The Emergence of Memory: Categorisation Far from Equilibrium

Andrew Wuensche

SFI WORKING PAPER: 1995-03-035

SANTA FE INSTITUTE

The Emergence of Memory;
Categorisation Far from Equilibrium

Andrew Wuensche

Santa Fe Institute
and University of Sussex
School of Cognitive and Computing Science
wuensch@santafe.edu
andywu@cogs.susx.ac.uk

contact address:
48 Esmond Road, London W4 1JQ
tel 0181 995 8893 fax 0181 742 2178
100020.2727@compuserve.com

Abstract

The brain is a vastly complex dynamical system. Cognitive functions such as memory, learning, volition, and consciousness seem to emerge as high level properties from the activity of billions of neurons sparsely connected into complex networks. What is it about networks that allows such emergence to happen?

I will argue that the basic level of emergence is a network's ability to categorise its space of possible patterns of distributed activation. State space is not just categorised by attractors. Categorisation also occurs far from equilibrium, within the long transients leading to attractor cycles, giving a role to chaotic dynamics. A network self-organises its state-space into a *basin of attraction field*, a space-time abstraction representing the network's latent distributed memory. When networks link up with other networks, the system is able to use memory in individual fields to provide the components for the emergence of higher level cognitive states. This paper examines the idea of basins of attraction in the context of a simple yet powerful neural model - random Boolean networks.

To appear in "Towards a Scientific Basis for Consciousness", eds, S.R.Hameroff, A.W.Kaszniak, A.C.Scott, MIT Press.

Introduction

How does memory emerge in a simple artificial neural network? I will try to provide an answer in the context of a network model first proposed by Ross Ashby (1952) in his classic book *Design for a Brain*. This network is a discrete dynamical system of sparsely connected cells updating in parallel according to each cell's Boolean function. Usually known as a *random Boolean network* following Stuart Kauffman (1969, 1993), the system can also be described as a *disordered* cellular automaton, with arbitrary non-local connections and different rules at each site.

Any artificial network model is necessarily a huge oversimplification compared to biological networks. However, Boolean functions, in contrast to threshold functions, arguably capture some essence of the logic implicit in the complex topology of each neuron's dendritic tree and synaptic microcircuitry. A Boolean function could be implemented in hardware by a combinatorial circuit, a sort of artificial dendritic tree. If the emergence of memory can be demonstrated in such simple networks, perhaps insights may be gained by analogy to the vastly more powerful processes that occur in animal brains. Memory is a relative concept, meaning the creation of useful categories for an organism's adaptive behaviour out of the plethora of sensory inputs arriving at neural sub-networks, and the even greater internal traffic between sub-networks. The idea of networks of sub-networks in a sort a sort of nested hierarchy is of course another necessary oversimplification, because the boundaries of sub-networks are unclear.

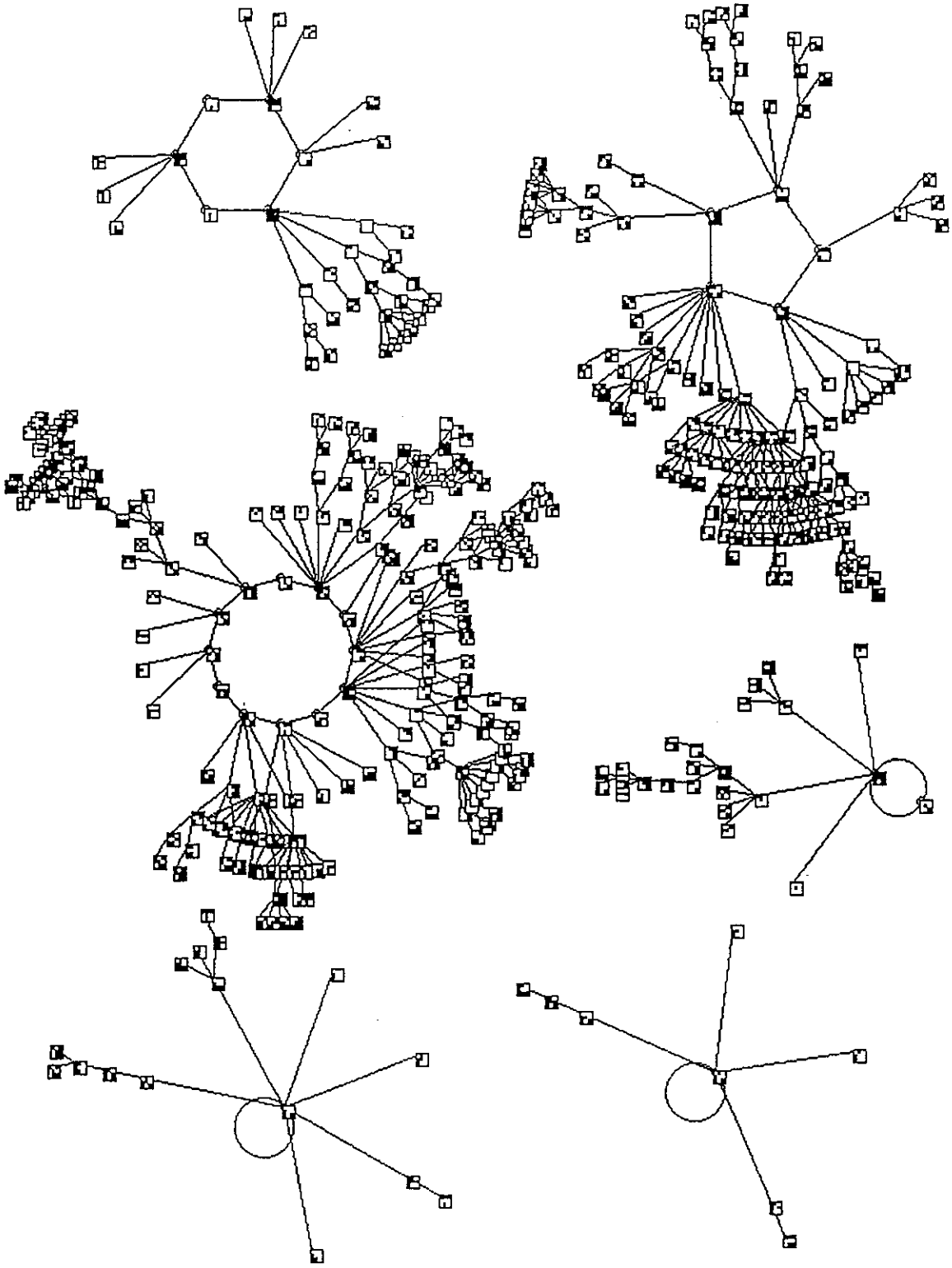
To be useful for adaptive behaviour, memory-categories should fall naturally into hierarchies of categories of sub-categories; they should be highly reliable yet easily changed to permit learning; and access should be extremely fast.

Recent work in unravelling the global dynamics of cellular automata (Wuensche and Lesser 1992), and more generally of random Boolean networks (Wuensche 1994a) puts into sharp focus a known intrinsic property of these networks. They organise state-space, the space of all possible patterns of network activation, into specific maps of connected states, the *basin of attraction field*. Directed graphs representing these objects can be calculated, and drawn by computer graphics.

Separate assemblies of connected states are *basins of attraction*. Typically, the connection topology is branching trees, representing non-equilibrium states, rooted on attractor cycles representing the equilibrium states available to the network. State space is not just categorised by attractors. Subtrees within individual basins, formed by merging trajectories leading to attractor cycles, produce hierarchies of sub-categories. The basin topology is characterised by long transients, verging on dynamics analogous to chaos in continuous systems. The vast majority of state-space is typically *far from equilibrium*, being distant in time from equilibrium represented by attractors.

Figure 1.

The basin of attraction field for a random Boolean network (N=9, K=4). The 2^9=512 states in state space, shown as patterns on a 3×3 grid, are organised into 6 basins, with attractor periods ranging from 1 to 12. The number of states in each basin is: 39, 191, 234, 24, 16, 8. The network's wiring/rule scheme is shown in figure 3. The length of transition arcs has no significance; it follows a graphic convention for clarity. Time proceeds inward from garden-of-Eden states, then clockwise around the attractor cycle, the only closed loop in the basin. The computer diagrams were generated with the author's software.



The idea that state-space is partitioned by attractors is the generally accepted paradigm for "content-addressable" memory in artificial neural networks, following Hopfield (1982) and others. In large recurrent networks, especially biological networks, a very large (probably astronomical) number of steps through state-space would typically be required for the system to settle at the equilibrium of one of its attractors. Explaining memory just by attractors thus poses the difficulty of the long time needed to reach attractors in large networks, whereas reaction times in biology are extremely fast. This problem is overcome by the realisation that categories occur in subtrees, far from equilibrium.

Random Boolean networks have a vast parameter space, and a correspondingly vast space of possible basin of attraction fields. Perhaps any basin field structure is possible given the appropriate parameters. Learning algorithms for random Boolean networks have been developed for changing the basin of attraction field (Wuensche 1994a). States at any location can be re-assigned as predecessors of other states in a single step by adjusting either connections or Boolean functions. Adding states implies learning, removing states implies forgetting. This might allow the sculpting of the basin of attraction field to approach any desired configuration. The process of learning and its side effects is made visible if the resulting basin of attraction field or fragment is reconstructed.

The detailed structure of a biological neural sub-network has found its form, and thus its basin of attraction field, by evolution, development and learning in the context of networks of sub-network which maintain each other far from equilibrium. Within this vastly complex dynamical system higher level cognitive properties based on lower level memory categories are able to emerge.

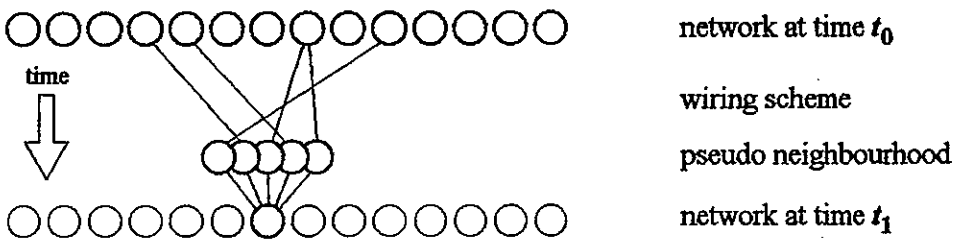


Figure 2.
Random Boolean network architecture. Each cell in the network synchronously updates its value according to the values in a pseudo neighbourhood, set by single wire couplings to arbitrarily located cells at the previous time-step. Each cell may have a different wiring/rule scheme. The system is iterated. Cells are arranged in a row for convenience only; their positions may be arbitrary.

Random Boolean Networks

Random Boolean networks may be viewed as *disordered* cellular automata. A homogeneous rule and coupling template, giving a regular space, make cellular automata appropriate models in physics. Deviating from either or both of these constraints by degrees progressively degrades coherent space-time patterns and emergent complex structures such as *gliders*, characteristic of cellular automata (Wuensche 1994a, 1994b). Different arbitrary couplings and rules at each cell give a vastly greater parameter space, and thus behaviour space, than cellular automata. The various symmetries and hierarchies that constrain cellular automata dynamics, such as shift invariance and

the conservation of rotational symmetry (Wuensche and Lesser 1992) no longer apply. It might be conjectured that any arbitrary basin of attraction field configuration is possible given the right set of parameters.

Figure 2 illustrates the architecture of a random Boolean network. A cell's value can be just 0 or 1, though in principle this could be extended to more values. A global state of a network of N cells is its pattern of 0s and 1s at a given moment. Each cell synchronously updates its value in discrete time steps. The value of a cell at time t_1 depends on its particular Boolean function applied to a notional or *pseudo* neighbourhood, size K. Values in the neighbourhood are set according to single wire couplings to arbitrarily located cells in the network at time t_0. The system is iterated. The system's parameters consist of a list specifying the function and pseudo neighbourhood wiring for each cell. To maintain a given basin of attraction field the parameters remain fixed. Changes to parameters would change the field.

There are 2^K permutations of values in a neighbourhood of size K. The Boolean function (equivalent to a cellular automata rule) can be written as a rule table, or look up table, with 2^K entries, specifying the output of all neighbourhood permutations. By convention (Wolfram, 1983) this is arranged in descending order of the values of neighbourhoods. For example, the rule table for rule 30 (K=3) is,

	111	110	101	100	011	010	001	000	... neighbourhoods
rule table ...	0	0	0	1	1	1	1	0	... outputs (0 or 1)

The total number of distinct rule tables, the size of rule space = 2^{2^K}. The number of alternative wiring schemes for one cell = N^K. The number of alternative wiring/rule schemes, S, that can be assigned to a given network turns out to be vast even for small networks, and is given by,

$$S = \left(N^K\right)^N \times \left(2^{2^K}\right)^N$$

for example, for a network where N=16 and K=5, $S = 2^{832}$

Random Boolean network architecture is in many ways similar to weightless neural networks (Alexander, Thomas, and Bowden 1984). Classical neural network architecture uses weighted connection and threshold functions. A random Boolean network may be regarded as a discrete generalisation of a sparsely connected classical neural network. Connections with higher weights may simply be replaced by multiple couplings, and the threshold function applied. However, a threshold function is a tiny sub-class of the 2^{2^K} possible Boolean functions.

Basins of attraction

Cellular automata and random Boolean networks are both examples of discrete deterministic dynamical systems made up from many simple components acting in parallel. The pattern of network activation at a given time is the network's state. State-space is the space of all 2^N possible patterns. Each state has just one successor, though it may have any number of predecessors (known as pre-images), including none. Any state imposed on the network will seed a determined sequence of states, known as a trajectory. Though determined it is usually unpredictable.

In a finite network any trajectory inevitably encounters a repeat. When this occurs the system has entered and is locked into a state cycle known as the *attractor*, a condition of dynamical equilibrium. The portion of a trajectory outside an attractor is a transient. Many transients typically

lead to the same attractor. Because states can have multiple pre-images transients can merge, and will typically have a topology of branching trees rooted on the attractor cycle (though this may be a stable point - an attractor cycle with a period of 1). The set of all transient trees plus their attractor make up a basin of attraction. The separate basins that make up state space (though there may be just one) is the *basin of attraction field* , a mathematical object in space-time constituting the dynamical flow imposed on state space by the network.

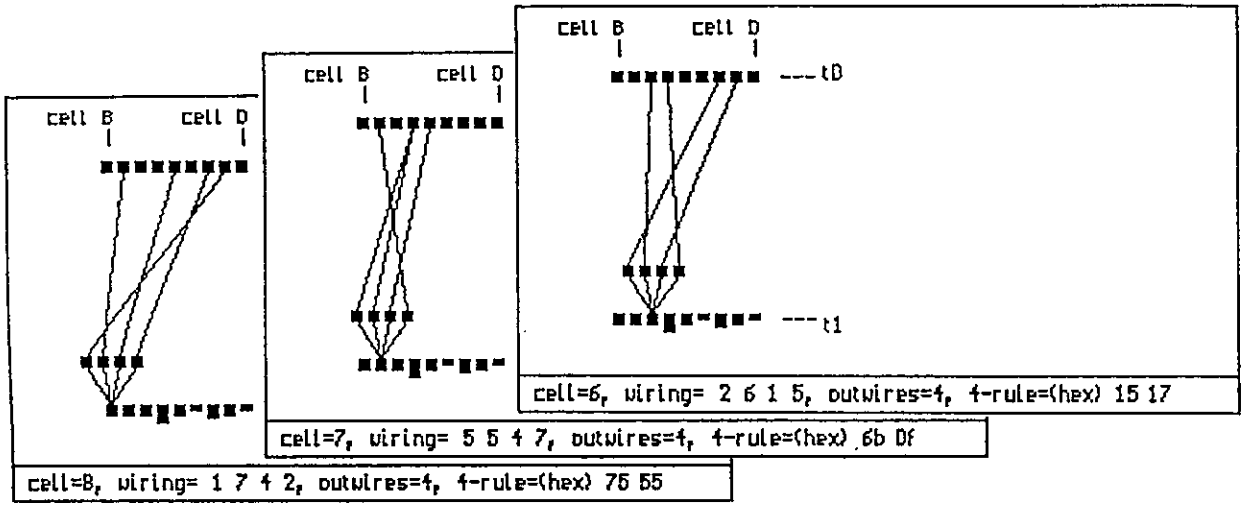
Computing transient trees or sub-trees, and basins of attraction, poses the problem of finding the complete set of pre-images of any global state. The trivial solution, exhaustive testing of the entire state space, rapidly becomes intractable in terms of computer time as the network's size increases beyond modest values. Algorithms have recently been devised, however, for directly generating pre-images, giving an average computational performance many orders of magnitude faster than exhaustive testing. A *reverse algorithm* for one-dimensional cellular automata was introduced in (Wuensche and Lesser 1992). and a *general* direct reverse algorithm for random Boolean networks in (Wuensche 1994a).

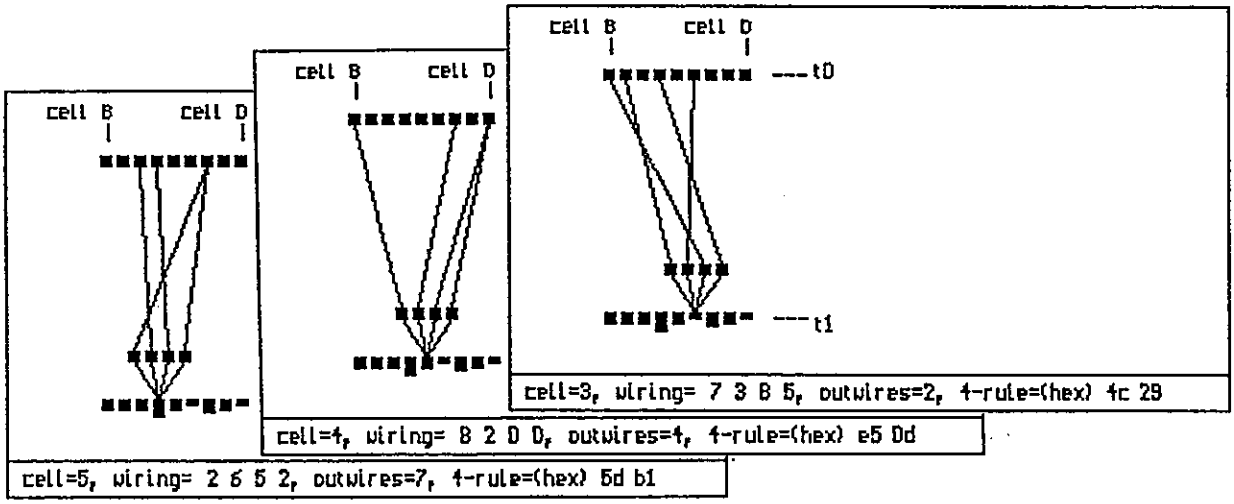
Basins of attraction are portrayed as computer diagrams in the same graphic format as presented in (Wuensche and Lesser 1992). States are represented by nodes, or by the state's binary, decimal or hex expression at the node position. Nodes are linked by directed arcs with zero or more incoming arcs but exactly one outgoing arc (*out-degree*=1). Nodes with no pre-images, thus no incoming arcs, represent so called *garden-of-Eden* states. Typically, the vast majority of nodes in a basin of attraction lie on transients outside the attractor cycle equilibrium condition, and the vast majority of these are garden-of-Eden states. Transient states that are not garden-of-Eden categorise the subtree of which they form the root. Figure 1 shows a typical basin of attraction field of a random Boolean network (N=9, K=4). The 2^9=512 states in state space are shown as patterns on a 3×3 grid. The network's wiring/rule scheme is shown in figure 3.

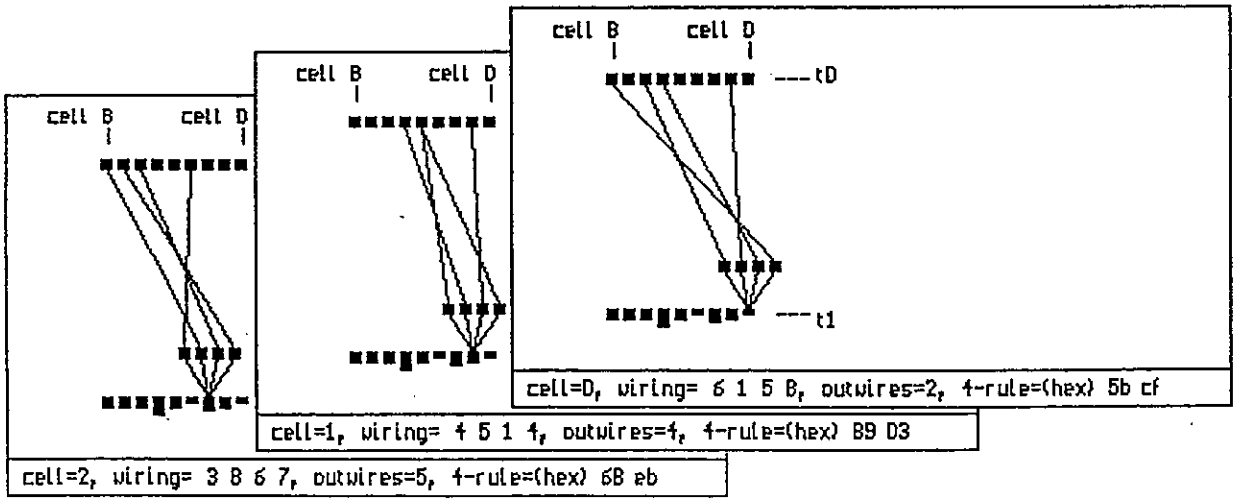
Memory, far from equilibrium

Memory far from equilibrium along merging transients may answer a basic difficulty in explaining memory by attractors in biological neural networks. A view of the brain as a complex dynamical system made up of many inter-linked specialised neural sub-networks is perhaps the most powerful paradigm currently available. Sub-networks may consist of further sub-categories of semi-autonomous networks, and so on, which contribute to re-setting or perturbing each other's dynamics. A biological neural sub-network is nevertheless likely to be extremely large; the time required to reach an attractor from some arbitrary global state will probably be astronomical. This has been demonstrated with a simple K=5 random Boolean network with 150 cells (Wuensche 1994a). Even when an attractor is reached, it may well turn out to be a long cycle or a quasi-infinite chaotic attractor. The notion of memory simply as attractors seems to be inadequate to account for the extremely fast reaction times in biology.

Figure 3.
The wiring/rule scheme of a random Boolean network (N=9, K=4); its basin of attraction field is shown in figure 1. The wiring specifies couplings from the pseudo neighbourhood of each cell (at time t_1) to cells at the previous time-step, t_0. Different heights of cells at t_0 indicate their number of output wires. Each cell's rule is shown in hex. The parameters were chosen at random.


cell B
cell D
tD
t1
cell=6, wiring= 2 6 1 5, outwires=4, 4-rule=(hex) 15 17
cell=7, wiring= 5 5 4 7, outwires=4, 4-rule=(hex) 6b 0f
cell=8, wiring= 1 7 4 2, outwires=4, 4-rule=(hex) 75 55


cell B
cell D
tD
t1
cell=3, wiring= 7 3 8 5, outwires=2, 4-rule=(hex) 4c 29
cell=4, wiring= 8 2 0 0, outwires=4, 4-rule=(hex) e5 0d
cell=5, wiring= 2 6 5 2, outwires=7, 4-rule=(hex) 5d b1


cell B
cell D
tD
t1
cell=0, wiring= 6 1 5 8, outwires=2, 4-rule=(hex) 5b cf
cell=1, wiring= 4 5 1 4, outwires=4, 4-rule=(hex) 89 03
cell=2, wiring= 3 8 6 7, outwires=5, 4-rule=(hex) 68 eb

A discrete dynamical system with synchronous updating categorises its state space reliably along transient trees, far from equilibrium, as well as at the attractors. A network that has evolved or learnt a particular global dynamics may be able to reach useful memory categories in a few steps, possibly just one. Moreover, the complex transient tree topology in the basin of attraction field, makes for a much richer substrate for memory than attractors alone, allowing hierarchies of memory sub-categories.

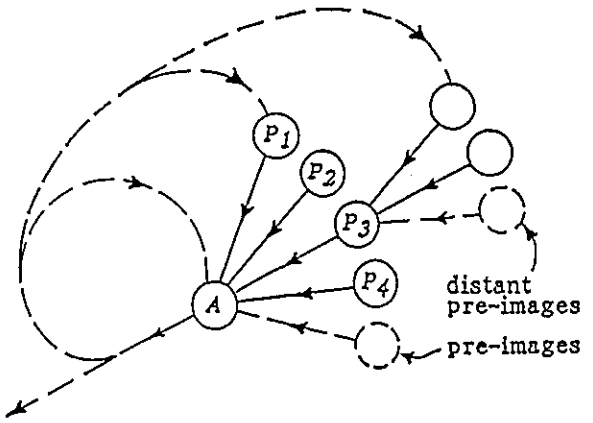
A biological model

A random Boolean network may serve as a model of a semi-autonomous patch of neurons in the brain whose activity is synchronised. A cell's wiring scheme models that sub-set of neurons connected to a given neuron. Applying the Boolean function to a cell's pseudo neighbourhood models the non-linear computation that a neuron is said to apply to these inputs to determine whether or not it will fire at the next time-step. This is far more complex than a threshold function (Shepherd 1990). The biological computation may depend on the precise topology of the dendritic tree, its microcircuitry of synaptic placements and intrinsic membrane properties. Networks *within* cells based on the cytoskeleton of microtubules and associated protein polymers may be involved, suggested by Stuart Hameroff (1987) as the neuron's "internal nervous system". There appears to be no shortage of biological mechanisms that could perform the role of a Boolean function.

A network's basin of attraction field is implicit in its wiring rule/scheme. In a sub-network of biological neurons it would be implicit in the wetware. If sub-networks are linked as components in a super-network, indirect feedback could maintain each sub-network's dynamics in the outer branches of transient trees, far from equilibrium. Suppose a new state is imposed on sub-network n_1 by another sub-network or sensory input. This seeds a determined trajectory in n_1, which opens up a sequence of categories to which the seed state belongs. *Recognition* of these categories would involve another sub-network, n_2, activated by the axons of neurons in n_1. Recognition of categories in n_2 would involve a third sub-network n_3. After just a few steps, n_1's dynamics might be reset to a new seed, shifting the start point in its basin of attraction, by the axons of another sub-network belonging to this hypothetical super-network. Recognition in this system is automatic because all trajectories in n_1, n_2, n_3 etc are inevitable given unchanged neuronal architecture. *Recall* is more difficult because a seed with the right association needs to be supplied. Folk psychology bears this out. I may know "Humphrey Bogart", but "its on the tip of my tongue". Recalling the name to say it may be hard. If I hear "Humphrey Bogart" mentioned, I will recognise it instantly, effortlessly and automatically.

Learning new behaviour implies amending the basin of attraction field by adjusting the wiring/rule scheme, analogous to some physical change to the wetware's neural architecture and synaptic function. Learnt behaviour often requires no mental effort; take driving a car. Such well rehearsed semi-conscious behaviour is delegated to a sort of "automatic pilot". The network's dynamics acts in the same way as for recognition, effortlessly and automatically. Learning to drive, on the other hand, requires mental concentration. After all, physical changes in the brain must somehow be induced. By this argument, a conscious state coincides with the desire to learn or behave creatively, or the actual process of making the changes. The mechanism for this is unknown. A sub-network may be able to alter the parameters of another. Hameroff suggests that cytoskeletal functions may provide retrograde signalling (analogous to back propagation) which may reconfigure intra-neuronal architecture (Hameroff *et all*, in press).

Figure 4. States P_1,P_2,P_3,... etc may be learnt as pre-images of the state A. Distant pre-images of A may also be learnt, for instance as pre-images of P_3. Learning A as a pre-image of itself creates a point attractor. Learning A as a distant pre-image of itself creates a cyclic attractor. If A is learnt as the pre-image of some other state in the basin of attraction field, the states flowing into A, its transient sub-tree, may be fully or partially transplanted along with A.



Learning Algorithms

Learning and its side effects in random Boolean networks show up as changes to the detailed structure of the basin of attraction field, and its computer graphic representation. In networks too large to allow basins, or even fragments of basins, to be computed, the principles would still apply.

Learning algorithms that enable a random Boolean network to learn new transitions from experience (and also to forget) are set out in detail in (Wuensche 1994a). Before learning starts, a wiring/rule scheme must already be in place. If the relevant transitions in the basin of attraction field are already close to the desired behaviour, the side effects of learning will be minimised. The initial wiring/rule scheme would ideally be pre-evolved from a population of wiring/rule schemes with a genetic algorithm.

Suppose we want to make the state P_1 in figure 4 the pre-image of state A. This entails correcting the mismatches between P_1's actual successor state and A. This can be done in one step by either of two methods, adjusting the network's wiring or rule scheme. The two methods have very different consequences. Correcting a mismatch by adjusting the rule scheme is achieved by changing a specific bit in each Boolean function at mismatched cells. The procedure is bound to succeed, and it turns out that there is no limit to the number of pre-images of a given state that can be learnt by this method with no risk of forgetting previously learnt pre-images of the state. There will be side effects elsewhere in the basin of attraction field. On the other hand, in correcting a mismatch by rewiring, success is not certain, though there are many alternative wire move options. Previously learnt pre-images of the state may be forgotten, besides other side effects. Side effects may be positive because similar pre-images are likely be learnt by default, so the network is able to generalise.

Re-wiring has a much greater effect on basin structure than mutating the rule scheme, but in either case the stability of basin structure is noteworthy. Using these methods, point attractors, cyclic attractors and transient sub-trees can be can be created. Transient sub-trees are sometimes transplanted along with the repositioned state, indicating how learnt behaviour can be re-applied in a new context. Forgetting involves *pruning* pre-images and transient sub-trees, and is achieved by the inverse of the method for learning. Since it is sufficient to create just one mismatch in order to forget, the side effects are minimal as compared with learning.

Conclusions

The emergence of novel structures by the interaction of many low level components underlies complex adaptive systems, in particular the emergence of life by the self-organisation of matter. Can this approach extend to the emergence of cognition and consciousness, and if so what are the low level components from which cognition emerges? I have suggested that these components are the hierarchical categories implicit in the dynamics of neural sub-networks, and that memory emerges when sub-networks combine and maintain each other far from equilibrium. Memory may in turn provide the substrate for unconstrained emergence producing higher level cognitive states.

The basin of attraction diagrams of random Boolean networks capture a simple network's capacity for distributed memory. The diagrams demonstrate that a complex hierarchy of categorisation exists within transient trees, far from equilibrium, providing a vastly richer substrate for memory than attractors alone. In the context of many semi-autonomous weakly coupled networks, the basin field/network relationship may provide a fruitful metaphor for the mind/brain.

Acknowledgements

I am grateful to colleagues at the Santa Fe Institute, the University of Sussex and elsewhere for discussions and comments.

References

Ashby,W.R. 1952. *"Design for a Brain: The Origin of Adaptive Behaviour"*, Chapman & Hall, London.

Alexander,I., W.Thomas and P.Bowden. 1984. *"WISARD, a radical new step forward in image recognition"*, Sensor Review 120-4.

Hameroff,S.R. 1987. *"Ultimate Computing: Biomolecular Consciousness and NanoTechnology"*, North Holland, Amsterdam.

Hameroff,S.R., J.E.Dayhoff, R.Lahoz-Beltra, S.Rasmussen, E.M.Insinna and D. Koruga. In press. *"Nanoneurology and the Cytoskeleton: Quantum Signaling and Protein Conformational Dynamics as Cognitive Substrate"*, in "Behavioral Neurodynamics", K.Pribram and H Szu, eds., Pergamon Press.

Hopfield,J.J. 1982. *"Neural networks and physical systems with emergent collective computational abilities"*, Proceedings of the National Academy of Sciences 79 2554-2558.

Kauffman,S.A. 1984. *"Emergent properties in random complex systems"*, Phisica D, vol 10D , 146-156.

Kauffman,S.A. 1993. *"The Origins of Order: Self Organization and Selection in Evolution"*, Oxford University Press, New York.

Shepherd,G.M., ed. 1990. *"The Synaptic Organization of the Brain"*, Oxford University Press, New York.

Wolfram,S. 1983. *"Statistical Mechanics of cellular automata"*, Review of Modern Physics, vol 55, no 3 601-64.

Wuensche,A., and M.J.Lesser. 1992. *"The Global Dynamics of Cellular Automata: An Atlas of Basin of Attraction Fields of One-Dimensional Cellular Automata"*, Addison-Wesley, Reading, Mass.

Wuensche,A. 1994a. *"The Ghost in the Machine: Basin of Attraction Fields of Disordered Cellular Automata Networks"*, in "Artificial Life III", C.G.Langton, ed., Addison-Wesley, Reading, Mass., 465-501.

Wuensche,A. 1994b. *"Complexity in One-D Cellular Automata: Gliders, Basins of Attraction and the Z Parameter"*, Working Paper 94-04-026, Santa Fe Institute, Santa Fe, N.M.