

A Neurocomputational Model of Anticipation and Sustained Inattentive Blindness in Hierarchies

Anthony F. Morse, Robert Lowe, and Tom Ziemke

COIN Lab, Informatics Research Centre, University of Skövde
SE-541 28 Skövde, Sweden
{anthony.morse, robert.lowe, tom.ziemke}@his.se

Abstract. Anticipation and prediction have been identified as key functions of many brain areas facilitating recognition, perception, and planning. In this chapter we present a hierarchical neurocomputational model in which feedback, effectively predicting or anticipating task-relevant features, leads to sustained inattentive blindness. A psychological experiment on sustained inattentive blindness in human subjects is simulated to provide visual input to a hierarchy of Echo State Networks. Other parts of the model receive input relevant to tracking the attended object and also detecting the unexpected object, feedback from which is then used to simulate engagement in the task and compared to results obtained without feedback, simulating passive observation. We find a significant effect of anticipation enhancing performance at the task and simultaneously degrading detection of unexpected features, thereby modelling the sustained inattentive blindness effect. We therefore suggest that anticipatory / predictive mechanisms are responsible for sustained inattentive blindness.

Keywords: Enaction, Anticipation, Prediction, Neurocomputation, Reservoir Systems, Association, Sustained Inattentive Blindness, Neural Modelling, Cortical Hierarchies.

1 Anticipation and Prediction; From Neuroscience to Cognition

In work published elsewhere [1, 2] we presented a model of sustained inattentive blindness in which predictive feedback enhanced performance at the feedback relevant task while degrading performance at other tasks. Furthermore by systematically varying the similarity between target and unexpected stimuli we were able to replicate human data showing that the size of the sustained inattentive blindness effect is reduced as similarity is increased. Somewhat surprisingly we also found that in our model, the size of the sustained inattentive blindness effect is also reduced as similarity decreases beyond a half way point. This prediction has yet to be confirmed in human data but provides an example of the classic U shaped curve where detection of the unexpected object is most likely if it is very similar or very dissimilar to the tracked object. Central to this model was the use of a predictive feedback signal which was artificially provided. It was therefore not clear where this predictive feedback would originate from, just that given this kind of feedback the inattentive blindness effect was present. In this chapter we extend that original work providing a

minimal hierarchy of simulated cortical micro-columns making explicit how and why such predictive feedback signals can be generated and used.

Neural mechanisms for prediction and anticipation are thought to be pervasive throughout much of the brain and are thought to vary greatly in instantiation in different anatomical structures. For example, Downing and others [3-5] provide a review of neurocomputational models of the cerebellum, the basal ganglia, the neocortex, the hippocampus, and the thalamocortical loop. In each of these models, and reflecting the underlying neuroscience, the anticipatory systems differ both in their mechanics and functional significance, which ranges from the anticipation of relational and sequential data in neocortical models, to the prediction of reward stimuli in the cerebellum and basal ganglia. Hawkins and Blakeslee [6] have suggested that “the cortex’s core function is to make predictions” [p. 113] and that “all predictions are learned by experience... if there are consistent patterns among the inputs flowing into your brain, your cortex will use them to predict future events” [p. 120]. This places prediction centrally in our understanding of brain function. To understand how this predictive process works, Hawkins depicts the neocortex as a macro hierarchy of homogeneous micro-circuits or structures, meaning that the same basic unit is repeated throughout the neocortex but the way they are wired together and thus their influence on each other varies from region to region. The hierarchical view of the cortex highlights the interplay between low levels (i.e. regions of cortex close to sensory input such as area V1 in the visual cortex) responding to fast changing ‘features’ and higher levels (such as the inferotemporal or prefrontal cortex) responding to more invariant, larger scale, and slower changing things such as faces and objects. Here, according to Hawkins, top-down signals indicating expectations or predictions, of which low level features will be active, not only fill in sensory gaps, but facilitate the higher level response and differentiate expected from unexpected bottom-up activity. In this chapter we develop such a model but highlight a different aspect of top-down predictive signals, specifically that they can be used to ‘tune’ input filters facilitating recognition of the anticipated features with consequences for the recognition of non-anticipated features.

Accounts of cognition rooted in underlying predictive or anticipatory neuroscience are becoming increasingly common. Gallese and Lakoff [7], for example, cite extensive neuroscientific evidence for predictive / anticipatory circuitry in the sensorimotor systems of the brain and go on to propose a sensorimotor theory of conceptual knowledge based therein. The point to note here is that prediction and anticipation are not simply ancillary functions of the brain but are central to its proper functioning and their existence is heavily supported by neuroscientific data. Such theories all have in common the idea that feedback or spreading activation from active regions, serves the role to anticipate, predict, and prime other related structures as a significant (and by some theories necessary) part of the cognitive perceptual process. In many ways this is not a new idea and has much in common with spreading activation models in early connectionism [8] and also psychological theories of associationism [9]. Unlike those early models and theories, which were largely disembodied, the current focus on embodied and situated cognition or enactive perception provides a different perspective from which to view these problems.

“...Human evolution has indeed led to increasingly complex forms of behaviour. However, these behaviours are not simply triggered from genetically determined mechanisms. Rather they are the outcome of the gradual formation of internal representations during the lengthy process of ontogenetic development.” [10][p. 144].

While the adult neocortex contains many function or behaviour specific areas, as Karmiloff-Smith [10] shows with extensive fMRI data, these functions are not genetically pre-specified but are globally processed in children and gradually formed into these areas by a process of active re-structuring according to relationships in external stimuli. Taking Hawkins [6] perspective this re-structuring is a matter of modifying the connectivity and thus influence of the already hierarchically connected cortical micro-columns. By comparison genetic disorders such as Williams Syndrome (WS), “appear to follow a deviant developmental pathway” [10][p. 151] whereby genetic alteration changes the adaptive mechanisms in the brain that would otherwise have led to the formation of these modules, as a result of which brain structuring by experience follows a different path. Thus what is different is not some genetic predisposition to develop such and such a capacity but rather the mechanics of the re-structuring process itself.

While the plasticity of the cortex is well recognized, the extent of this re-structuring is brought into focus by the work of Sharma et al [11] who performed a number of experiments in which the auditory nerve and the optic nerve were cut and crossed over in infant ferrets. The ferrets subsequently developed visual orientation modules in the auditory cortex (these are structures only ever previously discovered in the visual cortex, and never normally present in the auditory cortex). This suggests that rather than the visual and auditory cortex being specialized for one type of processing, they are actively structured around the input streams they receive. This is a radical view to take but one for which there is a great deal of evidence. Clearly evolution has played a part in specialising regions of cortex but rather than providing innate structures, we suggest rather that evolution has refined the relative sizes of the various unimodal and polymodal regions involved in processing different sensory streams and their innate hierarchical structure. As an example, in primates the visual areas of the cortex are significantly larger than the auditory areas and so the auditory cortex is more limited in the structures it can construct than the visual cortex is. To some extent this can be seen as division of resources so that in our case greater resources are given to vision than to audition.

Complementing this perspective is Mountcastle’s [12] view of the cortex as performing the same operation everywhere. While most neuroscientists are highlighting the differences (functional or anatomical) between different regions of cortex, Mountcastles’ analysis suggested that the cortex consists of the same basic unit, the cortical micro-column, everywhere one looks. While there may be some variation in micro-columns in different regions they are to a large extent very similar throughout the cortex. This is the basic unit of the hierarchy that we model herein and we shall return to the cortical micro-column in section 5.

This aim of this chapter is to investigate and model some of the less obvious consequences of prediction or anticipation in hierarchical structures and their effect on an account of cognitive processes. To provide an example we further develop a cortical microcircuit model of sustained inattentional blindness already reported elsewhere [1, 2] implementing a hierarchical structure based on the literature just introduced. The resulting models are demonstrated to preserve the sustained inattentional blindness effect. In the next section we introduce a cognitive theory of perception and work back to expose possible mechanisms able to implement such an account, in the hope that these match to some extent those neuroscientific mechanisms just introduced.

We then, in Section 3, highlight some of the computational difficulties in implementing these cognitive theories in an embodied and situated agent. Section 4 departs somewhat from this discussion and introduces psychological literature on the phenomena of sustained inattentional blindness, which is the target of the modelling experiments provided in Section 5. Finally Section 6 provides some discussion and conclusions to this work.

2 Perception and Action; From Cognition to Neuroscience

The central claim of Noë's [13] enactive theory of perception "is that our ability to perceive not only depends on, but is constituted by, our possession of... sensorimotor knowledge." [p. 2], where sensorimotor knowledge "is implicit practical knowledge of the ways movement gives rise to changes in stimulation." [p. 8]. This means that, sensorimotor knowledge is not simply factual knowledge about a domain but is intimately about the dynamic relationship between an agent and its environment. Mastering this dynamic relationship is manifest in the ability to predict the sensory consequences of actions, which in turn, according to Noë is constitutive of perception. Under this view the distinction between skills and knowledge is collapsed [14], as Maturana and Varela [15] put it, "all doing is knowing, and all knowing is doing". For example, to perceive something as a round plate is to exercise a particular skill predicting how sensory contact with the plate will change as one moves a little this way or a little that way. Such perception can be mistaken and such mistakes 'pop out' when these predictions are invalidated by further experience (consider the Ames distorted room illusion [16]). Such predictive ability would seem essential to any embodied agent interacting with a complex environment; for example, simulating or anticipating the effects of possible actions for evaluation in an agent-centred way (e.g. [17-20]). Here theories such as the simulation hypothesis suggest the re-use of, primarily, sensory pathways. The idea is that sensory data is processed as normal but rather than producing an overt motor response, the response is used to generate a prediction of what the next sensory input would be had that response been overt. By projecting this predicted sensory state back into the sensory areas the process can re-use the existing circuitry to iterate the prediction process further and further ahead in time. For more detail on the biological basis of the simulation hypothesis see the chapter on this subject by Svensson et. al. also in this book.

For Noë [13], the ability to make predictions is simply the application of sensorimotor knowledge which comes from finding "pattern[s] in the structure of sensorimotor contingency" [p. 103], i.e. patterns in the relationships between our actions and sensations. While this may seem to the uninitiated trivial to implement there are significant problems in the application of such pattern recognition, as we will discuss in Section 3. Going beyond Noë's formulation we would draw a distinction between 'shallow' perception of experience, which could result from merely recognising the relationship between optic flow and turning the head, and 'deep' perception of a world consisting of objects and affordances. 'Deep' sensorimotor knowledge requires the recognition of profiles of change and provides a means to recognize Gibsonian [21] affordances, i.e. "to perceive... is to perceive structure in sensorimotor contingencies. To see that something is flat is precisely to see it as giving rise to certain

possibilities of sensorimotor contingency [*to see its affordances*]. To feel a surface as flat is precisely to perceive it as impeding or shaping one's possibilities of movement" [13] (*italics added*) [p. 105]. For any agent then, skilled action is practical knowledge, the mastery of which can be achieved through identifying the contextual regularities between action and sensory perception. Here we can see a strong link between Noë's account of perception, Gallese and Lakoff's [7] theory of conceptual knowledge, and Hesslow's [19] simulation hypothesis.

This formulation of the enactive approach not only suggests that robots or agents could learn to "perceive an idiom of possibilities for movement" [13] [p. 105], but also suggests that such a capacity is both particularly amenable to the integration of other cognitive phenomena, such as planning (for which prediction would seem necessary [18, 19, 22]), and significantly different from mainstream representational theories (e.g. [23, 24]). Although evolution obviously has a role in this, it seems reasonably clear that much of our own world knowledge and skills are either derived from, or heavily shaped by, our life-time experiences. Learning sensorimotor relationships from experience and then using these learnt relationships for prediction is the target of our neurocomputational model.

While Gallese and Lakoff's [7] theory of conceptual knowledge differs from Noë's, the general form of both theories of cognition based on sensorimotor prediction have much in common. Having now briefly reviewed some of the biological (section 1) and the philosophical (section 2) theories of sensorimotor cognition, we turn in the next section to the computational problems any implementation of these perspectives faces.

3 Circularity, Regularity, and Time, but Not in That Order...

Acquiring the ability to predict changes in sensory streams that result from either motor actions (simulated or real), or from temporal aspects of our environment and embodiment requires an ability to learn temporal sequences and relational information. The latter would seem to be explained by known plasticity in biological nervous systems, roughly approximated by Hebbian plasticity [9] 'what fires together wires together'. Such plasticity is highly suggestive of the idea that co-occurring features, presumably indicated by the activity of sub-sets of neurons would lead to those neurons 'wiring' together with the result that activity in one set (caused by the presence of the relevant feature in sensory data) would lead to a spreading activation via the new wiring resulting in activation of the other set of neurons. Such ideas are well developed in localist connectionist models which are able to replicate a great deal of psychological data from a variety of phenomena such as; semantic and associative priming [25], cued recognition, name processing and lexical processing [26], semantic and repetition priming [27-29], face recognition and visual prosopagnosia [30], classical and operant conditioning [29, 31], and many many more. Though such connectionist modelling is often far abstracted from computational neuroscience the basic principles of spreading activation (rather than localism) would seem to be highly plausible biological mechanisms capable of accounting for a great deal of psychological phenomena. As Page states: "I make no claim to be the first to note each of these properties; nonetheless, I believe the power they have in combination has either gone unnoticed or has been widely underappreciated." [32][p. 450]

The power of such models to explain embodied cognition is highly limited by their assumption of localist representation which bypasses important questions such as where these representations come from and what from they take. Indeed for many the idea of an internal representation is a Cartesian regression with no explanatory power whatsoever, as Harvey puts it: “The gun I reach for whenever I hear the word ‘representation’ has this engraved on it: ‘When *P* is used by *Q* to represent *R* to *S*, *who is Q and who is S?’” If others have different criteria for what constitutes a representation, it is incumbent on them to make this explicit” [33]. While we do not wish to delve into the representation debate here it is clear that localist representation is neither present in neurobiology nor forthcoming in modelling without introducing significant design bias. Detectors can often be achieved by supervised training or complex statistical analysis, however the design choice of which features to detect will always be a limiting factor on such approaches. Restricting modelling methods to the more biologically plausible varieties of plasticity raises a serious problem concerning the separation of features in data streams. This is *the problem of marginal regularity*, as Kirsh [34] puts it; “what if so little of regularity is present in the data that for all intents and purposes it would be totally serendipitous to strike upon it? It seems ... that such a demonstration would constitute a form of the poverty of stimulus argument” [p. 317]. As anyone working in the fields of robotics, machine learning, or pattern recognition will tell you, this ‘what if’ is in fact the case for the vast majority of data and especially for the kinds of things that we humans seem to perceive so effortlessly.

The *problem of marginal regularity* results from the fact that most interesting features of a data stream, be it sensory or whatever, are not explicitly represented in that data stream, they are in fact relational by which we mean that their presence is indicated by specific relationships between the ‘bits’ of the data streams. For example the image of a cup is distributed over many pixels or retinal cells, the majority of which take very similar values for a range of different images. What changes between each image is the relationships between the pixels; however, these relationships can change also for the same cup seen from different angles or different distances or in different parts of the image. Under traditional static conceptions of vision and data analysis such problems often seem insurmountable even given implausible supervision (c.f. [35]).

Enactive sensorimotor theories such as that proposed by Alva Noë [13, 36] and introduced in section 2 explicitly reject such a static view of perception and propose that rather than looking for statistical regularities in snapshot images, that instead we search for patterns of contingency between actions and sensory streams. That is to say we no longer identify environmental features such as objects and affordances by the sensory regularities they provide, as we know these are not fixed anyway, instead we look for regularities in the ways that sensory streams change over time relative to our actions. Learnt or otherwise acquired profiles of such changes is what Noë refers to as sensorimotor knowledge, and the application of this, in prediction, Noë argues is the basis of perception. Similarly Gallese and Lakoff [7] argue that the association of sensory and motor areas allows for the simulation of actions leading to the prediction of sensory consequences. Such stimulation of sensory and motor areas with predicted rather than actual sensory data also conforms with the simulation hypothesis [19] and various other theories of cognition.

While sensorimotor theories would seem to provide a way out of the problem of marginal regularity, as we have argued elsewhere [37], the profiling of such changes still requires consistent tracking over time before the profiles can be learned from which consistent tracking is supposed to follow. Thus the proposed solution can easily become circular. To clarify this point with an example, if I know this sub section of sensory data right now is a cup (or whatever) then I need to be able to track the cup over some changes in order to make a profile of those changes in order to be able to track the cup. In part this circularity comes from not completely letting go of the static snapshot perspective, the way out is to recognise from experience that ‘this’ set of changes over time happens sometimes in coincidence with performing action X, AND that ‘this other’ set of changes consistently coincides with performing action Y, when the original set coincide with action X. Thus there are consistent clusters of relationships and actions that happen together, when they do occur we can infer the presence of some external object or more abstractly an external situation or event.

The enactive solution we propose then is to find patterns in the dynamics of sensory and motor streams over time; however, time introduces another important problem to the modelling of sensorimotor perception. *The temporal problem* as a variation of the *credit assignment problem* results from the fact that the result of an action may not be immediate, and may in fact result from a sequence of actions. The problem then is how to discover which subset of the actions performed is actually responsible for this sensory change. As a minimal requirement then *the temporal problem* necessitates the inclusion of some form of memory such that past events in the sequence are available to take part in the formation of profiles of change and the marginal regularity problem requires some consistent tracking of features as our sensory contact with them changes over time. Thus we argue for the inclusion of memory and the transformation into warped high dimensional spaces to maximise the availability of features from which to construct temporal profiles leading to perception of features and affordances in the environment. In Section 5 we will develop a model that does exactly this, but first, in the next section we introduce sustained inattentional blindness as this will be a bi-product of our predictive models.

4 Sustained Inattentional Blindness

Our perception of the world around us is subject to manipulation, even to the extent that we can be experientially blind to highly salient and temporally extended events. We can even be unaware of things we are looking directly at. Simons and Chabris [38] note that “we perceive and remember only those objects and details that receive focused attention.” p. 1059. Though it is not entirely clear in this context what attention is, similar claims have been made by many researchers e.g. [13, 36, 39]. This effect is demonstrated most startlingly in an experiment on sustained inattentional blindness by Simons and Chabris [40] in which human subjects watch a video showing two intermingled groups of people, one dressed in white and the other in black, each passing a basketball between members of their own group. Subjects are asked to count how many times the ball is passed by one particular group (either those dressed in white or those dressed in black depending on which condition the subject is in). Somewhat surprisingly, many “observers fail to notice an ongoing and highly salient

but unexpected event...[a] Gorilla walked from right to left into the live basketball passing event, stopped in the middle of the players as the action continued all around it, turned to face the camera, thumped its chest, and then resumed walking across the screen” p. 1069. Observers in this study “were consistently surprised when they viewed the display a second time, some even exclaiming, ‘I missed *that!*?’ ” p. 1072. In this, and other psychological experiments, the effect of similarity between the attended (the team the subject is watching), distracter (the other team) and unexpected objects (the gorilla) has been systematically varied showing that close similarity between the attended and unexpected objects reduces the occurrence of inattentional blindness [40-43]. For example subjects attending to the team dressed in white were more likely to miss the black-haired gorilla than subjects attending to the team dressed in black. Most et al [44] vary the luminance of the attended and unexpected objects showing that increasing similarity (in terms of luminance) decreases the likelihood of failing to detect the unexpected object. In a different task Koivisto and Revonsuo [43] ask subjects to count how many times balls of one colour bounce off the edge of a computer screen, while balls of a different colour also bounce around the screen (see figure 1 below). In this task the unexpected object appears on the left of the screen and travels across it until it exits on the right. Subjects engaged in the counting task often miss the unexpected object and thereby display sustained inattentional blindness. In a number of experiments Koivisto and Revonsuo [43] systematically vary the number of distracter objects and their similarity to the attended objects showing that (a) distracter objects have little or no effect and that (b) sustained inattentional blindness can occur even in the absence of any distracters. For simplicity sake it is this scenario, with no distracters that we will focus on here.

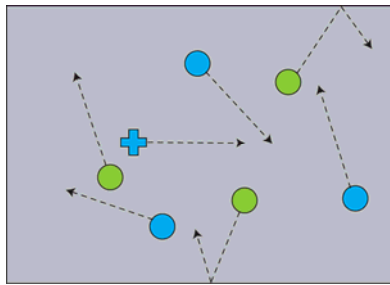


Fig. 1. Illustration of Koivisto & Revonsuo’s task. Human subjects count how many times the green (lighter) balls bounce, while ignoring the blue (darker) balls. The unexpected object, here a blue cross moves across the screen, often undetected.

5 Modelling a Cortical Hierarchy

Following discussion of the connectivity between cortical microcolumns in different regions of cortex in section 1, and noting that others model cortical microcolumns in far more detail than they are treated herein, we are not the first to suggest that dynamic reservoirs such as those found in Liquid State Machines or Echo State Networks, capture many of their properties. In fact the Liquid State Machine originates

from attempts to model the neuroscientific data produced by Markram et al. [45] and Gupta et al. [46] on the cortical micro-columns of rat somatosensory cortex. While the number of cells in a microcolumn and density of each kind vary between regions, certain features of microcolumns remain constant. Of particular interest here is that micro-columns are sparsely internally interconnected, that they are observed not to be chaotic, and not to implement stable attractor dynamics. Shepherd [47] and Douglas & Martin [48] provide neuroscientific examples highlighting the stereotypical circuitry found in these structures and the wide range of tasks they seem to be involved in. While there is a great deal more structural detail to the cortical micro-column, some of which is currently being modelled elsewhere, Maass et al. [49-51] provide an abstract model using random connectivity between ‘leaky integrate and fire’ neurons with both static and dynamic synapse models, the parameters of which are based on Gaussian distributions of the data from Markram et al. and Gupta et al.’s experiments. The resulting model, the Liquid State Machine (LSM) has been shown to possess a number of important computational properties, acting as an analogue memory, and as a recursive kernel function [49-51]. Simpler models preserving these computational aspects of the cortical micro-column have been developed such as the Echo State Network (ESN) [52-55] which we implement here. While the ESN model is somewhat removed from the anatomical detail of the cortical micro-column, we view it as a useful abstraction preserving particular properties of the underlying neuroscience while providing significant advantages in terms of lower computational cost.

To implement an ESN we generate a random valued and sparse (30%) randomly connected weight matrix for 100 neurons. The randomly achieved weight matrix \mathbf{W} of the resulting network is restricted to have a spectral radius of less than one, i.e. $|\lambda_{\max}| < 1$, where λ_{\max} is the eigenvalue of \mathbf{W} which has the largest absolute value, which guarantees a null state attractor. Similarly $|\lambda_{\min}| > -1$ where λ_{\min} is the smallest absolute eigenvalue of \mathbf{W} . As the separation property is also preserved the ESN is here viewed as a computational simplification of the biologically derived LSM architecture. Unlike Jaeger [52-55], who uses a CTRNN, we update the neurons according to simple discrete time dynamics using the following standard equations.

$$\mathbf{a}_i = \sum \mathbf{y}_j \mathbf{w}_{ij}$$

$$\mathbf{y}_i = \frac{\mathbf{I}}{\mathbf{I} + e^{-\mathbf{a}_i}}$$

Next we turn to the uncontroversial and far from new idea that the cortex is loosely structured in a hierarchical manner, at least with respect to major pathways of inter-micro-column connectivity. For example micro-columns in area V1 of the visual cortex are heavily connected to area V2 (about half the size of V1), and then to V4 (again about half the size) and from there to the Inferotemporal-cortex (a larger area receiving input from several different modalities). The computational properties of the ESN rely in part on the input being significantly smaller than the size of the ESN, thereby forcing an expansion of dimensions in the data stream and fostering the linear separation of relational features in that input stream. Stacking such systems into a hierarchy presents a problem in that the output from a reservoir is the size of that

reservoir, and feeding the output from several into one would require increasingly large reservoirs at each level of the hierarchy in order to preserve a dimension expansion at each step. The solution we adopt here is to autonomously compress the state of each ESN using a Self-Organizing Map (SOM) [56]. Each dimension of the map can be used as an output so that we effectively output the address of the winning SOM unit.

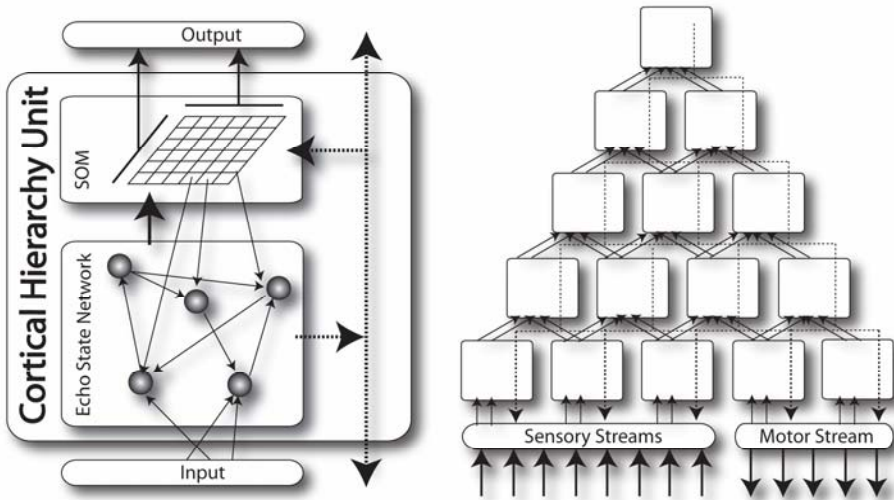


Fig. 2. (Left) showing the internal structure of each unit in the cortical hierarchy. Input flows into an ESN which is then classified by a SOM. The SOM projects back into the ESN and the address of the winning SOM unit is passed on as output. Additionally associative connections are learned between the ESN and the SOMs of other units. (Right) the overall structure of the whole hierarchy. Neighbouring units pass their outputs to the same unit in the next layer of the hierarchy. Associative connections are between the ESN of one unit and the SOMs of the units that that unit connects to.

This approximates a principle component analysis passing variance information on to the next layer. While we fully recognize that this dimension reduction throws away a great deal of information, the reduced output then combines with the output from neighbouring units as input to an ESN at the next level of the hierarchy where the same process is repeated again (see fig 2). This allows for the discovery of new input features, specifically those relying on relational properties between the parts of the input stream kept separate at all previous levels. SOM units then also provide normal input via sparse random connectivity back into the ESN driving that SOM.

As discussed in section 1 the development of functional regions of cortex results from plasticity refining and altering the connectivity between microcolumns within connected regions of cortex. This plasticity is modelled here using perceptrons autonomously trained to predict, from the activity of one ESN, what state connected

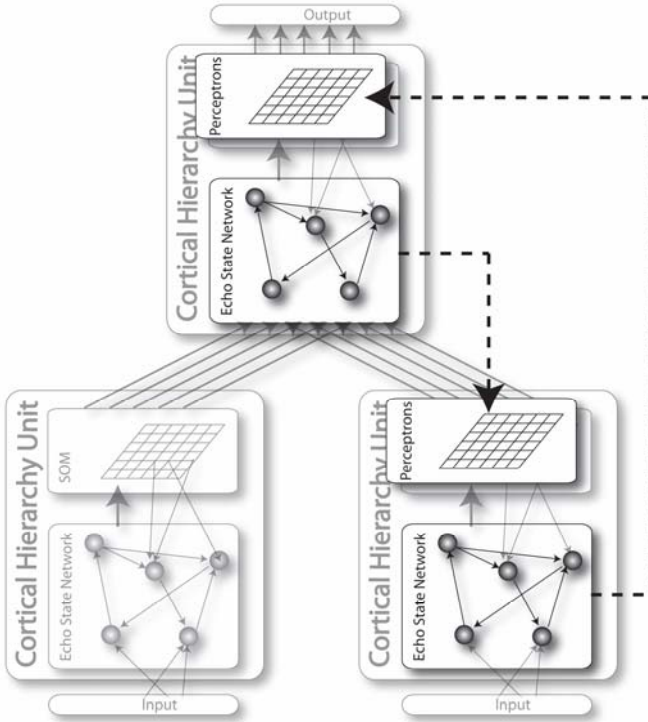


Fig. 3. Highlighting the plasticity between connected columns. Perceptrons are trained to predict the state (SOM) of one column, from the ESN activity of other connected columns.

columns are in. Thus based on the information available in one column, predictions are made as to what the likely state of other columns, having other information available to them, is likely to be (see Figure 3). The output and weight changes to each perceptron were calculated using the following standard formulas.

$$y_i = \sum y_j w_{ij}$$

$$\Delta w_{ij} = \alpha (y_i - target_i) y_j$$

5.1 Experimental Setup

Following a model based account of sustained inattentional blindness presented in [1, 2], we here provide an extension of the same experiment using the hierarchy model just described. We constructed a small hierarchy with three cortical units in the first level, two in the second level, and one in the third level. In the experiments detailed herein we simplified Koivisto and Revonsuo's task in the following ways. Firstly we removed all distracter objects as the number or presence of distracter objects was found not to significantly alter the extent of sustained inattentional blindness in experiments carried out in [43]. Secondly we reduced the number of attended objects to 1 so as to simplify the modelling task. The visual area was then divided up

into a 4 x 4 grid and the average green or blue pixel values of each cell provided two inputs respectively from each cell. This provided a total of 32 inputs to an ESN at every time step. The task here is to constantly track the vertical aspect of the direction of the attended simulated object, i.e. is the blue ball moving upward or downward (ignoring left and right velocity).

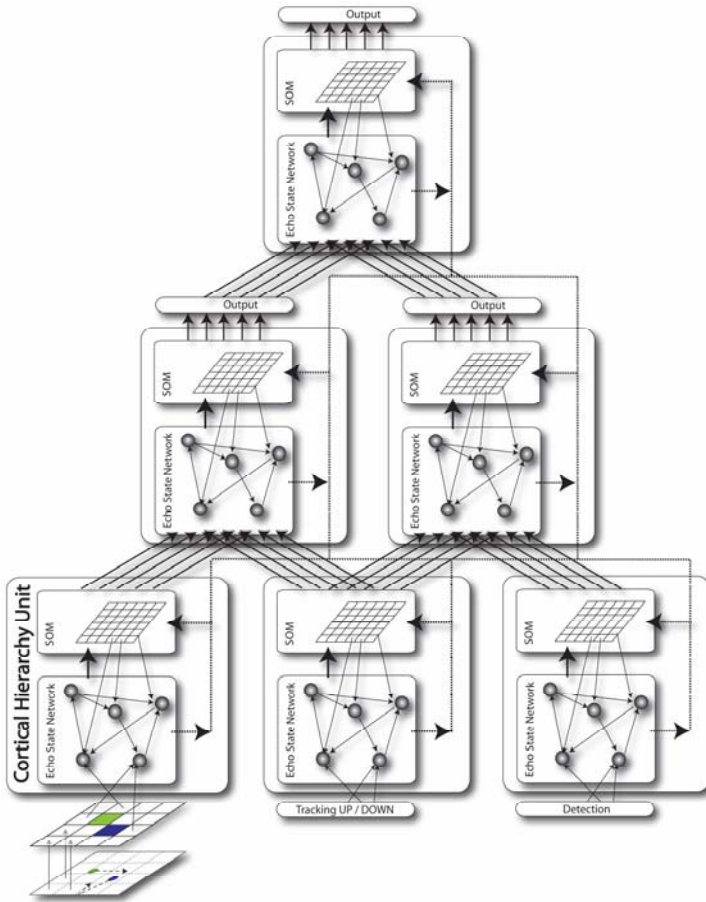


Fig. 4. The 6 unit hierarchy. Visual input is averaged over a 4 by 4 grid and then passed as input to the first cortical hierarchy unit. Tracking information and detection of the unexpected object are fed as input to the 2nd and 3rd units. Activity flows up the hierarchy driven by this input. Activity also flows down the hierarchy via the learnt associative connections, see text for a full explanation.

Our two conditions are then whether we allow top down activity to reach the unit receiving the image as input. In the first stage of the experiment we provide input for a total of 10000 time steps during which the model is ‘conditioned’, in that the model indirectly learns the association between the different inputs. During this stage we

also record the SOM activity of the tracking and detection units and note a strong correlation between the input state and the SOM output. In stage 2 we enter a testing phase in which the tracking and detection input is removed but the visual input continues. We record the SOM activity of the tracking and detection units, recording a correct output at every time step that the observed SOM activity corresponds with the noted correlation of what the actual tracking and detecting input should be. Thus if the association has been learnt then the hierarchy should reproduce the same activity in the tracking and detecting unit SOMs even in the absence of that input. During stage 2 the perceptron learning was also disabled. We repeated the whole experiment with feedback to the first unit enabled and disabled.

5.2 Results

As can be seen from figure 5 below, performance at tracking was considerably improved by feedback from the hierarchy. This shows that top down predictive information aids discrimination in correlated tasks. As can be seen from figure 6, performance at detection is hindered by the same feedback that improves tracking performance. We conducted a repeated measures ANOVA on this data and found significant main and interaction effects. Where the effect of feedback on tracking had a probability of $p < 0.001$, the effect of feedback on detection had a probability of $p < 0.05$ all interaction effects had a probability of $p < 0.001$. As can be seen from fig 5 and 6, feedback improved performance at tracking while degrading performance at detection, thus we have the sustained inattentive blindness phenomena. During engagement in a difficult task, predictive feedback is necessary to produce adequate performance at that task but has the effect of making detection of an otherwise detectable object less likely.

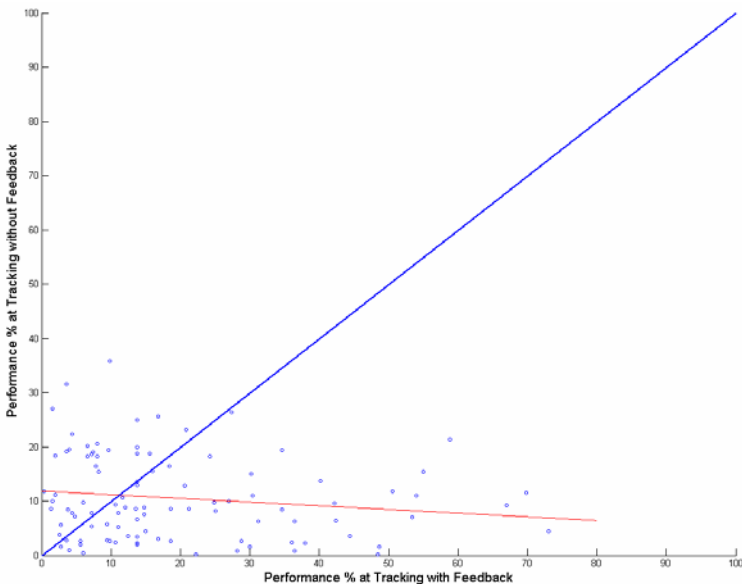


Fig. 5. Scatterplot showing the hierarchies performance at tracking with and without feedback. Note that the scale here is adjusted so that 0% indicates chance levels of performance.

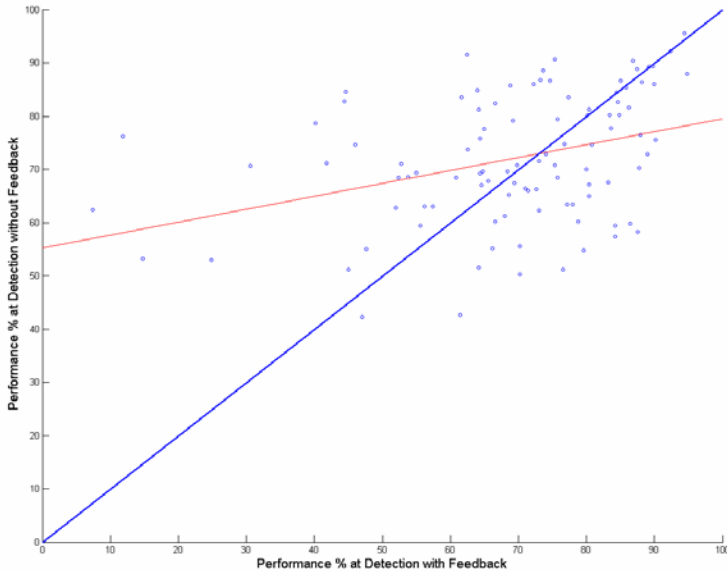


Fig. 6. Scatterplot showing the hierarchies performance at detecting the unexpected object with and without feedback. Note that the scale here is adjusted so that 0% indicates chance levels of performance.

6 Discussion, Whats Going on?

The model presented here draws on biological and philosophical theories of the relation between sensorimotor knowledge discussed in sections 1 and 2 to provide a scalable model of sensorimotor learning, one by-product of which is the inattentional blindness phenomenon. In this model, the role of each micro-column, is to identify relational patterns in and over time, from the activity of the input streams or columns in the layer below. Having identified these patterns, a spreading activation between layers provides anticipatory input from above and classification from below. The hierarchical model presented learns correlations between its inputs, without introducing supervision bias, allowing for the top down spreading of activation to aid, or hinder, the identification of relational features patterns and sequences. This is closely related to the simpler model of inattentional blindness presented in [1, 2] however we here provide a minimal hierarchical implementation to clarify and provide a more plausible account of where the feedback comes from. This model is very closely linked with the enactive account of sensorimotor perception, in as much as experienced correlations between input streams are learned and then provide the basis for a spreading activation providing prediction or anticipation of what the unobserved or missing input could be. This is however, not done directly to the input but rather manifests in the form of tuning input filters (ESN's) to improve separation of the anticipated features at the cost of less separation of other features.

As a tentative explanation of why the sustained inattentional blindness phenomena is observed in this model we can analyse the effect of feedback on a single ESN.

Clearly the addition of new inputs to an ESN correlated with some feature, changes the attractor landscape of the ESN and moves the trajectory in state space (for a fuller analysis of this see [2]). However the Euclidean distance between input with that feature and input without that feature is enlarged by the presence of the correlated input. This means that the pointwise separation between ESN states following input with or without those inputs is also greater. This typically leads to enhanced separation of the streams and facilitates linear separation by a perceptron. This accounts for the improved performance at tracking (see figure 7 upper graph). As for degrading performance at detection we can see that an input image with one ball on it provides a different level of input than an image with two balls on it (one being the unexpected stimulus). This difference in input magnitude provides further separation of the following ESN states facilitating normal detection of the unexpected object. With the presence of an uncorrelated feedback input, however, this separation is distorted leading to loss in performance at detecting (see figure 7 lower graph).

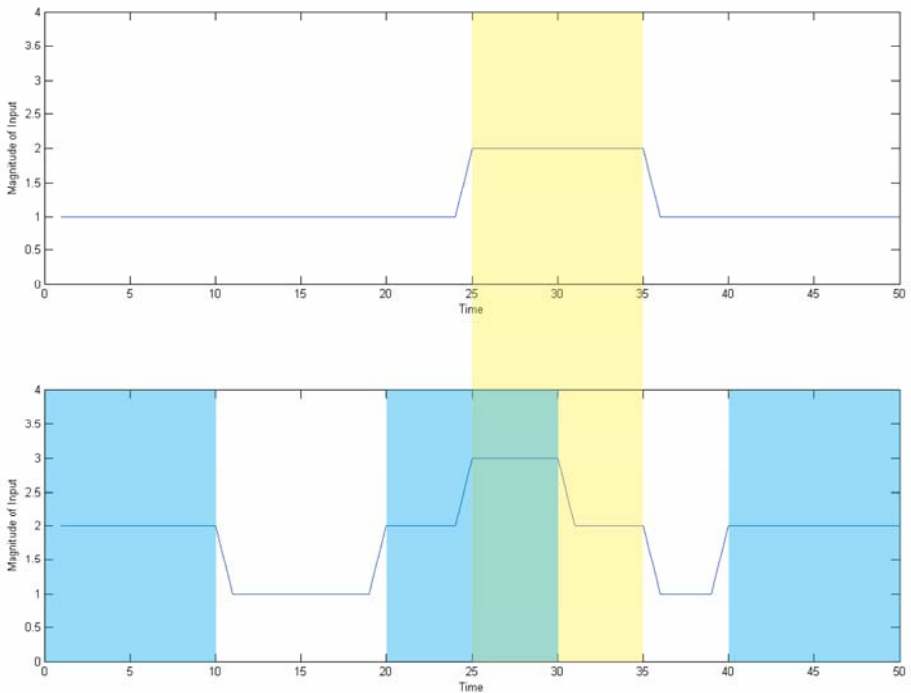


Fig. 7. The magnitude of the input to the visual ESN. In the case of tracking, the magnitude of input remains constant whether the ball is moving up or down, however, when the unexpected object appears the magnitude of input goes up thus the point-wise separation of these states is increased (**Top figure**). When feedback from tracking is made available, the magnitude (assuming correct feedback) as the ball moves up or down is different (e.g. 0-10 vs 10-20 in the **lower figure**), however the separation of the uncorrelated unexpected input is reduced (25-35 in the lower figure).

Acknowledgements

This work was supported by a European Commission grant to the project “*Integrating Cognition, Emotion and Autonomy*” (IST-027819, www.iceaproject.eu), as part of the European *Cognitive Systems* initiative.

References

1. Morse, A.F.: Neural Models of Prediction and Sustained Inattentional Blindness. In: NCPW11 the Neural Computation and Psychology Workshop. World Scientific, Oxford (in press)
2. Morse, A., Lowe, R., Ziemke, T.: Manipulating Space: Modelling the Role of Transient Dynamics in Inattentional Blindness. *Connection Science* (accepted)
3. Downing, K.L.: Neuroscientific implications for situated and embodied artificial intelligence. *Connection Science* 19, 75–104 (2007)
4. Downing, K.L.: Predictive models in the brain. *Connection Science* (in press)
5. Fleischer, J.G.: Neural correlates of anticipation in cerebellum, basal ganglia, and hippocampus. In: Butz, M.V., Sigaud, O., Pezzulo, G., Baldassarre, G. (eds.) *ABiALS 2006*. LNCS, vol. 4520, pp. 19–34. Springer, Heidelberg (2007)
6. Hawkins, J., Blakeslee, S.: *On Intelligence*. Times Books (2004)
7. Gallese, V., Lakoff, G.: The brain’s concepts: The role of the sensory-motor system in reason and language. *Cognitive Neuropsychology* 22, 455–479 (2005)
8. Rumelhart, D.E., McClelland, J.L.: *The PDP Research Group*, vol. 2 (1986)
9. Hebb, D.O.: *The Organization of Behavior: A Neuropsychological Theory*. John Wiley & Sons, Chichester (1949)
10. Karmiloff-Smith, A.: Why Babies’ Brains Are Not Swiss Army Knives. In: Rose, H., Rose, S. (eds.) *Alas, poor Darwin*, pp. 144–156. Jonathan Cape, London (2000)
11. Sharma, J., Angelucci, A., Sur, M.: Induction of visual orientation modules in auditory cortex. *Nature* 404, 841–847 (2000)
12. Mountcastle, V.B.: An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System. In: Edelman, Mountcastle (eds.) *The Mindful Brain*. MIT Press, Cambridge (1978)
13. Noë, A.: *Action in Perception*. MIT Press, Cambridge (2004)
14. Morse, A., Ziemke, T.: Cognitive Robotics, Enactive Perception, and Learning in the Real World. In: *CogSci 2007 - The 29th Annual Conference of the Cognitive Science Society*, pp. 485–490. Erlbaum, New York (2007)
15. Maturana, H., Varela, F.: *The tree of knowledge*, revised edition. Shambhala, Boston (1992)
16. Gehringer, W.L., Engel, E.: Effect of ecological viewing conditions on the Ames’ distorted room illusion. *Journal of Experimental Psychology: Human Perception and Performance* 12, 181–185 (1986)
17. Clark, A., Grush, R.: Towards a Cognitive Robotics. *Adaptive Behavior* 7, 5 (1999)
18. Grush, R.: The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences* 27, 377–396 (2004)
19. Hesslow, G.: Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Sciences* 6, 242–247 (2002)
20. Kunde, W., Elsner, K., Kiesel, A.: No anticipation–no action: the role of anticipation in action and perception. *Cognitive Processing* 8, 71–78 (2007)

21. Gibson, J.J.: *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston (1979)
22. Barsalou, L., Breazeal, C., Smith, L.: Cognition as coordinated non-cognition. *Cognitive Processing* 8, 79–91 (2007)
23. Fodor, J.: *The language of thought*. Harvard University Press, Cambridge (1975)
24. Fodor, J., Pylyshyn, Z.: Connectionism and cognitive architecture: A critical analysis. *Connections and Symbols*, 3–71 (1988)
25. Bruce, V., Burton, A.M., Craw, I.: Modelling face recognition. *Philosophical Transactions of the Royal Society, B* 335, 121–128 (1992)
26. Burton, A.M., Bruce, V.: Naming Faces and Naming Names: Exploring an Interactive Activation Model of Person Recognition. *Memory for Proper Names* (1993)
27. Bruce, V., Burton, M., Carson, D., Hanna, E., Mason, O.: Repetition Priming of Face Recognition. *Attention and Performance XV: Conscious and Nonconscious Information Processing* (1994)
28. Morse, A.F.: Psychological ALife: Bridging The Gap Between Mind And Brain; Enactive Distributed Associationism & Transient Localism. In: Cangelosi, A., Bugmann, G., Borisyuk, R. (eds.) *Modeling Language, Cognition, and Action: Proceedings of the ninth conference on neural computation and psychology*, pp. 403–407. World Scientific, Singapore (2005)
29. Morse, A.F.: *Cortical Cognition: Associative Learning in the Real World*. D Phil. Thesis, Department of Informatics, University of Sussex, UK (2006)
30. Burton, A.M., Bruce, V., Hancock, P.J.B.: From pixels to people: A model of familiar face recognition. *Cognitive Science* 23, 1–31 (1999)
31. Morse, A.F.: Autonomous Generation of Burton's IAC Cognitive Models. In: Schmalhofer, Young, Katz (eds.) *EuroCogSci 2003, The European Cognitive Science Conference*. LEA Press (2003)
32. Page, M.: Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences* 23, 443–467 (2001)
33. Harvey, I.: Untimed and Misrepresented: Connectionism and the Computer Metaphor. *AISB Quarterly* 96, 20–27 (1996)
34. Kirsh, D.: From connectionist theory to practice. In: *Connectionism: Theory and practice*. Oxford Univ. Press, New York (1992)
35. Clark, A., Thornton, C.: Trading spaces: Computation, representation, and the limits of uninformed learning. *Behavioral and Brain Sciences* 20, 57–66 (1997)
36. O'Regan, K., Noë, A.: A sensorimotor account of visual perception and consciousness. *Behavioral and Brain Sciences* 24, 939–1011 (2001)
37. Morse, A.F., Ziemke, T.: Cognitive Robotics, Enactive Perception, and Learning in the Real World. In: *CogSci 2007 - The 29th Annual Conference of the Cognitive Science Society*. Erlbaum, New York (2007)
38. Simons, D.J., Chabris, C.F.: Gorillas in our midst: Sustained inattentive blindness for dynamic events. *Perception* 28, 1059–1074 (1999)
39. Rensink, R.A., O'Regan, J.K., Clark, J.: To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science* 8(5), 368–373 (1997)
40. Simons, D., Chabris, C.: Gorillas in our midst: Sustained inattentive blindness for dynamic events. *Perception* 28, 1059–1074 (1999)
41. Most, S., Simons, D., Scholl, B., Jimenez, R., Clifford, E., Chabris, C.: How Not to Be Seen: The Contribution of Similarity and Selective Ignoring to Sustained Inattentive Blindness. *Psychological Science* 12, 9–17 (2001)

42. Most, S., Scholl, B., Clifford, E., Simons, D.: What you see is what you set: Sustained inattentional blindness and the capture of awareness. *Psychological Review* 112, 217–242 (2005)
43. Koivisto, M., Revonsuo, A.: The role of unattended distractors in sustained inattentional blindness. *Psychological Research* 72, 39–48 (2008)
44. Most, S.B., Simons, D.J., Scholl, B.J., Jimenez, R., Clifford, E., Chabris, C.F.: How Not to Be Seen: The Contribution of Similarity and Selective Ignoring to Sustained Inattentional Blindness. *Psychological Science* 12, 9–17 (2001)
45. Markram, H., Wang, Y., Tsodyks, M.: Differential signaling via the same axon of neocortical pyramidal neurons. *National Acad. Sciences* 95, 5323–5328 (1998)
46. Gupta, A., Silberber, G., Toledo-Rodriguez, M., Wu, C.Z., Wang, Y., Markram, H.: Organizing principles of neocortical microcircuits. *Cellular and Molecular Life Sciences* (2002)
47. Shepherd, G.M.: A basic circuit for cortical organization. *Perspectives in Memory Research*, 93–134 (1988)
48. Douglas, R.J., Martin, K.A.C.: Neuronal Circuits of the Neocortex. *Annu. Rev. Neurosci.* 27, 419–451 (2004)
49. Maass, W., Natschlag, T., Markram, H.: Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations. *Neural Computation* 14, 2531–2560 (2002)
50. Maass, W., Natschlag, T., Markram, H.: A Model for Real-Time Computation in Generic Neural Microcircuits. *Advances in Neural Information Processing Systems* 15 (2003)
51. Maass, W., Natschlag, T., Markram, H.: Computational models for generic cortical microcircuits. *Computational Neuroscience: A Comprehensive Approach* (2003)
52. Jaeger, H.: The echo state approach to analysing and training recurrent neural networks. *German National Institute for Computer Science* (2001)
53. Jaeger, H.: Short term memory in echo state networks. *German National Institute for Computer Science* (2001)
54. Jaeger, H.: Tutorial on Training Recurrent Neural Networks, Covering BPTT, RTRL, EKF and the echo State Network Approach. *GMD-Forschungszentrum Informationstechnik* (2002)
55. Jaeger, H.: Adaptive Nonlinear System Identification with Echo State Networks. *Neural Information Processing Systems, NIPS* (2002)
56. Kohonen, T.: The self-organizing map. *Neurocomputing* 21, 1–6 (1998)