Appendices

1 Sep 07

Contents

Appendix 0. Links (an overview)

Anigrafs is a framework for exploring the realm of minds. Minds, however, cannot be treated entirely in isolation from other realms of scientific exploration. Links must be present to the Natural world, to physical embodiments, including "brain" machinery that implements the wishes of various mental organisms. Furthermore, uncertainties among and within mental organisms, as well as imperfections in any associated physical system will create sources of noise in the choice aggregation process. Appendix 1 begins with the most obvious link between the anigraf mind and brain: the tally machine. Appendix 2 follows with data on the effects of noise on outcomes, either from imperfections in the tally machine, or by uncertainties in the domain model $\mathbf{M}_n$. Although these may be the most obvious links to realms adjacent to minds, there is another, perhaps more difficult connection. This is between the Natural world and the Anigraf models. Specifically, the Anigraf's intrinsic knowledge must be grounded in the Natural world. Similarities among alternatives are not arbitrary; they arise in part from regularities in behaviors. Organisms with different sensory inputs must find correlations among their observations, and these, in turn, must jibe with the correlations observed by other anigraf systems. These aspects are an important key to any cognitive architecture, and are summarized in Appendix 3. Nature's regularities also impact evolutionary processes. Appendix 4 illustrates some consequences with regard to optimality in natural selection. Finally, Appendices 5 and 6 provide technical details about the phase plots and asymptotes for top-cycles.

Appendix References:

[1] R. McN. Alexander, *Elastic Mechanisms in Animal Movement*, Cambridge, 1988.
[2] J. Barbour. *The End of Time*. Oxford, 1999.
[3] J. C. Borda. Memoire sur les elections au Scrutin. Histoire de l'Academie Royal des Sciences, 1784.

[4] Marquis de Condorcet. Essai sur l'application de l'analyse a la probabilite des decisions rendue a la pluralitedes voix, Paris, 1785.
[5] C. R. Darwin. *On Origins of Species by means of natural selection*…John Murray, Lond. 1859.
[6] R. J. Dufin, E. L. Peterson, and C. Zener *Geometric Programming*. Wiley, NY. 1967
J. Feldman. The structure of perceptual categories. Jrl Math. Psych. 41, 145-170, 1997.
[7] S. J. Gould & N. Eldredge. Punctuated equilibria: the tempo and mode of evolution reconsidered. Paleobiology 3, 115-151, 1977.
[8] J. J. Hopfield and D. Tank. Computing with neural circuits: a model. Science 233, 625-633,1986.
[9] R. Kasturirangan; Multiple scales in small-world network. MIT AI Memo 1663, 1999.
[10] H. L. Lanhaar. *Dimensional Analysis and Theory of Models*. Wiley & Sons, 1951.
[11] E. N. Lorentz. Deterministic non periodic flow. Jrl. Atmos. Sciences 20, 130-141, 1963.
[12] B. B. Mandelbrot. *The Fractal Geometry of Nature*. Freeman & Co., 1982.
[13] W. Maas. On the computational power of winner-take-all. Neural Computing. 12, 2579-2635, 2000.
[14] D. Marr. A theory of cerebellar cortex. J. Physiol. (Lond.) 202, 437-470, 1969.
[15] T. A. McMahon. *Muscles, Reflexes, and Locomotion*. Princeton Univ. Press. 1984.
[16] T. J. Pedley. *Scale Effects in Animal Locomotion. Academic*, 1977.
[17] M. Planck. *Intro. To Theoretical Physics*, Art. 28. 1932 ; Theory of Heat Radiation (trans. M. Massius, based on lectures delivered in Berlin in 1906/7. Dover, NY 1959.
[18] T. J. Purtell. Graph model for Artificial Intelligence. Adv. Undergrad. Project; Dept. Elec. Engr. & Comp. Sci., MIT, May 2004.
[19] D. Richards, B. D. McKay and W. Richards. Collective Choice and mutual knowledge structures. Adv. in Complex Systems. 1, 221-236, 1998.
[20] D. Richards. Strategic persuasion and the manipulation of knowledge structures in Social Choice. Presented at the Amer. Pol. Science Assoc. Meeting, Washington, D.C., Sept. 2000.
[21] W. Richards, B. D. McKay and D. Richards. Probability of Collective Choice with shared knowledge structures. Jrl Math. Psychol. 46, 338-351, 2002.
W. Richards, Seung, H.S. & G. Pickard (2006) Neural Voting Machines. *Neural Networks* **19**, 1161 – 1167**.**
[22] W. Richards, H. R. Wilson, and M. A. Sommers. Chaos in Percepts? Biol. Cyber. 70, 345-349, 1994.
S. Roweis & L. K. Saul. Non-linear dimensionality reduction by linear embedding. Science 290, 2323 – 2326, 2000.
[23] D. Saari,. *Geometry of Voting*. Springer-Verlag, Berlin 1994.
E. Saund. Dimensionality reduction using connectionist networks. IEEE Trans. Pat. Anal. & Machine Intel., 11, 304 –314, 1989.
[24] K. Sims. Evolving Virtual Creatures. Computer Graphics (Proc. Siggraph'94) pp 15-22, 1994.
J. Tenenbaum, V. deSilva, & J.C.Langford. A global geometric framework for non-linear dimensionality reduction. Science 290, 2319 –2323, 2000.

[25] D'Arcy Thompson. *On Growth and Form*, 2nd ed. Cambirdge Univ. Press. 1968.
J. A. Wheeler. On the nature of quantum geomettrodynamics. Annals of Phys. 2, 604 – 614, 1975.
[26] R. Wildes & W. Richards. Recovering Material Properties from Sound. Pp.356-364 in *Natural Computation*, W. Richards, Ed. MIT Press, 1988.
[27] E. O. Wilson. *Insect Societies*. Harvard University Press, Cambridge. 1971
[28] X-H Xie, R. Hahnloser, and H. S. Seung. Learning winner-take-all competitions between groups of neurons in lateral inhibiting networks. Adv. Neural. Info. Proc. 13, 350-356, 2001.
[29] H.P. Young, Optimal Voting Rules. Jrl. Econ. Perspectives 9, 51-64, 1995.
[30] C. Zener. A mathematical aid in optimizing engineering designs. Proc. Nat. Acad. Sci. 47, 537, 1961.
[31] C. Zener. *Elasticity and Anelasticity of Metals*. Univ. Chicago Press, 1948.

Neural Tally Machines

Just as there may be many different kinds of channels to establish correlations among the goals of a society of mental organisms, so may there be a host of different mechanisms for conducting any given tally procedure. If we relax the Condorcet constraint, then even within the same society, other procedures may be favored for certain tasks. For example, at a more reflexive level, the tally may be Plurality; when a modicum of alternative, second choices need to be aggregated, the Borda Count may suffice. However, when the social system strives for the most optimal tally in the context of correlated choices, the Condorcet procedure should be used. A neural system engaged in information aggregation and decision-making might therefore use any one of these possibilities.

1 Plurality Voting

Winner-take-all is a plurality procedure, easily recast as a simple voting machine. Let there be n alternative choices a_i with v_i of the voters preferring alternative a_i. The inputs to the n nodes in a neural network will then be the number of voters v_i sharing the same preference for a winner. The outcome is

$$\text{winner\_Plurality} = \text{argMax}(i)\ \{v_i\}\ . \qquad (1)$$

The dynamics of a suitable recurrent network is described elsewhere (Xie, et al, 2001).

2 Borda* Method

The classical Borda Count (Borda, 1784) includes second (or higher) ranked opinions, by assigning n-1 points to a voter's first choice, n-2 points to his second, and generally n-k points to the k-th ranked alternative. In the case of ties in a sequence of rankings, points that would otherwise be assigned to alternatives are averaged, and the average is given to each member in that set of ties. The maximum of these weighted sums is the winner.

Here we simplify the classical Borda Count by using only first and second choices. Furthermore, we assume that the alternative choices are related by a model $\mathbf{M}_n$ that is held in common by all voters. (This is the anigraf.) Each voter's ranking of alternatives is now not arbitrary, but is also reflecting information about choice relationships. Note that the effective role of $\mathbf{M}_n$ is to place conditional priors on the choice domain.

Although the shared model $\mathbf{M}_n$ has typically been represented as a graph, $\mathbf{G}_n$, it is more convenient to use the matrix $\mathbf{M}_{ij}$ where the entry "1" indicates the presence of the edge ij in $\mathbf{G}_n$ and 0 otherwise. For the graphical model of illustrated in Fig 1, we would have:

$$\mathbf{M}_{ij} = \begin{matrix} 0\,1\,1\,0\,0 \\ 1\,0\,1\,0\,0 \\ 1\,1\,0\,1\,0 \\ 0\,0\,1\,0\,1 \\ 0\,0\,0\,1\,0 \end{matrix} \qquad (2)$$

For simplicity, we assume that the edges of $\mathbf{M}_n$ are undirected, meaning that if alternative a_1 is similar to alternative a_2, then a_2 is equally similar to a_1. However, directed edges require only a trivial modification to the scheme.

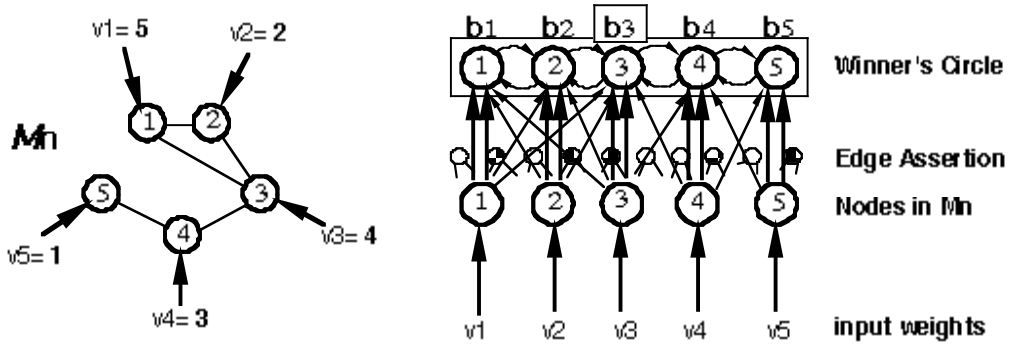

Figure 1: A Borda* Count network for the domain model $\mathbf{M}_n$ at left. The input weights v_i are analogous to the number of voters who favor alternative a_i, which is represented as node i in the network. Each input node projects to all nodes in another layer called "the winner's circle." The direct projection to the node similarly labeled has weight two, as indicated by the bold double arrows. The input nodes also project with weight one to the remaining n-1 nodes in the winner's circle, but some of these are inhibited by interneurons in the intermediate edge assertion layer. The role of inhibition is to delete the weight-one projections to the winner's circle if there are no edges in the model $\mathbf{M}_n$. The nodes in the winner's circle then carry out a WTA computation on the selected inputs. (Not all recurrent connections are shown in the winner's circle.)

With $\mathbf{M}_n$ expressed as the matrix $\mathbf{M}_{ij}$ we can include second choice opinions in a tally by defining a new voting weight v^*_i as

$$v^*_i = 2\, v_i + \mathbf{\Sigma}_j \mathbf{M}_{ij}\, v_j \qquad (3)$$

where the first-choice preferences are given twice the weight of the second-ranked choices, and third or higher ranked options have zero weight. Note this is a revised Borda method because two points are given to the voter's first choice, and one point is given to ***every*** second choice. (Justification for the revision is given in Richards et al, 2006.) The outcome is then

$$\text{winner\_Borda*} = \text{argMax}(i)\ \{v^*_i\} \qquad (4)$$

A neural network that executes this Borda* tally is shown in Fig 1. At the top (winner's circle) is a standard winner-take-all (WTA) layer. Below this is an edge-assertion layer that enforces the shared domain model $\mathbf{M}_{ij}$. The *i*th WTA node receives a synapse of strength 2 from the *i*th input node, which in turn is driven by v_i. This twice weighted input to a WTA node is depicted by heavy double arrows. The *i*th WTA node also receives synapses of strength $\mathbf{M}_{ij}$ from the *j*th input**.** These inputs are depicted by the slimmer arrows, which are active only if there is an edge in $\mathbf{M}_n$. This selection is carried out by the intermediate edge assertion layer, which will require ${}_nC_2$ neurons, each representing whether or not there is an edge in $\mathbf{M}_n$ between nodes i and j. Accordingly the inputs from v_i to v_j in the winner's circle are inhibited if edge $[v_i,v_j]$ is not in $\mathbf{M}_n$. The highlighted WTA node 3 is the Borda winner for the inputs v_i given in the model $\mathbf{M}_n$. Note that the more common winner-take-all plurality procedure would pick node 1.

3 Rank Vector

A disadvantage of ***any*** Borda method is that a weighting on voter's preferences is imposed depending on rank. In the simple Borda* model using only first and second choice preferences, a weight of 2 was given to the first choice and 1 for the second. Represent this bias as the rank vector [2, 1, 0], or, equivalently as the normalized form [1, 0.5, 0], where the 0 is the weight

applied to all preferences ranked after second choices. Then it is clear that the rank vector for the Plurality method is [1, 0, 0]. But we could also invent another rank vector [1, 1, 0] that would weight the "Top Two" choices equally. More generally, a normalized rank vector will have the form [1, b, c] with $0 < b < 1$ and $c = 0$ for our simplified preference rankings. But now we see that the outcome of a Borda* procedure will depend on the choice for b,. To avoid specifying values for b, alternatives can be compared pairwise, as proposed by Condorcet (1785.)

Definition: *let* $\boldsymbol{d_{ij}}$ *be the minimum number of edge steps between vertices i and j in* $\mathbf{M}_n$, *where each vertex corresponds to the alternatives* a_i *and* a_j *respectively.*

Then the Condorcet score S_{ij} between alternatives a_i and a_j is given by

$$S_{ij} = \Sigma_k \, v_k \, \text{Sign}[\, d_{jk} - d_{ik}] \qquad (5)$$

with the sign positive for the alternative a_i or a_j closer to a_k. (Note that if a_i or a_j are equidistant from a_k, then Sign=0 and the voting weight v_k does not contribute to S_{ij}.) Again a maximum is imposed on the value of d_{ij} of 2, which means that third or higher ranked alternatives do not enter into the tally.

A Condorcet winner is then defined by

$$\text{winner\_ Condorcet} = \mathbf{i} \text{ iff } S_{ij} > 0 \text{ for all } j \text{ such that } j \neq i. \qquad (6)$$

Hence the Condorcet winner is a true majority outcome. But, whereas there is always a Borda* winner in the absence of ties, a unique Condorcet winner is not guaranteed. Failures occur when there is a social cycle among the top three or more alternatives.

Definition: *A* ***top cycle*** *is present (and hence no unique winner) if there is some set of alternatives such that* a_i *beats* a_j , a_j *beats* a_k, *and* a_k *beats* a_i , *and every alternative not in the top cycle is beaten by at least one alternative in the top cycle.*

Thus, although the Condorcet tally also utilizes information about relations among alternatives in the choice domain, it has two disadvantages over the Borda* method. First, there is no guaranteed winner (which may be appropriate if one seeks a clear consensus), and second, because pairwise comparisons are required there is a greater computational cost.

4 Condorcet Network

To reduce the computational complexity, the insight is to choose a special subgraph of $\mathbf{G}_n$, namely $\mathbf{g}_k$, with $k << n$. Conceptually, the subgraph chosen is a ridge in the landscape of Borda* weights. The ridge consists of the k nodes in $\mathbf{G}_n$ with the highest Borda* scores (i.e. Eq. 3). This choice is based on the observation that for connected random graphs of fixed edge probability, with weights chosen from a uniform distribution, there is near 100% likelihood that the Condorcet winner will be among those alternatives with the top five Borda* scores. (See Richards et al, 2006 for details.)

4.1 Specifics for the subgraph $\mathbf{g_k}$.

Let the Borda* rank vector be [2,1,0] as before, with the Borda* scores v^*_i for each vertex i in $\mathbf{G}_n$. Without loss of generality, label the vertices in $\mathbf{G}_n$ by the rank order of their Borda* score, with vertex $i = 1$ having the largest score. In cases where the Borda* scores are tied, simply choose the indexing arbitrarily among the tied vertices to create a total order.

Definition: $\mathbf{g}_k$ *is the* ***spanning subgraph*** *of* $\mathbf{G}_n$ *containing the vertices with the top* k *Borda scores.*

4.2 Sketch of Neural Network

The objective is to obtain a crude sense of the complexity of a plausible neural network that calculates a winner for $\mathbf{g}_k$. There are three design challenges: (1) finding $\mathbf{g}_k$, (2) computing the winner for each pairwise comparison, and (3) to determine which alternative (node) beats all others. To find the nodes in $\mathbf{g}_k$ that constitute the maximum Borda* ridge, use a scheme similar to that shown in Fig. 1. Typically, the vertices with the highest Borda* scores have the largest vertex degrees and hence the greatest connectivity.

The second challenge is to implement ${}_kC_2$ pairwise comparisons [Eq. 5]. The scheme requires noting whether the vertices being compared are adjacent or not. Consider first the case where two vertices in $\mathbf{g}_k$ are not adjacent in $\mathbf{G}_n$ (and hence also not adjacent in $\mathbf{g}_k$.) Then we simply need to sum the weights of the neighbors to each vertex, plus the weight of the vertex itself, and then compare these two weight sums to determine the pairwise winner. Note that this is equivalent to using the Top-Two rank vector [1,1,0] for each vertex, and then picking that vertex with the largest score. If the two vertices being contested are adjacent, however, then note that the weight of each vertex will be added to the score of the competing vertex. Hence the weights of the vertices themselves will be cancelled if the [1,1,0] rank vector is used. To correct for this cancellation, the weights of each vertex in the comparison must be doubled when vertices are adjacent (Richards et al, 2002). This is the Borda* rank vector [2,1,0]. Hence, when calculating each pairwise Condorcet score, the rule is to use a Top-Two rank vector [1,1,0] when vertices are non-adjacent in $\mathbf{G}_n$, and to use the Borda* rank vector [2,1,0] when adjacent. This requires making explicit whether or not the ${}_kC_2$ edges in $\mathbf{g}_k$ are adjacent or not in $\mathbf{G}_n$.

The third challenge is to determine that node or vertex beating all others. As will be seen below, this can be handled easily by a logical AND of the WTA outputs from each pairwise comparison.

Fig. 2 depicts a $\mathbf{g}_k$ network with six layers. The computation is carried out as follows:

(i) the k-maximum Borda* ridge (nodes in $\mathbf{g}_k$) is given, as well as the neighborhood sums in $\mathbf{G}_n$ for each node in $\mathbf{g}_k$.

(ii) activate the ${}_kC_2$ set of nodes in an "edge assertion" layer to make explicit which edges in $\mathbf{M}_n$ are present in $\mathbf{G}_n$ (as was done previously in the Borda* network.) Note that this activation controls inputs to both the neighborhood sums layer, as well as to nodes in the comparator layer.

(iii) for every node in $\mathbf{g}_k$, find the sum of the weight of vertex i and its neighbors (neighborhood sums.) Again, this step is analogous to that used in the Borda* network of Fig. 1, except at this stage the sum uses the "Top-Two" rank vector [1,1,0]. The second input for the weight of vertex i itself will be added in step (iv) depending upon edge connectivity in $\mathbf{G}_n$.

(iv) project the activity of [1,1,0] neighborhood node sums onto each member of a pair of nodes in the "comparator" layer that has the same vertex label.

(v) project the weight of vertex i itself onto the appropriate member of all pairs in the comparator layer, but only if the two vertices in the comparator layer are adjacent in $\mathbf{g}_k$ (as controlled by the edge assertion activations in (ii).)

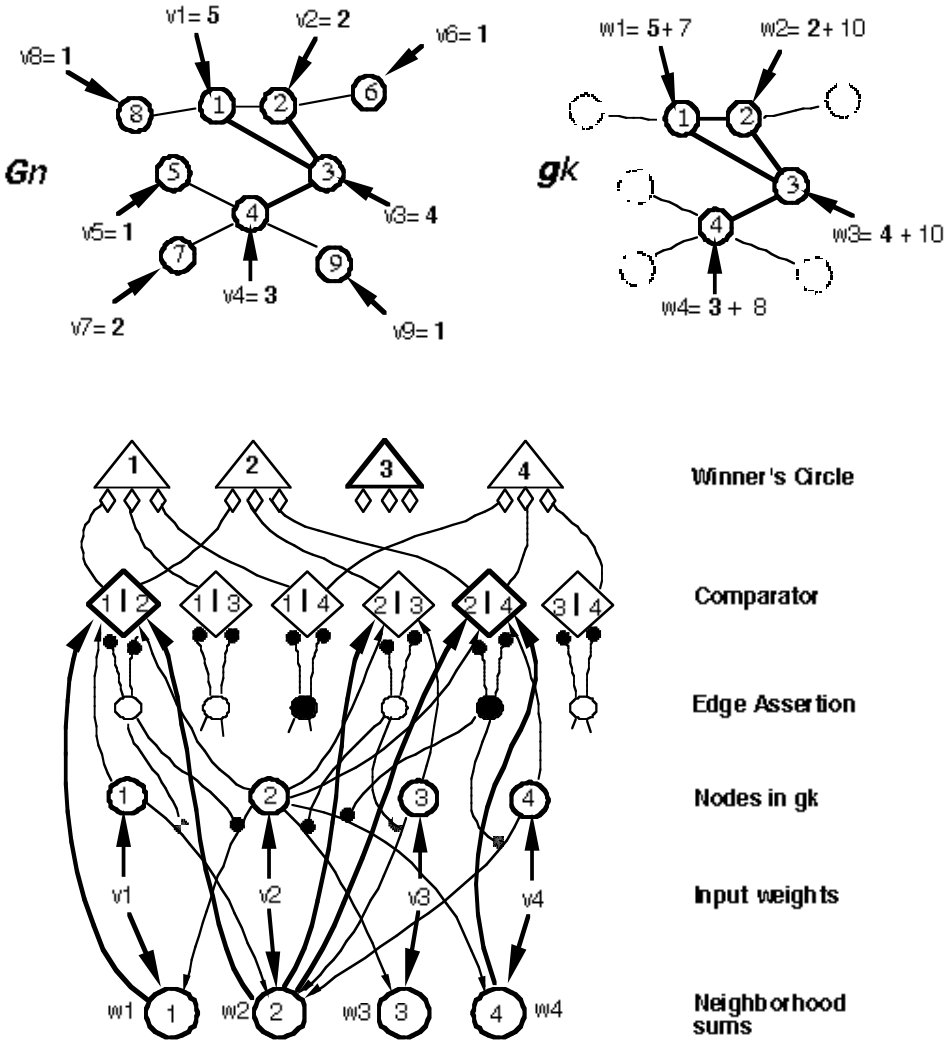

Figure 2: In the upper left is a simple domain model $\mathbf{G}_n$ relating nine alternatives, with input weights v_i. The reduced version is depicted at the upper right as the subgraph $\mathbf{g}_k$ with 4 nodes that correspond to the nodes in $\mathbf{G}_n$ having the four highest maximum Borda* scores. These nodes in $\mathbf{g}_k$ continue to receive weights from nodes adjacent in $\mathbf{G}_n$, indicated by the unlabeled dashed circles. These same nodes, with their augmented weights w_i, are shown in the network below as the "neighborhood sums", namely the inputs v_i to the layer labeled "nodes in $\mathbf{g}_k$." The scheme then proceeds analogous to the Borda* network, where each node in $\mathbf{g}_k$ sends its input weight to a set of ${}_nC_2$ comparator nodes. Between these two layers is, as before, another set of ${}_nC_2$ interneurons that makes explicit through inhibition whether an edge in the domain model $\mathbf{G}_n$ is present or not. If nodes are adjacent in $\mathbf{g}_k$ when weights are being compared, then the weight of the $\mathbf{g}_k$ node must be doubled. This is the role of the back-projections of the interneuron "edge assertion" layer onto the neighborhood sums (note small black circles depicting inhibitory synapses.) Finally, the loser of each pair-wise comparison in the second highest layer (diamonds) sends an inhibitory signal to its namesake in the winner's circle (triangle), which shuts down this node, as it cannot be the Condorcet winner. If all nodes in the winner's circle become inactivated, then there must be a top-cycle.

(vi) use a WTA procedure to select the winner of each pairwise comparison in the comparator layer, and send either a "0" (loser) or "1" (winner) signal into the appropriate node in the "winner's circle".

(vii) do a logical "AND" of the inputs to each of the k nodes in the winner's circle. If there is a unique winner, then only one node will remain active. If there is no such unique winner, then there is either a tie or a top-cycle.

Note that although there are only k nodes in the winner's circle, in the comparator layer there will be a much larger set of roughly 2 x ${}_kC_2$ depending upon the tiling of neurons. This comparator layer, and also the comparable edge assertion layer, are the critical components that govern the size of the network. If the diameter of $\mathbf{G}_n$ is very large, the connectivities required become too distant. Some hint of this problem is given in Fig. 2 for $k = 4$. This depiction also makes clear that neither $\mathbf{G}_n$ nor $\mathbf{g}_k$ appear explicitly as graphs. Rather, the connectivity is represented by the filled nodes that indicate whether the vector [2, 1, 0] or [1, 1, 0] should be applied to the paired comparison in the comparator module. This representational form has the obvious benefit that weighted edges, i.e. correlations among alternatives, can be easily be incorporated by allowing analog, rather than binary inhibition by the "edge assertion" nodes in layer 3 (small circles.)

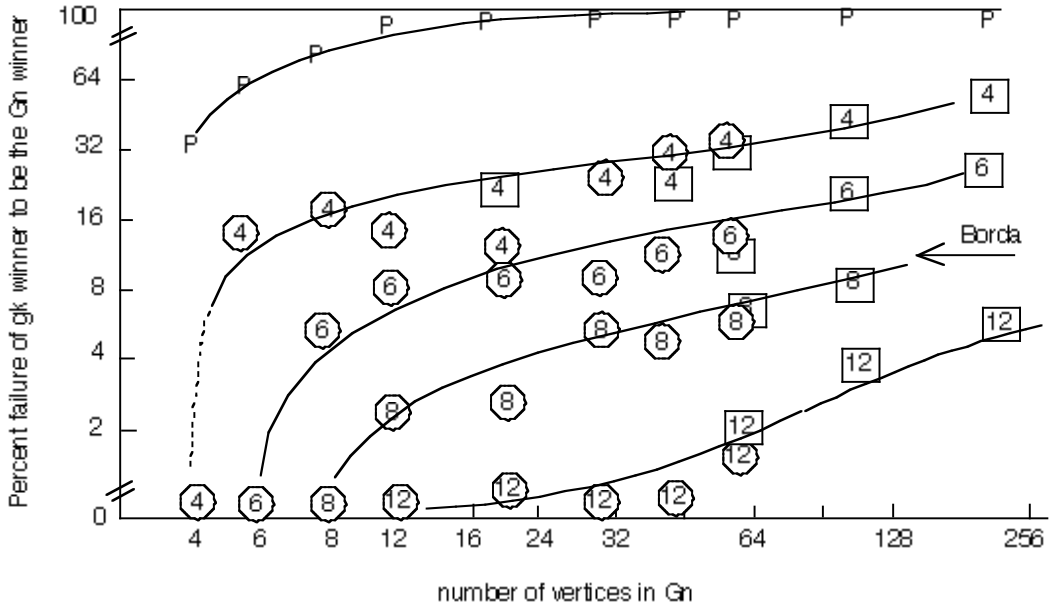

Fig. 3: Winners for $\mathbf{g}_k$ compared with winners for $\mathbf{G}_n$. The numbers along the curves indicate the values of k. Note that for $k > 6$, the smaller subgraph $\mathbf{g}_k$ does a very good job of finding the true Condorcet winner – roughly the same as the Borda* winner (arrow at right.). Note that the Plurality winner (P), which is the maximum weight node in $\mathbf{G}_n$, is rarely the Condorcet or Borda* winner.

5 Success of gk

Figure 3 shows the success rate of the k-Condorcet procedure for graphs of size n < 250, with different choices for k. The models $\mathbf{M}_n$ used were connected random graphs with edge probability 1/4. (Similar or better results are found for higher edge probabilities.) A set of weights on the nodes was chosen from a uniform distribution. Each point is based on a minimum of 100 trails, with more trials included so that the percent standard errors would be less than 5% of the ordinate values, or 0.5% if the ordinate is less than 5%. The ordinate of Fig.

3 shows the failure rate of $\mathbf{g}_k$ to yield the same Condorcet winner as $\mathbf{G}_n$, (Also shown is the failure of the Plurality method (P) and the Borda* count, compared with the Condorcet choice. Regardless of *n,* the Borda* and Condorcet winners differ only about 10% of the time, as indicated by the arrow at right. A small fraction of this percent is due to top cycles in $\mathbf{G}_n$. Likewise, an important factor for different winners for $\mathbf{g}_k$ and $\mathbf{G}_n$ is the presence of additional top cycles in $\mathbf{g}_k$. In other words, when $\mathbf{g}_k$ picks a winner, this winner is almost certainly the $\mathbf{G}_n$ winner (98+% for $k > 8$.) The approximation by $\mathbf{g}_k$ is thus conservative: there are few false positives; instead no winner is chosen, unlike the Borda* count.

6 Biological Implications

The Plurality method is the simplest to implement, the Borda* next, with the Condorcet network being the most complex. For both the Condorcet and Borda* procedures, $O(n^2)$ computations are required for dense networks, with the limiting factor for the Condorcet calculation being the Borda* ridge. From a biological perspective, using parallel computations, both of these implementations would be rather trivial and rapid even for large n. This observation favors using the proposed Condorcet network because of its benefits over the somewhat simpler Borda* scheme. First, the Condorcet network finds correlations among the most significant choices when the subgraph $\mathbf{g}_k$ is chosen. This design thus gives the network the potential to learn priors on important correlations. Second, it has a clear rejection strategy during the learning phase, namely the presence of top-cycles. No other method mentioned above has this kind of built-in feedback mechanism because all others always output a winner. Finally, it is not inconceivable to see the potential of such a layered network design in primitive cortical areas, even for the aggregation of rather simple features.

Noise & Uncertainty

1 Preliminaries

Our worst-case scenario is when n agents vote their first choice but otherwise choose alternatives arbitrarily, ignoring the shared model $\mathbf{M}_n$. Then in a Condorcet tally the likelihood of no-winner will rise rapidly toward 100% as n increases (See text Fig 1.2.) On the other hand, if the shared model $\mathbf{M}_n$ is respected most of the time, or by most of the agents, then for large n, the chance of a no-winner will be reduced, roughly in proportion to the percent of haphazard votes (Richards, 2004.) Such haphazard voting that includes all choices other than an agent's first choice borders on the malicious. There are two other types of noise incursions that are more reasonable: (1) when the alternatives are third or fourth ranked and (2) where there are individual departures from the shared model $\mathbf{M}_n$. Unlike random voting for all less preferred alternatives, the behavior of these other two noise incursions is somewhat unexpected.

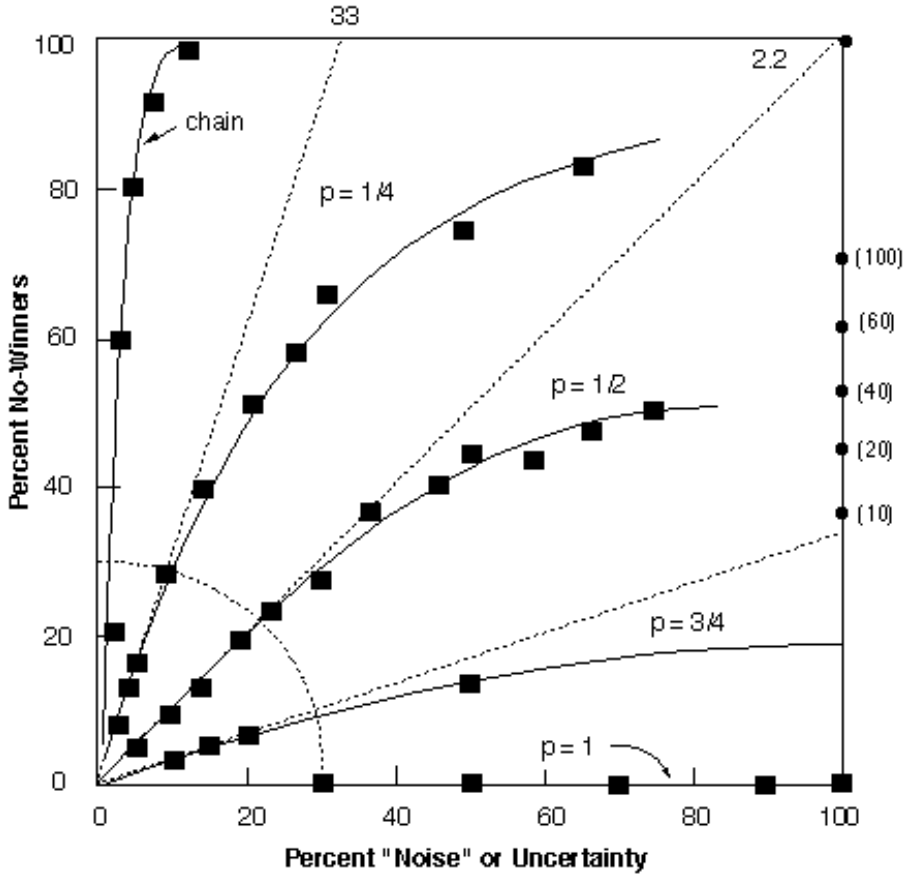

Fig. 2.2. Haphazard votes for alternatives not immediately similar to ideal point (i.e. non-adjacent vertices in Gn.) Dat are for 40 vtx random graphs with edge probability, p, as indicated.

Fig 2.1

2 Uncertainty in low ranked alternatives

When an agent ponders a choice between two alternatives that are third or higher ranked choices – i.e. not the neighboring vertices in $\mathbf{M}_n$ - one would expect less

discrimination among alternatives, and hence more uncertainty. Ideally, in these circumstances, the agent should punt on a choice, not electing one alternative over the other. Here, however, we explore the effect on outcomes if a haphazard vote is made. The results are illustrated for 40 vertex random graphs having different edge probabilities, with weights on nodes were chosen from a uniform distribution.

In Fig. 2.1 each curve shows the probability of no-winner as the percent of haphazard vote increases. Note the significant effects of the edge probability for $\mathbf{G}_n$. For highly connected random graphs, noise is ineffective as expected, because covering of nodes eventually becomes guaranteed. For sparse graphs, however, such as chains, an almost trivial amount of noise or uncertainty can create a high probability of no-winners.

We also see a striking correlation between the edge probability of $\mathbf{G}_n$ and the asymptotic slope of the relation between no-winners and noise as the noise approaches zero (dashed lines). The slope is (1-p)/p for edge probability p, which can be derived by considering the chance that noise will overturn the noise-free winner in each pairwise contest.

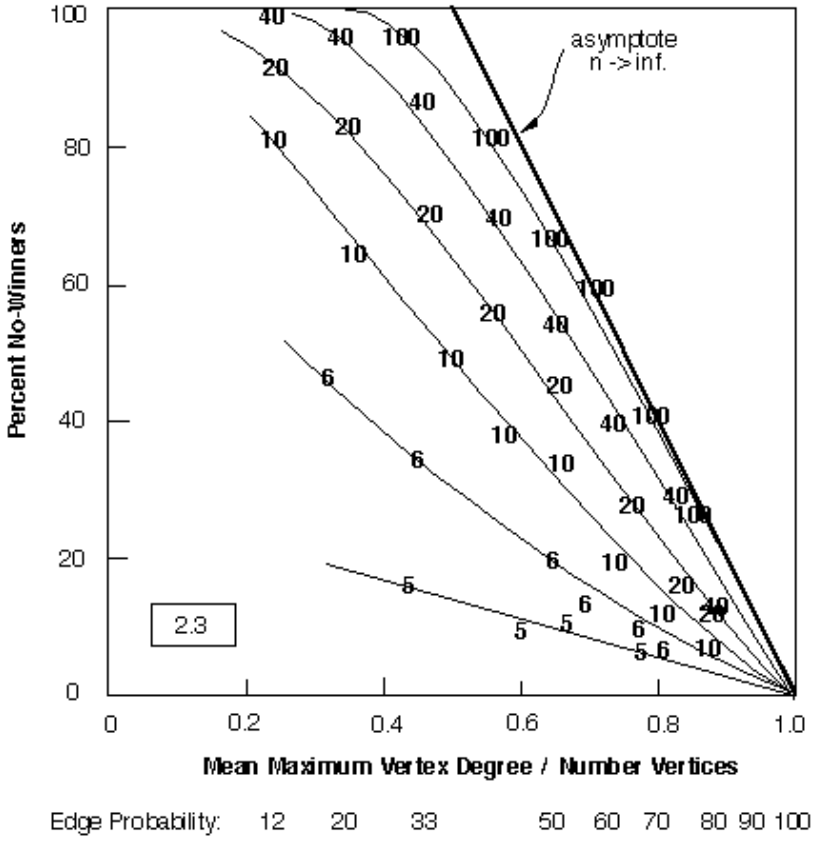

Fig. 2.3. Same data as in fig. 2.2, replotted to show asymptotic behavior for large n.

Fig 2.2

Figure 2.2 provides another plot of a portion of the same data, Here, the abscissa is the mean maximum degree of $\mathbf{G}_n$, normalized by the number of vertices n. For all points shown, the noise is fixed at the maximum of 100%, provided that the two contested

alternatives are not adjacent to the agent's first choice. An asymptotic limit seems to be present as n increase (bold line).

3 Model Variations

In the ideal case, agents use a shared model $\mathbf{M}_n$ of the domain to rank their preference orderings of alternatives. But practically, there may be variations of $\mathbf{M}_n$ .where the ranking of alternatives by some agents depart from the accepted model. A simple example is all edges in $\mathbf{G}_n = \mathbf{M}_n$ are not bi-directional. Another case is when some agents omit edges in $\mathbf{M}_n$, whereas others do not. The effect on outcomes from these aberrations is quite different from changing weights on alternatives. Figure 2.3 illustrates by plotting percent agreement on winners versus the size of $\mathbf{M}_n$ (number of alternatives.)

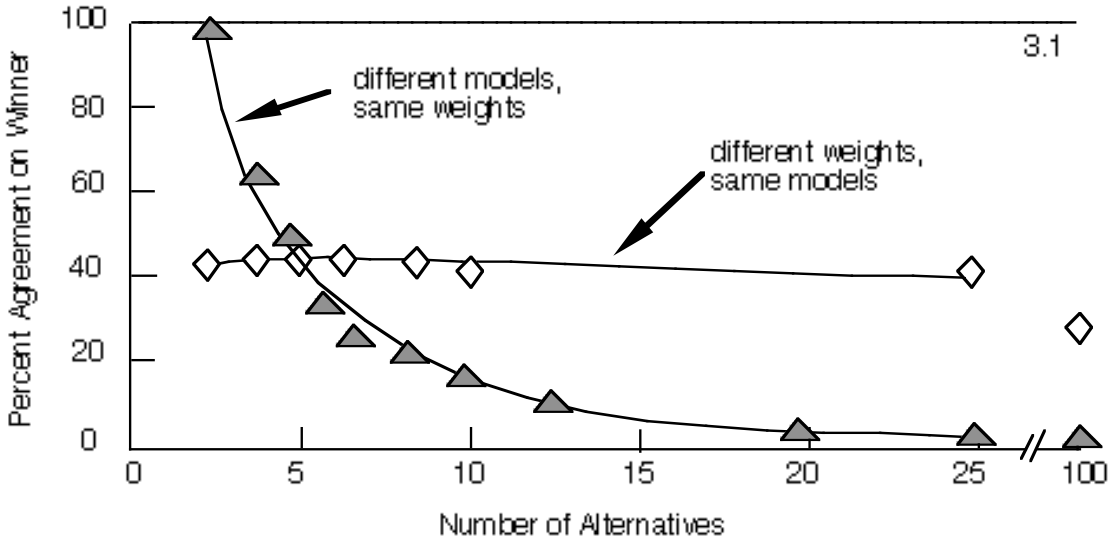

Fig 2.3

To illustrate the different effects of alterning the domain model versus changing weights, two cases are considered: (1) the domain model is altered, but the weights remain unchanged, and (2) the weights on nodes are altered, but the model is the same. Again, as typical, weights vertices in $\mathbf{M}_n$ are chosen from a uniform distribution, and the graphical model $\mathbf{M}_n$ with n vertices is a random graph with edge probability of one-half, and with all edges bi-directional. The open diamonds show that when the domain model is held fixed, but a second set of weights are chosen: There is little change in the percent of agreement in outcomes, which remains roughly constant at 40% for $n < 30$. In other words, the structure of the model dominates the outcomes. In contrast, as shown by the gray triangles, when the weights are held fixed, but applied to two different random models there is a dramatic fall in agreement between the two models. (Note that this is an asymptotic condition of text Fig. 8.3.) Manipulations of the domain model thus can have significant effects on outcomes, and these effects behave quite differently from noise introduced by changing weights on vertices. (Indeed, these two factors are largely independent over a range of edge probabilities for small graphs.) In other words, if $n > 6$, belief revision has a much more powerful effect on achieving agreement than does voting strength.

4 Top-Cycles for non-ideal Domain models

Further insight into the difficulty of finding agreement when there are individual violations in the shared model $\mathbf{M}_n$, one can explore the probability of top-cycles after altering bi-directional edges in $\mathbf{G}_n$ (equivalently $\mathbf{M}_n$).

Let e_{ik} be the directional edge linking vertex v_i to v_k, and e_{ik} be the opposite directional edge from v_k to v_i, If agent (i) associated with vertex v_i drops a similarity relation to v_k, then edge e_{ik} is deleted and the remaining directional edge from vertex v_k is e_{ki}. In this process, only vertex v(i) has been altered, or, equivalently, only agent (i) has changed his preference orderings.

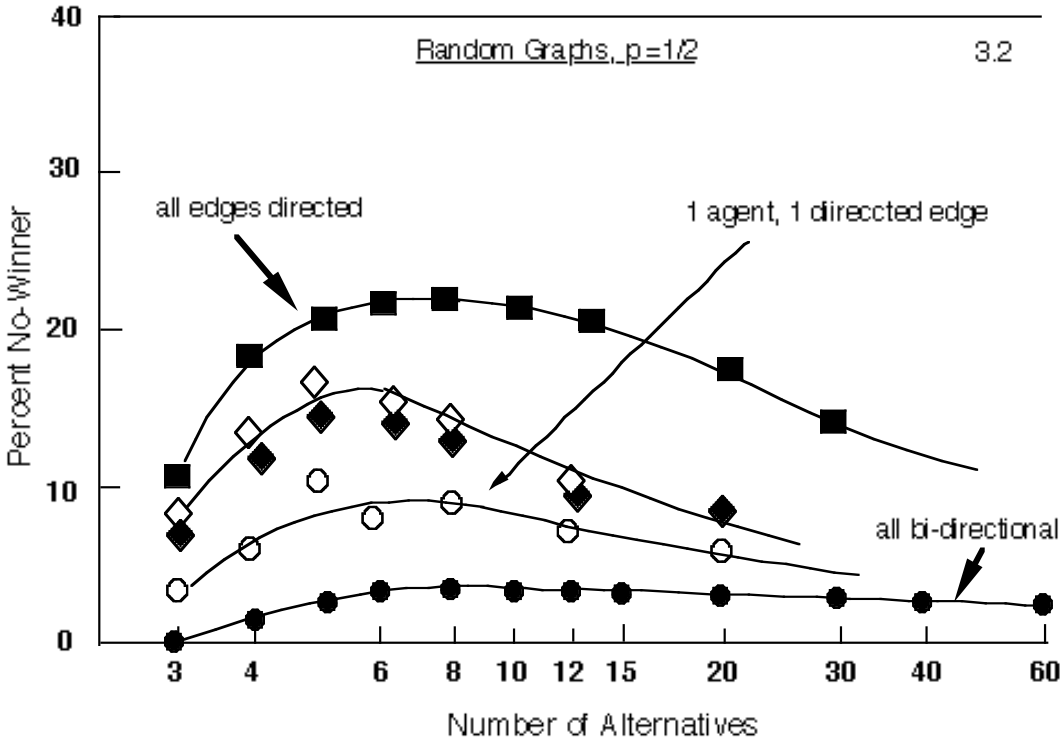

Fig 2.4

The top curve in Fig. 2.4 (filled squares) shows the probability of top-cycles when all agents rearrange their edges in $\mathbf{M}_n$, choosing new neighbors from a uniform distribution of (n-1) vertices. (Hence for p = 1/2, about one-half (0.4) of the links between vertices will be bi-directional.) For this condition, note the maximum of roughly 20% compared with only 4% of top cycle outcomes for the ideal bi-directional $\mathbf{M}_n$ (lowest curve.) Significantly, unlike random noise on alternative weights, as the number of alternatives becomes large, the odds for no unique winner become small.

Between these two cases of all bidirectional or mostly directed edges in $\mathbf{M}_n$ are shown two other, less extreme "miss-matched" conditions. The solid diamonds show the probability of no consensus if all agents rearrange only one relationship in the global domain model by changing an existing bi-directional edge to directed. In a complementary case, only one agent rearrange *all* edges, creating random miss-matches with the original domain model (open diamonds.) Not surprisingly, the effect of this one rogue agent on disturbing the probability of consensus is roughly equivalent to all agents mismatching one relationship.

Intrinsic Knowledge

"Nature wears an infinite grin. If only we could see it!" -- P. Gunkel

1. Modes

Shared knowledge about a domain plays a key role in Anigrafs. This knowledge is based on the expected correlations among states in the world. Preference orderings for agents and mental organisms are derived from these relations. Although similarities among goals and actions have been highlighted, analogous relations are present in sense data, and are used by mental organisms to interpret these data. For example, visual systems have a variety of information streams such as motion, color, texture, shape, 3D depth, etc., all of which must be interpreted and aggregated. Other organisms might be focused on interpreting acoustic information; yet others are analyzing tactile, olfactory, vestibular or kinesthetic data. Both within and between these various modalities, different types of objects and events will lead to different kinds of correlations among such multi-modal data. The validity of the correlations among these inputs dictates the accuracy of interpretations and categorizations of world states.

Definition: *A **Natural Mode** is a cluster of quasi-independent regularities where the measurement of one regularity is highly predictive of measurements of other regularities belonging to that cluster, in the given context.*

In other words, useful observations are those that reflect some modal regularity of Nature. These provide a basis for similarities among different modalities, thereby linking these modalities. Modal effects become even more prevalent when creatures with predictive models of Nature can mold their world to advantage. In turn, this altered world presents new kinds of resources that may favor new creature designs, further modifying that world, A strong coupling develops between natural forces, creatures and objects in their environment, creating a web of linked regularities among object properties, driving evolutionary designs (See Appendix 4.) The works of Man provide a most compelling example. We build houses and cities at our scale, with standards for buildings (the size of doors, steps, windows, door-knobs heights, room size, even locations …..) Our "rules of the road" lead to very constrained and predictive traffic patterns, not just for flow, but also for speed, stop-and-go, drop-off locations, and facilities for parking. Regularities also appear at all other space-time scales explored by Man.

2. Allometry

In the realm of biology, the essence underlying Nature's designs have led to striking patterns of structure and form (Thompson, 1912,1968; Huxley, 1932). One of the simplest regularities is scaling, where one object or pattern is a larger version of another. More generally, similarities are affine, and transcend species. The underlying cause (seen by models of classical physics – e.g. allometry) is the effect of force fields such as gravity on materials, leading to elastic transformations that are unequal in radial and axial directions. Consider a common case where cylinder-like objects (limbs) are supporting a large ellipsoidal mass. For example, legs supporting animal bodies, or trunks supporting trees. If the body dimension is proportional to that of the supporting limb, then proportionately thicker limbs are needed to support the greater body masses (McMahon, 1984). Thus, in

Fig 3.1, for one member of each pair of objects, the legs are thinner, implying that they are supporting the *smaller* of the two objects. Over a thousand-fold range, the relation between limb diameter (D) and its length (L) is approximately $D \alpha L^{2/3}$. The recognition of these kinds of regularities is intrinsic knowledge that gives analogical reasoning

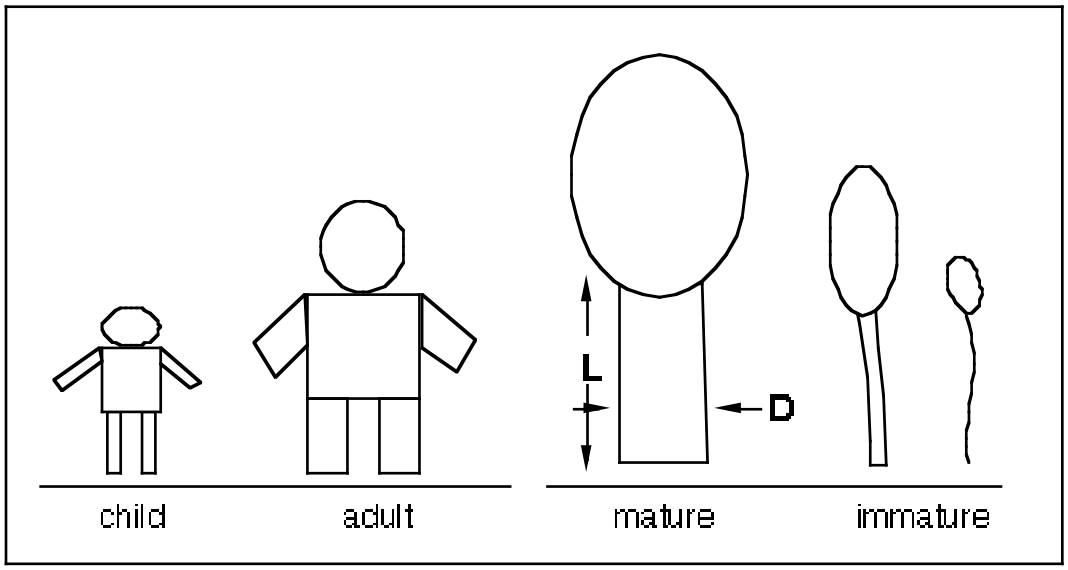

Fig. 3.1: Small trees or children tend to have thinner supporting limbs, with respect to the body size.

its power. Given one tree or mammal that follows Nature's design, the relative dimensions of all such trees or animals of different size immediately become apparent. This is elastic scaling, with the single parameter of interest being a dimensionless number, Young's modulus, or its dynamic, complex equivalent, "internal friction" (Zener, 1948.)

3.2 Cross Modal Correlations

3.2.1 Vision

Observations of shape identify a species. If a creature knows roughly the range of sizes of a species, say mouse or elephant, then the limb to body size-ratios indicate rough size within that species. However, observations of movements of limbs themselves can also provide independent estimates of size. When animals walk or branches sway in the wind, we expect the period of oscillation to be proportional to $L^{1/2}$ because this corresponds to simple pendular movements. To a first approximation, regardless whether the limb is jointed or flexible without joints, the result is similar except for a scale factor (Fig. 3.2.) One can immediately infer the length of a limb (and hence object size) from the period of sway. A dimensionless scaling can be derived if the period is normalized to one's own period of stride. Then, the length of the observed pendulum is the square of the normalized period times one's own limb length. When the limb is a rooted supporting structure, we have an inverted pendulum. However, a similar rule applies, but with an added (dimensionless) scaling constant depending on the stiffness of the material, i.e. Young's modulus.

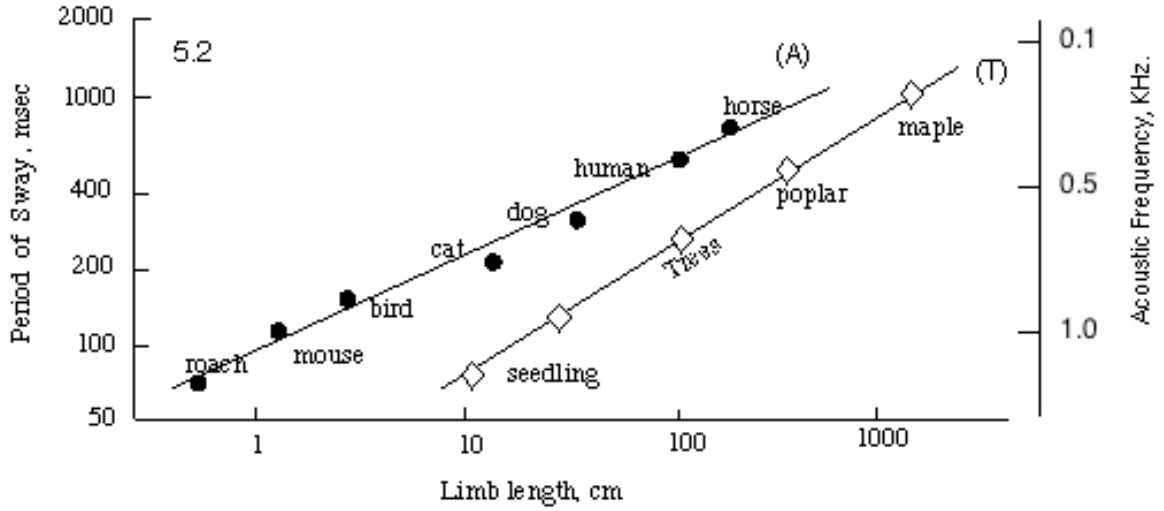

Fig. 3.2: Larger animals (or trees) typically have longer legs (or trunks). Efficient (pendular) movement dictates the period of limb motion. On the right ordinate is a scale showing a related regularity for animals, namely the pitch of their utterances.

3.2.2 Sound

The roar of a lion is more threatening to humans than the squeak of a mouse. The pitch of speech sounds is constrained by the size of vocal tracts, which can be approximated by a cylindrical tube. The longer the tube, the lower the pitch, as shown by the scale on the right ordinate of Fig. 3.2. This is the third regularity associated with mammalian terrestrial animal forms at the human scale. There are many others, such as the sound of footfalls or the tapping or scratching when animals forage. If the sound has a temporal pattern and the object is visible, the period is also obvious in the visual modality. A striking case is human speech, where the lip movements are highly correlated with sound patterns. Indeed, these correlations can be used to categorize the speech sounds, which otherwise could not be partitioned as clearly (Coen, 2006.)

3.3 Dimensionality Reduction

Regularities underlying modal correlations offer an efficient and powerful characterization of observations. Rather than constructing a very high dimensional space from the set of all kinds of perceptual and sensory data, events that qualify as belonging to Natural Modes can be evaluated in a much lower dimensional manifold (Richards & Jepson, 1992; Feldman, 1997.) These manifolds highlight the characteristic properties of the event space, identifying the truly independent parameter set (Saund, 1989; Tenenbaum et al, 2000; Saul & Roweis, 2000) For example, the relations for affine shape similarity illustrated in Fig. 3.1 are tied to the elastic properties of the materials. In passing, we note that these same properties also govern the acoustic ringing and damping of struck cylinders (Wildes & Richards, 1988) further reinforcing the important role of Young's modulus in multimodal correlations.

In the space of types of animal gaits and movements, we can identify another modal property. When creatures move through their environments, ballistic pendular movements are very efficient, with minimal rise and fall of the body. However, aggressive defensive actions may require

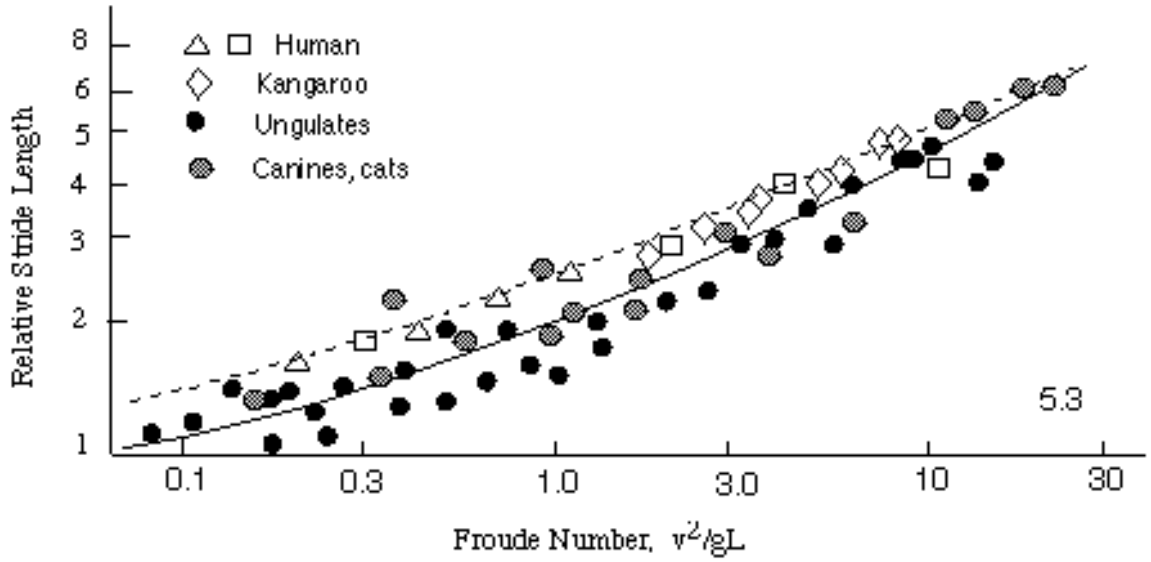

Fig. 3.3 Dynamic Similarity (adapted from Alexander (1988).

movements having greater energy expenditures, such a jumps, or leaps and bounds where the limbs exert large forces against gravity (mgh), and also apply kinetic energy ($mv^2/2$) to generate speedier forward motion. As shown by Alexander (1988), dynamically similar patterns of movements can be characterized by the ratio of these two quantities, namely $v^2/2gh$, which is Froude's number, Both Froude's number and Young's modulus (like Reynolds' number for turbulent flow) are dimensionless. Similar dimensionless numbers can also be found for types of bird flight, or swimming styles of creatures ranging in size from Euglena to Dolphins (Pedley, 1975.)

Although somewhat abstract, dimensionless parameterizations of observations have an elegance noted by others (e.g. Planck, 1906; Wheeler, 1975). In some sense they are "purer", because the mode of observation has been avoided. Like Platonic shapes, dimensionless parameters seem closer to Nature's essence. They break the anthropomorphic bonds of a creature's embodiment.

Modal Co-Evolution

When varieties of creatures interact with each other, and with their environments, an evolutionary process emerges that strives to optimize benefits to all parties (Darwin, 1859) Here we use anigraf constructions to show that achieving a unique optimal state is unlikely because there are a large number of suboptimal fixed points terminal states. Movement from one of these states to another one would presumably need a major catastrophic event (Gould & Eldredge, 1977).

1. Set-up

Objects are modeled by two classes of labeled graphs, which are pairwise constrained in their relationships. One class are triangle-free graphs with at least one, cycle. These will represent the creatures. Let the largest ring in each of these graphs constitute the creature's "body"; limbs will be edges ending in a vertex of degree one (See Fig. 4.1a) The second class of object will be fruit-bearing trees, represented by three types of tree graphs. as shown at the top of the figure. The potential for constructing a triangle by adding only one edge to a tree graph will indicate its fruit-bearing capability. The ability to form the *same* labeled triangle in the triangle-free graphs will show that the creature can nurture and benefit from that particular fruit. Add the further initial constraint that potential triangles can be formed in a creature only if at least one of the identified vertices has degree one. Given this world, what pairs of trees and creatures will be optimal ?

2. Evolution

Consider a creature having 5 vertices (upper left.) Thus, the only possible creature, given our current constraints, is one with a square "body" with one limb segment. This "limb" will form one edge of a potential triangle "hand" that can gather fruit. At first, one might imagine that for this creature, the best design for a fruit-bearing "tree" would be "bush" with one root node and four limbs, each with a vertex of degree one. However, a "tree" with only three free ends of vertex degree one is almost equally optimal (not shown.) This result follows because our matching is with labeled graphs, and the labels on the creature's limb must match the labels on the tree branch before we can test for the same "potential triangle" in both tree and creature. With proper design and labeling of both the tree and creature, there will be 100% certainty in forming one of two possible triangles from vertices with the same labels in each, with no possibility of miss-matched triangles. However, with a 5-vertex "bush", there is the potential for six triangles to be constructed, leading to many triangle constructions that are "miss-matched". In figure 4.1 (upper left) the percent successful matching triangles is shown for the best solution pairs where the probability of a miss-match for the selected trees is nil. These percentages provide a measure of the "fitness" of a particular creature-tree pairing. Each column shows the fitness values for the different trees, together with a constraint on allowable triangle construction. The first column requires the edge added to the tree to be incident to vertices of degree one; the second column requires that the edge forming the triangle in the tree be incident to one vertex of degree two; the third column applies both of these constraints simultaneously, namely the added edge must link vertices of degree one and two. These constraints should be regarded simply as "freezing" some property that relates the creature to the tree (such as a hand is related to the size of a doorknob, or to the diameter of a stick needed for leverage.)

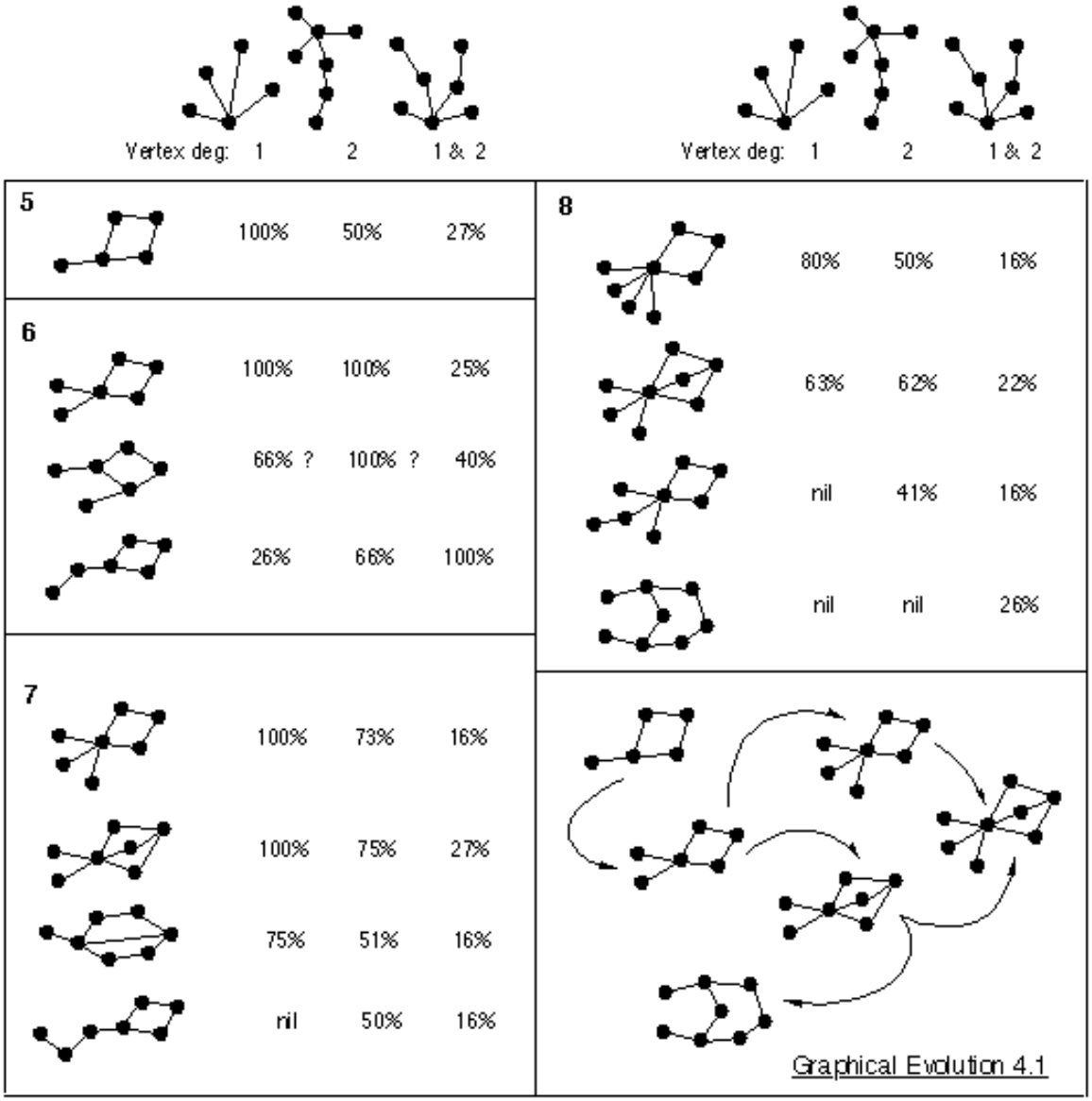

Fig. 4.1: Four species of anigrafs are illustrated in the panels labeled 5 – 8. Three plant forms are shown at the top. The percentages indicate successful match-ups between animals and plants (see text.) The panel at the lower right shows evolutionary paths for the simplest anigraf (5), with the paths ending at two fixed points.

3. Observations

Scanning these percentages, we see that creatures with 5 to 8 vertices (as indicated by the numbers in the upper left corner of each panel), several creatures may have comparable success rates in matching their forms to the three varieties of trees. Although not shown, a similar finding is found for the adaptability of tree to creatures..

Observation 1: Optimizations in creature-tree designs typically have many solutions with roughly equivalent "fitness" measures.

This observation has already been reported by Sims (1994]) in simulations of creatures that locomote. The impact is that when co-evolution occurs within a very large set of possible objects chosen rather haphazardly, it will be difficult to predict which solutions will appear first, potentially "locking out" other solutions. Furthermore, as anigrafs increase in size, the likelihood of solutions that are satisfactory and sufficient for creature's needs will increase. Even with only 10 vertices, there are over 5000 unlabelled triangle-free graphs and over 100 unlabeled trees, or over 500,000 possible pairs. With our set of labeled graphs that require "tighter" matches, finding the best pairing in a random search is almost hopeless. Hence suboptimal solutions are almost surely certain to dominate.

Observation 2: Optimizations in the coevolution of designs that involve multiple parametric constraints applied in succession typically will not reach the same "satisficing" solution obtained if the same constraints were applied in parallel.

This point is about the non-linearity of an evolutionary process. Its impact is that in the evolution of designs, the "fittest" solutions will depend on past developments – i.e. those creatures and resources that have already "locked out" paths to better matches. For the anigraf world, this "lockout" begins almost immediately. In the last panel (lower right) of we show the evolution a creature design that optimizes the simultaneous presence of both degree one and degree two vertices in potential triangle construction. At the second stage of 7-vertex graphs, we have two equally-fit solutions. These in turn lead to two quite different 8-vertex creatures (and trees.) Clearly, as these anigrafs coevolve further, more branching will occur because more "solutions" will have fitness scores that are roughly equivalent. The situation is much akin to the evolution of languages. Although it may be possible to rationalize the specific evolutions, predicting the form that eventually dominates becomes impossible.

As anigraf complexity increases, with further properties of both creatures and resources being introduced, the number of relationships between creatures and their resources should increase, but the strength of their correlations may diminish. This result is suggested by the percentages in the third column of Fig. 4.1: it becomes increasingly harder to avoid extraneous constructions that do not specifically match both members of the pair.

Observation 3: As creature-resource evolution continues, the modal regularities in their domain will become "less sharp", like a web with no clear focal point.

The impact is that a creature's world becomes more and more difficult to predict *precisely*. However, the *form* of a creature's models for its world could still remain accessible. As the creature gains more and more control over its world, the relationships among objects will reflect the creature's models, which in turn are constrained by its design. Modal regularities are reinforced. Although the specifics of these regularities may not be predictable, categorical forms set by the inherent structure of a creature's models can influence the more abstract, categorical nature of the modal web of regularities in the context. In these circumstances, The realm of Mind may be more accessible than that of Nature.