



Available at

[www.ElsevierComputerScience.com](http://www.ElsevierComputerScience.com)

POWERED BY SCIENCE @ DIRECT®

Neurocomputing 56 (2004) 329–343

---

---

NEUROCOMPUTING

[www.elsevier.com/locate/neucom](http://www.elsevier.com/locate/neucom)

# Predictions of a model of spatial attention using sum- and max-pooling functions

Fred H. Hamker\*

*Allgemeine Psychologie, Psychologisches Institut II, Westf. Wilhelms-Universität, Fliednerstrasse 21, Münster 48149, Germany*

Received 2 August 2002; received in revised form 4 April 2003; accepted 9 September 2003

---

## Abstract

Assuming a convergent projection within a hierarchy of processing stages stimuli from different areas of the receptive field project onto the same population of cells. Pooling over space affects the representation of individual stimuli, and thus its understanding is crucial for attention and ultimately for object recognition. Since attention, in turn, is likely to modify such spatial pooling by changing the competitive weight of individual stimuli, we compare the predictions of sum- and max-pooling methods using a model of attention. Both pooling functions can account for data investigating the competition between a pair of stimuli within a V4 receptive field; however, our model using sum-pooling predicts a different tuning curve. If we present an additional probe stimulus with the pair, sum-pooling predicts a bottom-up bias of attention, whereas the competition for attention using max-pooling is robust against the additional stimulus.

© 2003 Published by Elsevier B.V.

*Keywords:* Attention; Spatial pooling; V4; Vision

---

## 1. Introduction

Increases in receptive field size along the ventral pathway of the visual system are assumed to facilitate location-invariant object recognition [17,10,28,40,32]. Many feature detectors with smaller receptive field sizes at different spatial positions must converge onto cells with large receptive fields. Typically a sum over all afferents is assumed. Hubel and Wiesel [17] suggested that the response of a “complex cell” can

---

\* Corresponding author. Division of Biology, 139-74, California Institute of Technology, Pasadena, CA 91125, USA.

*E-mail addresses:* [fhamker@uni-muenster.de](mailto:fhamker@uni-muenster.de), [fred@klab.caltech.edu](mailto:fred@klab.caltech.edu) (F.H. Hamker).

be generated by a pooled response from several “simple cells”. This idea of location invariance was first exploited by the Neocognitron from Fukushima [10]. Recently, a max-like pooling has been proposed to establish more robust shift invariant feature detectors [32]. The idea is that a max-function results in a sharp tuning curve around the encoded stimulus, whereas a summation smears the information from different locations.

Regarding the competition for attention, we have also used a max-pooling function [14,15]. It has previously been suggested within the framework of biased competition that competitive interactions in the receptive field are responsible for attentional phenomena [7]. Given a target stimulus among several identical distractors, a sum-pooling would enhance cells encoding the distractors, which would result in a preference of distractors. As an emergent consequence of using sum-pooling, Humphreys and Müller suggested a *SEarch via Recursive Rejection* model that tends to repeatedly reject strongly represented distractors until the target is detected [18]. Such a strategy has not been confirmed by findings, yet. A max-pooling would prevent multiple distractors placed within the receptive field from adding up their feature weight and dominating the competition.

It seems that different constraints, one for robust object recognition and the other for an efficient attentional processing, lead to a max-like pooling mechanism. To shed light on this convergence, we discuss the predictions of a model of attention, which is consistent with the above-mentioned biased competition framework of attention [7], using a max-pooling and a  $\sum$ -pooling function.

## 2. Competition and attention

Findings in attention experiments indicate that attention and spatial pooling are related to each other. For example, V4 neurons that encode a target stimulus show a strong increase in activity compared to neurons that encode a distractor stimulus, but only when the target and the distractor are both within the cells receptive field [5]. Why is the receptive field (RF) a central-processing resource? When more than one object is located within the classical receptive field, ambiguities emerge. A neural population receiving input from two spatial positions can give priority to one, the other or to neither. Spatial attention is known to suppress the influence of the unattended stimuli [26] and thus assigning processing priority for the computation in the following stages. The general idea is that stimuli compete within the classical receptive field, which is biased by some form of spatial attention [7]. Recent single cell and fMRI recordings shed light on the neural process of this competition [21,23,25,30,31,33,36–38].

Our model was designed to replicate the experimental data of McAdams and Maunsell [24] and of Reynolds et al. [31]. McAdams and Maunsell found that if only one stimulus is presented, the effect of spatial attention on the population response is a multiplicative gain increase over the entire population. Reynolds, Chelazzi and Desimone observed competitive interactions by placing two stimuli (reference and probe) within the receptive field of a V4 neuron (Fig. 1a). They found that when spatial attention was directed away from the receptive field the response to both stimuli was a weighted average of the responses to the stimuli presented in isolation. In brief, if the reference

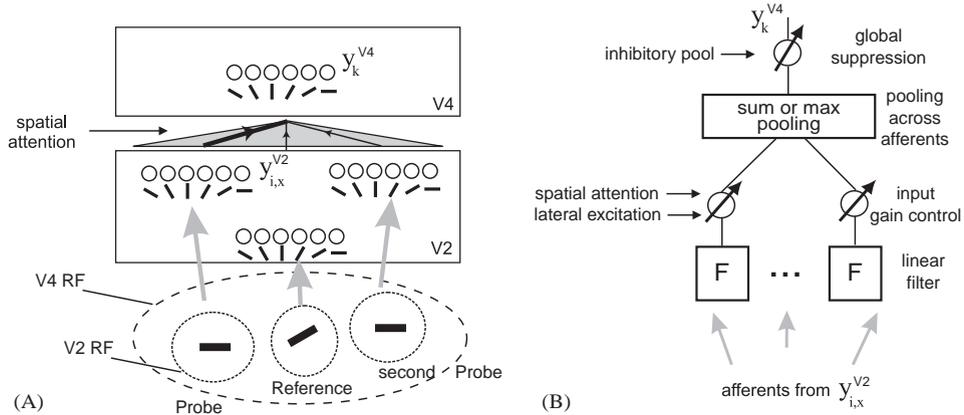


Fig. 1. Competition among stimuli within a receptive field on the basis of a population code. (A) Convergent projection of several source populations  $y_{i,x}^{V2}$  at the locations  $x$  onto a target population  $y_k^{V4}$ . Spatial attention implements an input gain control prior to spatial pooling. Each population is implemented as a laterally interconnected circuit. The presentation of stimuli and the deployment of spatial attention follows the conditions from the experiment of Reynolds et al. [31]. They measured the responses to (1) the reference stimulus, (2) the probe stimulus, (3) the pair, while the monkey attended somewhere else, (4) the pair, while attending to the location of the reference stimulus and (5) the pair, while attending to the location of the probe. To investigate the competition of two equal stimuli against another one, we propose to add a second probe to the pair. (B) Schematic diagram of how afferents determine the cell's output. Prior knowledge, that is spatial attention and the lateral input, can each enhance the effective input gain. The afferents are then pooled by either a linear summation or by a nonlinear maximum-like rule. Inhibitory pooling performs a normalization and provides competitive interactions among active populations. Each population consists of 11 cells.

elicits a high firing rate and the probe a low firing rate, then the response to both is in between. Attending to the location of one of the stimuli biases the competition towards the attended stimulus.

In order to differentiate between both pooling mechanisms, we added a second probe to the pair. Our model predicts that the  $\sum$ -pooling induces a bottom-up bias that interferes with spatial attention. Thus, according to our model, the attentional competition among stimuli using a max-pooling function is more robust against the number of stimuli.

### 3. Model of attention in V4

We present a model that describes how spatial attention assigns a processing priority to optimize object recognition. The interactions in V4 are modeled using a population code approach. One influential and often-used approach for modeling cooperative and competitive interactions is based on an additive activation function [1,16,2,27]. However, a comparison with neural data is difficult, since unlike real cells, units with less input are very quickly suppressed. These models typically show a winner-take-all

behavior, i.e. they converge to a single activity hill. The shape of the hill is, in general, determined by the weights and the position of the hill depends only on the input. An encoding of multiple stimuli is difficult to explain by such codes. To overcome this limitation, Grossberg [12] introduced a shunting activation function with automatic gain control to reduce the sensitivity to the input range. A normalization term found in the shunting model was also used in additive activation function models [20]. Nevertheless, both types offer no rule regarding how to incorporate top-down input. I now suggest a related model which, however, allows the balanced representation of multiple stimuli including a definition of their modulation by top-down stimuli.

The V2 layer contains at each location  $x \in \{1, 2, 3\}$  a population of 11 cells each encoding a different orientation of a bar. The V4 layer is simulated by a single population of the same size on which the V2 populations project. We consider feedforward, lateral excitatory and inhibitory input and spatially organized attentional feedback as relevant factors (see [29]), but do not model feedback from IT. Each V4 cell receives a weighted input from each V2 population (Eq. (1)) at different locations  $x$  within the V4 RF (Fig. 1B). Excitatory lateral connections (Eq. (2)) and spatial attention (Eq. (3)) increase the input gain. After the gain control stage, spatial pooling is applied (Eq. (4)). We consider two alternative pooling functions: max-pooling and  $\sum$ -pooling. Inhibition among V4 cells is modeled by an inhibitory pooling among all cells in the population (Eq. (5)). The final response is then determined by a differential equation (Eq. (6)), which describes the change through time of a model V4 cell.

Since we map orientation-selective cells in V2 onto orientation-selective cells in V4, the weights  $w_{ik}$  (Eqs. (1)–(3)) of the filter  $F$  are simply determined by a Gaussian kernel  $w_{ik} = g_k(\|\mathbf{u}_k^{V4} - \mathbf{u}_i^{V2}\|)$  (with  $w_{ii} = 1.3$  and  $\sigma^2 = 0.1$ ), according to the similarity of the preferred stimulus of a cell in the source population  $\mathbf{u}_i^{V2}$  to the preferred stimulus of a cell in the target population  $\mathbf{u}_k^{V4}$ . That is, a projection is maximal from a cell  $i$  with a preferred orientation (e.g.,  $\mathbf{u}_i^{V2} = 30^\circ/180^\circ$ ) in V2 to a cell  $k$  in V4 which has the same preferred orientation ( $\mathbf{u}_k^{V4} = 30^\circ/180^\circ$ ). The projection weight to cells with other preferred orientations decreases. The afferents  $a_{i,x}$  of V4 are driven by the feedforward term  $y_{i,x}^{V2} g_k(i)$ . The lateral weights  $w_{kj}^l$  are computed from a Gaussian (with  $w_{ii}^l = 0.3$  and  $\sigma = 1$ ).

A neural interpretation of the Bayesian inference theory is used for input gain control, as suggested by Koechlin et al. [22]. Bayesian inference is an approach to statistics in which all forms of uncertainty are expressed in terms of probability. The priors are meant to capture the beliefs about aspects of the scene before seeing the actual input. After presenting the input a posterior distribution is obtained, which takes account of both the prior and the input. From this posterior distribution, predictive distributions for future observations can be computed. It has been suggested that the firing rates  $y_i$ , assumed as stochastic, implicitly code the posterior probability distribution  $p(V|\omega)$  of the variable  $V$  given the available information  $\omega$ , and the set of input tuning curves or receptive fields  $f_i(V)$  [42]. Although we do not explicitly decode posterior distributions from firing rates, this interpretation allows an intuitive explanation of the effect of prior knowledge. An increase of the input gain, for example, if we attend to a certain location or to a specific feature, enhances the probability of an object being detected. We consider the top-down spatial attention signal  $A_x = 3$  (its

value is chosen to achieve a fit with the experimental data in the pair condition) and the lateral input  $w_{kj}^L y_j^{V4}$  as prior knowledge (Eqs. (2) and (3)). It has been shown that a multiplication of the neural input (not the neural activity) with the prior is a good approximation of this concept [22]. Our additional simulations have shown that the definition of the lateral input as prior knowledge is not essential for the findings reported here. An additive influence on the cell or no lateral interactions lead to similar results. However, the multiplicative effect of the top-down signal on the input is essential for our results, since it implements an enhancement of the effective gain. Although the exact mechanism is still not known, such an input gain control is supported by findings that indicate that spatial attention increases neural sensitivity towards the input signal [23,31,9] and that feedback connections can rapidly facilitate responses to stimuli, but do not drive cells without bottom-up activation [19]:

$$a_{i,x}^F = w_{ik} y_{i,x}^{V2} + N, \quad (1)$$

$$a_{i,x}^L = w_{ik} y_{i,x}^{V2} \cdot \sum_j w_{kj}^L y_j^{V4} + N, \quad (2)$$

$$a_{i,x}^A = w_{ik} y_{i,x}^{V2} \cdot A_x + N. \quad (3)$$

$N$  is a noise term that leads to variations in the transmission from one cell to another. Each connection has an independent noise term. As in the gain field approach [35], we multiply the attentional gain signal with the input of the cell (Eq. (3)). Additionally, we add up the feedforward  $a^F$ , the lateral  $a^L$  and the attention term  $a^A$  (see Eq. (6)) after applying the pooling (Eq. (4)). Thus, provided a cell is sensitive towards a given stimulus in its receptive field, the cell fires even when attention is directed elsewhere. Attention, in this model, induces a preferred processing, but does not operate in an all or none fashion. This is consistent with recent results [24,31,5].

The pooling function  $f$  defines the influence of the afferents  $a_{i,x}, x \in \{1, 2, 3\}$  on the cell  $k$ , where  $x$  is the spatial location of a bar in V2:

$$I_k^F = f(a_{i,x}^F), \quad I_k^L = f(a_{i,x}^L), \quad I_k^A = f(a_{i,x}^A); \quad f = \max_{i,x} \vee f = \sum_{i,x}. \quad (4)$$

The normalization of responses [12,22,3] is performed by a shunting inhibition which inhibits the cell  $k$  depending on the overall activity within the population  $\sum_j y_j^{V4}$  and on its own activity  $y_k^{V4}$  (Eq. (5)). Thus, vigorously firing cells receive more inhibition than less active cells. This guarantees that the entire distribution of the population and not just one activity hill can encode meaningful information. To induce more competition among active populations, a small fraction ( $w_{inh}^f = 0.1$ ) of the population activity inhibits the cell independent of its own activity. Such an inhibition term is necessary to simulate the competitive interactions among active populations if more than one stimulus is within the receptive field. Here, we model the temporal process of this competition explicitly. However, the final cell response (steady state) is similar to a standard normalization model (e.g. [3]), in which the linear response of every cell is divided by a number that grows with the activity of a large number of cortical cells, except that our steady-state response would be reduced by  $w_{inh}^f/w_{inh}$ . The

baseline inhibition  $By_k^{V4}$  prohibits an overly high baseline resulting from the noise  $N$  (Eqs. (1)–(3)):

$$I_k^{\text{inh}} = (w_{\text{inh}} y_k^{V4} + w_{\text{inh}}^f) \sum_j y_j^{V4} + By_k^{V4}, \quad w_{\text{inh}} = 1.6, \quad w_{\text{inh}}^f = 0.1. \quad (5)$$

The cell's response is computed by a differential equation:

$$\tau \frac{d}{dt} y_k^{V4} = I_k^F + I_k^L + I_k^A - I_k^{\text{inh}}. \quad (6)$$

Simulations with different pooling functions were performed with the same equations and parameters. The effect of spatial attention in our model is a direct enhancement of the input and an indirect suppression that acts after pooling. Essential parameters are the strength of spatial attention  $A_x$  and the inhibitory weights  $w_{\text{inh}}^f$  and  $w_{\text{inh}}$ .

#### 4. Results

The finding that attention implements a multiplicative gain increase if only one stimulus is presented [24,38] is replicated by both the  $\sum$ - and max-pooling functions (Fig. 2). To work out the predictions of the alternative pooling functions, we now observe the effect of different attentional conditions, when a reference, a probe or both stimuli are presented. In order to determine the effect of adding a probe on the whole population, we computed a selectivity value and interaction index for each cell and each condition (Fig. 3A) as done by Reynolds et al. [31] (please refer to the Appendix for details). Our model reliably describes the neural interactions of two stimuli within the RF of V4 cells. A slope of 0.5 indicates that reference and probe are equally well represented by the population. The small positive  $y$ -intercept signifies a slight overall increase in activity when presenting the additional probe, regardless of selectivity. Attending to the probe increases the slope, indicating the greater influence of the attended probe over the population. Attending to the reference reduces the slope, signifying the greater influence of the attended reference stimulus. Attention, in general, enhances the overall response within the population, which is observed by the greater upward shift of the sensory interaction index as compared to the attend away condition. Before the stimulus appears, the model shows a baseline increase of individual cells when attention is directed towards a location, which replicates findings in V4 [23,31] (Fig. 3B).

The model using each pooling function matches with the experimental data measuring the attention effect over the entire population. However, a look at the shape of the population response reveals that the  $\sum$ -pooling of afferents leads to a uniform distribution around the average stimulus (Fig. 4). The information of both stimuli is hidden within a broad tuning curve. In fact, the cell with the highest activity is optimally tuned for a stimulus orientation that is not presented to the cell population. However, only the cells with almost identical responses to reference and probe (selectivity = 0) clearly indicate the pooling mechanism. A very careful analysis of the shape of the tuning curve would be required in order to favor one pooling function against the other.

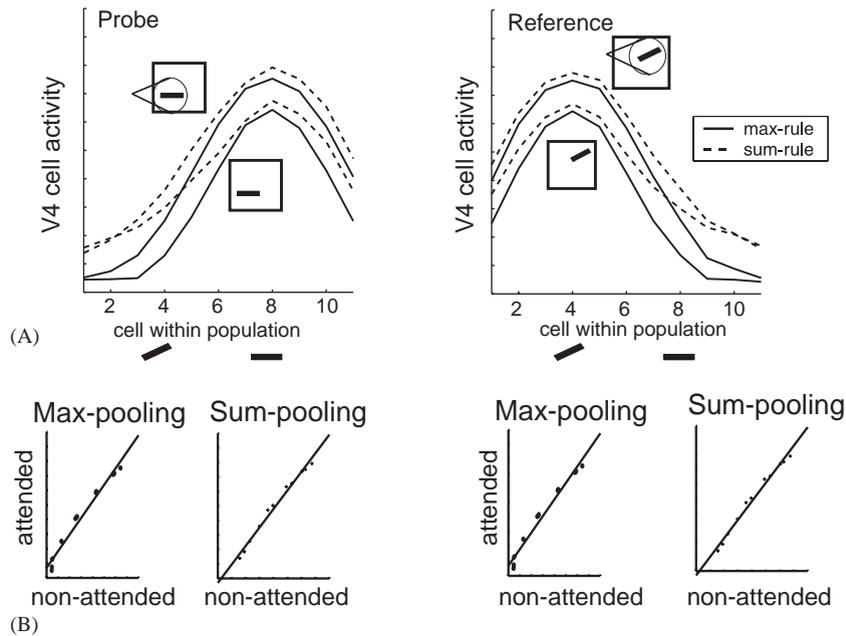
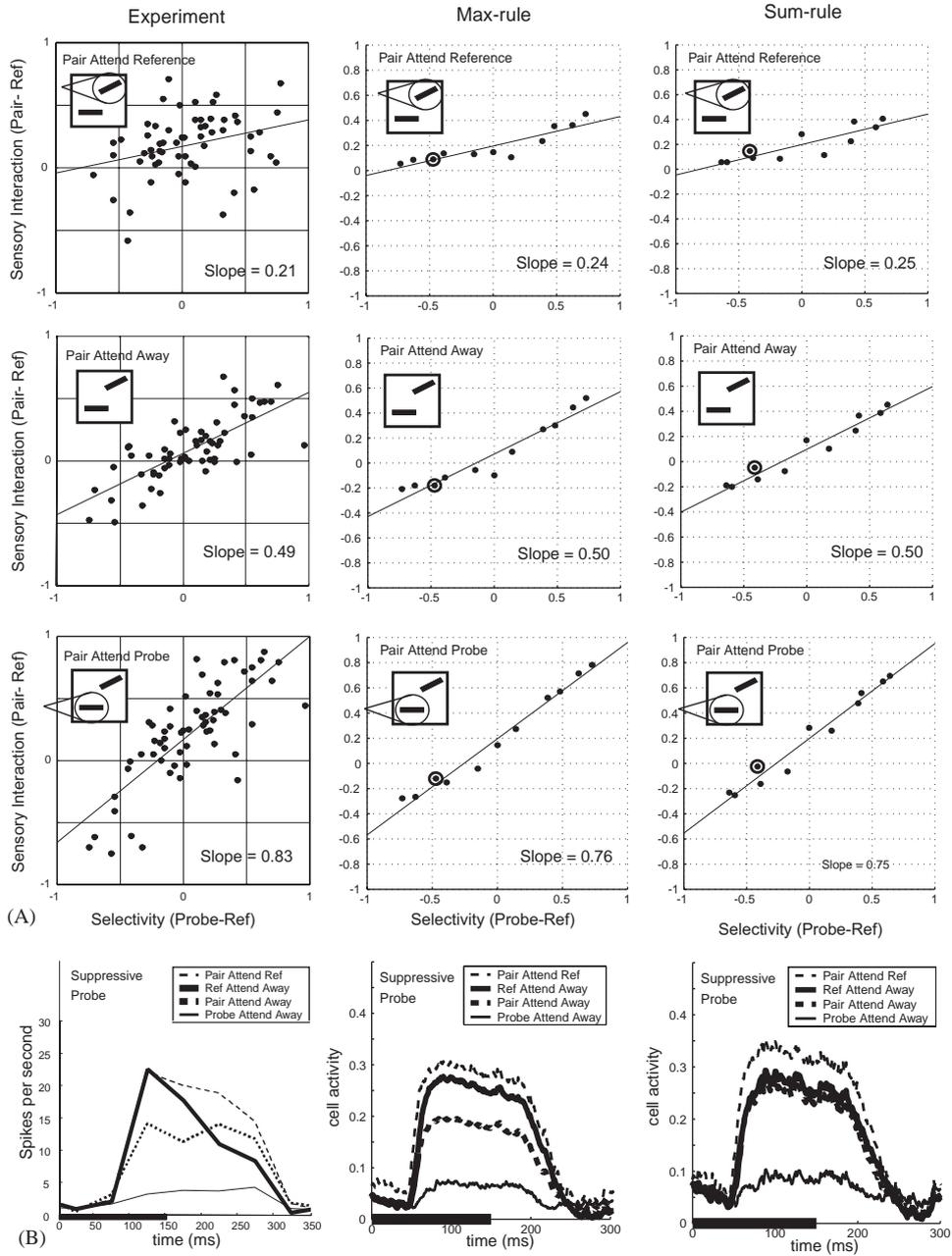


Fig. 2. Population responses to a single stimulus 100 ms after stimulus onset. (A) Spatial attention increases the effective gain. A sum-pooling function leads to a broader tuning curve. The reference stimulus elicits the highest response at cell 4, whereas the preferred stimulus of cell 8 is the presented probe. (B) Responses of the non-attended case are plotted against the attended case (see [24]). The multiplicative gain increase is not impaired by the pooling function used (both have a  $\text{CorrCoeff} = 0.99$ ). With the same parameters, the max-pooling results in a 13% increase (slope = 1.13), whereas the sum-pooling results in approx. 26% increase (slope = 1.26) of the population response by attention.

We now simulate an experimental approach using three stimuli. Since the  $\sum$ -pooling mechanism results in an input gain increase of cells which are sensitive to both stimuli presented in isolation (Fig. 4), an additional bias for cells encoding the probe must occur by adding an additional probe to the pair (Fig. 1A). In contrast, a max-pooling would ignore that two equal probe stimuli are presented to the population.

Our simulation results confirm that there is no change in the sensory interaction index when adding an additional probe to the pair using a max-pooling function (Fig. 5A). Our model with the  $\sum$ -pooling function, however, predicts that the response can be biased towards the probe by placing more than one probe in the receptive field (Fig. 5A). In the "attend away" condition, the population response is now dominated by the two probe stimuli, although attention is directed elsewhere. Even attending to the reference can only achieve an equal representation within the population. The population response (Fig. 6) completes this view. The insufficient separation and the broad shape in the attend reference condition hardly allows a detection of the reference stimulus within a following stage.



## 5. Discussion

We used a model of attention to explore the possible effects of pooling on the response of cells under different attentional conditions. Alternately using spatial max-pooling and  $\sum$ -pooling, our model accounts for the finding that if the receptive field contains just one stimulus attention results in a multiplicative gain increase, as observed in MT, MST and V4 [38,24]. If two stimuli are presented within the receptive field, the model using each pooling function can reproduce the data of Reynolds et al. [31]. According to the model, those attention effects in V4 can be explained by an input gain increase prior to pooling and additionally by an indirect inhibition among active populations after pooling has occurred. The fact that the population response to the attended stimulus is increased more than the response to the unattended is suppressed also explains the observed upward shift towards a higher sensory interaction index. However, the sensory interaction index only estimates the average competitive weights of the stimuli. Pooling already plays a role as indicated by the shape of the tuning curves (Fig. 4). A max-pooling function tries to keep the information of both stimuli separate within the joint representation, whereas the  $\sum$ -pooling function tends to merge both stimuli into a uniform, but broader distribution.

We have seen that  $\sum$ -pooling results in an increase of a cell's firing rate to a pair of stimuli if the cell is sensitive to both the reference and the probe. When the reference stimulus appears together with two equal probe stimuli, a  $\sum$ -pooling of afferents predicts that the two identical probe stimuli will bias the population response as if the probe stimulus would have been selected by spatial attention. In comparison to the pair condition, a max-like pooling function predicts no change.

If we define that an efficient processing of stimuli does indicate the content of the receptive field and the saliency of the stimuli independent of its number, then our model with a max-like pooling function is more efficient than our model with the  $\sum$ -pooling function. However, our simulations cannot completely rule out the  $\sum$ -pooling function. Even though we made sure that the model accounts for relevant findings, attention can be more complex than the mechanisms captured by our model. A biologically plausible max-pooling function could be implemented by a winner-take-all (WTA) among all

---

Fig. 3. Influence of attention on the sensory interaction by using the max- and sum-pooling compared to the data of single-cell recordings in V4 (modified from [31]). (A) For each cell, its selectivity index is plotted over its sensory interaction. A selectivity value of 0 indicates identical responses to reference and probe in isolation, a positive value a preference towards the probe and a negative preference towards the reference. An interaction index of 0 signifies that the cell is unaffected by adding a probe. Positive values indicate that the cell's response to the reference is increased by adding a probe and negative values signify a suppression by the probe. Both pooling functions account for the data, which measures the overall population response. However, the activity of some cells strongly depends on the pooling function. With a  $\sum$ -pooling, the cell with identical responses to reference and probe (selectivity = 0) is enhanced by addition of a probe, whereas it is suppressed when using a max-pooling function. (B) The course of activity of single cells. The temporal course of the observed cell using the  $\sum$ -pooling (circled in the population response) shows that the cell is hardly suppressed by adding a probe, although it is much more selective to the reference stimulus. The bar indicates the stimulus presentation period.

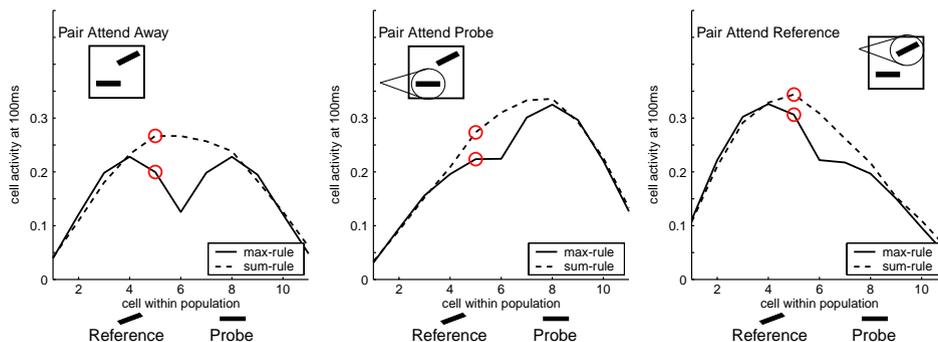


Fig. 4. Influence of attention on the response of cells within the population 100 ms after the presentation of the pair. Each stimulus activates a number of cells. The competition among stimuli occurs due to inhibitory pooling within the population. Attending to one location increases the gain of the attended population, whereas the other population is slightly suppressed. The  $\sum$ -rule predicts that cells which are sensitive to both of each stimulus presented alone gain advantage when both stimuli are shown. The circle indicates the cell shown in Fig. 3B.

afferents. This would not only require that all other inputs are suppressed, but also that the saliency of the input signal is maintained, that is, a signal representing a stimulus with low saliency should not be transformed into high saliency. A number of WTA networks have been investigated regarding their realization of maximum operations [41] including spike timing [39,34].

How does our model relate to previous models? Along with the experimental data, Reynolds et al. [31] demonstrated that a feedforward shunting model [12] can account for their findings [7]. In their model, competition among cells occurs due to feedforward inhibition from V2 cells onto V4 cells. Spatial attention was implemented by increasing the excitatory and inhibitory connections originating from the attended location. Our approach is similar, but competition occurs after pooling. It is based on lateral short-range excitatory and long-range inhibitory connections within the population, which is in accordance with findings in V4 [8,29]. With respect to pooling, our simulations suggest that an additional max-like competition among afferents would prevent a bias due to the fact that multiple distractors add up their competitive weight. Such a case was not discussed in earlier models of attention and the model of Reynolds et al. [31] predicts a bias similar to our model using the  $\sum$ -pooling. Comparisons with other models that are able to simulate the interactions among stimuli under different attentional conditions [13,6] are difficult, because neither a quantitative comparison with the experimental data discussed here nor an investigation of spatial pooling was made.

Similar to the claim of Riesenhuber and Poggio [32], but here from the viewpoint of attention, we suggest that max-pooling is more efficient than  $\sum$ -pooling. Since object recognition requires an interpretation of a V4 population response at a subsequent stage (e.g. IT or prefrontal areas), the mechanism of spatial attention should increase the competitive weight of the attended stimulus such that it dominates the population response regardless of other stimuli in the same RF. In so far our results also indicate that object recognition is facilitated by a max-pooling function. However, Riesenhuber

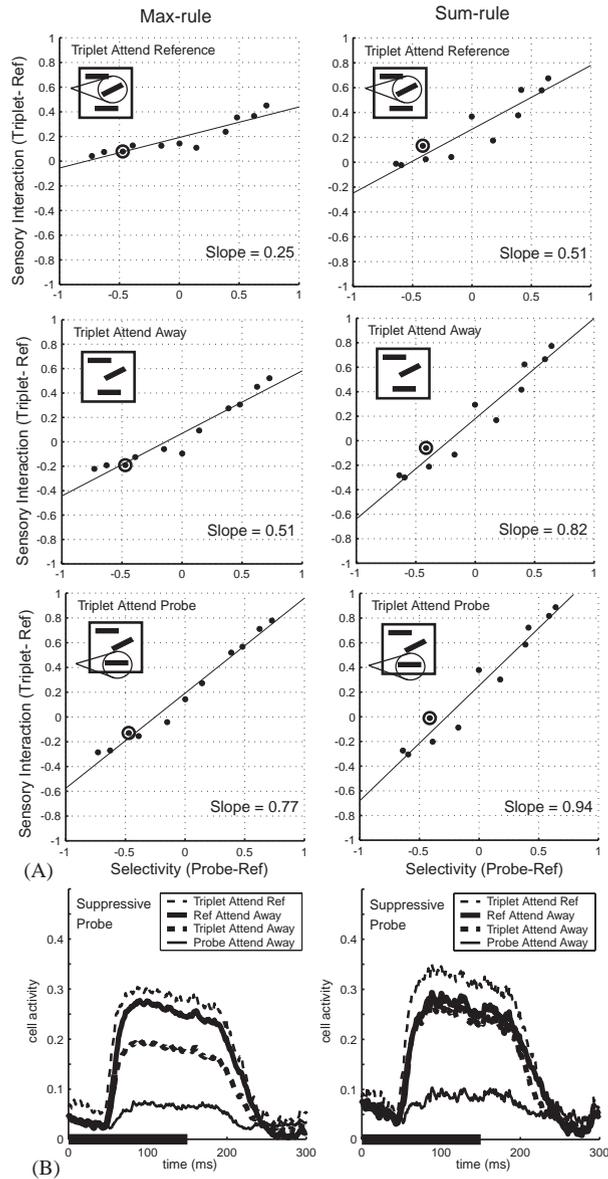


Fig. 5. Influence of attention on the sensory interaction when presenting a triplet. (A) The  $\Sigma$ -rule predicts a bottom-up bias in the attend-away condition due to the fact that the responses of the same stimuli at the different locations in the RF add up. The effect is similar to as if the probe is attended. Thus, competition can be biased towards the probe by presenting a second probe with the pair. (B) The course of activity of the circled cell shows almost no suppressive effect to the reference stimulus when the probe is presented simultaneously.

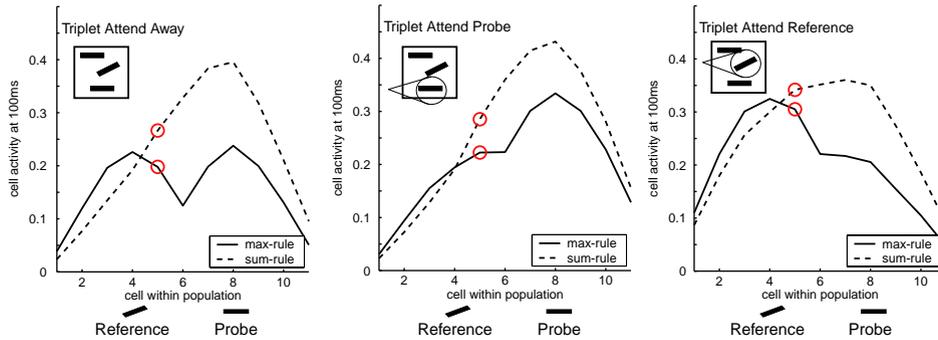


Fig. 6. Influence of attention on the response of cells within the population 100 ms after the presentation of the reference and two probe stimuli. Due to the bias of the additional probe, the reference stimulus is hidden in the population response using the  $\Sigma$ -pooling. The circle indicates the cell shown in Fig. 5(B).

and Poggio [32] compared max-model neurons to linear model neurons. Linear neurons are biologically implausible. Here, we have compared the effects of different pooling functions on populations of non-linear neurons. A pure max-function as proposed by Riesenhuber and Poggio [32] is inconsistent with the discussed competitive effects. It would predict no change of a cell's response on a pair compared with the response of the preferred stimulus in isolation. Experiments, however, revealed an averaging of the individual responses [31,23,30]. For example, Recanzone, Wurtz and Schwarz measured the interactions between the responses to two moving objects within a single receptive field in MT and MST [30]. They found that averaging, compared to summation, multiplication, or vector addition of the stimulus responses best modeled the data. However, a recent study of Gawne and Martin [11] compared the responses of V4 cells with a max-, a linear-, and a weighted average model. They found that the max- and weighted average model somehow fit the data, with a slight advantage for the max-model. A potential weakness of this study is the missing control of attention. The stimuli are presented for approximately 282 ms. Such a long presentation time elicits attentional effects. Although they initially train the monkey to fixate, it is not clear if the monkey maintains attention at the fixation point.

The problem with the current experimental data is that several other models can be consistent with the findings. Note from Fig. 3b that our model in case of max-pooling is not a true weighted average model. In addition, we have shown that our model using either sum- or max-pooling is consistent with the data of Reynolds et al. [31]. Fig. 3b illustrates that our model with sum-pooling is also consistent with the data of Gawne and Martin [11] since the inhibition forces the response to the pair to be not higher than the response to the reference. Using just two stimuli does not allow to dissociate between sum- and max-pooling on the population level given normalized neurons. It would require a thorough analysis of the tuning curve, which is difficult due to the noisy responses. Thus, we suggested a paradigm using three stimuli to better dissociate pooling from normalization. From the viewpoint of our model, a max-like pooling avoids a bottom-up bias for attentional competition in the presence of multiple stimuli.

Concluding, our model provides a good fit with the experimental data in V4 showing a multiplicative gain increase [24] and competitive interactions among stimuli in different attentional conditions [31]. Current experimental data do not allow to dissociate the subtle differences between the different models and do not even rule out sum-pooling, assuming that the cells within the population are inhibited by the overall population activity. Our model suggests that pooling across the receptive field is based on a max-like operation, since this avoids a bottom-up bias independent of the number of stimuli.

## Acknowledgements

This research has been performed at Caltech. I am grateful to John Reynolds for providing the data showing the attention effects on V4 cells (Fig. 3A). I am pleased for extensive discussions with Jamie Mazer and Rufin VanRullen. I thank Christof Koch, Dirk Walther, Brad Motter, Max Riesenhuber and Tomaso Poggio for helpful comments on an earlier manuscript. This work was supported by DFG HA2630/2-1 and in part by the ERC Program of the NSF (EEC-9402726).

## Appendix A

### *Simulation and data analysis*

Input stimuli are encoded as a population determined by a Gaussian distribution. Such populations were presented to the model for 150 ms (simulation time). For realistic experimental conditions, we delayed the input for 30 ms to account for the time a stimulus needs to reach V2. Since V1 cells typically fire very strongly in the beginning and then decrease in sensitivity, we use a short-term synaptic depression  $S_i$  (similar to [4]) of the input  $I_i^{V2} S_i$ :

$$\tau_S \frac{d}{dt} s_i = I_i^{V2} - s_i, \quad S_i = (1 - d s_i), \quad d = 0.35. \quad (\text{A.1})$$

We would like to emphasize that our simulation, on the basis of a population code, has the advantage that all the cells in the network are used to compute the overall response. The population response is not computed from repeated simulations.

### *Data analysis*

The data of the model V4 cells were analyzed in the same manner as the experimental data [31]. For each cell, in each condition, the average firing rate was computed beginning 70 ms after stimulus onset and ending at 250 ms:

$$\bar{y}_i = \text{mean}(y_i(t)). \quad (\text{A.2})$$

All responses are normalized by dividing by the highest firing rate  $y_{\max}$  observed in all stimulus configurations (reference, probe, pair) and in any attentional condition (attend

away, attend probe, attend reference):

$$\tilde{y}_i = \bar{y}_i / y_{\max}. \quad (\text{A.3})$$

The selectivity and interaction values are obtained by computing the difference for each cell in each condition. All 11 cells in the population are used in the analysis. For the selectivity index (SE), this yields into

$$SE_i = \tilde{y}_i^{\text{probe}} - \tilde{y}_i^{\text{reference}} \quad (\text{A.4})$$

and for the interaction index (SI) into

$$SI_i = \tilde{y}_i^{\text{pair}} - \tilde{y}_i^{\text{reference}}. \quad (\text{A.5})$$

For a comparison with real data in the attend-away condition, we chose the results computed from cells whose response changed significantly when attention was directed to the probe stimulus (see [31]).

## References

- [1] S. Amari, M.A. Arbib, Competition and cooperation in neural nets, in: J. Metzler (Ed.), *Systems Neuroscience*, Academic Press, San Diego, 1977, pp. 119–165.
- [2] D. Amit, *Modelling Brain Functions*, University Press, Cambridge, 1989.
- [3] M. Carandini, D.J. Heeger, J.A. Movshon, Linearity and normalization in simple cells of the macaque primary visual cortex, *J. Neurosci.* 17 (1997) 8621–8644.
- [4] F.S. Chance, S.B. Nelson, L.F. Abbott, Synaptic depression and the temporal response characteristics of V1 cells, *J. Neurosci.* 18 (1998) 4785–4799.
- [5] L. Chelazzi, E.K. Miller, J. Duncan, R. Desimone, Responses of neurons in macaque area V4 during memory-guided visual search, *Cereb. Cortex* 11 (2001) 761–772.
- [6] S. Corchs, G. Deco, Large-scale neural model for visual attention: integration of experimental single-cell and fMRI data, *Cereb. Cortex* 12 (2002) 339–348.
- [7] R. Desimone, J. Duncan, Neural mechanisms of selective attention, *Annu. Rev. Neurosci.* 18 (1995) 193–222.
- [8] R. Desimone, S.J. Schein, Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form, *J. Neurophysiol.* 57 (1987) 835–868.
- [9] P. Fries, J.H. Reynolds, A.E. Rorie, R. Desimone, Modulation of oscillatory neuronal synchronization by selective visual attention, *Science* 291 (2001) 1560–1563.
- [10] K. Fukushima, Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, *Biol. Cybern.* 36 (1980) 193–202.
- [11] T.J. Gawne, J.M. Martin, Responses of primate visual cortical V4 neurons to simultaneously presented stimuli, *J. Neurophysiol.* 88 (2002) 1128–1135.
- [12] S. Grossberg, Contour enhancement short term memory, and constancies in reverberating neural networks, *Stud. Appl. Math.* 52 (1973) 217–257.
- [13] S. Grossberg, R. Raizada, Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex, *Vis. Res.* 40 (2000) 1413–1432.
- [14] F.H. Hamker, *Visuelle Aufmerksamkeit und lebenslanges Lernen im Wahrnehmungs-Handlungs-Zyklus*, Ph.D. Thesis, Technische Universität Ilmenau, Germany, 1999.
- [15] F.H. Hamker, The role of feedback connections in task-driven visual search, in: D. Heinke (Ed.), *Connectionist Models in Cognitive Neuroscience*, Springer, London, 1999, pp. 252–261.
- [16] J.J. Hopfield, Neurons with graded response have collective computational properties like those of two-state neurons, *Proc. Natl. Acad. Sci. USA* 81 (1984) 3088–3092.
- [17] D. Hubel, T. Wiesel, Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat, *J. Neurophysiol.* 28 (1965) 229–289.

- [18] G.W. Humphreys, H. Müller, Search via recursive rejection (SERR): a connectionist model of visual search, *Cogn. Psychol.* 25 (1993) 43–110.
- [19] J.M. Hupé, A.C. James, P. Girard, S.G. Lomber, B.R. Payne, J. Bullier, Feedback connections act on the early part of the responses in monkey visual cortex, *J. Neurophysiol.* 85 (2001) 134–145.
- [20] D. Jancke, W. Erlhagen, H.R. Dinse, A.C. Akhavan, M. Giese, A. Steinhage, G. Schöner, Parametric population representation of retinal location: neuronal interaction dynamics in cat primary visual cortex, *J. Neurosci.* 19 (1999) 9016–9028.
- [21] S. Kastner, P. De Weerd, M.A. Pinsk, M.I. Elizondo, R. Desimone, L.G. Ungerleider, Modulation of sensory suppression: implications for receptive field sizes in the human visual cortex, *J. Neurophysiol.* 86 (2001) 1398–1411.
- [22] E. Koechlin, J.L. Anton, Y. Burnod, Bayesian inference in populations of cortical neurons: a model of motion integration and segmentation in area MT, *Biol. Cybern.* 80 (1999) 25–44.
- [23] S.J. Luck, L. Chelazzi, S.A. Hillyard, R. Desimone, Mechanisms of spatial selective attention in areas V1, V2 and V4 of macaque visual cortex, *J. Neurophysiol.* 77 (1997) 24–42.
- [24] C.J. McAdams, J.H. Maunsell, Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4, *J. Neurosci.* 19 (1999) 431–441.
- [25] E.K. Miller, P.M. Gochin, C.G. Gross, Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus, *Brain Res.* 616 (1993) 25–29.
- [26] J. Moran, R. Desimone, Selective attention gates visual processing in the extrastriate cortex, *Science* 229 (1985) 782–784.
- [27] H. Nakahara, S. Wu, S. Amari, Attention modulation of neural tuning through peak and base rate, *Neural Comput.* 13 (2001) 2031–2047.
- [28] D.I. Perrett, M.W. Oram, The neurophysiology of shape processing, *Image and Vision Comput.* 11 (1993) 317–333.
- [29] D.A. Pollen, A.W. Przybyszewski, M.A. Rubin, W. Foote, Spatial receptive field organization of macaque v4 neurons, *Cereb. Cortex* 12 (2002) 601–616.
- [30] G.H. Recanzone, R.H. Wurtz, U. Schwarz, Responses of MT and MST neurons to one of two moving objects in the receptive field, *J. Neurophysiol.* 78 (1997) 2904–2915.
- [31] J.H. Reynolds, L. Chelazzi, R. Desimone, Competitive mechanism subserve attention in macaque areas V2 and V4, *J. Neurosci.* 19 (1999) 1736–1753.
- [32] M. Riesenhuber, T. Poggio, Hierarchical models of object recognition in cortex, *Nat. Neurosci.* 2 (1999) 1019–1025.
- [33] E.T. Rolls, M.J. Tovee, The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field, *Exp. Brain Res.* 103 (1995) 409–420.
- [34] G.A. Rousset, S.J. Thorpe, M. Fabre-Thorpe, Taking the MAX from neuronal responses, *Trends Cogn. Sci.* 7 (2003) 99–102.
- [35] E. Salinas, L.F. Abbott, Invariant visual responses from attentional gain fields, *J. Neurophysiol.* 77 (1997) 3267–3272.
- [36] T. Sato, Effects of attention and stimulus interaction on visual responses of inferior temporal neurons in macaque, *J. Neurophysiol.* 60 (1988) 344–364.
- [37] E. Seidemann, W.T. Newsome, Effect of spatial attention on the responses of area MT neurons, *J. Neurophysiol.* 81 (1999) 1783–1794.
- [38] S. Treue, J.H. Maunsell, Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas, *J. Neurosci.* 19 (1999) 7591–7602.
- [39] R. VanRullen, Visual saliency and spike timing in the ventral visual stream, *J. Physiol. (Paris)*, in press.
- [40] G. Wallis, E.T. Rolls, Invariant face and object recognition in the visual system, *Prog. Neurobiol.* 51 (1997) 167–194.
- [41] A. Yu, M.A. Giese, T. Poggio, Biophysiological plausible implementations of the maximum operation, *Neural Comput.* 14 (2002) 2857–2881.
- [42] R.S. Zemel, P. Dayan, A. Pouget, Probabilistic interpretation of population codes, *Neural Computation* 10 (1998) 403–430.