. Gu, & E. Shen (Eds.), *Advances in cognitive neurodynamics*. Berlin: Springer.
- Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble. A theory. *Journal of Neuroscience*, *16*, 2112–2126.