Neural Network Modelling of Autism

Recent developments in autism research (M. F. Casanova, ed.),
pp. 100–134, Hauppauge, New York: Nova Science Publishers, Inc., 2005. In press.

Lennart Gustafsson,
Luleå University of Technology,
S-971 87 Luleå, Sweden
ph. +46 920 491321
fax +46 920 492082
Lennart.Gustafsson@sm.luth.se

Andrew P. Papliński
Computer Science and Software Engineering
Monash University, Vic 3800, Australia
ph. +61 3 9905 3242
fax +61 3 9905 5146
app@csse.monash.edu.au

December 20, 2004

Abstract

Neural network models comprise models of neurons, both their firing characteristics and their plasticity characteristics and of the connectivity between the neuron in a neural circuit. The models are used in computer simulations and, because of the size and complexity of the central nervous system, are greatly simplified in all these aspects. A brief overview of such models is presented. A number of attempts to explain aspects of autism, in particular poor generalization and good discrimination, have been made, employing neural network models. These attempts are briefly presented and the biological relevance of the models is discussed. Finally, simulations of learning processes subjected to abnormalities in attention shifting using self-organizing neural networks are presented. It is shown that when attention shifting is restricted by familiarity preference exact learning of objects with little variability occurs, arguably a characteristic of autism, and how early intervention in this learning process can result in normal learning of a broad range of objects.

Key Words: Neural networks, autism, modelling

1 Introduction

A great deal is known about neurons, their internal structure and processes, how they are interconnected and how these connections change over time. This knowledge makes it possible to calculate with some degree of accuracy the efferent (output) signal along a neuron's axon, given the afferent (input) signals that reach synapses on the dendritic tree. Given that we can calculate the activity of a single neuron and knowing how neurons are connected to each other we can also, in principle, calculate the overall neural activity in an area of e.g. neocortex.

The human central nervous system (CNS) contains a large number of neurons and many more connections between these neurons. These numbers vary greatly between individuals and at stages of any individual's life. For neocortex figures of more than ten billion neurons (10^{10}) and more than a hundred thousand billion (10^{14}) connections, or synapses, are often given. It is, however, not necessary to take all neural activity in neocortex into account in order to say something interesting about the activity within one area or a smaller neural structure within an area. This is because neocortex is compartmentalized into specialized areas and has several levels of organization.

Even if not all the neural activity in neocortex has to be taken into consideration when we wish to calculate the activity in a small part of neocortex the number of neurons is still staggering — tens of thousands of neurons in a square millimeter. And a square millimeter is indeed a very small part

of neocortex. It is obvious that we cannot manage simultaneous input/output relations among many thousand neurons in any other way than by using a computer.

Some researchers use parallel processing supercomputers and can study nervous systems of simpler organisms, e.g. lampreys, see Lansner et al. (?, ?), on a high organizational level or even in their entirety. Other researchers design models which have been very greatly simplified in all respects — synapse function, synapse plasticity, connectivity and number of neurons — and hope to say something interesting about mammals and especially humans from these models. The validity of the results depends, of course, on how judiciously the models are designed. The ultimate task of running a model of the entire human CNS with all its complexities at all levels is beyond the capacity of any supercomputer at the present, but there is no compelling reason to believe it will remain so for all future.

The first goal in this chapter is to acquaint the reader with some standard models of biological neural circuits. These models are sometimes called artificial neural networks, but a shorter form is gaining usage and we will call them neural networks.

The second goal of this chapter is to discuss some attempts that have been made to use such neural networks to explain autism. We believe it is wise at this time to pursue the study of autism with different neural network models. If, as there is ample reason to believe, see e.g. Gillberg and Coleman (?, ?), there are multiple etiologies for autism, it is entirely plausible that several different neural network models will remain relevant as our understanding of autism widens and deepens.

We agree with Douglas and Martin (?, ?) who state that "... it would be rash to press their [models which differ in important respects from biology] analogy to cortical circuits too far. Nevertheless, the potential usefulness of network models that are biologically based cannot be overestimated." It should be understood, however, that the use of neural networks will not provide "proof" of anything, but that a number of hypotheses or theories can be generated and these in turn can be subjected to neurophysiological and neuropsychological tests. Neural networks can also be used to test existing hypotheses and to strengthen or weaken them. Neural networks are emerging as a useful tool in research, but are not a final judge.

Have neural networks proven their worth in any biological context? Yes, results obtained from

neural networks have shown a remarkable likeness with experimentally found neural activity in animal sensory cortices, see e.g. Ritter et al. (?, ?).

Our presentation of neural networks is on a conceptual level. We have found a mathematical treatment to be beyond the scope of this chapter and instead we give references to suitable texts for the reader who wishes to pursue a study of neural networks.

2 A generic neuron's input/output characteristics and its modelling

There are many different types of neurons in the CNS but most neural networks have only one neuron model — a generic neuron is modelled. In Figure 1 such a generic neuron is shown. There is a dentritic tree where afferent signals reach the neuron through a number of synapses. From the cell body or soma, an axon will carry one efferent signal to the dendritic trees of many other neurons. This efferent signal is the result of the cell's processing of all the afferent signals. The signal has the form of trains of spikes of voltages as illustrated in Figure 2. There is a minimum frequency of such spikes — obviously there cannot be fewer than zero spikes per second — and there is a maximum frequency since there is a refractory period after a spike during which a new spike cannot be generated. A neuron is said to be firing when it produces trains of spikes of voltages.

(Figure here)

(Figure here)

The simplest model of a single neuron is shown in two possible forms in Figure 3. Each synapse is a junction where an afferent signal enters a dendrite. The afferent signals are aggregated along the dendrite to form the postsynaptic activity. The postsynaptic activity is typically a weighted sum of the afferent signals. This postsynaptic activity is then limited in the neuron body to ensure that the efferent signal stays between some minimum and maximum values. Often, but not always, the signals are modelled as voltage levels rather than as frequencies of spikes of voltages. These levels of voltages are simply the averages of the spikes — a high frequency train corresponds to a high voltage levels. All neuron models used in the attempts to model autism use voltage levels rather than spikes of voltages. This is done in order to simplify the models, of course. A critical question will then

(Figure here)

be: does this mean that these models will loose the possibility to display characteristics which are fundamentally important in autism? There is no definitive answer to this question today.

3 Neurons connected into circuits and neural network models

Neurons form circuits with different characteristics in different parts of the CNS. In cerebral cortex the neurons form a thin sheet, of typically a tenth of an inch thickness. The sheet consists of six layers with different types of neurons in different layers. The interconnections between neurons, both local and far-reaching, are arranged so that the different layers provide different types of connectivity.

The anatomy of cortex is known in much detail but in most neural networks no attempt has been made to model all the complexities of the cortical sheet. Some neural networks yield results where the neural activity show striking resemblance with that measured in animal experiments, whereas others are not designed to yield such similarities. It seems obvious that for modelling biological neural circuits, the interconnections in the neural network should be modelled on biological neural circuits. Models may, and for practical reasons must be, simplified, but what is modelled should be biologically motivated.

The rest of this section contains some mathematical terminology that might be unknown to some readers. The following sections on modelling autism with neural networks will be intelligible without a full understanding of this terminology.

There are two basic building blocks of the neural networks which have been employed in modelling autism: a layer of neurons, and a competitive layer. A layer of neurons is presented in Figure 4. The layer is created from single neurons as shown in Figure 3. There are p afferent signals connected to synapses of m neurons. Each synapse stores one parameter, a weight w_{ij}, which is used to form a postsynaptic signal as a weighted sum of the afferent signals. (Figure here)

A competitive layer is shown in Figure 5. In a competitive layer the efferent signals are fed back to the competing neurons and form the local self-excitatory and lateral inhibitory connections. The self-excitatory connection are marked with "1" and the inhibitory connections with "$-\alpha$". As the result of the competition the output of the neuron with the largest input signal ("the winner") is enhanced, (Figure here)

whereas all other outputs are suppressed.

Interconnection of layers of neurons forms a multilayer feedforward neural network or multilayer perceptron (MLP). In Figure 6 two views of a two-layer perceptron are shown. Multilayer feedforward neural networks are often called backpropagation networks due to a popular learning algorithm. (Figure here)

Self-Organizing Maps (SOMs) are competitive neural networks in which neurons are organized in a one- or two-dimensional lattice (grid) representing the **feature space**. In Figure 7 we present an example of a self-organizing map consisting of $m = 12$ neurons in which the **input space** is 3-dimensional $(p = 3)$ and the **feature space** is two-dimensional. The first section of the network is a distance-measure layer of neurons consisting of $m = 12$ dendrites each containing $p = 3$ synapses excited by 3-D input signal vectors $\mathbf{x} = [x_1 \ x_2 \ x_3]$ and characterised by the weight vector $\mathbf{w}_i = [w_{i1} \ w_{i2} \ w_{i3}]$. The distance-measure layer calculates the distances d_i between each input vector $\mathbf{x}$ and every weight vector $\mathbf{w}_i$. This distance information, $(d_1, \ldots, d_m)$ is passed to the competition layer, the MinNet in Figure 7, which calculates the minimal distance $d_k = \min d_i$ in order to establish the position of the winning neuron k. The competition is implemented through the lateral inhibitive and local self-excitatory connections between neurons in the competitive layer. In addition, every neuron is located at $l = 2$-dimensional lattice and its position is specified by an l-dimensional vector $\mathbf{v}_i = [v_{i1} \ v_{i2}]$. (Figure here)

The synaptic weight vectors, $\mathbf{w}_i$, and the vectors of topological positions of neurons, $\mathbf{v}_i$, are grouped into the weight and position matrices, W, V, respectively.

4 Learning is modification of synapses

Perceptions, thoughts, emotions, motor action are all manifestations of neural activity, the firing of many neurons. Learning is the process by which the neurons change their properties so that new or modified patterns of neural activity become possible. It is the efficacies or strengths of the synapses which define the possible neural activities and it is changes in these synapses which constitutes learning.

What could be the mechanism that changes synapses in such a constructive manner as to make

learning possible? Hebb suggested in 1949 (?, ?) in a famous statement that

> "When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic changes take place in one or both cells such that A's efficiency as one of the cells firing B, is increased."

It has since been shown that the strengthening of synapses is not sufficient to make learning possible, there must also be a mechanism which weakens synapses. A modified statement which takes this into account is: When the firings of two neurons are correlated then a synapse connecting them is strengthened but if the firings are not correlated then a synapse connecting them is weakened.

Learning according to this statement is called Hebbian learning and it has been demonstrated in the CNS. In order to employ Hebbian learning in a computer simulation it must be formulated in mathematical language. Such formulations have reached a high degree of sophistication but we do not need to go into these details here.

During the 1990's it became clear that a modification of a synapse is not only dependent on the activity of two neurons (the presynaptic and the postsynaptic neuron) but also on the activities of neurons in a neighborhood. A diffusive agent, nitric oxide, was found to be emitted from firing neurons and spread in a neighbourhood, strengthening some synapses and weakening others in that neighbourhood, see e.g. (?, ?).

Hebbian learning is a time consuming form of learning and much more efficient strategies for learning in a neural network have been designed. Again a question arises: will the results of simulations of learning be valid for biological learning if other, more efficient strategies for synapse strength modifications are employed? Kohonen (?, ?) has presented some computational results that answer this question in the affirmative, but caution should be exercised.

5 Learning in neural networks

There are three main paradigms for learning in neural networks, supervised learning, unsupervised learning or self-organization, and reinforcement learning. Arguably the last two are of biological relevance. It is, however, the first two that have been employed in modelling of autism.

In supervised learning a stimulus is entered to a neural network and the resulting output is measured and compared to a target, or desired output, for that particular stimulus. If, as can be expected particularly in an early phase of the learning process, the output is not close to the target, the synapse weights of the neural network are altered to bring the output closer to the target. An algorithm lets a computer calculate suitable changes of all the synapses. This process is repeated thousands of times with different stimuli. Even though this may be an effective learning process it is difficult to see a biological parallel to this calculation of synapse weights by the computer.

In unsupervised learning, or self-organization, the network just strengthens (compare Hebbian learning!) the tendencies present in the network in its initial state and seemingly magically adapts to the stimuli, meaning that the weights of synapses for one neuron or one group of neurons will adapt to and match the characteristics of one category of stimuli. A neuron's output is maximized when its input is the stimuli which it has adapted to. Other categories of stimuli will have other neurons adapted to them. The resulting neural networks are often called feature maps. Feature maps are important in sensory cortices.

In reinforcement learning a resulting output from a neural network is evaluated from some response from the environment as "good or bad". The neural network changes its synapse weights in some non-deterministic way and if the result is good than these changes will be strengthened, and if they are bad they will weaken. This learning has biological relevance and was understood in psychology before neural networks had been conceived. Thorndike's "law of effect" from 1911 (?, ?) offers this formulation of reinforcement learning:

> "Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have their connections with that situation weakened, so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond."

6 Cohen's model of autism

Cohen was first (?, ?, ?) to present explanations of characteristics in autism, based on the theory of neural networks. If a neural network has an excess of neurons and synapses for a given task it will learn every required response to the presented inputs exactly but will perform poorly on inputs which are slightly different than the learned inputs. Overfitting to learned inputs is the opposite of generalization and renders a neural network useless. A correctly dimensioned neural network has a sufficient number of neurons and synapses to learn the required response to the presented inputs but only in an approximate way. The advantage that stems from this is that it will also give good responses to inputs which have not been presented during learning but are similar to the ones which have been learned. Such a correctly dimensioned neural network has the capacity to learn to generalize well, a fundamentally important quality of neural networks. In Figure 8 the neural network's response to inputs presented during training and "new" inputs are shown. (Figure here)

Noting that parts of the brain, notably the amygdala and hippocampus, have been found to contain more neurons in brains from some individuals with autism than in normal cases, Cohen hypothesizes that this can cause autism with its characteristics of exact learning of facts and poor generalization. Cohen also argues that these characteristics in turn may also cause a demand for sameness, a cardinal feature in autism.

The case illustrated in Figure 8 has a one-dimensional input and a one-dimensional response or output. This is of course because this case can be easily visualized. Realistic situations which entail many-dimensional inputs and outputs cannot be visualized but it is known from neural network theory that a neural network will loose its capacity to generalize if is over-sized also in these more realistic situations.

Cohen uses a multilayer feed-forward backpropagation network such as the one presented in Figure 6 and suggests that even though these networks "are not similar in all respects to biological nervous systems, they mimic some of their properties and may help to explain the properties of real nervous systems." Testing Cohen's hypothesis on other artificial neural network structures more similar to biological nervous systems, in particular to those present in the amygdala and the hippocampus,

is needed.

7 Gustafsson's model of autism

Starting from a statement by Hermelin (?, ?) that "autistic children do not tend to integrate current experience with schemas and representations stored from previous sensory impressions", Gustafsson (?, ?) presented another model based explanation of autism. Cortical feature maps, specifically characterized by narrow neural columns, is argued to explain basically the same characteristics in autism, good discrimination but poor generalization skills, that Cohen had focussed.

Mountcastle (?, ?) had much earlier stated that (referring to cerebral cortex): "Whatever the level in the processing hierarchy, and the particular task there is a common architecture: the neural columns (mini- and macrocolumns)". It has been established that neurons in a minicolumn have similar but not identical sensitivities to stimuli. Gustafsson argues that narrow columns with fewer than normal neurons would be responsive to a narrower than normal range of stimuli and therefore exhibit good discrimination at the cost of poor generalization.

There was no experimental support for this idea in 1997, but in 2002 Casanova et al. (?, ?) reported that they had found an abnormal columnar organization (narrow and many minicolumns) in autism. What could cause this abnormal columnar organization?

The artificial neural network Gustafsson discusses in his arguments from 1997 is a self-organizing map with lateral excitatory and inhibitory feedbacks (both biologically motivated) in which synapses change according to Hebb's law. In such one-layered neural networks groups of neurons all of which are active upon presentation of one class of stimuli, such as a phoneme in speech, also emerge as a result of self-organization. These neural groups have the same function, albeit much simplified, as the neural columns in cortex. In the following we will assume that these one-layered neural groups may represent the six-layered neural columns in cortex and, in agreement with common usage, call them neural columns.

The width of these neural columns emerging in the model depends on the balance of excitatory and inhibitory effects as illustrated in Figure 9. It has long been known that (?, ?) too little excitatory

effect and too much inhibitory effect both result in narrow columns. (Figure here)

It is, however, not clear whether any such imbalance exists in autism. One argument against high inhibitory effects raised by Casanova et al (?, ?) is the relatively high co-morbidity of autism with epilepsy. Increasing the level of GABA is a well-established therapy against epileptic seizures. GABA is an inhibitory neurotransmitter and one would expect this to aggravate autism, if the hypothesized excessive lateral inhibition would hold. This is, however, not the case. In the case of co-morbid autism and epilepsy the hypothesized excessive lateral inhibition is obviously not convincing. The hypothesis might of course hold in the majority of cases of autism with no co-morbid epilepsy.

It is possible, of course, that the simple neural network model chosen by Gustafsson with its emergent neural columns does not adequately represent the columnar structure with both mini- and macrocolumns in cortex. If this is so, then causes for the narrow minicolumns in autism might not be possible to obtain from neural network theory, applying this model. See also (?, ?) for a discussion on this issue.

Searches for a genetic linkage to autism have shown that it is not a “single-gene disorder” but rather that multiple, possibly interacting, genes are involved in causing autism, see (?, ?). Therefore it is reasonable to search for complementary explanations for narrow minicolumns. Two such alternative/complementary explanations have been presented by Gustafsson (?, ?, ?)

8 Are narrow neural columns in autism an effect of a serotonin abnormality?

Although no strong candidate-genes for autism have been found, the linkage studies have indicated a serotonin transporter gene as the most consistent genetic linkage to autism (?, ?). A serotonin abnormality in the CNS has also been found in autism. In early development children with autism have been found to have a lower capacity to produce serotonin than normal children, but maintain this capacity while it declines in normal children (?, ?).

Serotonin has more than one role in the CNS, an early role being in synaptogenesis in sensory

cortices (?, ?, ?). The barrel fields in rats are decreased if serotonin levels are lowered (?, ?).

If serotonin plays a similar role in early development in humans, then the reported initial low capacity for producing serotonin could contribute in causing narrow cortical columns and thus conceivably autism.

We have good reason to believe that a serotonin abnormality contributes in causing autism and the narrow neural columns implicated in autism. The genetic studies suggest that we should try to find complementary mechanisms. Neural network theory proves helpful in this search.

9 A neural network theory finding: insufficient nitric oxide causes narrow neural columns

It is well established that nitric oxide plays an important role in synaptic plasticity, both for long term potentiation and long term depression, see e.g. (?, ?). There are also results that suggest that nitric oxide is important in the metasynaptic columnar organization of cortex, (?, ?).

The influence of nitric oxide on synaptic plasticity has been included into the mathematical models of synaptic modification, employed in neural network modeling. Simulations with such models have shown that a stable neural columnar structure emerges when the neural network self-organizes (?, ?, ?). Mathematical analysis of such self-organization has yielded as a correlate that the width of the neural columns depends on the production of nitric oxide during self-organization — low levels of nitric oxide results in narrow neural columns (?, ?, ?).

Results from two simulations of self-organization with nitric oxide as a mediator are shown in Figure 10. All factors except production of nitric oxide, were the same in the two simulations. The result from theory on the dependence of the neural column width on the level of nitric oxide is clearly illustrated. (Figure here)

It has been hypothesized by Gustafsson (?, ?) that insufficient production of nitric oxide could cause the narrow neural columns implicated in autism. Further aspects of this hypothesis are discussed in the following.

It is well known that vision is relatively spared in autism. This could then, according to the hypothesis, be explained if nitric oxide does not play a role in the columnar organization in visual cortex and there are results that suggest this to be the case (?, ?). In animal experiments it has also been directly demonstrated that visual discrimination and also other visual capacities are not affected by inhibition of nitric oxide (?, ?).

Nitric oxide has in animal experiments been demonstrated to have effects also in regard to epilepsy, but the effects are surprising: nitric oxide has been found to be both a proconvulsant and an anticonvulsant, see e.g. (?, ?, ?, ?). Nitric oxide has three different origins in the CNS, neuronal nitric oxide synthase (nNOS), endothelial nitric oxide synthase (eNOS), and inducible nitric oxide synthase (iNOS). Selective inhibition of nNOS makes it more difficult to induce epilepsy but inhibition of eNOS makes it easier to induce epilepsy. This would indicate that insufficient nNOS could contribute in causing autism without co-morbid epilepsy while insufficient eNOS could contribute in causing autism with co-morbid epilepsy. iNOS appears to be similar to nNOS in this respect.

10 Does columnar structure in neural networks represent columnar structure in cortex?

There is presently no definitive answer to the question what might cause narrow minicolumns in autism. In neural networks of the self-organizing kind neural columns do emerge, both as a result of lateral excitatory and inhibitory feedback and as a result of the influence of a diffusive messenger, assumed to be nitric oxide. It is, however, not certain that this emergence of neural columns models the columnar structure in cortex. The latter contains two levels, the minicolumns and the macrocolumns.

In the self-organizing neural networks discussed above there is no given initial structure. The columnar structure emerges to fulfill the functional requirements to represent different classes of the stimuli presented to the network.

In cortex there is an initial structure in the form of ontogenetic columns (?, ?), not stimulus-driven but prenatally formed. If the functional minicolumns are the same as the ontogenetic columns, then

the synaptic plasticity of the neurons in the minicolumn allows for tuning the minicolumn to a given stimulus, but the minicolumn itself would not be an emergent structure, as the columns are in the neural network models.

Can the results obtained from a model which is not initially constrained by grouping neurons together in columns be of interest for understanding the development of cortex and the activities in minicolumns? One obvious possibility is that the development of a neural column of the model represents the tuning of the neurons in the minicolumn to a particular stimulus with some limited variability. This is all the more natural an interpretation since both the lateral inhibitory feedback surrounds the minicolumn (?, ?) and the activity of nitric oxide is prominent along the minicolumnar periphery (?, ?).

A narrow neural column in the model would correspond to the case where few neurons in the minicolumn get tuned to the stimulus. The total neural activity from such a minicolumn as a response to a stimulus would then be lower than normal. The total number of neurons in the minicolumn would set an upper limit to the width of the neural columns formed in the model — larger widths could not then be interpreted in a meaningful way.

A correspondence between the neural network column and the cortical macrocolumn should not be ruled out, however. The basket cells are inhibitory interneurons that reach across a macrocolumn and the diffusion of nitric oxide and its sphere of influence is estimated to reach a distance exceeding 150 mm (?, ?) i.e. covering several minicolumns. Both these properties of cortex can be included in a self-organizing neural network and result in neural columns much wider than the minicolumn, possibly corresponding to the macrocolumn. The shape of the macrocolumns of sensory cortices may be determined by information from sensory inputs rather than being prenatally formed as the minicolumns, for a review see (?, ?). In this respect the macrocolumns of cortex correspond to the neural network columns.

It should be noted that (?, ?) argue that the increased number of minicolumns in autism may have its origin in a disruption in the early prenatal development and that the larger number of minicolumns rather than their width is important in causing autism.

11 McClelland's model of autism

In 2000 McClelland, like Cohen and Gustafsson before, takes hyperspecificity, or poor generalization in autism as the starting point for a discussion drawing on insights from design of neural networks that can group similar objects into one category when they are sufficiently close but distinguish between objects which should form different categories (?, ?).

The balance between generalization and discrimination is of paramount importance in the design of neural networks. McClelland suggests that one technique employed in the design of neural networks to facilitate discrimination, conjunctive coding, might be utilized also in the central nervous system. McClelland further suggests that

> "in the brains of children with autism, they may be predisposed to use an excessively conjunctive form of neural coding [...] This could prevent them from exploiting overlap in cases where overlap leads to the useful ability to generalize. Instead, it would leave the child with the ability to learn associations to particular, specific inputs and without the ability to extend what they have learned to other similar experiences."

McClelland does not suggest any biological mechanism which would provide for this conjunctive coding and identifying such a mechanism will require further work, preferably employing neural networks which more closely model parts of the central nervous system than the traditional feed-forward networks discussed by McClelland do.

12 The temporal binding deficit hypothesis of autism

It has been known for more than a hundred years that cerebral cortex has functionally specialized areas. These areas have extensive two-way connections and simultaneous neural activity in several areas may constitute e.g. the neural response to a complex object. It is then of course necessary that the connections between the areas involved function properly. In 2002 Brock et al. suggest that the different areas may work well seen as single entities but do not function properly together and

therefore do not integrate different properties of, e.g. a complex object, well (?, ?). This would offer a neural explanation of the weak central coherence theory of autism, proposed in 1989 by Frith (?, ?).

The hypothesis is not based on neural networks, rather on biological and behavioural arguments, but the authors compare this hypothesis with the three models based on neural networks presented above and stress that whereas these models are relevant within areas, i.e. they are local models, the binding deficit hypothesis is of a global character. This hypothesis needs to be investigated experimentally but it is of a kind which could also be advantageously studied by employing neural networks.

13 Learning under attention shifting restrictions shows autistic characteristics

Attention abnormalities, among them attention shifting abnormalities, are common in autism. The nature of the restricted attention shifting is not agreed upon, however. One school of thought, led by Courchesne (?, ?, ?, ?, ?, ?, ?), suggests that there is a general attention shifting impairment caused by the deficit in number of Purkinje cells in the cerebellum, an almost universal finding in autism. A number of researchers, among them Dawson et al. (?, ?), Pascualvaca et al. (?, ?) and Minshew et al. (?, ?), hold a different opinion — there is a higher executive function impairment which restricts attention shifting. Some of these authors argue that people with autism tend to not shift attention to a source of stimuli which they expect to be novel, they are influenced by novelty avoidance or familiarity preference. This of course can be seen as a consequence of the insistence on sameness which Leo Kanner in 1943 (?, ?) found to be a prominent feature of autism.

Gustafsson and Papliński have translated this scientific debate into a test, employing neural networks (?, ?, ?). They present stimuli from two different sources to a Kohonen (?, ?) self-organizing map such as presented in Figure 7, which during learning adapts to the stimuli from one or both sources. It should be stressed that the neural networks are assumed to be completely adequate for the task they are presented with, it is only the attention shifting which shows any abnormality.

The simulation employs three different modes of attention shifting when a stimulus is presented

by the source momentarily not attended to. Normal learning is understood to mean that attention is then unconditionally shifted — normal learning is understood to be novelty seeking. Attention will shift but only with a low probability in the case of a general attention shift impairment.

The case of attention shifting restricted by familiarity preference is more complicated. In the beginning of the learning process both sources are unfamiliar to the self-organizing map and attention is shifted unconditionally. Then as familiarity develops the shifting will show a statistical bias towards the most familiar source. Finally, if both sources become familiar to the map, attention will shift unconditionally.

The results have so far, without exception, yielded maps of the same character for normal learning and learning when attention shifting is restricted by a general attention shifting impairment. Learning when attention shifting is restricted by familiarity preference yields maps characterized by detailed learning of the source which exhibits the least variability among its stimuli. This is detailed learning in narrow fields, a learning with arguably autistic characteristics.

This should not be seen as a proof that the general impairment in attention shifting hypothesis is invalid, there are many more comprehensive simulations that must be done before such a statement could be warranted.

In the simulations presented below the two sources contained stimuli, or objects, that were for one source a number of animals (mammals, birds, fish, reptiles) and for the other source a number of felines. Naturally the feline source has much less variability than the general animal source. In Figure 11 we show the resulting maps consisting of 16 neurons, arranged in a 4×4 matrix, and how they have adapted to the different stimuli. Each stimulus is shown at the neuron which has adapted best to that stimulus. Since there are more stimuli than neurons in this case, some stimuli will share a neuron. (Figure here)

In Figure 11**a** a typical map resulting from normal learning and from learning with a general attention shift impairment is shown. This map shows a very economic use of the neural capacity to represent all stimuli. Stimuli which are not similar to any others, like the gray whale and the anaconda have “their own” neurons with exact adaptation (the number after each stimulus is a measure of the adaptation with a smaller number being better). Stimuli which are similar, like the great cats share

one neuron. Some neurons have not been identified with any stimuli. This is a common situation — some neurons are “dead”, i.e. they have not adapted to any stimulus, while others have but not as well as those identified with the stimuli.

In Figure 11**b** a typical map resulting from learning with attention shifting restricted by familiarity preference is shown. Here the map has ceased shifting attention to the source containing general animals, devoting almost all the neural capacity to the feline source, learning its stimuli very well, in this case exactly.

Even though these simulations employ neural networks to yield the maps shown above, there is also a rule driven part of the system which is simulated. The attention shifting rules are written to agree in character with experimentally verified behaviour but they are not modelled as neural networks. In autism attention shifting is, of course, also caused by neural activity. A complete neural model of autism based on attention shifting restricted by familiarity preference will demand a non-trivial research effort.

14 Simulating a scheme for early intervention in autism

The process that leads to exact learning of the source with the least variability when attention shifting is restricted by familiarity preference can be observed — when an attention shift to the alternate source is rejected we understand that that source is the least familiar of the sources.

It is then conceivable that we can overexpose the neural network to stimuli from that source to compensate for this relative unfamiliarity. This compensatory action will very possibly have to be repeated but can lead to a map which learns both sources with a resulting map which is identical to that obtained in normal learning. The learning process is, however, entirely different.

In Figure 12 we show two diagrams, representing different aspects of the same learning process. In the top diagram we see how the number of attention shifts initially grows linearly and then grows less rapidly as familiarity preference starts to play a role in the learning process. Finally the number of attention shifts again grows linearly, as rapidly as in the initial stage, this is after the neural network has become familiar with both sources. (Figure here)

The intervention activity which makes it possible for the neural network to develop into a normal map is shown in the lower diagram in Figure 12. The probability for presenting a new stimulus from the source with general animals is in the initial stage constant and such that each individual stimulus has an equal chance for presentation, regardless of source. When familiarity with the feline source start to manifest itself through fewer attention shifts to the source with general animals, then the probability for presenting a stimulus from the source with general animals is immediately and drastically changed. After some such large changes of probability have passed the neural network has become familiar with both sources and no more intervention is necessary.

15 Suggestions for further reading

A reader who is willing to face some mathematics can find a great many books and journal papers on neural network theory. Some of this literature is devoted to the study of biological nervous systems and to "mind modelling". We will here make two suggestions which for the most part demand high school mathematics or first year college calculus and algebra.

Neural Networks and Brain Function by Rolls and Treves (?, ?) gives an overview of different neural network architectures and discusses their usefulness in explaining function in different parts of the CNS, such as sensory cortices, hippocampus etc.

Rethinking Innateness: A Connectionist Perspective on Development (Neural Networks and Connectionist Modeling) by Elman et al. (?, ?) argues strongly for the importance of the environment, as taken in by the senses, in the development and organization of cerebral cortex and uses neural networks broadly in the argumentation.

Acknowledgements

This work is part of a cooperation between Luleå University of Technology, Luleå, Sweden and Monash University, Australia. We wish to express our appreciation to the financial support from Stiftelsen Einar och Ragnhild Lundströms minne, Sweden. We would like to thank Professor Manuel Casanova, Juan Trippe and Andrew Switala from University of Kentucky for helpful discussions.

List of Figures

1 A generic neuron 21

2 Train of spikes of voltages 22

3 Two representations of a single neuron: left — dentritic form, right — signal-flow form. 23

4 A layer of neurons 24

5 Two views of a competitive layer of neurons. 25

6 Various representations of a Two-Layer Perceptron 26

7 A 2-D SOFM with $p = 3;\ m = [3\ \ 4];\ l = 2$. 27

8 Left figure shows a good fit to the test set and mediocre fit to training set. Right figure shows a poor fit to the test set but perfect fit to the training set. 28

9 Left figure shows "normal" activity column resulting from good balance of excitatory and inhibitory lateral feedback. Right figure shows narrow activity column resulting from poor balance — weak excitatory/strong inhibitory lateral feedback 29

10 Narrow and wide neural columns driven by the nitric oxide level. In the left figure the nitric oxide level is low and in the right it is high. 30

11 The feature maps developed in the **a**. novelty seeking, and **b**. attention shifting restricted by familiarity preference learning modes. The shaded ovals represent the network response to a test animal 31

12 Early intervention. The pA probability is the probability that the next stimulus presented will be chosen from the source containing general animals. The lower row of rejection states indicates when attention shifting to the source containing general animals has been rejected and the upper row of rejection states indicates when attention shifting to the source containing felines has been rejected. 32

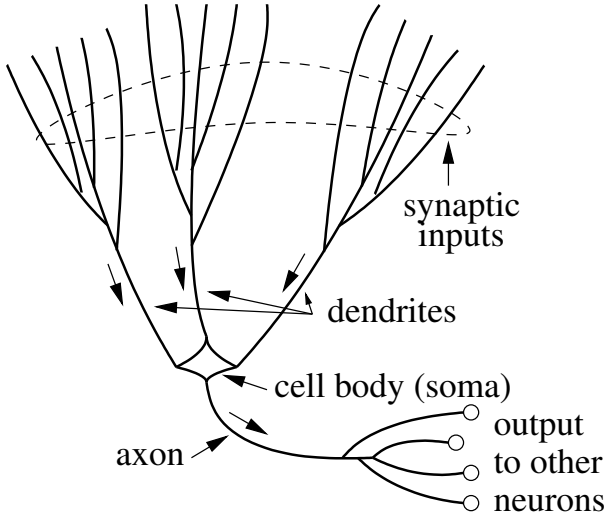


Figure 1: A generic neuron

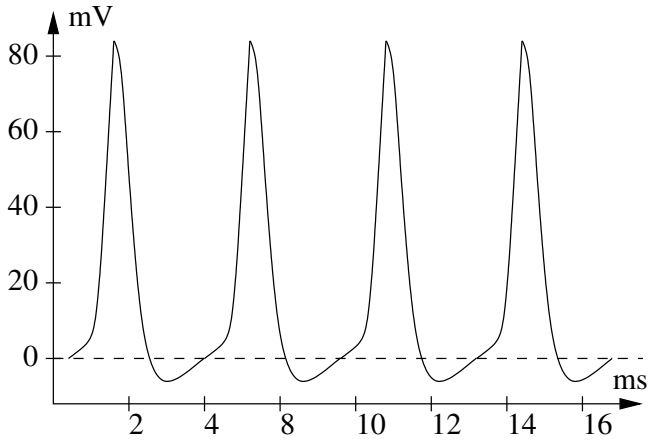


Figure 2: Train of spikes of voltages

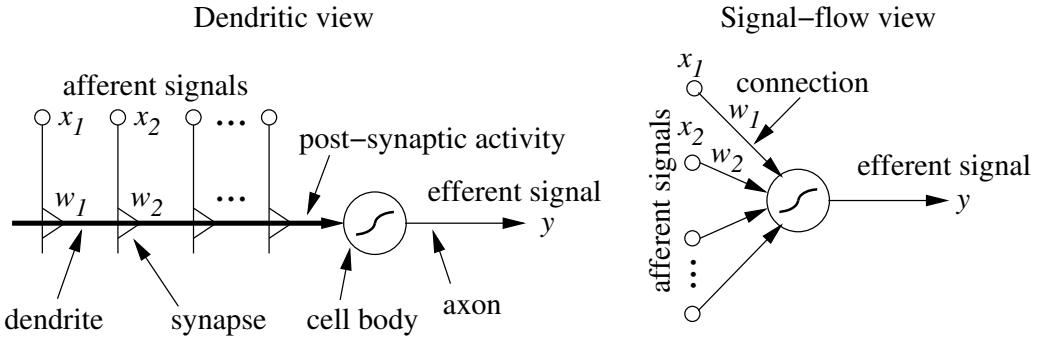


Figure 3: Two representations of a single neuron: left — dentritic form, right — signal-flow form.

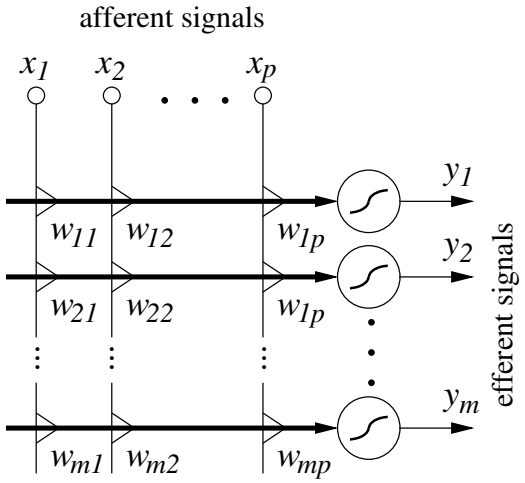


Figure 4: A layer of neurons

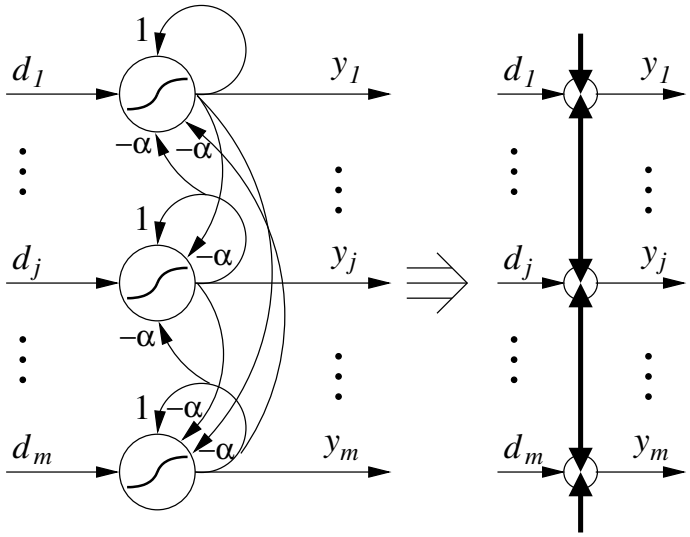
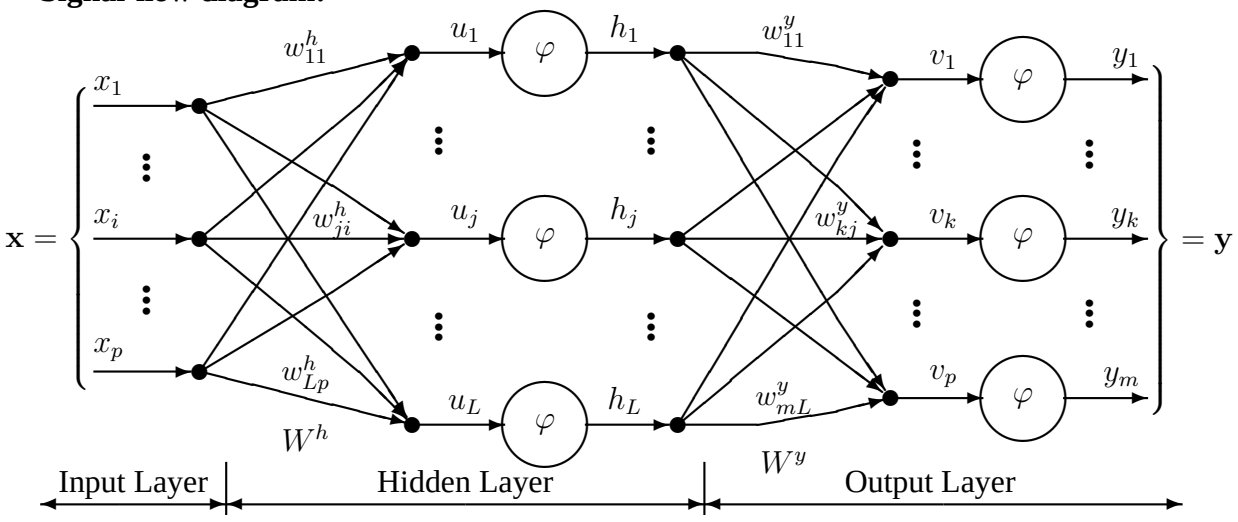


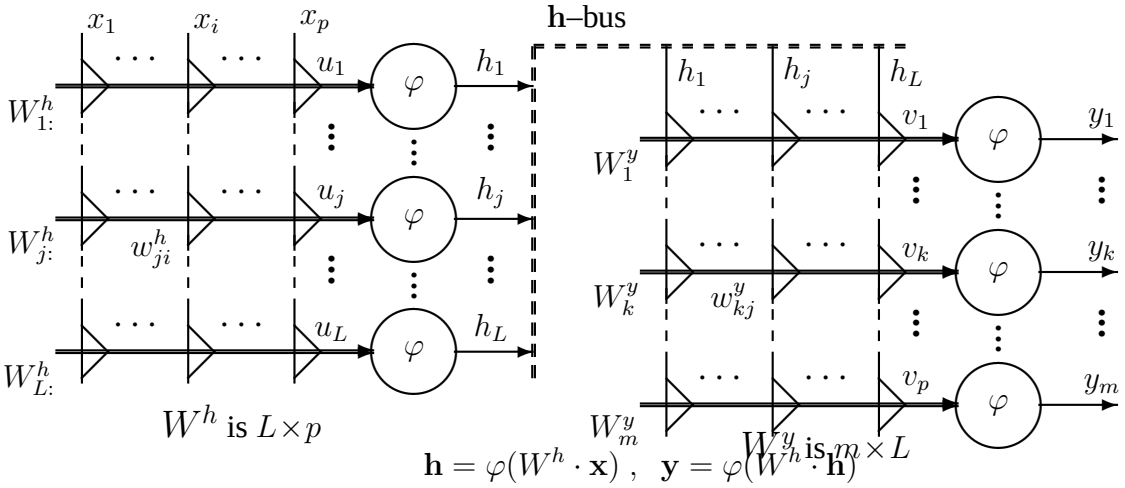
Figure 5: Two views of a competitive layer of neurons.

Signal-flow diagram:



$$u_j = W^h_j \cdot \mathbf{x}\ ; \quad h_j = \varphi(u_j)\ ; \quad v_k = W^y_k \cdot \mathbf{h}\ ; \quad y_k = \varphi(v_k)$$

Dendritic diagram:



$$\mathbf{h} = \varphi(W^h \cdot \mathbf{x})\ , \quad \mathbf{y} = \varphi(W^y \cdot \mathbf{h})$$

Figure 6: Various representations of a Two-Layer Perceptron

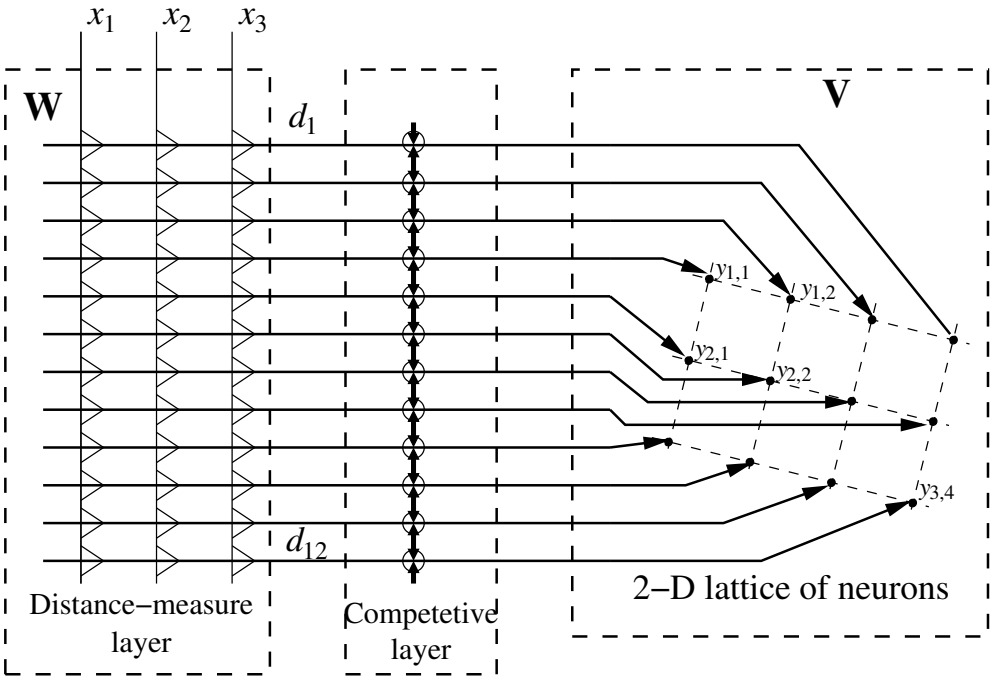


Figure 7: A 2-D SOFM with $p = 3;\ m = [3\ \ 4];\ l = 2$.

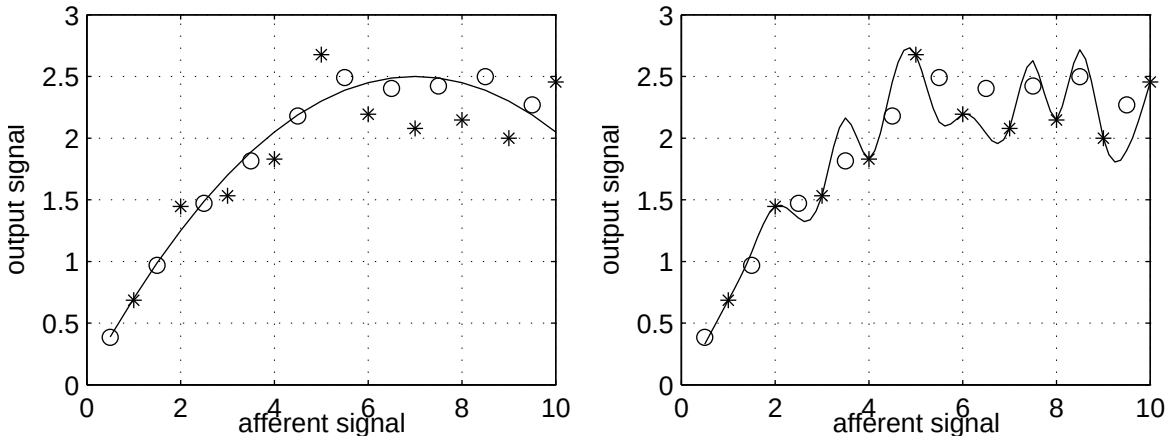


Figure 8: Left figure shows a good fit to the test set and mediocre fit to training set. Right figure shows a poor fit to the test set but perfect fit to the training set.

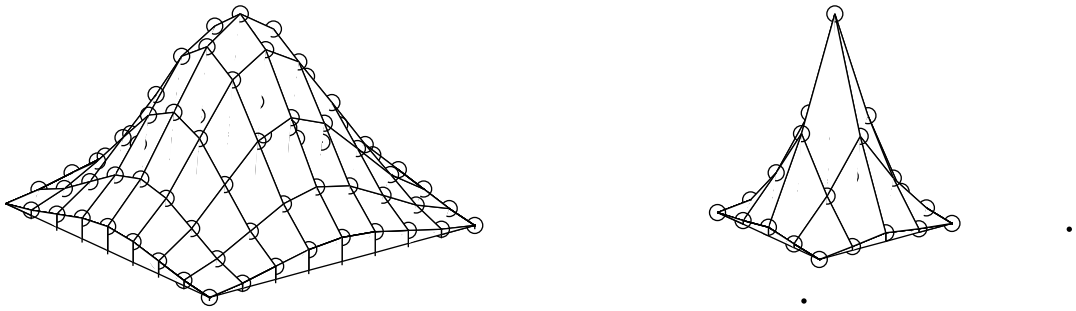


Figure 9: Left figure shows “normal” activity column resulting from good balance of excitatory and inhibitory lateral feedback. Right figure shows narrow activity column resulting from poor balance — weak excitatory/strong inhibitory lateral feedback

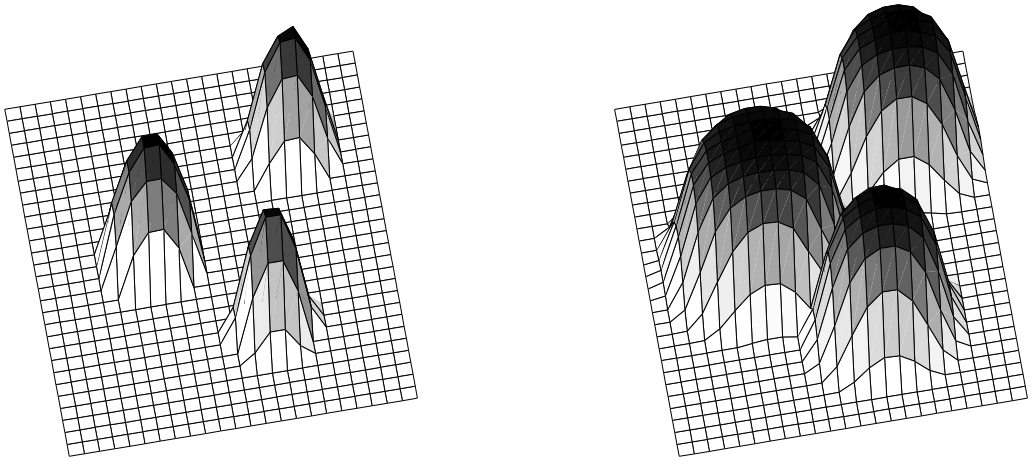


Figure 10: Narrow and wide neural columns driven by the nitric oxide level. In the left figure the nitric oxide level is low and in the right it is high.

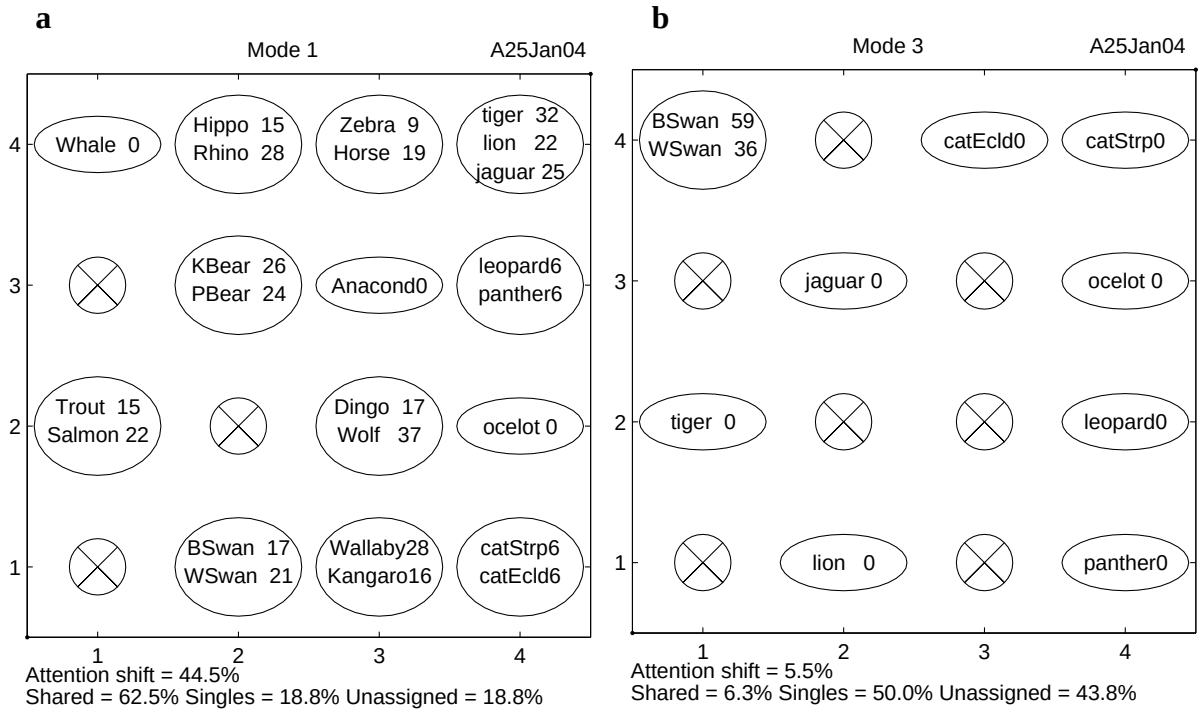


Figure 11: The feature maps developed in the **a**. novelty seeking, and **b**. attention shifting restricted by familiarity preference learning modes. The shaded ovals represent the network response to a test animal

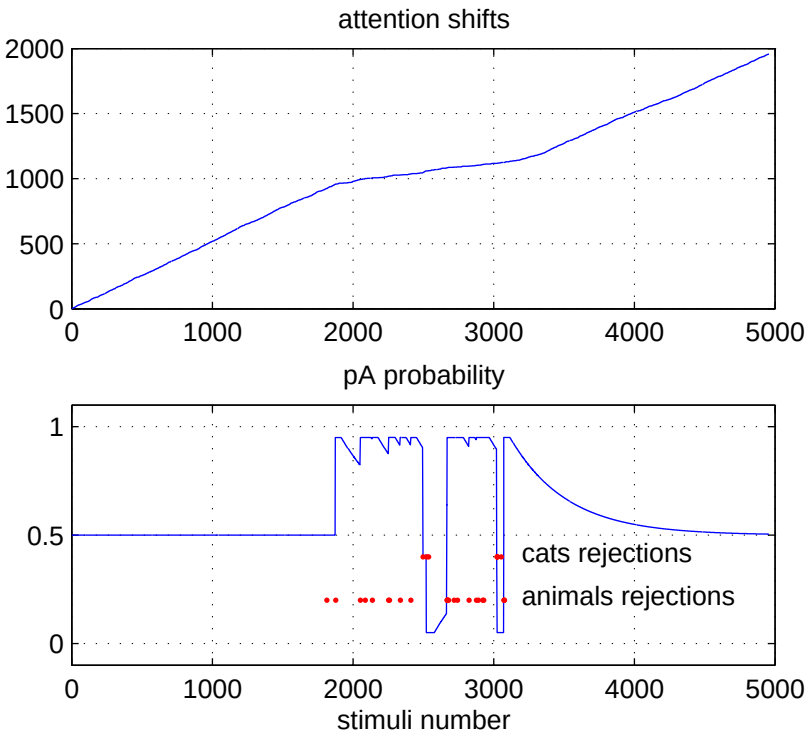


Figure 12: Early intervention. The pA probability is the probability that the next stimulus presented will be chosen from the source containing general animals. The lower row of rejection states indicates when attention shifting to the source containing general animals has been rejected and the upper row of rejection states indicates when attention shifting to the source containing felines has been rejected.