Attractor Nets, Series I:

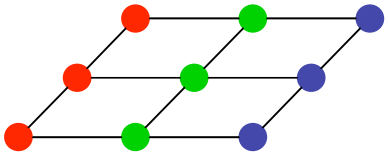
Notes Toward a New Theory of Mind, Logic and Dynamics in Relational Networks

William L. Benzon

These notes explore the use of Sydney Lamb's relational network notion for linguistics to represent the logical structure of complex collection of attractor landscapes (as in Walter Freeman's account of neuro-dynamics). Given a sufficiently large system, such as a vertebrate nervous system, one might want to think of the attractor net as itself being a dynamical system, one at a higher order than that of the dynamical systems realized at the neuronal level. A mind is a fluid attractor net of fractional dimensionality over a neural net whose behavior displays complex dynamics in a state space of unbounded dimensionality. The attractor-net moves from one discrete state (frame) to another while the underlying neural net moves continuously through its state space.

163 11th Street, Basement
Hoboken, NJ 07030
201.217.1010
bbenzon@mindspring.com

Attractor Nets, Series I



Systematic exploration of how things look if we use Sydney Lamb's notation to represent high-level activity of large many-layered neural nets partitioned into 100s and 1000s of quasi-independent regions.

I Introduction

This document consists of working notes on a new approach to thinking about minds, both animal and human. As such these notes are primarily for the purpose of reminding me of what I have been thinking on these matters. They are not written, alas, in a way designed to convey these ideas to others.

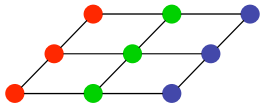
Thus, they are variously rambling, inconsistent, vague, incomplete, and, no doubt, wrong-headed in places.

I would recommend that interested readers go through the notes in this order:

1. "Simple Animal" (the third section of notes, **III**)
2. "Lamb Notation" (the second section of notes, **II**)
3. "Minds in Nets" (**VII**)

While that is not the order in which I wrote the notes, the "Simple Animal" notes do a bit better job on the basic issues than the "Lamb Notation" section, which I had written first. Perhaps one should then read the rather long final section "Minds in Nets" (VII), which suggests some of the larger implications of this conception. If energy holds, I suggest the section on "Consciousness and Control" (VI). The section "Brief Notes" (V) is just that, and is dispensable.

The section on "Assignment" (IV) discusses one particular construction in some (not entirely satisfactory) detail.



II Lamb Notation for Attractor Logic

I'm exploring the notion that we can use Sydney Lamb's relational network notion for linguistics to represent the logical structure of complex attractor landscapes — an attractor net. If we're going to think about human cognition in terms of the complex dynamics of the brain, considered as a neural net having a phase space of very high dimensionality, we're going to need a way of thinking about processes in a topology of thousands of attractors related to one another in complex ways. I think Lamb's notation may be a way of representing the topology of such an attractor landscape.

As is typically the case in such matters, we've got multiple levels of modeling/representation. In this case, two. The bottom level is the standard world of complex dynamics with its systems of equations which are used to analyze and build computer simulations of neural nets and to analyze neural and behavioral data. I have nothing to say about the details of this level. I'm interested in the upper level, where we're trying to represent the relationship between the large number of attractors in a rich neural network. We can't think about this structure by examining the underlying system of equations, nor even by examining such computer simulations as we are currently capable of running.

I am imagining, in the large, that we have a neural net where we can make meaningful distinctions between microscopic, mesoscopic, and macroscopic processes. Roughly speaking, I see Lamb's notation as a way to begin thinking about the relationship between mesoscopic and macroscopic processes.

These notes are of a preliminary and very informal nature. I'm just trying things out to see what's involved in doing this.

Some Basics

The following diagram depicts three neuro-functional areas (NFAs) that are reciprocally connected with one another as indicated. These NFAs may well be the mesoscopic patches of neuropil that Freeman has studied. These interconnections are massively parallel, as in the nervous system.

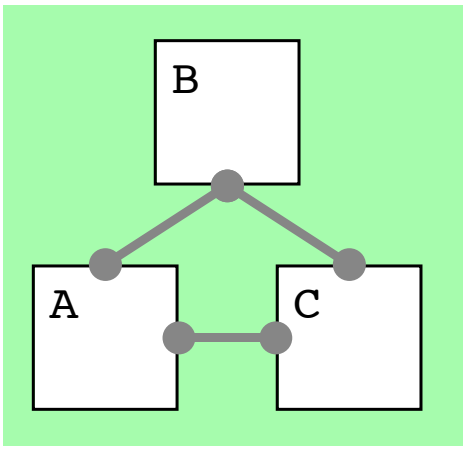


Figure 1: Three NFAs

Each NFA has many attractors arrayed in what is called an landscape. The state of each NFA is, however, dependent on the states of the NFAs to which it is connected. Let us imagine that when the system is "framed",[1] each NFA is in some basin of attraction. What happens, then, when some NFA receives input that knocks it from its current attractor basin? Depending on the nature of the connections and so forth, some or all of the others may be disturbed as well. The system will meander until each NFA has settled into a net set of mutually compatible basins; that is, until the system is once again framed. The idea is to use a relational network as a way of representing the attractor basins and their mutual relationships, their compatibilities.

The following diagram represents an OR relationship between attractors in A with respect to an attractor in B:

---

1 "Equilibrium" is the term I had originally used, but it probably is not a good term to use. We are dealing with systems that typically operate far from (physical) equilibrium. It is not clear just what term to use, but the notion is that the NFAs are each in some particular basin of attraction. Perhaps we could say that the system has become *framed* or *in-frame*, where the term suggests a single frame from a motion picture at the moment the film is momentarily stopped at a frame actually projected onto the screen. Freeman thinks of consciousness in this way, where a "frame" of consciousness is a hemisphere-wide coherent state lasting between 100 msec 200 msec. Consciousness is then a succession of such frames (see page 47 ff.).

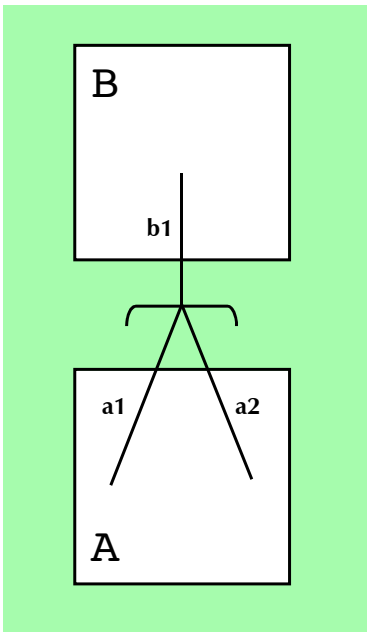


Figure 2: Logical OR

Let us interpret that to mean that attractor b1 is compatible with either a1 or a2. It is inherent in the nature of the system, of course, that the attractors of an NFA are incompatible with one another. [Note that this diagram assumes reciprocal connectivity between NFAs A and B, as indicated in Figure 1.]

Note that in this notation the nodes are relationships while the arcs between them are substantive entities [see note * starting on page 11]. This is quite different from most relational nets that have been used in the cognitive sciences; in these nets the nodes have indicated substantive entities while the arcs were relationships between those entities. In this interpretation of Lamb's notation, the arcs represent attractors in some NFA. The relationships are logical operators and so embody interactions between NFAs. Given the nature of the underlying dynamical system, we might want to think of these operators as embodying some form of fuzzy logic. Given a sufficiently large system, such as a vertebrate nervous system, one might want to think of the attractor net as itself being a dynamical system, one at a higher order than that of the neurons themselves.

Similarly, the following diagram means the b1 is compatible with either c1 or c3:

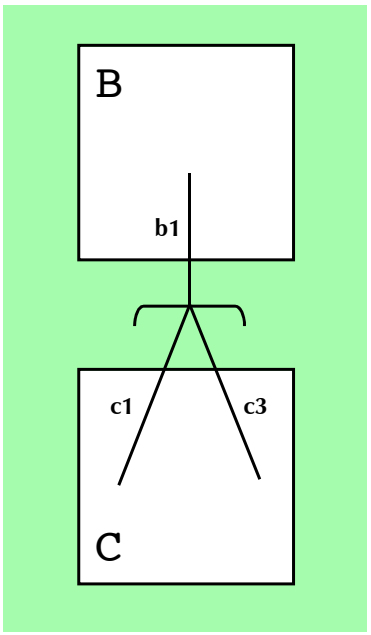


Figure 3: Another example of logical OR

The following diagram illustrates the AND relation:

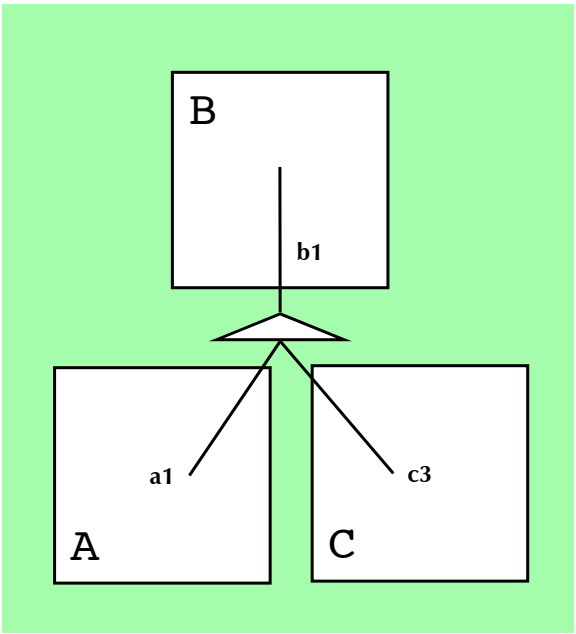


Figure 4: Logical AND

Let us interpret this to mean that b1 requires both a1 and c3.

As it stands, these notions are not adequate, for we are interested in process, how these NFAs move from one state to another. What makes this tricky is that all these NFAs have

activity at all times, to some degree or another. Informally it seems to me that, at any given moment, the activity of some NFAs is more or less in the background while that in others is more or less in the foreground. Further, at any given interval, the states of some foreground or background NFAs are more or less fixed while the states of others are allowed to vary (until they reach frame with one another and with those whose state has been fixed).

Continuing informally, let us assume that, at this level of analysis, we can think of the system as moving in discrete jumps, which we can call frames (the metaphor derives from motion pictures). This is compatible with some of Freeman's recent remarks on consciousness. Thus the state at frame one is followed by some other state at frame two:

F1 → F2

Considering Figure 2 (an OR). Let us assume that the state of B is fixed by virtue of interactions not depicted in the above diagram. We have:

a1+b1 → a2+b1
OR
a2+b1 → a1+b1

Obviously this can be simplified, perhaps to something like:

b1 | a1 → a2
OR
b1 | a2 → a1

Where b1 is understood to provide the context for the transition following the vertical line.

In the case of Figure 4 (AND) we have these possibilities:

b1 | a1 → c3

b1 | c3 → a1

b1 | ai,cj → a1,c3
(or perhaps: b1 | b1 → a1,c3)

a1,c3 | a1,c3 → b1
(or perhaps: a1,c3 |bj → b1)

One of the things we need to think about is the creation of a

new attractor in some NFA. This would come about when it has now attractor that is readily compatible with the set of attractors currently fixed in the NFAs with which it is interacting strongly at the moment. I have nothing to say about this, but it is obviously a very important issue.

Paradigmatic Structure

I'd now like to think about a more or less "real" example. I'm interested in paradigmatic structure, the type of structure built with what is often called the IS-A or ISA relation, e.g. a dog is a beast.

Let us begin with the following diagram:

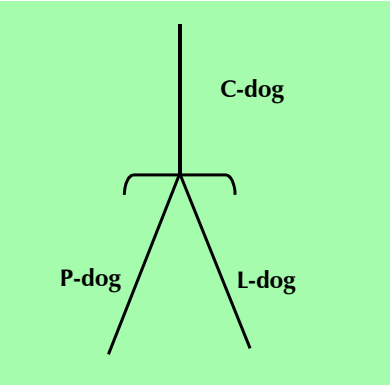


Figure 5: The concept of a dog

What it says, more or less, is that the conceptual dog (C-dog) can be aroused by either the perceptual dog (P-dog) or the lexical dog (L-dog). C-dog is an attractor in an NFA that is part of the systemic system (in the language of Hays 1981). P-dog is an attractor in some NFA that is linked to a perceptual system — visual (one sees a dog), auditory (one hears a dog), olfactory (one smells a dog), and so forth. L-dog is an attractor in some NFA that is part of the language system; it is, more or less, a word. C-dog is thus neither a word nor a percept. It is a *concept* and, in this depiction, a rather minimal one at that. (At the moment I don't want to worry about all the conceptual "stuff" that can be associated with C-dog.) This is a bit different from anything I've considered before but, for various hard-to-conceptualize reasons, I rather like it.

Figure 6 shows just a bit of paradigmatic structure:

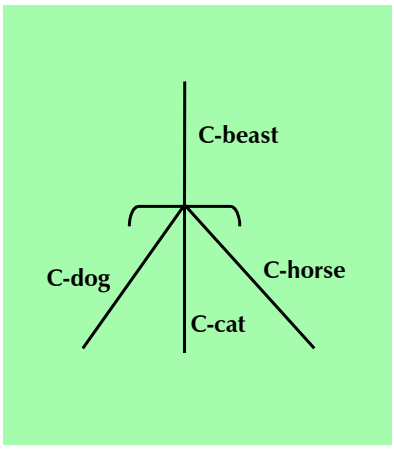


Figure 6: Beast

This means that a C-beast can be aroused by a C-dog, a C-cat, or a C-horse (but not a C-seahorse). More needs to be said about this — in particular, are all these concepts attractors in the same NFA? — but I'll leave it alone for the moment. I offer, for further reflection, Figure 7:

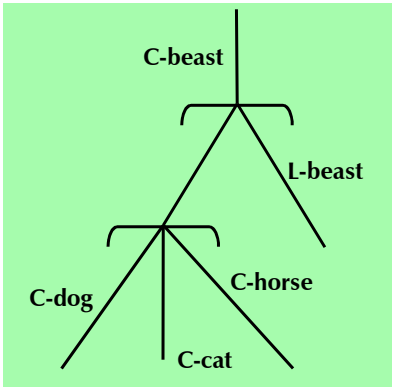


Figure 7: Beast, again

In the next diagram, Figure 8, I depict the conceptual beagle. I have reasons for this peculiar construction, but I do not wish to consider them in detail here and now.

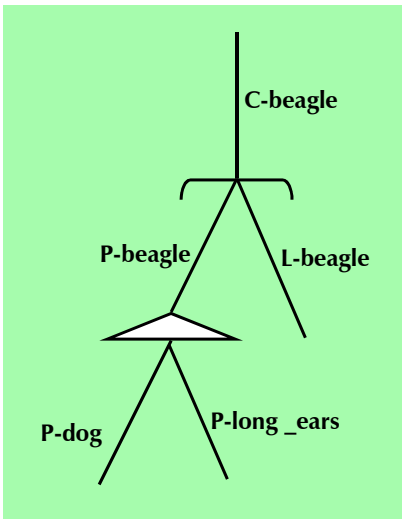


Figure 8: Beagle

What I am asserting in the lower left is that the perception of a dog that is specifically a beagle entails the perception of a generic dog (P-dog) plus the obligatory (AND) perception of some feature or set of features that is specifically diagnostic of beagles, in this case represented by P-long_ears. P-dog can be recognized in a single glance; but the recognition of specific kinds of dogs is more complex. This requires some inspection of the P-dog to acquire more information, thought not necessarily very much more. What ever the case, as with C-dog, the C-beagle can be aroused by either P-beagle or L-beagle.

(I can further imagine that one might recognize particular individual dogs, regardless of breed, at a single glance. These are dogs that one knows quite well and are part of one's social world. But that's a different matter.)

What emerges from this bit of thinking is that the relationship between words and their meanings is not so simple as most of us have imagined. For the most part, we have imagined a more or less self-contained world of concepts, which is linked to a more or less self-contained world of percepts on the one hand, and to a more or less self-contained world of words on the other. The representation of paradigmatic relationships between beagle, dog, and beast have been imagined to exist entirely within the worlds of concepts and percepts: a beagle IS-A dog IS-A beast. Each of these concepts is, in turn, linked to the domain of words by a simple and transparent relationship (with provisions of polysemy). I am less and less inclined to such a view. I think that the *representation* of paradigmatic relationships is all

but impossible without language.

Comments

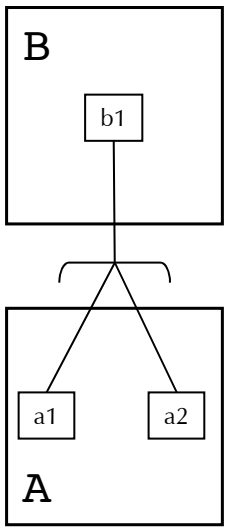
Lamb and his followers have developed a fairly rich description of phonology, morphology, and syntax, expressed in this notation. Their account of semantics is rather less richly developed. This, of course, is true of linguistics in general. Phonology, morphology, and syntax all seem to involve a relatively circumscribed set of entities, which can, however, enter into an unbounded set of combinations. Semantics, by contrast, seems to involve a rather unbounded set of entities (which can, in turn, under into an unbounded set of combinations). In this situation phonology, morphology and syntax have seemed more amenable to analysis, both individually and collectively.

And yet, language and meaning reside in pretty much the same kind of neural tissue. Neither has unbounded neural resources dedicated to it, though the neural resources are considerable. I thus believe that semantics is, in fact, almost as limited as the rest of language. But it needs to be understood properly. I believe that the framework Hays established in *Cognitive Structures* (1981) lays the basis for that. That framework, however, needs considerable reworking along lines suggested in part by Benzon and Hays (1988). The thrust of these current notes is that, by using Lamb's notation as a way of representing the structure of a large attractor net (where the nets operates more or less as in Freeman 1999, 2000), we can make headway.

As I've noted, Lamb has considerably more than a notation. He's got a rich theory of language, and one that he has already begun to relate to brain structure. If I am correct, then we may be in a position to explicate that relationship in a more detailed way. For the scheme I've sketched above places fairly strong constraints on how one relates a relational net of conceptual and linguistic entities to the 2D topology of the neorcortex.

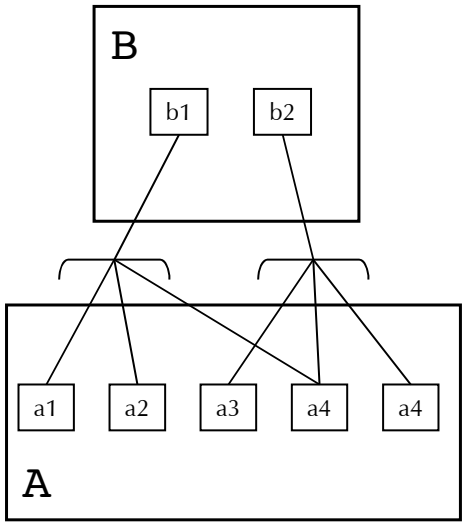
Note

*But of course, one can represent attractors as nodes by doing this:



In this notation, we have two kinds of nodes, attractors and logical operators. The meaning or significance of an attractor is a function of the NFA to which it belongs; given my affection for Pribram's neural holography, I tend to think of attractors as "neural holograms." The arcs are all of the same type.

This notation suggests, in turn, the possibility of doing this:



It's all a matter of convenience. As Lamb is at considerable pains to argue, the only real "content" of a relational net is at the periphery, where it connects to the world. Internally, the net is nothing but relationships. In the case of a vertebrate nervous system, the net as a whole has four interfaces, input and output to the external world, and input

and output to the internal milieu.

References

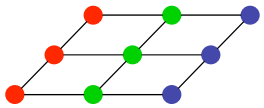
Benzon, W. L. and D. G. Hays (1988). "Principles and Development of Natural Intelligence." *Journal of Social and Biological Structures* 11: 293-322.

Freeman, W. J. (1999). *How Brains Make Up Their Minds*. London, Weidenfeld and Nicholson.

Freeman, W. J. (2000). *Neurodynamics: An Exploration in Mesoscopic Brain Dynamics*. London, Springer-Verlag.

Hays, D. G. (1981). *Cognitive Structures*. New Haven, HRAF Press.

Lamb, S. M. (1999). *Pathways of the Brain*. Amsterdam, John Benjamins B. V.



III Attractor Nets: Toward a Simple Animal

By *attractor net* (A-net) I mean the relationships among the attractors of a neural net. A net is said to be *stressed* if it is not at "equilibrium" or perhaps that should be "at frame." In general, stress is applied to an NFA from outside. When it is at frame it is in one of its attractor basins.

[**Note:** Recall the note at the bottom of page 4 about the term "equilibrium" being a misleading one.]

Homogenous Attractor Nets

An attractor net is said to be *homogenous* if all of its attractors are can be related to one another through logical 'or'. Thus:

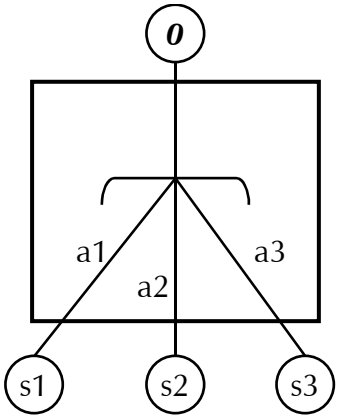


Figure 1A: Homogeneous Net

In this diagram the rectangle is some neural net while the superimposed graph is a homogeneous attractor net. The zero node indicates "equilibrium" of "at frame."[2] We can think of nodes s1, s2, and s3 as different patterns of stress on the network. Each pattern of stress is associated with a different path to frame, where the paths terminate at different points in the system's state space. Those points are attractors, a1, a2, a3, with which I have labeled the arcs. The stressors are

---

2 See footnote 1 above.

linked to the frame state through an ‘or’ connector (in Lamb’s notation).

What is stress in this context? Where the neurons are regulating the state of muscle fibers I am inclined to think of stress as the difference between the current state of the muscle fibers and the desired state. That is, it is *error*, in the sense that Powers’ uses the term in his control theory account of behavior.

If I do that on the motor side, then, I would like to do it on the sensory side as well. In effect, sensory systems are designed to respond to the difference between expected sensations and actual sensations. The general role of *preafference* in the regulation of behavior is in favor of this view. It is not clear to me how far such preafference extends toward the periphery. I believe there is some evidence in the auditory system that the preafference extends to the inner ear. In the visual system I believe there are efferent fibers in the optic nerves, though I’m not sure whether or not anyone has good ideas about what those optic efferents are doing. I don’t have any notions on the other sensory systems.

[Notice that how this notation transparently expresses Lamb’s observation that the only real “content” for such a network is at the periphery. Arc labels in the net are just a notational convenience that make it easier to read the diagram.]

Having said all that, I propose to use a slightly different notation in these notes, as follows:

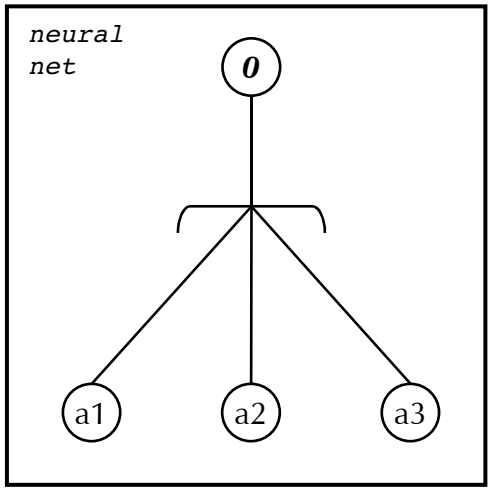


Figure 1b: Homogeneous Net

As a convenience we use nodes a1, a2, and a3 to represent

attractors. The arcs will be left unlabeled.

The danger of this notation is that it temps one to reify attractors into physical things, like neurons, or collections of neurons, or collections of synapses scattered about in some population of neurons. This temptation must be avoided.

Neural Interpretation, Mine & Lamb's

Now let us think about this at the neural level and compare this with what I take to be Lamb's neural interpretation of his notation. We've got a meshwork of tightly interconnected neurons. The net can be said to be at frame when, in Hays' formulation, inputs have been "accounted for." The net will have arrived at one of its attractor states and so aroused a gestalt that "absorbs" the externally induced stress, the input. This, of course, is some particular pattern of activity in the net. Each attractor state corresponds to a different pattern of neural activity.

An attractor node, then, represents some state, or set of states (the so-called attractor basin) of the neural net. The arcs connecting an A-node to the 0-node through the logical operator thus correspond to trajectories through the state space. Neither the attractors, the operators, nor the trajectories are physical things that one could discover though dissection and visual inspection. Rather, they the salient aspects of the topology of the net's phase space.

This is somewhat different from Lamb's interpretation. As I've indicated above, Lamb uses labels on his arcs where I use attractor nodes, He clearly thinks of his nodes, the logical connectors (which he calls *nections*, from connection) as collections of neurons, with the arcs being collections of axons. He offers thumbnail calculations of the number of neurons per nection (based on Mountcastle's work on cortical micro-anatomy), and so forth. Thus he uses his notation in a more concrete way than I am proposing.

If one thinks about my proposal, however, one might wonder what the neurons in a homogeneous A-net are connected to, other than one or another. For, as I have defined it, a homogeneous A-net corresponds to a single Lamb or-nection, nothing more. These patterns of neural activity don't seem to go anywhere, except to frame. I thus introduce the notion of a partitioned net, where each partition has an or-nection A-net. Partitions whose neurons are interconnected thus influence one

another's states; there are dependencies among their attractors.

Partitioned Nets

An attractor network is said to be *partitioned* if its attractors are in two or more sets such that an attractor from each set is required for the network as a whole to be at frame. Given the way the notion of an attractor is defined, this would seem to be an odd thing to happen; indeed, it would seem to be impossible, by definition. We must remember, however, that real neural nets almost certainly have a small world topology.

Every neuron is connected to each other neuron by at most only a small number of links. Some neurons are connected to one another directly (order 1); obviously, these neurons will have a strong influence on one another's states. Other neurons will be connected through a single intermediary, making two links between them (order 2); still others through two intermediaries (order 3); and so on. Partitioning might arise in situations where two or more sets of neurons are strongly connected within the set through connections of order N or less while almost all connections between neurons in different sets are greater than order N.

Let us begin with the following diagram:

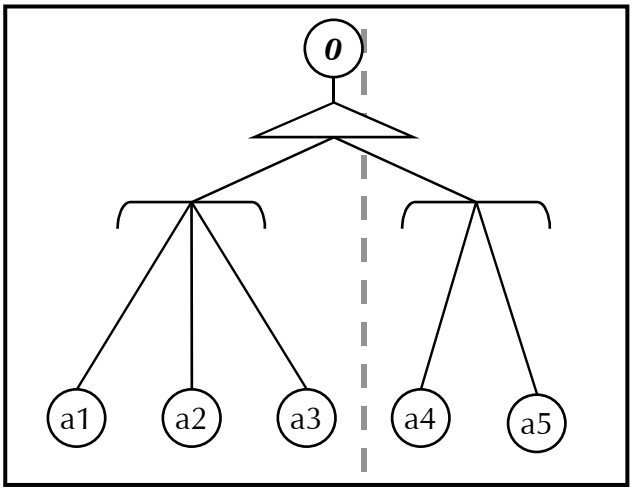


Figure 2: Partioned Net

We have two sets of attractors, a1, a2, a3, and a4, a5. Each of these sets is connected by an 'or' relation. The two sets are, in turn, connected by an 'and' relation. The notion is that the neurons that dominate attractors a1, a2, and a3 are

fairly closely coupled with one another, as are the neurons that dominate attractors a4 and a5. Neurons in these two sets are only weakly linked to neurons in one another, through relatively distant connections and through synapses with high thresholds.

We might prefer to represent this situation like this:

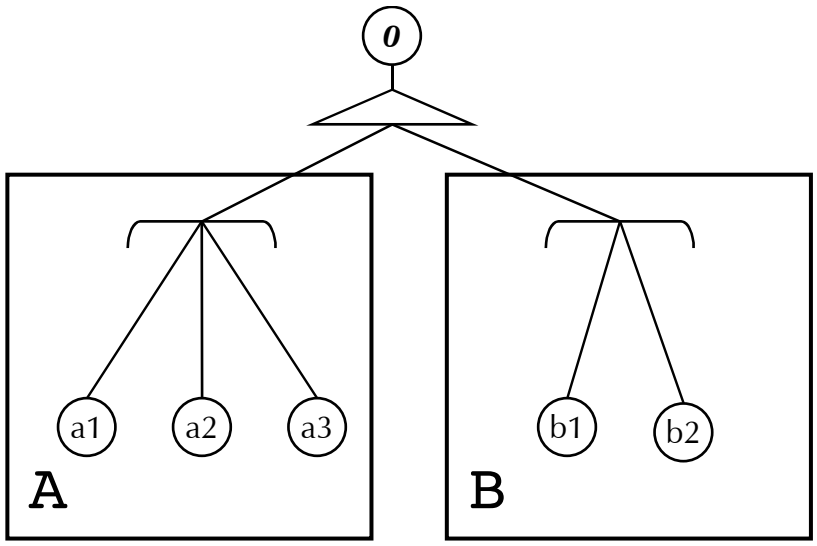


Figure 3: Partitions A and B

Partitions A and B consist of networks where the neurons are tightly connected to one another within the nets. There are also connections between the neurons in the two nets, but these are looser. We might want to think of these partitions as being Freeman's mesoscopic patches of neuropil.

We must be careful, however, because the nervous system is, on the whole, a rather fluid system, with some flux at every time scale. The work on cortical plasticity, for example, clearly indicates that functional divisions between cortical regions are not rigidly fixed. They are plastic on a time scale of hours or more. We might want to think of things being more like this:

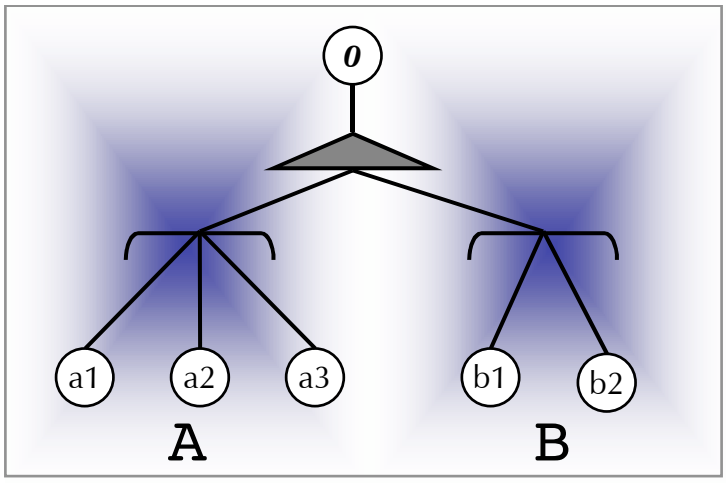


Figure 4: Fluid Partitions

Modal Differentiation

We know that the behavioral of neuropil varies according to biochemical ambiance. The following diagram shows attractors colored according to their biochemical affinity:

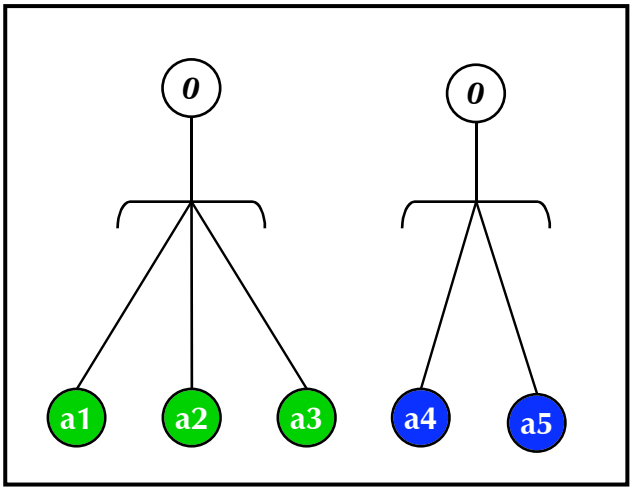


Figure 5: Attractors with different biochemical affinities

The next diagram shows some ways one might elaborate on the notion of biochemical coloring:

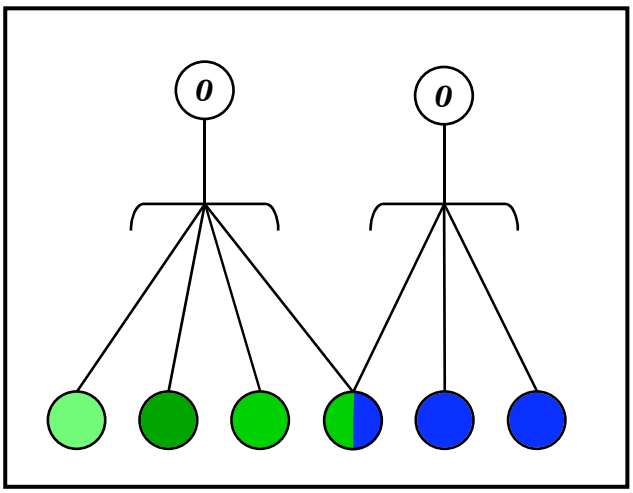


Figure 6: Color variations

On can imagine whatever interpretation of these colorings seems appropriate at this point. In thinking about biochemical coloring I have in mind, of course, Warren McCulloch's idea of behavioral mode.

A Simple Animal

The nervous system of a simple animal is connected both to the external world and to the animal's interior milieu:

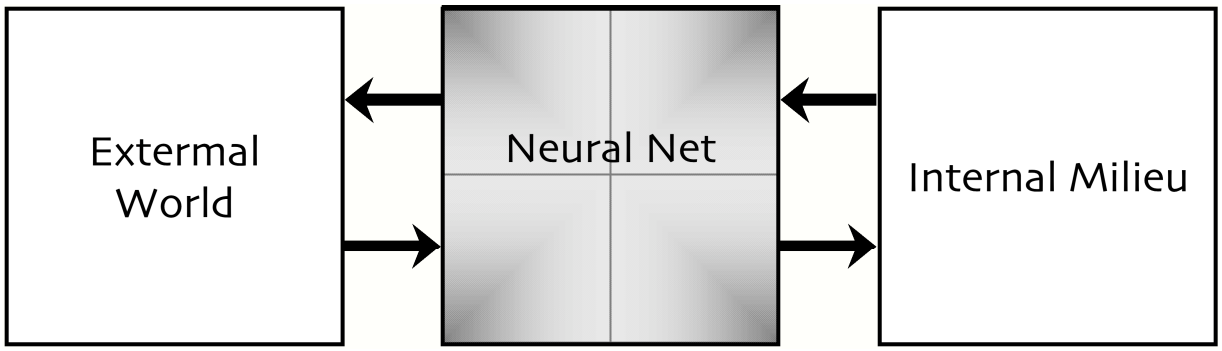


Figure 7: A Simple Animal

Its neural net is divided into at least four partitions. Each partition has two classes of neurons. One is linked to the world outside the nervous system while the other is linked only to other neurons. The latter neurons may be linked to neurons within the partition, neurons in other partitions, or both. These partitions are as follows, identified by their external links:

External Effectors: Produces effects in the external

world through coupling to the motor system.

External Sensors: Senses the state of the external world through sensors of various types.

Internal Effectors: Produces effects in the internal milieu by secreting chemicals or affecting the states of muscle fibers.

Internal Sensors: Senses the internal milieu through appropriate sensors.

The overall goal of the neural net is to keep its trajectory close to frame:

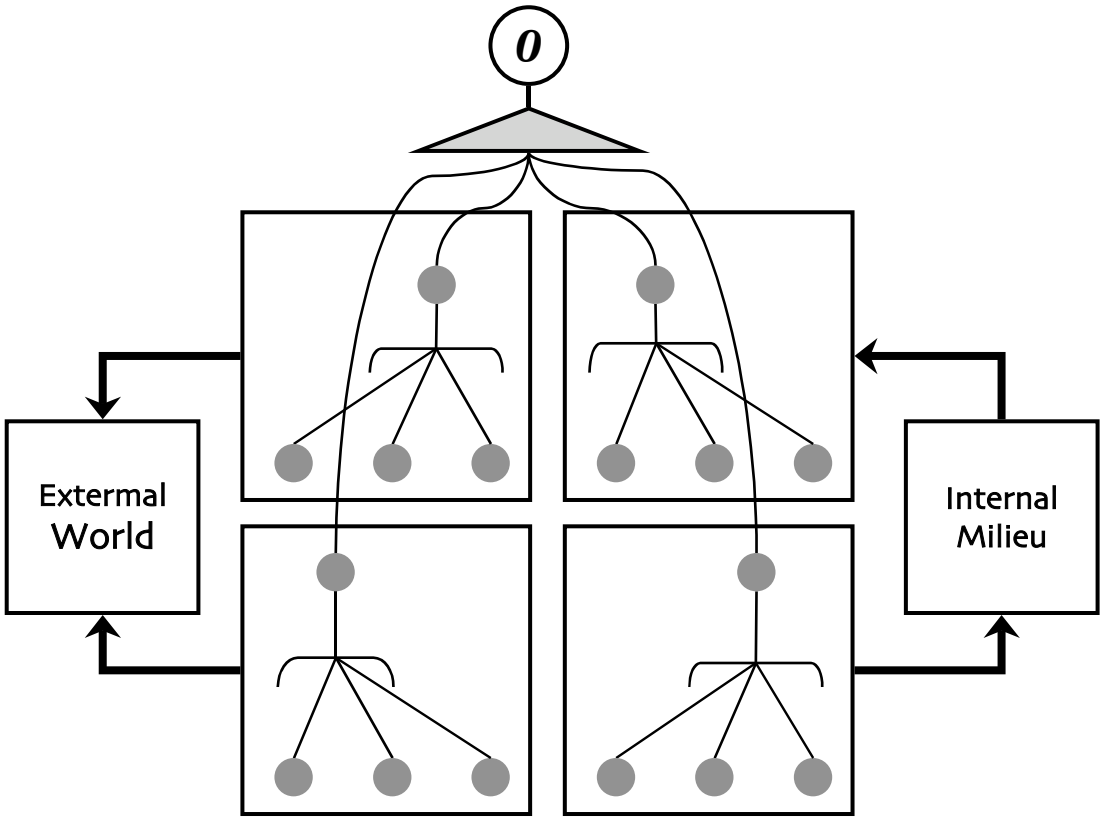


Figure 8: Life in the Net

This figure calls to mind some remarks that Powers has made about the top-level reference signal being zero.

At this point, I suggest, we have just about what we need to model the behavior of a very simple animal, such as a jelly fish or a nematode. Much of the analytic and modeling burden would, in fact, devolve upon the properties of the underlying

neural net. Keep in mind that individual neurons are themselves quite complex and at the edge of our current analytic and modeling tools.

I note that there is no explicit treatment of timing in the attractor net as presented so far. To be sure, there is the notion of a trajectory from stress to frame (and, I suppose, from one frame state to the next), but I submit that the timing of that evolution is to be handled entirely within details of the neural net. That timing need not be explicitly represented in the attractor net.

It is not, however, obvious to me that the attractor net can be entirely free of explicit timing. In our account of the structures of natural intelligence, Hays and I talked of on-blocks as primitive control devices. The notion is that, when some specified situation is sensed, an associated action is generated. I suspect we need at least that much explicit ordering at the level of attractor nets. I will leave that, however, as an exercise for a later time.

Scaling Up

What happens as the neural net gets larger and the animal evolves a more complex behavioral repertoire? Initially, the net can differentiate into more partitions and more modes. I suspect, however, that there are limits to how much growth can be accommodated that way.

Beyond that point, further growth leads to overgrowth and subsequent confusion. At that point, the way forward probably involves having the entire system differentiate into different *degrees* (Benzon and Hays 1988). It is not immediately apparent what the means. It might mean that an attractor net of the second degree has, as its states, combinations of first degree attractors. We can leave that, as well, for reflection at a later time.

References

Benzon, W. L. and D. G. Hays (1988). "Principles and Development of Natural Intelligence." *Journal of Social and Biological Structures* 11: 293-322.

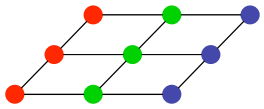
Freeman, W. J. (1999). *How Brains Make Up Their Minds*. London, Weidenfeld and Nicholson.

Kilmer, W. L., W. S. McCulloch and J. Blum (1969). "A Model of the Vertebrate Central Command System." *International Journal of Man-Machine Studies* **1:** 279-309.

Lamb, S. M. (1999). *Pathways of the Brain*. Amsterdam, John Benjamins B. V.

Merzenich, M. (1998). "Long-Term Changes of Mind." *Science* **282 40**(5391): 1062-1063.

Powers, W. T. (1973). *Behavior: The Control of Perception*. Chicago, Aldine.



IV Assignment

7.23.2003

The notion of assignment is closely related to that of a category error. Chomsky's famous example – *colorless green ideas sleep furiously* – is built on such errors. It makes no sense to assert that ideas are green, or that they can sleep. There is a mismatch here. The problem is that there are no *assignment* relations between the color domain and that of ideas, nor between that of such things as sleeping and that of ideas. Assignment, in this sense, gives cognition an ontological aspect. Assignment is about the "compatibilities" between objects in a domain.

Homogeneous Nets (Review)

We have already been introduced to the notion of a homogeneous net (above) as follows:

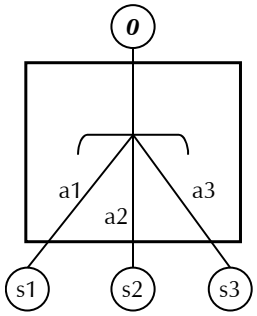


Figure 1: Homogeneous Net

This diagram can be taken to mean that, given stressors s1, s2, and s3, the net can achieve frame (0) by traveling trajectories to attractors a1, a2, and a3 respectively. This can be represented in a slightly more compact form as follows:

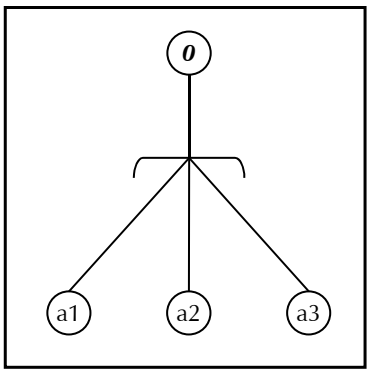


Figure 2: Homogeneous Net

As a convenience we use nodes a1, a2, and a3 to represent attractors. The arcs will be left unlabeled.

In brief, all the attractors of a given net, or partition of a net, are said to be related to one another through the ‘or’ relation. They are mutually exclusive alternatives.

Assignment, Concrete Object

Classically, when Aristotle asserted that things consist of form and substance, he was making an assertion about assignment. Concrete objects consist of a substance, such as stone or wood or metal, etc., and a form, such as round, square, bear-shaped, etc.

[Note: The formulations below are even more provisional than the rest of these notes.]

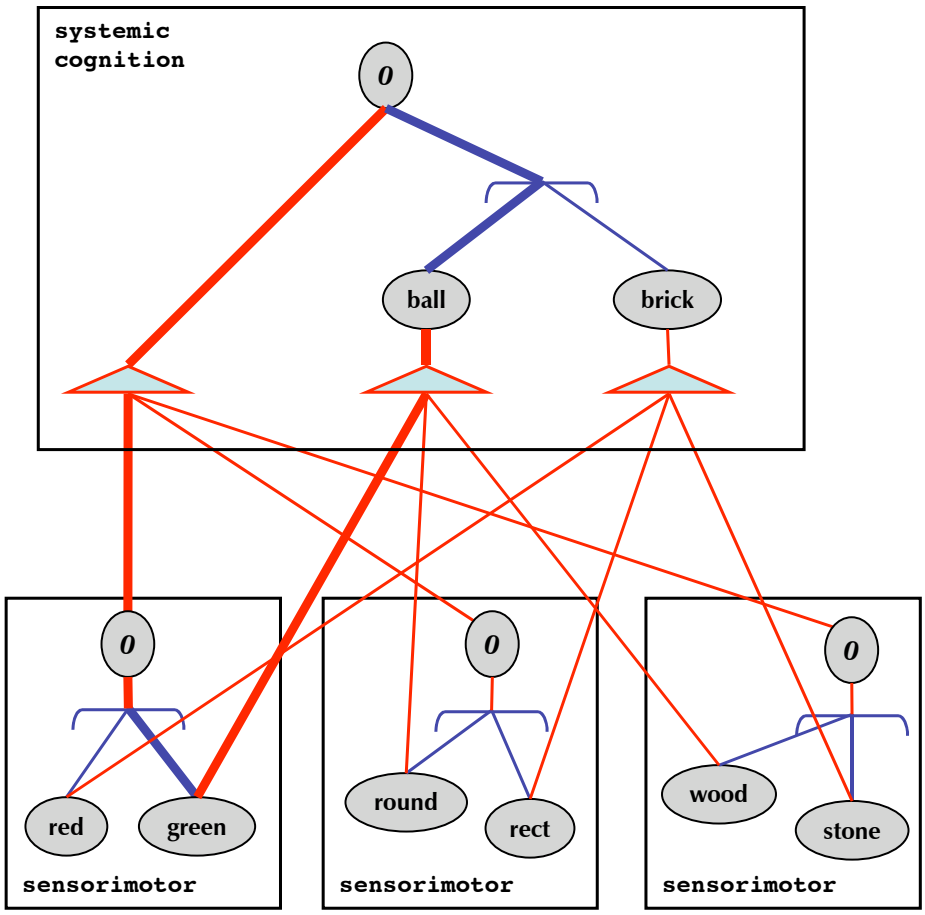


Figure 3: Assignment, concrete object

Figure 3 illustrates an attractor net representation of some of the assignment structure for a concrete object. At the bottom of the diagram we have the *sensorimotor* (SMS)domains of color, shape, and material, respectively. At the top we have the systemic (SYS) *cognitive* domain of concrete objects. Assignment linkages are depicted in red while the blue arcs indicate simple paradigmatic relationships. These colorings have no formal significance; I use them only to clarify the diagram.

In the object domain we have represented a ball, which is green, round, and wooden, and a brick, which is red, rectangular, and made of stone. The objects are linked to their constituent elements through 'and' relations. These objects are also linked to ***0*** through an 'or' relation. Notice that ***0*** is also linked the ***0***-nodes for the constituent domains. That linkage among ***0***-nodes is assignment.

[It is not obvious to me what this linkage among frame nodes means in terms of the connectivity in the underlying neural

nets. Which neurons are firing at frame depends, of course, on what the stressor is.]

Think of the structures depicted in Figure 3 as being of the sensorimotor degree; they are perceptual in nature. Figure 4 adds language into the mix and depicts a bit of the systemic degree:

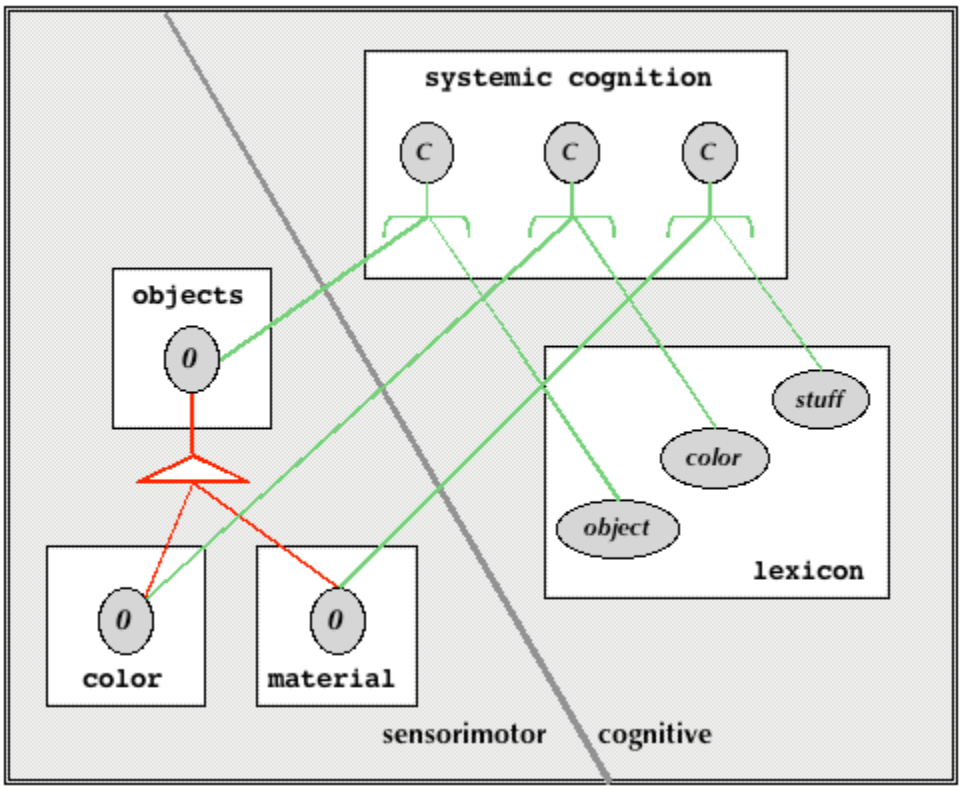


Figure 4: Some Simple Concepts

At the lower left we have the partitions for objects, color, and materials (I dropped shape to simplify the diagram). At the lower right we have the lexicon, containing word forms for /object/, /color/, and /stuff/. At the top we have a fragment of the systemic cognitive network where we see attractors for three elementary concepts. The meaning of those concepts, of course, is a function of their location in the relational attractor net. Notice that each is linked to a lexical attractor on the one hand, and the frame states of the perceptual domains on the other hand.

We can gloss the significance of these relationships in this way: Any attractor in the material domain can be designated by the lexical item /stuff/. And so on for the other perceptual domains the corresponding lexical items. Of course, each

perceptual attractor may also be designated by more specific lexical items (/stone/, wood/, etc.), but that is not depicted in this diagram. Notice that these conceptual or-relations function in the same way as those used in building the dog and animal paradigm in previous notes.

Abstraction and Process

The systemic degree is said to be higher than the sensorimotor degree. This is an asymmetrical relationship, one, I believe, that is comparable to the different strata in Lamb's stratificational grammar. In Lamb's theory, items on a lower stratum are said to *realize* items from a higher stratum. Thus, a perceptual dog can be said to realize C-dog, but also C-beast, or even C-animal.

In the formulation of *Cognitive Structures*, higher degrees perform various "services" for lower degrees. But they also abstract over the contents of those degrees. In fact, this is the key to their services.

What does it mean for a systemic C-node to be linked to the 0-state of some sensorimotor partition (or domain)? It means that that C-node can be aroused by any of active attractor in that partition. The C-node is thus an abstraction over the attractors in that partition. That the C-node is or-ed to a lexeme allows the lexeme to "fix" the attractor in the network. [Now what does that mean?] Without this "fixing" the C-node would have no content at all. We can imagine that when, for example, the /color/ lexeme arouses it's corresponding C-node (because you heard someone utter the word), it also activates the color partition. If you will, it stresses the partition "from above." As a result, one now seeks to identify the most salient color in the visual focus. That is, one seeks to develop an attractor in the color partition. When one emerges, the color partition is at it's 0-state with an active attractor now arouses the appropriate lexeme: "it's blue."

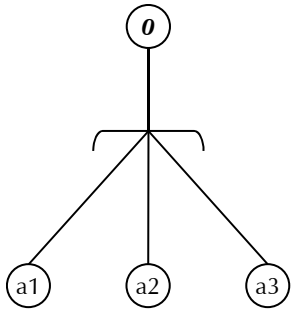
Thus we can begin to see how the various processes from *Cognitive Structures* might happen, and how the whole network has a servomechanical cast to it.

I imagine that shifting partitions from one mode to another — "stressing from above" etc. — is often chemically mediated. One chemical process might be equivalent to varying the temperature parameter in simulated annealing. [I wonder what the chemical signature of frame is?] In a large vertebrate nervous system this would be mediated by the reticular core.

V Brief Notes

Alternative Interpretation

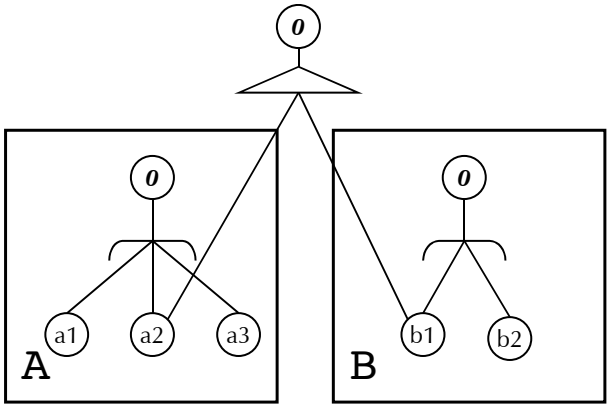
Consider a homogeneous net:



We can interpret nodes a1, a2, and a3 as representing three "neural holograms" stored in the neural net. Node 0 is then the mesoscopic patch of neuropil containing those holograms.

Partitioned Net

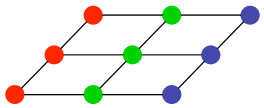
This is a revision of an earlier diagram.



This diagram asserts that both a2 and b1 are required.

In that earlier diagram (Figure 3 of the partitioned nets discussion under Simple Animal) the and-node was connected directly to the outputs of the or-nodes. That too is meaningful, but the meaning is different. That asserts that some attractor in A and some attractor in B is required.

That construction might be useful in stating syntactic requirements in a pattern where A and B make contributions. The construction immediately above can be used to assert some specific assertion over A and B.



VI Consciousness and Control

The topology of attractor nets is the structure of consciousness and control. Because an attractor net is constructed of logical relations (OR, AND) it is a control structure. As control is dispersed throughout the net, there is no need for a separate executive.

Consciousness

Freeman has suggested that consciousness:

1) is organized in hemisphere-wide global states

2) these states occur in discontinuous frames (the analogy is to motion pictures, not the AI notion)

3) the frame-rate is roughly 10Hz for a resting state and 7Hz when one is actively engaged in a task

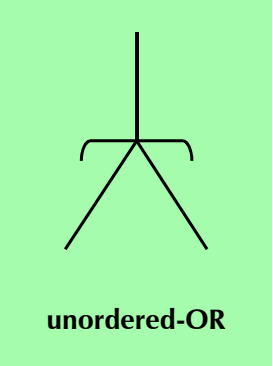
He has also suggested that consciousness functions in the manner of Baars' so-called global workspace.

In my current view, then, the so-called cognitive unconscious involves those processes that are internal to a partition, to a mesoscopic patch of neuropile. Consciousness involves switching between and among active attractors of mesoscopic patches.

Operators

Control requires logic. Lamb uses four logical operators, ordered and unordered AND, ordered and unordered OR. The operators can take more than two arguments.

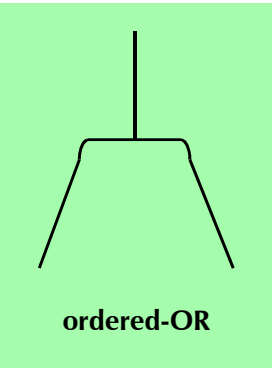
Selection: Unordered OR



In Lamb's notation, relations are ordered or unordered. The ordering is temporal. Ordered relations require their arguments to be filled in temporal sequence. Unordered relations do not.

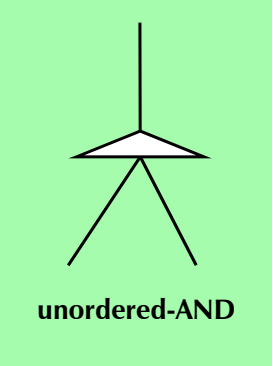
The basic use of the unordered-or relation is to indicate competing attractors in a net partition. When the partition incurs stress, its state will evolve along a trajectory until it arrives in an attractor basin. That trajectory will involve a number of bifurcations. The exact pattern of bifurcations will depend on the local circumstances. The details of these trajectories are invisible at the level of attractor logic and irrelevant to it. What matters is the attractor that has been activated, not detailed itinerary that led to that activation.

Selection: Ordered OR



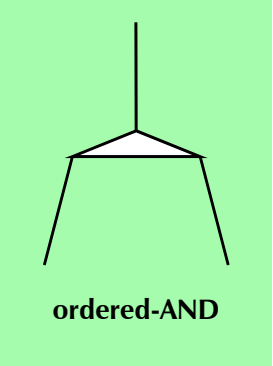
Lamb uses ordered-or to assert priority among acceptable alternatives: use the first one possible. One could use this, for example to represent salience, as in the bird paradigm, where the robin is the most salient kind of bird (at least for a certain population).

Combination: Unordered AND



The basic use of unordered-and is to indicate the strong codependency of particular attractors in different partitions. In the terms of the so-called binding problem in neuroscience, these attractors are bound to one another. They can be, are, or should be activated in the same frame of consciousness.

Combination: Ordered AND

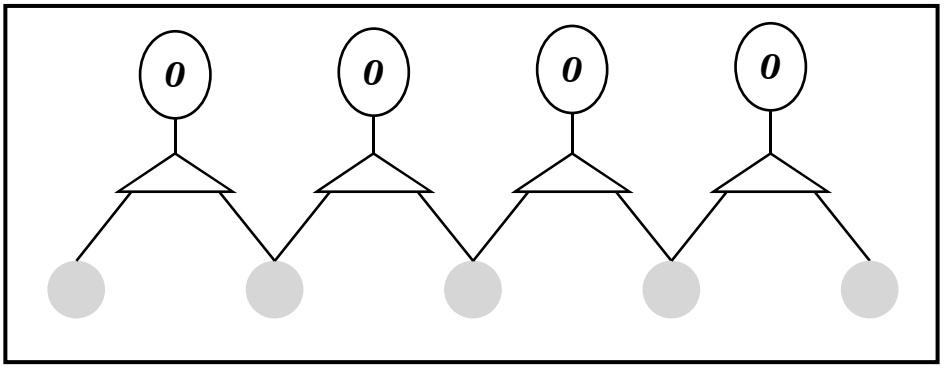


Where attractors are linked by ordered-and, they should be activated in the indicated order.

Ordered-and can be used for attractors in different partitions and for attractors in the same partition. Unordered-and cannot be used for attractors in the same partition for attractors cannot be co-activated in the same partition.

Where it is the case that attractors within the same partition keep being activated in succession, the partition might differentiate into two partitions. Whether or not this can happen will depend, at least in part, on whether or not the co-activated attractors can be divided into two (or more) disjoint sets such that co-activation occurs only between members of different sets.

One can imagine building a behavioral sequence by using a string of ordered AND's co-resident in a single partition:



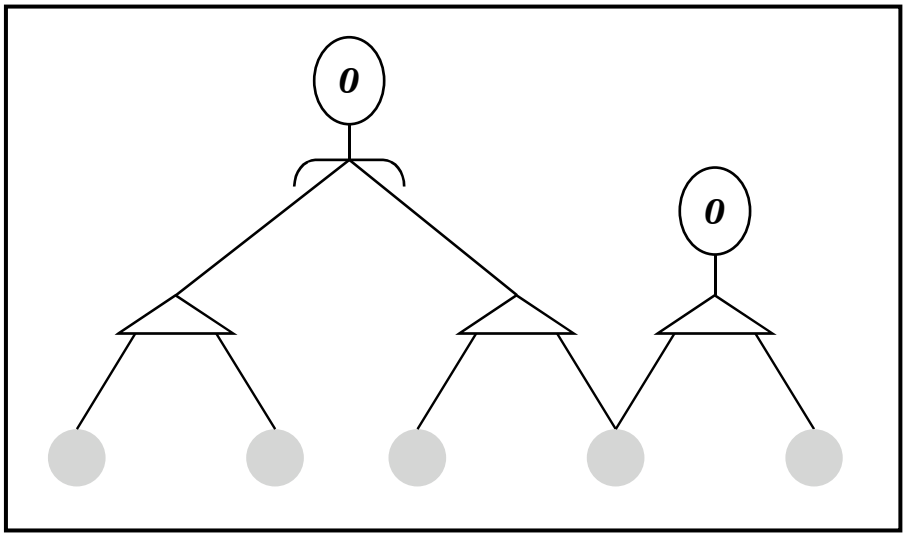
As intermediate attractors are connected before and aft we do not need any other structure to bind the sequence into a single unit. As depicted, however, the sequence will play out only from beginning to end. There is no intermediate entry point nor, for that matter, any intermediate stop, There is evidence that some behavioral patterns are, in fact, sequenced in this way.

I don't know whether or not there are any patches of neural tissue that function like this. It is one way to get the job done, and it is a job that the nervous system needs to perform.

For this to be a plausible scheme we need to have a reasonable interpretation of the attractor nodes in the above diagram. I would not necessarily expect them to have any explicit content. What we care about is that one seems to trigger the next one in the sequence. Think of a chain that advances one link at a time with each cycle of an oscillator. Perhaps the hippocampus (and related structures) directs the creation of these "links." In the rat these links are associated with positions in physical space, hence the hippocampus is conceived of as a cognitive map. In humans the links can be associated with almost anything. When the hippocampus is destroyed, the brain cannot create any new links. Thus, no new episodes, no new memories. The old links are still there and function as always, but that's it. Such links are thus a prerequisite for episodic structure. Their use, however, does not itself constitute episodic structure. Episodic structure implies storing "records" of episodes; one can remember paths through the environment without storing episodes.

I note that there are artificial neural nets that learn sequences. I do not know whether or not they do it like this.

One can imagine building more complex arrangements as well, for example:



Comment on Sequence and Control

Lamb presents sequencing constructions in *Pathways of the Brain* that are quite different from the concatenated ordered-ANDs I just used. At the moment I think the above constructions are sufficient, but, of course, I do not know that. If these somewhat simpler constructions are, in fact, adequate, that may be due to the rather different interpretation I place on the underlying neural activity.

I note further that, if these constructions are adequate, then the system has no need for a separate executive controller. The topological structure of the attractor net is itself a control structure. Because temporal activity is inherent in the basic processes of neural nets there is no need for any external system to "drive" activity forward.

Control is self-organizing and emergent.

Flow of Control

The system manages by exception. A control system (cf. Powers) pays attention to the difference between its actual input and its expected input. Freeman's account of the limbic generation of preafference implies that the nervous does just that. So let us think of this system as a Powers stack. The depth of the physical stack is set by neuroanatomy. But the power of recursive abstraction allows the depth of control to be

extended to indefinitely many effective levels. As I have not developed this notion, I'll say no more about it at present.

The system has on the order of 1000s of partitions. Think of them as supplying reference levels for some servo in the stack each working to resolve the difference between expected downstream input and actual downstream input. Where the difference cannot be resolved we have an error. Persistent error will "draw" a partition's activity into a frame of consciousness.

Conscious problem solving employs the sorts of processes Hays discussed in chapters 9 and 11 of CogStruct. Successive frames of consciousness execute successive steps in the appropriate process until one of three things happens:

1) the problem is solved

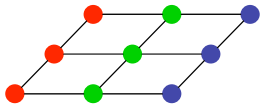
2) the task is abandoned without solution

3) reorganization (learning) takes place that allows a solution

In general, such a system works to solve problems at as low a level as possible. The lower the level, the more local the solution, and the more easily it is found. If a low level solution is not possible, then control must percolate up the stack where more resources are available.

Caveat 1: Problem-solving is not the only thing such a system does. Take musical performance. The performance is best when there are no problems whatever. Low-level problems, in contrast, threaten utter disaster. You do not want to be consciously thinking about what your fingers should be doing, etc. Etc.

Caveat 2: Note the mobility of consciousness as implied by the ordinary notion of the stream of consciousness. The mind often flits from thing to thing. The brain is capable of interleaving different streams of conscious activity in the same time interval.



VII Minds in Nets

Introduction

The notes in this section more or less complete the preliminary phase of these investigations. Appropriately enough, these remarks conclude with a definition of mind. Obviously a great deal more needs to be done, but these ideas seem pretty much to “close” the conceptual space.

In this section I want to discuss the general properties of A-nets, while interspersing this discussion with comments on Freeman’s work on perception, his work on consciousness, and on other matters as appropriate.

System Dimensionality

What’s the dimensionality of an attractor net?

We have several things to think about:

1. The **brain** considered as a more or less homogenous lump of tissue.

2. The **cortex** itself.

3. The connectivity of the **neural net.**

4. The dimensionality of the net’s **state space.**

5. The dimensionality of the **attractor space.**

(1) Considered as a lump of tissue, the brain is a complex form in 3D space.

(2) The cortex itself may be thought of as a 2D sheet that has been crumpled up to fit into the 3D cranial cavity. Much of the neural structure that interests us in laid out on this sheet. In particular, most of the partitions in the neural net can be thought of as a tessellation on this surface.

Note, however, that the functional tessellations do not seem to be rigidly fixed long-term structures. Work on cortical plasticity indicates that on the scale of hours or more these boundaries are labile. This is two or more orders of magnitude greater than the time scale of basic sensory, motor, and cognitive processes and so we may, for the purposes of analyzing those functions, consider the functional geometry to be fixed. Further we must consider the changes induced by changes in the neurochemical milieu, which can happen on the scale of seconds to minutes. One can imagine, in the abstract, that the tessellation pattern might vary appreciably from one neurochemical regime to another.

In these notes I am primarily concerned with neural processes taking place on a scale of milliseconds to seconds and minutes. Thus, with the exception of changes induced by learning, I am going to ignore these aspects of the neural network's fluidity. Thinking of A-nets as a general formalism, this suggests a distinction between *stationary* and *fluid* A-nets, where fluidity has to do with the stability of the network topology over time at a scale some orders of magnitude larger than the scale of basic A-net state transitions. For a network to be fluid either the underlying neural net must have the capacity for growth and development or there must be an external agency guiding the change. Animal nervous systems are obviously A-nets of the former kind. These issues must be addressed at some time.

(3) The neural net may have a connectivity arrayed in more than 3 dimensions (cf. V. Braitenberg, *Vehicles*, 1984, pp. 39 ff.). I'm not sure how important this dimensionality is. What is surely important is the small world topology of the net.

Neural nets may be of various *degrees*, in the language of Benzon and Hays, "Principles and Development of Natural Intelligence." *Journal of Social and Biological Systems* ***11***, 293 - 322, 1988. This has to do with details of the small world topology of these nets. In this sense, the human nervous system is of degree five while all other animals are of lesser degree. It would seem that a system with physical connectivity of degree five is unbounded in a way that is uncharacteristic of systems with lesser connectivity. For example, such systems are capable of cultural evolution. I will have to leave this for a later discussion.

Almost all of the constructions in these notes are of the first three degrees (modal, sensorimotor, and system) and do not involve the two highest (episodic and gnomonic). This does

not affect that basic matters at issue in these notes — the basic nature A-nets — but is obviously something that must be dealt with later on. I note that, inherent in this discussion, is the notion that structures of all degrees are constituted by the same kinds of basic process, The difference between a partition of degree two and one of degree five, for example, has nothing to with the nature of underlying neural substrate and its processes. The difference is a function of the location of the partition in the overall neural net.

(4) Regardless, the dimensionality of the neural net's state space is, for all practical purposes, infinite. This is true regardless of the net's degree. This dimensionality is strictly a matter of the nature of individual neurons and of the large number of them constituting real nervous systems.

This leaves us with the attractor net.

Dimensionality of the Attractor Net

The fact that the system's overall attractor landscape is a very complex surface in the state space doesn't necessarily mean that the attractor landscape is of the same dimensionality. By analogy, the cortex is a complex 3D volume of tissue, but its surface is 2D. I'm suggesting that the surface whose points consist only of the system's attractors is of (considerably) lower dimensionality than the state space.

Consider a single partition of the net independently of all others. At any moment it is in one of three conditions:

1. at some attractor,

2. moving on a trajectory toward an attractor, or

3. in a trajectory moving erratically in state space and not likely to settle into any attractor basin.

The first is what interests me. The attractors are mutually exclusive. Let us map them to the positive integers, thus suggesting that a single partition has a 1D attractor geometry. We might, for example, number the attractors according to the order in which they were first formed in the partition.

In situation 1 (above) the logical structure of this 1D attractor space is simple: exclusive OR. I leave the other two situations as exercises for some later date; this looks like

some kind of fuzzy logic may be called for.

Now let us consider two partitions coupled together so that neither can be at frame unless both are. How many dimensions does this system have? If they were independent of one another, then it would have two dimensions. But they are not independent. Hence it must have less than two dimensions. However, since either one alone has one dimension, I conclude that the dimensionality (K) of two coupled partitions is

$$1 < K < 2$$

In this case K has a fractional value. The logical structure of this K-dimensional attractor space requires AND relations over the attractors in the two partitions. [It is my understanding that *fractional* dimensionality does not necessarily imply *fractal* dimensionality. Nor, I should add, is it immediately obvious to me that K must be greater than 1; perhaps it can assume a fractional value between 0 and 1. I am not prepared to reason further on this.]

> Upon reflection: That argument seems a bit abrupt. It is surely not adequate nor do I know how to make an adequate one.
>
> In general, it seems unlikely to me that any given attractor in one partition is compatible with any of the attractors in the other. If that were the case, however, then I would not hesitate to say that the 2-partition system had a dimensionality of 2. That is one extreme case.
>
> But that seems unlikely to occur in reality. There are going to be mutual constraints on the attractors. Let us assume that partition A and B have the same number of attractors. has as many or more attractors that does partition B. In the extreme we might imagine that for each attractor in A there is only one attractor in B. That suggests a dimensionality of 1 for the coupled system. And if you buy that . . . . then a more flexible coupling would admit of a higher dimensionality, but not a dimensionality of 2.
>
> What happens, however, when one partition has more attractors than the other? Either some attractors in one partition will be mapped onto *more than one* attractor in the other or some attractors in one will not be compatible with *any* in the other. In this last case (which strikes me as one approach to thinking about psychopathology) does the system dimensionality drop below 1?
>
> So, perhaps it is the case that the dimensionality is between 1 and 2.
>
> For now, let us continue as before . . . .

The dimensionality of the attractor net for a neural net with

N partitions is

$$1 < K < N$$

K may have an integer value or it may be fractional, as it in the case where *N* = 2. The existence of a system dimensionality, *K*, that is less than the number of partitions in the system, *N*, is evidence of dependencies among the elements in the underlying neural net.

My guess is that the human brain has on the order of 1,000 to 10,000 NFAs, or partitions. The dimensionality of the attractor net is thus less than 10,000. It may in fact be *considerably* less, maybe on the order of 100 or of 10; I don't know. I do have some small reason to suspect that the A-net dimensionality is closer to 10 than to 100.[3] Still, as these things go, a dimensionality on the order of 10 is still way beyond anything we can visualize, and it is more than what Haken has in mind when he remarks that the key to understanding systems of very high dimensionality is to find low dimensional phenomena within them. But it's still considerably less than the unbounded dimensionality of the state space itself.

Process

On the basis of Freeman's current work on consciousness, let us assume that the underlying neural net is set-up so that the A-net can be said to move from one state to another in discrete jumps. This does not, of course, imply that the underlying neural net itself operates discretely. It is the A-net that is discrete, not the neural net. [Thus we have the

---

3 My suspicion is based on the graphs we used back in the Twin Willows sessions in Buffalo in the late 1970s. We have four basic edge types in the systemic degree: VAR, SYN, CMP, and ASN. Sub-graphs consisting of only a single arc type had to be trees. We can thus see that four systemic arc types corresponds to four dimensions. We had two other cognitive degrees, episodic and gnomonic, and they had the same four arc types. If we add a dimension each for the other degrees that brings us to six dimensions. We also have to add a dimension for the sensorimotor network; that gives us seven dimensions. Whether or not this reasoning is valid in this context is not something I can determine at the moment.

This suggests the possibility that the dimensionality of the A-net is, on the whole, relatively high. But for any given task, the A-net "collapses" into a low-dimensional net tailored to the task. (Perhaps it is assignment structure that take the lead in structuring the collapsed net.) Here we have the domain specificity of the evolutionary psychologists. But, instead of a fixed repertoire of genetically engineered "mental modules" we have a mechanism for creating an unbounded number and variety of task-specific A-nets on demand.

same dual analog-digital nature found in the neuron itself.]

Now let us consider Freeman's work on olfaction. He finds that the AM wave forms diagnostic of a given odorant change from one occasion to the next though they seem to have a basic family resemblance. The (rather small) portion of the cortical surface that Freeman monitors (for EEG) is, of course, connected to the rest of the brain. The odorant may be the same from occasion to occasion, and elicit the same peripheral activity in the sensors, but the overall neural context varies from occasion to occasion. Let us take that to indicate that the microscale activity of some NFA (neuro-functional area) in one of its attractor basins varies depending on the state of the other NFAs. That is to say, for each attractor within a partition, there are a variety of microscale states compatible with that attractor. Some of this variability can be attributed to the influence of other NFAs to which a given NFA is coupled.

Let us characterize the momentary state of the A-net by a state vector with one value for each partition in the net. With each state transition, the value of any position in the state vector can change form one attractor to another. There are at least two types of state vectors: frame vectors and transit vectors.

Let us imagine the brain is in a state where each partition is at frame; that is each partition is in one of its attractor basins. That state is an F—vector. What happens when we stress one of the partitions so that it is no longer at frame? That is, it is no longer in one of its basins of attraction.

What state is the A-net in now? That, it seems to me, is in some sense indeterminate. We can perfectly well characterize the state of the underlying neural net, but the state of the A-net as a whole is indeterminate. If we assign integer values to each attractor in the partition, and use those values for the A-net's state vector, how do we characterize the A-net's state when some one or many elements in the vector do not have integer values? What kind of values do they have?

In such a state, where one or more partitions is not at frame, we might say that the A-net is on a trajectory away from some one F-vector and toward another one. But there's no way to tell which one until it arrives there. After all, the partition could be anywhere in the state space of its underlying neural net. Being not-at-frame is a rather fuzzy and open-ended matter.

I figure that when one partition is knocked from frame, it will act so as to put stress on partitions to which it is closely connected. It is thus possible that they will, in turn, disturb other partitions and so on until the entire attractor net is thrashing about without ever settling into another F-vector. It is also possible that the original disturbed partition will return to frame at the same or a different attractor and leave the other partitions unperturbed. One cannot tell.

In general, however, I suspect that at any given moment some considerable group of partitions is at frame while the rest are not. The A-net's state vector at such moments thus has determinate values for many of its places, but not for all. Let us call such a vector a T-vector (transit vector). It seems reasonable to suppose that, on a time scale of minutes to hours, the brain spends much of its time moving from one T-vector to another. One can even imagine that the brain might become locked into some particular collection of T-vectors such that it never moves to any F-vector but simply keeps jumping around in the circuit.

[Q.: Is there a relationship between the fractional dimensionality of the A-net and the possibility of having its state defined by T-vectors?]

That is to say, there's no reason why there can't be more or less stable circuits of T-vectors. Or that a person can't jump around between two or more such circuits. (In the back of my mind I'm now thinking about anxiety and psychopathology.)

Further, we must consider the fact that people are normally open to the world. We experience things, sense them, move about. All this has various effects on the attractor net. An unexpected sight will knock some partitions from frame while a sought-after ice cream cone, for example, will allow others to return to frame. And so forth.

Dreams & Flow

Given all this, one might wonder whether or not the system is ever at frame and whether or not there is ever a regime in which it moves from one F-vector to another, moment by moment by moment. I suggest that the answer to both questions is "yes," more or less.

Consider REM sleep. During sleep the brain is all but disconnected from the external world. Some parts of the

nervous system are almost totally shut down. In REM, I suggest, the brain goes from one F-vector to another and can do so mostly because it is unconstrained by having to deal with the external world.

Notice that the brain is not, however, totally disconnected. If there is a sudden loud noise, for example, the system will instantly jerk to a new mode. The person will be awake and wondering what's going on.

OK that's during sleep. Is there any time that one moves from F-vector to F-vector while awake? That, I suggest, is what expressive culture is about, whether individually or collectively. When one is able to do this, that is flow, that is pleasure. The arts exist expressly to allow extended periods the A-not simply *flows* from one F-vector to another.

Consider Csikszentmihalyi on flow. He defines flow as a relationship between task difficulty and skill level (see Figure 1 below). If a task exceeds one's abilities by a large degree, one will be anxious. In the current model I take that to mean many T-vectors with a large percentage of the items in the vector being indeterminate; their partitions cannot reach frame. If one's abilities exceed task demands by a large degree, one will be bored — frame all around, ho hum. However, when task demands and abilities are well-matched, the task is interesting, and one performs it in a state of pleasant and absorbed flow — just enough tension in the system to give it a good ride.

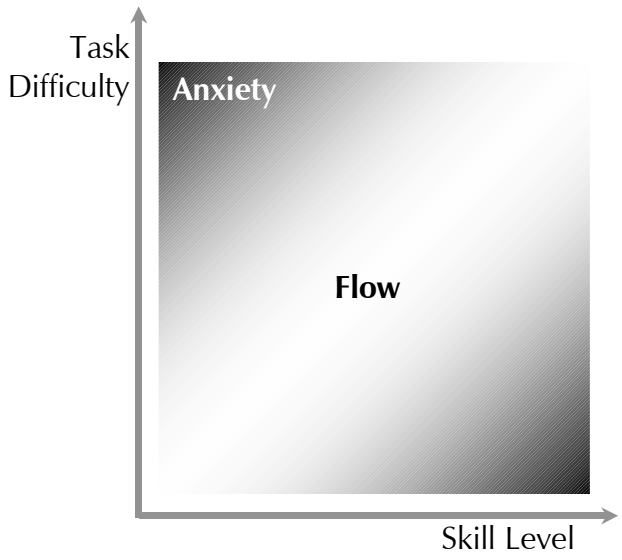


Figure 1: Flow

Given this relationship between demand and ability, it is obvious that, as we set about adding a skill to our repertoire, performing the ever-more familiar task will cease to engender flow and instead engender boredom. To regain that pleasing sense of flow we must set ourselves a more difficult task, one which challenges our ever-enlarging skill set.

Notice further that flow does seem to require some resistance; there has to be a challenge. Thus the flow experience cannot consist complete of F-vectors moment after moment. I suspect that some T-vectors are needed to trigger the more sophisticated cognitive and motor systems, which normally would come into play only when lower level systems cannot adequately cope. I further suspect, based on my experience as an improviser, that perturbations at high-level systems can be corrected by low-level changes — but that discussion requires more attention than is appropriate here.

Behavioral Mode

Finally, we need to think about behavioral mode in McCulloch's sense. He suggests that the brain **as a whole** is in one or another distinctly different behavioral mode during a given period of time; this is an aspect of the A-nets fluidity. It is not clear to me just how this affects the overall picture I've been presenting. It suggests that we need to think in terms of several different but related A-nets, one for each behavioral mode. What's the relationship between these different mode-specific A-nets? Do they correspond to the same underlying partitions or not? Etc. These strike as me both extremely important and extremely thorny questions. I want to leave most of that general discussion for later. But I would like to introduce say a bit more about mode in general and then discuss learning.

McCulloch suggested that the whole neuro-behavioral show is being "run from below" by the most primitive parts of the brain, by the reticular formation and its closely related structures. The RF has reciprocal connections throughout the brain and its various nuclei produce and distribute many (most, all?) of the neuromodulators so important to brain activity. The RF also has fairly quick and direct access to both the gut and the external world. In McCulloch's view the RF evaluated the join significance of the inner and outer state and, on that basis, committed the organism to this or that behavioral mode — such a feeding, exploration, courtship, sleep, and so forth.

Let us say that a *mode vector* contains an argument for each

partition in the A-net. That argument specifies the concentration of neurochemicals characteristic of each mode (at least those chemicals under control of the RF). When the RF has determined the most appropriate mode for the moment, it "calls up" the appropriate mode vector and configures the whole system to operate in the required mode. Each partition is parameterized and the RF regulates the overall behavioral characteristics of the system by adjusting those parameters.

Learning

Now we must consider learning. How does the A-net acquire new attractors? Through what mechanism does the A-net topology change on a time scale of hours to days, perhaps up through a life time?

[Here I'm thinking in terms of Wm Powers' discussion of reorganization in *Behavior: The Control of Perception*. But I will not review that discussion here.]

Let us imagine that some partition is thrashing about in a state of high stress. What would happen if we "spritzed" some neurochemicals into the distressed partition that then allowed it to assume different states, that changed the connectivity between the partition's neurons? Let us imagine that, in this situation, the partition's state continues to evolve but that its trajectory is no longer oscillating out of control. Instead it arrives at frame. At that point we give it a spritz with a different chemical and the spritz "fixes" the active synaptic states of the currently active neurons. The system has, in fact, learned something. It as acquired a new attractor. Henceforth the situation that had driven it to distraction on this occasion will no longer do so. Rather, it will drive the partition to its new attractor.

There is just one problem with this scenario — or, at any rate, one problem that is worse than any of the others — just who is this "we" who is doing this spritzing and how does it know to do it? For we DO NOT want to invoke a *deus ex machina*, a homunculus, to keep the whole shebang running.

I suggest that these are matters of behavioral mode and so are under the "sub rosa" regulation of the RF. It can somehow sense the state of some hunk of cortical tissue and influence that state by altering its chemical ambiance.

How would it sense cortical state? I have two suggestions, with no evidence I'm aware of to offer for either one:

1.) chemical sensitivity

2.) rhythm

Chemical sensitivity would have to work through connections from the cortex to the RF. The cortical afferents from some partition to the RF would, by their activity (high or low), signal the chemical state of that partition. Depending on that signal the RF could then deliver chemicals to that area. Here I'm assuming that a partition in distress would begin to deplete its compliment of critical modulators and somehow signal that to the RF. This doesn't strike me as being very plausible.

The other possibility is that the RF is sensitive to rhythmic activity of all sorts. As I have described it, a cortical region under tension would have a different signature from one under stress, and a region in active distress would have still a different signature. These patterns would be automatically conveyed to the RF via cortico-RF efferents. The RF would "dispense" chemicals as needed. I like this suggestion a bit better. If a one neural net cannot be exquisitely sensitive to the rhythms in another neural net, then just what CAN it do?

In either case, I'm assuming a topic mapping between RF structures and cortical areas such that fibers from cortex to RF are closely coupled with RF to cortex fibers projecting back to the same region. This seems like a plausible arrangement to me. We certainly don't want some RF-homunculus having to give routing instructions to its efferent neurons.

Finally, I note that among the simpler animals we find many that are rhythm virtuosi. They wriggle and sway and flap and buzz and swim, etc. These creatures are our evolutionary precursors.

[If dream states move from one F-vector to another, what does that imply about the role of dreams? How does this facilitate the consolidation of new knowledge, of memories?]

Comments on Freeman

Consciousness

Freeman tells me that consciousness moves from frame to frame in discrete jumps. In the terms of this discussion, it is moving from one A-net state-vector to another. I suspect what the fMRI "hot-spot" folks are seeing are regions where the

partitions are not at frame. The system as a whole is under stress, with most of the stress being concentrated in certain areas. So, those areas need more energy, yadda yadda yadda, and show up in the fMRI images as having more activity than other areas. But those other areas are not dead, they are just more or less at frame.

One would like to see the relationship between activity as measured by fMRI and consciousness as revealed by Freeman's EEG analysis.

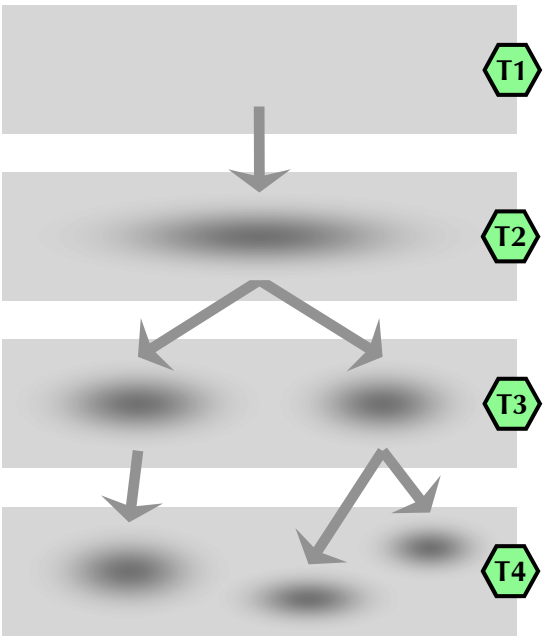
One would also like to see what Freeman's techniques reveal about sleep states, in particular, REM sleep. Does REM sleep jerk from frame to frame the way waking states do? — think of the saccades that are the tell-tale clue to the presence of REM sleep. The possibility of lucid dreams suggests this is likely. What about non-REM sleep states?

Freeman notes that the frame rate is faster when, in effect, the system is under low stress, than when it is stressed.

Learning

Freeman presents evidence that new attractors arise through a process of differentiation such that, with the emergence of a new attractor to an NFA, *all* of the attractors in the NFA are somewhat altered. This is consistent with other evidence of various sorts.

The following diagram depicts the process. At the top (T1) we have an NFA in its *pristine* state: it has no attractors and no attractor basins. Once the NFA has learned to discriminate one pattern from the background, an attractor forms (T2, second from the top).



Learning: Attractor Basins Bifurcate

Then another discrimination is acquired and we see two attractor basins (T3) in the NFA. At a later date one of these, in turn, differentiates into two attractors (T4) as another discrimination is learned.

Notice, in the first place, that this developmental process provides a rationale for mapping NFA's attractors to the positive integers. An attractor is simply assigned the ordinal number appropriate to its developmental order (there is a small problem here in that one has to decide which of two daughter attractors retains the number of the parent attractor; this seems to me a rather small problem).

In the second place, notice that this pattern of development displays a bifurcation structure similar to that of biological development — I'm thinking here of Waddington's epigenetic landscape.

It will be interesting to see how well this conception is born out in the developmental literature.

Dendritic growth and pruning vs. Hebbian learning: By way of an approach I observe that the former has the effect, first, of increasing the dimensionality of the neural net and then of decreasing it *under environmental influence*. Perhaps we can think of this guided pruning as diagonalizing through the net space, thus reducing its

dimensionality and giving us a “coherent” (i.e. now functionally specialized) NFA in which attractors can form through Hebbian learning.

Learning and Completeness: Is it possible to construct an A-net sufficiently powerful that it is, shall we say, Gödel-complete?

Once one has discovered a Gödel-axiom that is not in the system, one can simply add it to the system. There will, of course, *always* be other such axioms, and they can be added *in time*. With ordinary systems the discovery of G-axioms and their addition to the system is to be done from outside the system. That is not necessary for a sufficiently powerful A-net. These axioms can be noted as they arise and then added to the system.

Note that this has the effect of making *extension in time* intrinsic to the system. Time is not just some homogeneous “medium” in which events happen and are ordered. It is a resource available to the system.

Society and Culture

Humans are, as they say, social creatures. We are not isolated Cartesian nervous systems. The most important aspect of an individual’s environment is those other people with whom that individual interacts. In *Beethoven’s Anvil* I have made the case for the importance of rhythmically coupled interaction. I need not repeat that here. I simply note that when individuals are closely coupled with one another the effect is to bring their two nervous systems into the same dynamic regime.

Without this coupling all of the things we think of as being uniquely human would be impossible. But we should not, therefore, think that we can understand human society as one big collective A-net. That is not the case.

To be sure, I did argue that, in *certain circumstances*, we can think of a group as having a single mind. I still believe that to be the case. But those circumstances are special and we need to understand how they work, and how groups move to and from such states. That is a different matter entirely.

As Freeman has argued, we are unique individuals and our experience has meaning and significance that is unique to each individual and which is utterly private. The structure and

contents of one's attractor net is open to the population. All of symbolic culture has to do with the attractor nets of the individuals in a given society. This is where we need to begin thinking about cultural evolution and of gene-like elements in that evolution. It is these gene-like elements that allow social processes to bring about convergence among the A-nets of individual members of society.

We can thus think of cultural evolution as taking place in the A—nets of some population of individuals. Interaction among these individuals places mutual constraints upon the content and structure of their individual A—nets. It is somewhere between unlikely and impossible that any two individuals will ever have the A-nets that are exactly the same, even identical twins raised together in a small-scale society. But the A—nets of individuals within a culture must share a strong "family resemblance" if they are to interact with one another in a mutually beneficial fashion.

Mind: A Definition

It has been said that the mind is what the brain does. I do believe that. But I also believe that the value of that assertion depends on just what one thinks that the brain is doing. If one thinks that the brain is processing data in the manner of digital computers, then the assertion has little value, a tends to be identified with a belief in mental modules. The current approach is quite different.

By way of conclusion I offer the following definition:

A MIND is a fluid attractor net of fractional dimensionality over a neural net whose behavior displays complex dynamics in a state space of unbounded dimensionality. The A-net moves from one discrete state (frame) to another while the underlying neural net moves continuously through its state space.

Notice that this definition does not presuppose that only humans have minds. Other animals can, by this account, have minds as well. One can imagine classifying minds by the topology of their A-nets. (Cf. Benzon, W. L. and D. G. Hays (1988). "Principles and Development of Natural Intelligence." *Journal of Social and Biological Structures* 11: 293-322.)

Finally, just as neural nets are used to model non-neural

phenomena, so A-nets might well be used to model phenomena other than the behavior of nervous systems. I would think, for example, that they would be useful in considering biological systems, whether it be the properties of an ecosystem or the development of organisms from germ cells to adult forms. This does not, however, necessarily imply that these other phenomena are mental in nature.