Chapter 1

Scaling in Nervous Networks

Why are biological structures shaped or organized like they are? For example, why is the brain in the head, why is the cortex folded, why are there cortical areas, why are neurons and arteries shaped like they are, and why do animals have as many limbs as they do? Many aspects of morphology can be usefully treated as networks, including all the examples just mentioned. In this chapter I introduce concepts from network theory, or graph theory, and discuss how we can use these ideas to frame questions and discover principles governing brain and body networks.

The first topic concerns certain scaling properties of the large-scale connectivity and neuroanatomy of the entire mammalian neocortical network. The mammalian neocortex changes in many ways from mouse to whale, and these changes appear to be due to certain principles of well-connectedness, along with principles of efficiency (Changizi, 2001b). The neocortical network must scale up in a specific fashion in order to jointly satisfy these principles, leading to the kinds of morphological differences between small and large brains.

As the second topic I consider the manner in which complexity is accommodated in brain and behavior. Do brains use a "universal language" of basic component types from which any function may be built? Or do more complex brains have new kinds of component types from which to build their new functions?

The final topic concerns the nervous system at an even larger scale, dealing with the structure of the nervous system over the entirety of the animal's body. I show that the large-scale shape of animal bodies conforms to a quantitative scaling law relating the animal's number of limbs and the body-to-limb proportion. I explain this law via a selective pressure to minimize the amount of

limb material, including nervous tissue (Changizi, 2001a). That is, because we expect nervous systems to be "optimally wired," and because nervous systems are part and parcel of animal bodies, reaching to the animal's extremities, we accordingly expect—and find—the animal's body itself to be optimally shaped.

One feature connecting the kinds of network on which we concentrate in this chapter is that each appears to economize the material used to build the network: they appear to be volume optimal. It is not a new idea that organism morphology might be arranged so as to require the least amount of tissue volume [see, for example, Murray (1927)], but in recent years this simple idea has been applied in a number of novel ways. There are at least three reasons why optimizing volume may be evolutionarily advantageous for an organism. The first is that tissue is costly to build and maintain, and if an organism can do the same functions with less of it, it will be better off. The second reason, related to the first, is that minimizing tissue volume gives the organism room with which to pack in more functions. The third reason is that minimizing tissue volume will tend to reduce the transmission times between regions of the tissue. These three reasons for volume optimization in organisms are three main reasons for minimizing wire in very large-scale integrated (VLSI) circuit design (e.g., Sherwani, 1995); we might therefore expect organisms to conform to principles of "optimal circuit design" as made rigorous in the computer science fields of graph theory and combinatorial optimization theory (e.g., Cormen et al., 1990). ...and we might have this expectation regardless of the low level mechanisms involved in the system.

Y junctions

The first quantitative application of a volume optimization principle appears to be in Murray (1926b, 1927), who applied it to predict the branching angles of bifurcations in arteries and trees (e.g., aspen, oak, etc.). He derived the optimal branch junction angle (i.e., the angle between the two children) to be

$$\cos\theta = \frac{w_0^2 - w_1^2 - w_2^2}{2w_1 w_2},$$

where w_0, w_1 and w_2 are the cross-sectional areas of the junction's parent and two children. One of the main consequences of this equation is that, for symmetrical bifurcations (i.e., $w_1 = w_2$), the junction angle is at its maximum of 120° when the children have the same cross-sectional area as the parent segment, and is 0° when the children's cross-sectional area is very small. [Actually, in this latter case, the branch angle falls to whatever is the angle between

the source node of the parent and the termination nodes for the two children.] That is, when trunks are the same thickness as branches the branch angle that minimizes the volume of the entire arbor is $120°$. This is very unnatural, however, since real world natural arbors tend to have trunks thicker than branches. And, if you recall your experience with real world natural arbors, you will notice that they rarely have junction angles nearly as high as $120°$; instead, they are smaller, clustering around $60°$ (Cherniak, 1992; Changizi and Cherniak, 2000). *Prima facie*, then, it seems that natural arbors are consistent with volume optimality. Murray also derived the equation for the volume-optimal angle for *each* child segment relative to the parent, and one of the main consequences of this is that the greater the assymmetry between the two children's cross-sectional areas, the more the thinner child will branch at $90°$ from the parent. We find this in natural arbors as well; if there is a segment out of which pokes a branch at nearly a right angle, that branch will be very thin compared to the main arbor segment from which it came. Qualitatively, then, this volume optimality prediction for branch junctions fits the behavior of natural junctions. And it appears to quantitatively fit natural junctions very well too: These ideas have been applied to arterial branchings in Zamir et al. (1983), Zamir et al. (1984), Zamir and Chee (1986), Roy and Woldenberg (1982), Woldenberg and Horsfield (1983, 1986), and Cherniak (1992). Cherniak (1992) applied these concepts to neuron junctions, showing a variety of neuron types to be near volume-optimal; he also provided evidence that neuroglia, Eucalyptus branches and elm tree roots have volume optimal branch junctions. [Zamir (1976, 1978) generalized Murray's results to trifurcations, applying them to arterial junctions.]

Although it is generally difficult to satisfy volume optimality in systems, one of the neat things about this volume optimization for natural branch junctions is that there is a simple physical mechanism that leads to volume optimization. Namely, the equation above is the vector-mechanical equation governing three strings tied together and pulling with weights w_0, w_1 and w_2 [see Varignon (1725) for early such vector mechanical treatments]. If each of the three junction segments pulls on the junction with a force, or tension, proportional to its cross-sectional area, then the angle at vector-mechanical equilibrium is the volume-optimizing angle (Cherniak, 1992, Cherniak et al., 1999). Natural arbors conforming to volume optimality need not, then, be implementing any kind of genetic solution. Rather, volume optimality comes for free from the physics; natural arbors like neurons and arteries self-organize into shapes that are volume optimal (see, e.g., Thompson, 1992). In support of this,

many non-living trees appear to optimize volume just as well as living trees, from electric discharges (Cherniak, 1992) to rivers and deltas (Cherniak, 1992; Cherniak et al., 1999).

In addition to this physics mechanism being advantageous for a network to have minimal volume, natural selection may favor networks whose average path between root and leaves in the network is small—shortest-path trees—and one may wonder the degree to which this mechanism simultaneously leads to shortest-path trees. Such shortest-path trees are not necessarily consistent with volume optimization (Alpert et al., 1995; Khuller et al., 1995), but the near-volume-optimal natural trees tend to come close to minimizing the average path from the root. Considering just a y-junction, the shortest path tree is the one which sends two branches straight from the root; i.e., the junction occurs at the root. An upper bound on how poorly a volume-optimal junction performs with respect to the shortest-path tree can be obtained by considering the case where (i) the root and two branch terminations are at the three vertices of an equilateral triangle with side of unit length, and (ii) the volume per unit length (i.e., cross-sectional area) is the same in all three segments. The volume-optimal branch junction angle is 120° and occurs at the center of mass of the triangle. The distance from the root to one of the branches along this volume-optimal path can be determined by simple geometry to be 1.1547, or about 15% greater than the distance in the shortest-path tree (which is along one of the unit-length edges of the triangle). This percentage is lower if the relative locations of the root and branch terminations are not at the three vertices of an equilateral triangle, or if the volume per unit length of the trunk is greater than that of the branches (in which case the junction point is closer to the root). In sum, natural selection has stumbled upon a simple vector-mechanical mechanism with which it can simultaneously obtain near volume-optimal and near-shortest-path trees.

Multi-junction trees

The applications of a "save volume" rule mentioned thus far were for single branch junctions. Kamiya and Togawa (1972) were the first to extend such applications to trees with multiple junctions, finding the locally optimal multi-junction tree for a dog mesenteric artery to be qualitatively similar to the actual tree. Schreiner and Buxbaum (1993), Schreiner et al. (1994) and Schreiner et al. (1996) constructed computer models of large vascular networks with realistic morphology by iteratively adding locally volume-optimal y-junctions.

Traverso et al. (1992) were the first to consider modeling natural arbors, neural arbors in particular, with multiple junctions using the concept of a Steiner tree (Gilbert and Pollak, 1968) from computer science, which is the problem of finding the length-minimal tree connecting n points in space, where internodal junctions are allowed (i.e., wires may split at places besides nodes). Branch junctions in Steiner trees have angles of 120°, and Traverso et al. (1992) found that some sensitive and sympathetic neurons in culture have approximately this angle.

Most neurons and other natural arbors, however, have branch junction angles nearer to 60° or 70° (Cherniak, 1992; Changizi and Cherniak, 2000); the Steiner tree model is inadequate because it assumes that trunks have the same volume cost per unit length (i.e., same cross-sectional area) as branches, when it is, on the contrary, almost always the case that trunks are thicker than branches. To determine if natural trees have volume-optimal geometries a generalized notion of Steiner tree is needed, one that allows trunks to be thicker than branches. Professor Christopher Cherniak and myself invented such a notion and showed that axon and dendrite trees (Cherniak et al., 1999), coronary arterial trees (Changizi and Cherniak, 2000) and Beech trees (Cherniak and Changizi, unpublished data) optimize volume within around 5%, whereas they are around 15% from surface area optimality and around 30% from wire-length optimality. [The average values for the unpublished tree data are from eight 4-leaf Beech tree arbors, and are 4.63% ($\pm 3.21\%$) from volume optimality, 10.94% ($\pm 6.56\%$) from surface area optimality, and 26.05% ($\pm 12.92\%$) from wire length optimality (see Cherniak et al., 1999, for methods).]

The studies mentioned above concentrated on the morphology of individual natural trees, e.g., individual neurons. We may move upward from individual neurons and ask, How economically wired are whole nervous systems? This has been asked and answered in a variety of ways.

Larger scales

At the largest scale in the nervous system, Cherniak (1994, 1995) showed that animals with brains making more anterior connections than posterior connections should have, in order to minimize volume, the brain placed as far forward as possible; this explains why the brain is in the head for vertebrates and many invertebrates. Radially symmetric animals, on the other hand, are expected to have a more distributed neural network, as is the case (e.g., Brusca and Brusca, 1990, p. 87). In what is to date the most stunning conformance

to volume optimality, Cherniak (1994, 1995) showed that, of approximately forty million possible positions of the ganglia in *Caenorhabditis elegans*, the actual placement of the ganglia is the wire-optimal one. He also provides statistical evidence that the placement of cortical areas in cat (visual cortex) and rat (olfactory cortex) are consistent with the hypothesis that the total length of area-to-area connections is minimized.

Ruppin et al. (1993) show that each of the following facts about the brain decrease the overall required volume of the brain: (i) that gray matter is separated from white matter, (ii) that gray matter is a shell on the surface of the brain with white matter in the center (rather than vice versa), and (iii) that the gray matter has convolutions. Van Essen (1997) also argues that the convolutions of the cortex are a fingerprint of low wiring. Wire-minimization has also been used in a number of ways to explain local connectivity patterns in the visual cortex (e.g., stripes, blobs or patches) (Durbin and Mitchison, 1990; Mitchison, 1991, 1992; Goodhill et al., 1997; Chklovskii, 2000; Chklovskii and Koulakov, 2000; Koulakov and Chklovskii, 2001).

Well-connectedness

As we will see later, the neocortical network not only reveals principles of volume optimization, it also reveals principles of well-connectedness, where by that I refer, intuitively, to properties of the network which bear on how "close," in some sense, vertices are to one another. One way to measure neuron-interconnectedness is via the average percent neuron-interconnectedness of neurons, where a given percent neuron-interconnectedness of a neuron is the percentage of all neurons to which it connects. It has been recognized, however, that it is prohibitive to maintain an invariant average percent neuron-interconnectedness as the network size is scaled up (Deacon, 1990; Stevens, 1989; Ringo, 1991), because this requires that the average degree of a vertex [the *degree* of a vertex is the number of edges at a vertex] scales up proportionally with network size, and thus the total number of edges in the network scales as the square of network size. Average percent neuron-interconnectedness is an overly strong notion of neural interconnectedness, however. A weaker measure might be the characteristic path length, which I will call the *network diameter*, which is defined as, over all pairs of neurons, the average number of "edges" (i.e., axons) along the shortest path connecting the pair. Intuitively, network diameter measures how close—in terms of connectivity—neurons are to one another, on average.

For most random networks the diameter is "low." [In a *random network* each pair of nodes has the same probability of having an edge between them.] In particular, the network diameter is approximately $(\log N)/(\log \delta)$ (Bollobás, 1985, p. 233), where N is the network size and δ the average degree. [This also requires assuming $N >> \delta >> \log N >> 1$.] While keeping the average percent neuron-interconnectedness invariant requires that degree (the number of neurons to which a neuron connects) scale proportionally with network size, the network diameter can remain invariant even if the degree grows disproportionately slowly compared to the network size. Suppose, for example, that $N \sim \delta^b$, where $b > 1$; that is, the degree grows disproportionately slowly as a function of network size, but as a power law. [A *power law* is an equation of the form $y = ax^b$, where a and b are constants.] Then in a random network the diameter is approximately

$$\frac{\log(C\delta^b)}{\log \delta} = b + \frac{\log C}{\log \delta},$$

where C is a proportionality constant. In the limit of large brains, the network diameter is invariant, approaching b, even though the degree is scaling disproportionately slowly.

This observation is not particularly useful for biological systems, however, because few biological networks are random. However, in what has now become a seminal paper, Watts and Strogatz (1998) noticed that highly ordered networks can have random-network-like diameter by merely adding a small number of "shortcuts" connecting disparate regions of the network; they named such networks *small world networks*. A firestorm of papers have since been published showing that many natural networks appear to be small world, including nervous systems and social networks. Nature has, indeed, hit upon one useful kind of network, allowing highly ordered clustering, and yet low network diameter. With these ideas in mind, we will be able to see that the mammalian neocortex has low and invariant network diameter, and, in fact, it is approximately 2.

1.1 The mammalian neocortex

Mice, men and whales have brains, and as they are all mammals, their cortex mostly consists of something called the *neocortex*, which is found only in mammals. When researchers talk of the cortex, they are almost always referring to the neocortex. This is the part of the mammalian brain that excites

most researchers, as it appears to be the principal anatomical feature separating mammalian brains from other vertebrate brains; it is where the key to mammalian intelligence will be found. The neocortex consists of gray matter which lies on the surface of the brain, and white matter which is in the interior. The gray matter consists of neurons of many types, synapses, and glial cells. The white matter consists of axons which reach from neurons in one part of the gray matter to make synapses with neurons in another part of the gray matter. The neocortical gray matter is characterized by what appears to be layers as one moves from the surface of the gray matter inward toward the center, and six layers are usually distinguished. There are also distinct cortical areas at different locations on the cortical sheet, inter-area connections being made primarily via white matter connections, and intra-area connections being made primarily by local connections not traveling into the white matter.

Although all mammals have a neocortex, many of the basic properties of the neocortex change radically from small brains to large brains. The changes are so dramatic that one might justifiably wonder whether the neocortex in mouse and whale are really the same kind of organ. And if they are the same kind of organ, then why the radical changes? Are there some underlying properties that *are* being kept constant—these properties being the key ones, the ones that really define the neocortex—and the properties that change are changing for the purpose of keeping these key properties invariant? It is these questions we examine in this section. I will describe the ways in which the neocortex changes as it increases in size, and describe a theory (Changizi, 2001b) that aims to explain what the central features of the neocortex are, such that all these other properties must change as they do.

To understand how the neocortical network scales up in larger brains, we need to understand the notion of a *power law*. Power laws are of the form $y = ax^b$, where a and b are constants. a is a proportionality constant, and it often depends on the specific nature of the studied systems. For example, the volume of a cube is $V = D^3$, where D is the diameter; i.e., $a = 1$ and $b = 3$. The volume of a sphere, however, is $V = (4/3)\pi(D/2)^3$, which is $V = (\pi/6)D^3$; i.e., $a = \pi/6 \approx 0.5$ and $b = 3$. The volume of both cubes and spheres scale as the diameter cubed, but the proportionality constants are different. So long as the geometrical shape is similar, the proportionality constant will not change, and it is often appropriate to ignore it, writing $y \sim x^b$ to mean that y is proportional to x^b. Power laws are particularly appropriate for neocortical scaling [and in biology more generally (see Calder, 1996; Schmidt-Nielson, 1984)]. It turns out that many of the properties of the neocortex scale

Table 1.1: Measured scaling exponents for neocortical variables against gray matter volume V_{gray}. The measured exponents are in most cases acquired from scaling data against brain volume. To obtain exponents against V_{gray}, I have assumed that V_{gray} is proportional to brain volume. This proportionality is empirically justified, as measured exponents for V_{gray} to brain volume are near one: 0.983 *(Prothero, 1997a),* 0.982 *(Hofman, 1991),* 1.054 *(Hofman, 1989),* 1.04 *(Prothero and Sundsten, 1984),* 1.06 *(Frahm et al., 1982) and* 1.08 *(Jerison, 1982).*

Variable description	Variable	Measured exponent	References
# areas to which an area connects	D	0.30	Here
Neuron number	N	0.62	Jerison, 1973
		0.67	Passingham, 1973
Neuron density	ρ_{neuron}	-0.312	Prothero, 1997b
		-0.28	Prothero, 1997b
		-0.28	Tower, 1954
		-0.32	Tower, 1954
Number of areas	A	0.40	Changizi, 2001b
Thickness	T	0.092	Prothero, 1997a
		0.115	Prothero, 1997a
		0.129	Hofman, 1991
		0.197	Hofman, 1989
		0.08	Prothero and Sundsten, 1984
		0.17	Jerison, 1982
Total surface area	S	0.905	Prothero, 1997a
		0.893	Prothero, 1997a
		0.922	Prothero, 1997a
		0.901	Hofman, 1991
		0.899	Hofman, 1989
		0.89	Hofman, 1985
		0.91	Prothero and Sundsten, 1984
		0.91	Jerison, 1982
Module diameter	m	0.135	Manger et al, 1998
Soma radius	R_0	0.10	Changizi, 2001b
Axon radius	R_1	0.105	Shultz and Wang, 2001
Volume of white matter	V_{white}	1.318	Allman, 1999
		0.985	Prothero, 1997b
		1.28	Hofman, 1991
		1.37	Hofman, 1989
		1.31	Frahm et al., 1982

against gray matter volume as a power law. That is, if Y is the property of interest and V_{gray} is the gray matter volume, then it has been empirically found that $Y = aV_{gray}^b$ for some constants a and b. Ignoring the proportionality constant, we say that $Y \sim V_{gray}^b$. When we say how a neocortical quantity scales up, we can, then, just report the scaling exponent for it against gray matter volume. Table 1.1 shows the scaling exponents measured thus far for neocortex, and which I will explain below. I only show plots here if they have not yet appeared elsewhere in the literature (and this is only for the number of areas to which an area connects, and for module diameter).

Before presenting a theory of neocortical scaling, I begin by making some simplifying assumptions about the neocortical network. Because about 85% of neocortical neurons are pyramidal cells (Schüz, 1998) and only pyramidal cells appear to significantly change in degree of arborization from mouse to whale (Deacon, 1990), it is changes to pyramidal cells that must account for the decreasing neuron density. Accordingly, I will idealize the neocortical network to consist only of pyramidal neurons. Also, because most (over 90%) of the neocortical connections are from one part of neocortex to another (Braitenberg, 1978), the other neocortical connections are probably not the principal drivers of neocortical scaling; I will therefore concentrate on the cortico-cortical connections only, and I will assume that a single pyramidal neuron's axon can innervate only one area.

There are multiple principles shaping the neocortex, and we will see that the exponents are not all due to the same underlying explanation. There are, in fact, three central principles, and they are

- Economical well-connectedness.
- Invariant computational units.
- Efficient neural branching diameters.

The exponents each of these principles explains are shown in Table 1.2. I will take up each principle in turn, and derive the exponents which follow from it.

1.1.1 Economical well-connectedness

The principles

Consider that an area in the gray matter connects to other areas (i.e., it has neurons connecting to neurons in other areas). The fraction of the total number of areas to which it connects is called the *percent area-interconnectedness*. It

Table 1.2: The exponent predicted by my theory, along with the approximate value for the measured exponent. The exponents are partitioned into three groups, each which is explained by the principle stated above it.

Variable description	Variable name	Measured exponent	Predicted exponent
Economical well-connectedness ⇒			
- Number of areas to which an area connects	D	0.30	1/3≈0.33
- Neuron number	N	0.65	2/3≈0.66
- Neuron density	ρ_{neuron}	-0.3	-1/3≈-0.33
- Number of areas	A	0.40	1/3≈0.33
Invariant computational units ⇒			
- Thickness	T	0.13	1/9≈0.11
- Total surface area	S	0.90	8/9≈0.89
- Module diameter	m	0.135	1/9≈0.11
Efficient neural branching diameters ⇒			
- Soma radius	R_0	0.10	1/9≈0.11
- Axon radius	R_1	0.105	1/9≈0.11
- Volume of white matter	V_{white}	1.3	4/3≈1.33

seems a plausible and natural hypothesis that, for an area's efforts to be useful, it must make its results known to an invariant percentage of the areas in the neocortex. That is, suppose that for a mouse brain to work, each area must talk to about one tenth of the areas. Then, so the idea goes, in a whale brain each area must also connect to one tenth of the areas. Whether the percentage is one tenth or one half I do not know; the claim is only that what is good for the mouse in this regard is also good for the whale. This is recorded as the

> *Principle of Invariant Area-Interconnectedness,* which states that the average percent area-interconnectedness remains roughly constant no matter the total number of areas. (See Figure 1.1.)

There is some direct evidence for this hypothesis. From it we expect that the total number of area-to-area connections E should scale as the number of areas A squared; or $A \sim E^{1/2}$. Data exist for only two species—cat and macaque—but we may use disjoint proper subsets of each animal's neocortex as distinct data points. Figure 1.2 shows these data, where the relationship fits $A \sim E^{0.45}$, or closely fitting Hypothesis 1.

Moving to the second invariance principle associated with well-connectedness, areas are composed of many neurons, and thus a connection from one

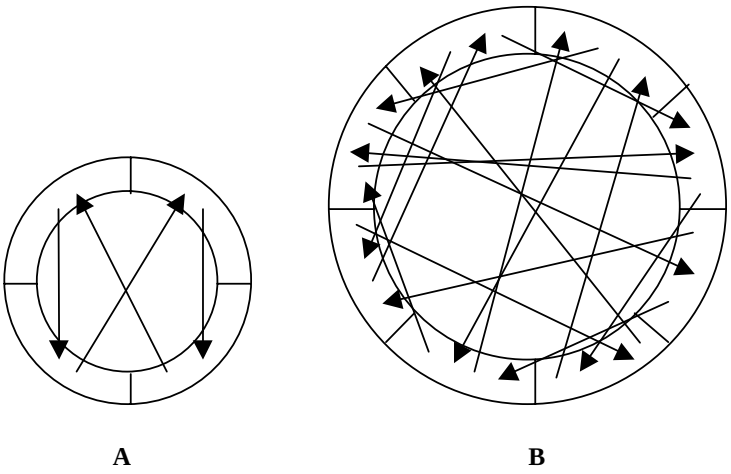


Figure 1.1: Illustration of invariant percent area-interconnectedness. The average percent area-interconnectedness in a small and large neocortex. The outer part of each ring depicts the gray matter, the inner part the white matter. Each neocortex has multiple areas. (a) Each of the four areas in this small neocortex connects to one other area. The average percent area-interconnectedness is thus $1/4$*. (b) Each of the eight areas in this large neocortex connects to two other areas. The average percent area-interconnectedness is thus* $2/8 = 1/4$*, the same as for the small brain.*

area to another is always from a neuron in the first area to a certain percentage of the neurons in the second area. We might call this percentage the *percent area-infiltration*. It is, again, natural and plausible to hypothesize that when an area tells another area about its efforts, it must tell a certain invariant percentage of the neurons in the area in order for the area to understand and appropriately respond to the information. That is, if white matter axons in mouse connect to roughly, say, one tenth of the number of neurons in an area, then, so the idea goes, in a whale brain each such neuron connects to one tenth of the neurons in an area. We record this as the

> *Principle of Invariant Area-Infiltration*, which states that, no matter the neocortical gray matter volume, the average percent area-infiltration stays roughly the same. (See Figure 1.3.)

I know of no data directly confirming this principle. It is here a hypothesis only, and it will stand or fall to the extent that it is economically able to account for the observed scaling exponents.

The two above invariance principles (invariant percent area-interconnectedness and invariant percent area-infiltration) concern the degree of well-connectedness of the neocortex, and we might summarize the pair of above principles by a single principle labeled the *Principle of Invariant Well-Connectedness*.

We have not yet made use of the idea of economical wiring, but I mentioned much earlier that the neocortical network appears to be driven, in part, by economical wiring. . . which leads us to the next principle. All things equal, it is advantageous for a nervous system to use less neural wiring, and as we saw at the start of the chapter many aspects of neuroanatomy and structural organization have been found to be consistent with such wire-optimization hypotheses. With this in mind we might expect that the neocortex would satisfy the Principle of Invariant Well-Connectedness, but that it would do so in a fashion sensitive to the connection costs. In particular, we would expect that the average number of neurons to which a neuron's axon connects—the *average neuron degree*, δ—will not be much greater than that needed to satisfy invariant well-connectedness. The reason for this is as follows: Connecting to more neurons requires a greater number of synapses per neuron, and this, in turn, requires greater arborization—more wire. In terms of scaling, this save-wire expectation can be weakened to the expectation that average neuron degree scales no faster than needed to satisfy invariant well-connectedness. I record this third principle as the

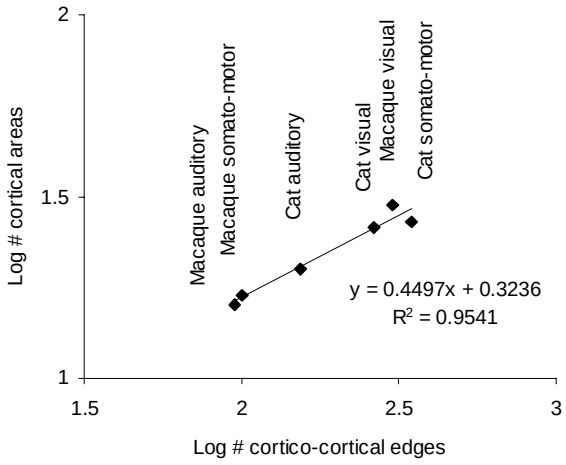


Figure 1.2: Logarithm (base 10) of the number of cortical areas versus logarithm of the number of area-to-area connections, for disjoint proper subnetworks within cat and macaque. Data points are as follows. Cat visual ($A = 26$, $E = 264$), cat auditory ($A = 20$, $E = 153$), and cat somato-motor ($A = 27$, $E = 348$) are from Scannell and Young (1993). Macaque visual ($A = 30$, $E = 300$), macaque auditory ($A = 16$, $E = 95$), and macaque somato-motor ($A = 17$, $E = 100$) are from Young (1993).

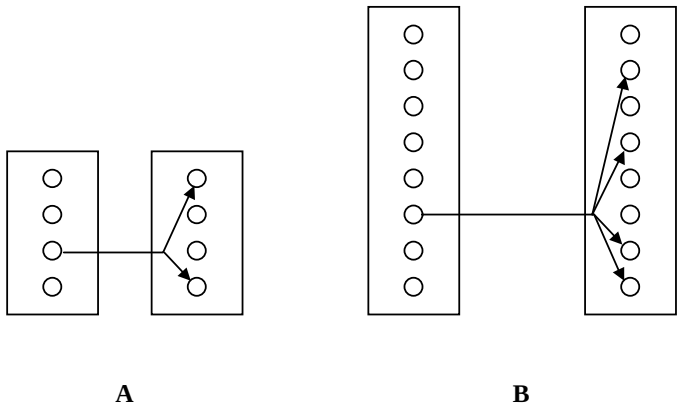


Figure 1.3: Illustration of the invariance of percent area-infiltration. The average percent area-infiltration for small and large areas. Each rectangle depicts an area, and each small circle a pyramidal neuron. (a) Each of these two areas has four neurons, and the left area connects via a pyramidal axon to two neurons in the right area. The percent area-infiltration is $2/4 = 1/2$*. The other neurons' connections are not shown. (b) Each of the two areas has eight neurons, and the left area connects to four neurons in the right area. The percent area-infiltration is* $4/8 = 1/2$*, the same as for the small area.*

> *Principle of Economical Wiring*, which states that the average neuron degree scales as slowly as possible consistent with an invariant well-connectedness.

Informally, the conjunction of these three above principles says that, no matter the neocortex size, an area talks to a fixed fraction of all the areas, and when an area talks to another area it informs a fixed fraction of the neurons in the area; furthermore, this is done in a volume-optimal manner. I will call the conjunction of these principles the *Principle of Economical Well-Connectedness*.

Scaling exponents derived from economical well-connectedness

Now let us consider the consequences of this principle. There are a few symbols we will need. First, recall that δ is the average neuron degree, defined as the average number of neurons to which a neuron's axon connects. Let A denote the total number of cortical areas, D the average number of areas to

which an area connects, and W be the average number of neurons in an area. The first invariance principle stated that the percent area-interconnectedness is invariant, and this means that D/A is invariant, i.e., $D \sim A$. The second invariance principle stated that the percent area-infiltration is invariant, and this means that δ/W is invariant, i.e., $\delta \sim W$. Since an area connects to D areas and each neuron in an area can connect to neurons in only one area, there must be at least D neurons in an area; i.e., $W \geq D$. The Principle of Economical Wiring stated that δ must scale up as slowly as possible given the other constraints. Since we have already seen that $\delta \sim W$, economical wiring therefore implies that W must scale up as slowly as possible given the other constraints. Since we already saw that $W \geq D$, we can now also say that W scales no faster than D, and thus that $W \sim D$. To sum up for a moment, we have now concluded the following proportionalities:

$$A \sim D, D \sim W, W \sim \delta.$$

By the transitivity of proportionality, it follows that all these are proportional to one another; i.e.,

$$A \sim D \sim W \sim \delta.$$

This will be useful in a moment. Now notice that the total number of neurons N is proportional to the number of areas times the number of neurons per area. That is, $N \sim AW$. But we already have seen that $A \sim W$, and thus $N \sim A^2$, and so $A \sim N^{1/2}$. In fact, it follows that all those four mutually proportional quantities are proportional to $N^{1/2}$. That is,

$$A \sim D \sim W \sim \delta \sim N^{1/2}.$$

In particular, we will especially want to remember that

$$\delta \sim N^{1/2}.$$

I have just related δ to N, and in this paragraph I will relate δ to V_{gray}. With both of these relationships we will then be able to say how N and V_{gray} relate. A combination of empirical and theoretical argument suggests that, to a good approximation, a pyramidal neuron connects to almost as many different neurons as its number of axon synapses (Schüz, 1998). We do not here need to be committed to a claim as strong as this. Instead, all we need is that the neural degree δ is proportional to the number of axon synapses per neuron s. [I am here using the fact that the number of axon synapses scales identically with

the total number of synapses per neuron. This must be true since every axon synapse is someone else's dendrite synapse.] Well, since

$$s \sim \frac{\rho_{synapse} V_{gray}}{N},$$

it follows that

$$\delta \sim \frac{\rho_{synapse} V_{gray}}{N}.$$

We may use the fact that $\rho_{synapse}$ is invariant [see Abeles, 1991, and also Changizi, 2001b for an explanation of this that is not presented here] to then say that

$$\delta \sim \frac{V_{gray}}{N}.$$

Thus far, we have seen that $\delta \sim N^{1/2}$, *and* we have seen that $\delta \sim V_{gray}/N$. But then it follows that

$$N^{1/2} \sim V_{gray}/N.$$

Solving for N we can finally conclude that

$$N \sim V_{gray}^{2/3}.$$

It also, of course, immediately follows, that

$$\rho \sim V_{gray}^{-1/3}.$$

And, since $D \sim A \sim N^{1/2}$, we also conclude that

$$D \sim V_{gray}^{1/3}$$

and

$$A \sim V_{gray}^{1/3}.$$

These scaling relationships are very close to the measured ones in Table 1.1. The number of cortical areas increases in bigger brains, then, not because of some kind of pressure to have more specialized areas, but because by not increasing the number of areas the network would become decreasingly well-connected, or would no longer be economically wired. [There are other theories hypothesizing that cortical areas may be due to issues of economical wiring, including Durbin and Mitchison (1990), Mitchison (1991, 1992), Ringo (1991), Jacobs and Jordan (1992) and Ringo et al. (1994).] Note that this theory also predicts that the number of neurons in an area, W, scales with gray matter volume with exponent $1/3$.

Invariant network diameter of 2

I will now show that the Principle of Invariant Well-Connectedness has the remarkable consequence that the neocortical network has an invariant network diameter of around 2. (See start of this chapter for the definition of network diameter.) How may we compute its network diameter? The neocortex is certainly not a random network, so we cannot straightforwardly use the network diameter approximation for random networks discussed in the introduction of this chapter. But recall the notion of a small world network introduced there: because pyramidal neurons usually make long range connections, or "shortcuts," via the white matter, the neocortical network is almost surely a small world network, and thus would have a network diameter nearly as low as that for a random network, namely approximately $\log(N)/\log(\delta)$. [This also requires that $N >> \delta >> \log N >> 1$. For the mammalian neocortex this is true; $N \approx 10^7$ to 10^{11}, $\delta \approx 10^4$ to 10^5 (Schüz, 1998).] The scaling results described thus far have informed us that $N \sim \delta^2$. The network diameter is, then,

$$\Gamma \approx \frac{\log(C\delta^2)}{\log \delta} = 2 + \frac{\log C}{\log \delta},$$

where C is a proportionality constant. That is, for sufficiently large V_{gray}, the neuron degree δ becomes large and thus the network diameter Γ approaches 2; in the limit there are on average only two edges—one neuron—separating any pair of neurons. A rough estimate of the constant C can be obtained by comparing actual values of the neuron number N and the average neuron degree δ. For a mouse, $N \approx 2 \cdot 10^7$ and $\delta \approx 8,000$ (Schüz, 1998), so the constant $C \approx N/\delta^2 = 0.3$. Common estimates for human are around $N \approx 10^{10}$ and $\delta \approx 50,000$ (Abeles, 1991), making the constant $C \approx 4$. What is important here is that these estimates of C (i) are on the order of 1, and (ii) are well below the estimates of δ. Thus $(\log C)/(\log \delta) \approx 0$ and the network diameter is approximately 2. As a point of comparison, note that the network diameter for *C. Elegans*—the only nervous system for which the network diameter has been explicitly measured—is 2.65 (Watts and Strogatz, 1998); its network diameter computed via the random network approximation is 2.16. This suggests the conjecture that a network diameter around 2 is a feature common to all central nervous systems.

1.1.2 Invariant computational units

To derive the scaling exponent for the thickness of the gray matter sheet and the total surface area, it suffices to note another invariance principle to which the neocortex appears to conform. It is the

> *Principle of Invariant Minicolumns*, which states that the number of neurons in a "minicolumn"—which is a neuroanatomical structure lying along a line through the thickness of the gray matter, from pia to white matter—is invariant. (An invariance principle of this form was first put forth by Prothero (1997a).)

The motivation is that, independent of brain size, these minicolumns are fundamental computational units, and that more "computational power" is achieved by increasing the number of such units, not by changing the nature of the fundamental computational units themselves. Evidence exists for this invariance from Rockel et al. (1980). [Rockel et al. (1980) mistakenly concluded that the surface density was invariant, but the latter could only be concluded if the number of neurons under, say, a square millimeter of surface was invariant. This, however, is not the case (Haug, 1987). See Prothero (1997b) for a cogent resolution to this issue.] The line density along a line from pia (the outside boundary of the neocortex) to white matter (the inside boundary), λ, scales as

$$\lambda \sim \rho^{1/3} \sim (V_{gray}^{-1/3})^{1/3} = V_{gray}^{-1/9}.$$

Since the number of neurons along this line is invariant, the sheet must be thickening, namely

$$T \sim V_{gray}^{1/9}.$$

It follows immediately that

$$S \sim V_{gray}^{1-1/9} = V_{gray}^{8/9}.$$

These are very close to the measured exponents, as shown in Table 1.1. The gray matter surface area scales more quickly than $V_{gray}^{2/3}$, then—and thus becomes convoluted—for two principal reasons. First, it is because the neuron density is decreasing—and this, in turn, is because the number of synapses per neuron is increasing in order to economically maintain satisfaction of the Principle of Economical Well-Connectedness. Second, it is because the pia-to-white-matter structure of the cortex remains the same (e.g., same number of neurons in a minicolumn, same number of layers) across mammals. If, instead, the number of neurons along a thin line through the cortical sheet increased in

larger brains, the new neurons would not have to spread only along the surface, but could spread into deeper regions of the gray matter; the surface would then not have to scale up so quickly. This, however, would require modifying the basic, uniform structure of the gray matter every time the brain was enlarged; it would demand inventing new basic computational architectures in each brain, whereas by keeping the structure the same, larger brains can work with the same "primitive computational units" as in smaller brains.

A related issue concerns modules found in the neocortex, such as columns, blobs, bands, barrels and clusters. They are intermediate features, smaller than cortical areas, and larger than minicolumns. The simplest hypothesis is that modules conform to the following invariance principle,

> *Principle of Invariant Modules*, which states that the number of minicolumns in a module is invariant.

The motivation is similar to that for the Principle of Invariant Minicolumns. If this principle holds for neocortex, then from the neuron density decrease it follows that the diameter of a module (when measured along the cortical surface), m, should scale as $V_{gray}^{1/9}$. Manger et al. (1998) measured module size across mammals varying over four orders of magnitude in brain size, and one may compute from their data that the exponent is 0.135 (see Figure 1.4), or very close to the predicted $1/9$. The number of neurons in a module therefore appears to be independent of brain size.

1.1.3 Efficient neural branching diameters

Murray's Law

As far as I know, every kind of tree in nature has thicker trunks when the trunk supports more leaves (Cherniak, 1992; Cherniak et al., 1999; Changizi and Cherniak, 2000). The same is therefore expected of neurons—the soma being the ultimate trunk of a neuron—and certainly appears to be the case (e.g., Cherniak, 1992; Cherniak et al., 1999). But in exactly what quantitative manner do we expect trunk diameter to scale as a function of the number of leaves in the tree? Many natural trees conform to a relationship called Murray's Law (1926a), which says that, for any two depths, i and j, in a tree, the sum of the cubes of the diameters at depth i is identical to the sum of the cubes of the diameters at depth j. So, for example, the cube of a trunk diameter must equal the sum of the cubes of its daughter segment diameters. Murray's Law is expected to apply for any tree where (i) there is laminar fluid flow, and (ii)

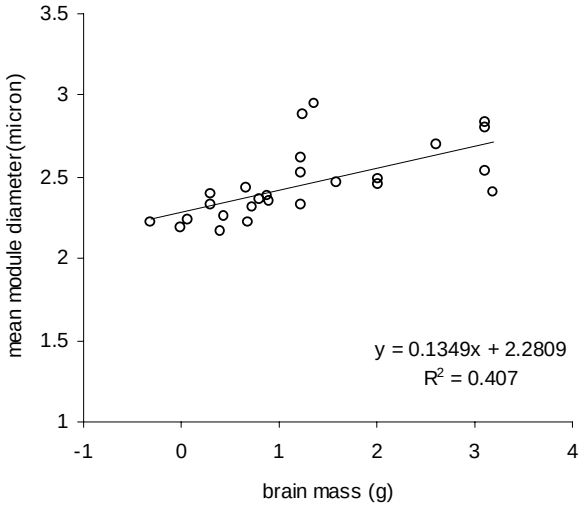


Figure 1.4: Logarithm (base 10) of the mean module diameter versus logarithm of brain size. Data from Manger et al. (1998).

the power required to distribute the fluid is minimized. In fact, it is well known that there is fluid flow in neural arbors (Lasek, 1988), and that the fluid flow is laminar follows from the facts that fluid flow in pipes of diameter less than one millimeter tends to be laminar (Streeter and Wylie, 1985) and that neural arbors have diameters on the micron scale. Murray's Law, in fact, appears to apply to neural trees, as shown in Cherniak et al. (1999). I record this principle as the

> *Principle of Efficient Neural Branching Diameters*, which states that neural segment diameters are set so as to maximize power efficiency.

Soma and axon radius

From this principle—i.e., from Murray's Law—we may derive the expected scaling relationship between trunk diameter, t, and the number of leaves in the tree. Murray's Law states that the trunk diameter, t, cubed should be the same value as the sum of the cubes of all the diameters of the leaf segments in the tree. Let s be the number of leaves in the tree, and d be the diameter of each

leaf segment. Then the relationship is,

$$t^3 = sd^3.$$

Given that the leaf segments in neurons—i.e., synapses—do not vary in size as a function of brain size, we may conclude that

$$t^3 \sim s.$$

[West et al., 1997, use elaborations on ideas like this to derive metabolic scaling exponents. Their arguments require space-filling, fractal-like networks, whereas the argument here does not require this. Murray himself back in 1927 might well have noted this scaling feature.] That is, trunk diameter—whether we treat the soma or the source axon as the trunk—of a neuron scales as the 1/3 power of the number of synapses in the neuron. From earlier developments we know that $s \sim V_{gray}^{1/3}$, and thus we may derive that

$$R \sim V_{gray}^{1/9},$$

where I am now using R for trunk radius rather than trunk diameter (since they are proportional). Measured scaling relationships conform well to this, for both soma (or neuron body) radius and for the radius of a white matter axon (see Table 1.1). [Note that Murray's Law states that $t^3 = b_1^3 + b_2^3$, where b_1 and b_2 are the two daughter branch diameters of a trunk with diameter t, and thus, in general, trunk diameter $t \approx 2b$, where b is the average daughter diameter, and thus $t \sim b$. This is why it is justified to treat soma and axon radius as proportional.]

White matter volume

Finally, there is the issue of white matter volume, V_{white}. White matter volume is composed entirely of myelinated (and some unmyelinated) axons from pyramidal neurons sending cortico-cortical connections. Thus, white matter volume is equal to the total number of white matter axons, $N_{whiteaxon}$, times the volume of a white matter axon, $V_{whiteaxon}$. That is,

$$V_{white} = N_{whiteaxon} V_{whiteaxon}.$$

All we need to do is to figure out how these two quantities scale with gray matter volume.

There must be one white matter axon for every neuron, and thus $N_{whiteaxon} \sim N$, and so $N_{whiteaxon} \sim V_{gray}^{2/3}$. The volume of a white matter axon, $V_{whiteaxon}$, is proportional to the length, L, of the axon times the square of its radius, R_1; i.e.,

$$V_{whiteaxon} \sim LR_1^2.$$

White matter axons travel roughly a distance proportional to the diameter of the white matter, and so $L \sim V_{white}^{1/3}$. Also, we saw just above that $R_1 \sim V_{gray}^{1/9}$. Thus,

$$V_{whiteaxon} \sim V_{white}^{1/3}(V_{gray}^{1/9})^2,$$

and so

$$V_{whiteaxon} \sim V_{white}^{1/3}V_{gray}^{2/9}.$$

Recalling that $V_{white} = N_{whiteaxon}V_{whiteaxon}$, we can now combine our conclusions and get the following.

$$V_{white} = N_{whiteaxon}V_{whiteaxon},$$

$$V_{white} \sim [V_{gray}^{2/3}] \cdot [V_{white}^{1/3}V_{gray}^{2/9}].$$

Now we just need to solve for V_{white}, and we can then conclude that

$$V_{white} \sim V_{gray}^{4/3},$$

very close to the measured exponents, as shown in Table 1.1.

White matter volume scales disproportionately quickly as a function of gray matter volume because of the increasing axon radius, and this, in turn, is due to the satisfaction of Murray's law for efficient flow. The exponent would fall from $4/3$ to 1 if axon radius were invariant.

1.1.4 Wrap-up

It is instructive to summarize the principles that appear to govern the neocortex.

1. *Efficiency Principles*
    - *Efficient Neural Diameters*: neural diameters are set for maximum power efficiency for the distribution of materials through the arbor.
    - *Economical Wiring*: invariant well-connectedness is achieved in a volume-optimal manner.
2. *Invariance Principles*

- *Invariant Well-Connectedness*
  - *Invariant Area-Interconnectedness*: the fraction of the total number of areas to which an area connects is invariant.
  - *Invariant Area-Infiltration*: the fraction of the number of neurons in an area to which a white matter axon connects is invariant.
  - (And these lead to an invariant network diameter of 2.)
- *Invariant Computational Units*
  - *Invariant Minicolumns*: the number of neurons in a minicolumn is invariant.
  - *Invariant Modules*: the number of minicolumns in a module is invariant.

Why are these principles advantageous for the neocortex? The answer is obvious for the two efficiency principles. Invariant well-connectedness is useful, lest larger brains have networks that become more and more widely separated, in terms of the average minimal path length between neurons. It is less obvious why a network would maintain invariant computational units. In the next section this will be taken up in more detail, where we will see that in a wide variety of network—including neocortex—there appears to be scale-invariant "functional units," and I will show that this is to be expected if network size is optimized. The basic idea underlying the argument can be seen here in Figure 1.5. Thus, that the neocortex has invariant computational units is derivable from a network optimization principle. This allows us to simplify the above so as to state the least number of principles from which it is possible to explain neocortical scaling.

1. *Efficiency Principles*
   - *Efficient Neural Diameters*: neural diameters are set for maximum power efficiency for the distribution of materials through the arbor.
   - *Economical Wiring*: invariant well-connectedness is achieve in a volume-optimal manner.
   - *Optimal Network Size*: network size scales up no more quickly than "needed" (see next section), from which invariant computational units are derivable.
2. *Invariant Well-connectedness Principles*
   - *Invariant Area-Interconnectedness*: the fraction of the total number of areas to which an area connects is invariant.
   - *Invariant Area-Infiltration*: the fraction of the number of neurons in an area to which a white matter axon connects is invariant.
   - (And these lead to an invariant network diameter of 2.)

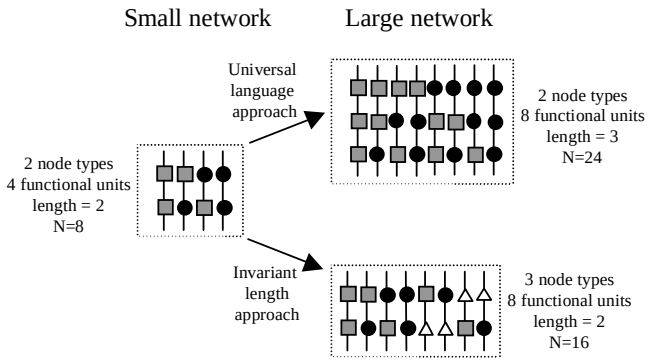


Figure 1.5: There are broadly two ways to increase complexity in networks, as we will discuss in Section 1.2. Top. *In the "universal language approach" there are a fixed number of node types with which all functional units are built. In this case, networks with greater numbers of functional unit types must accommodate the new function types by having longer functional units. In the figure, the small network under the universal language approach begins with 4 functional unit types, each of length 2, and a language of 2 node types; the total number of nodes required for this is 8. In order to accommodate 8 functional unit types in the larger network, the length of functional units must be increased since the expressive power for length-2 functions has already been exhausted. The larger network, which has 8 functional units, has functional units of length 3, and the total number of nodes required is 24.* Bottom. *Consider now, in contrast, the "invariant length approach" to complexity increase. Since functional units have an invariant length in this approach, in order to achieve greater numbers of types of functional units, new node types must be invented. In the figure the small network is identical to the small network under the universal language approach. The large network under this approach has, as in the universal language case, 8 functional unit types. However, to achieve it one new node type must be added. The total number of nodes in the larger network is 16, which is smaller than the 24 nodes in the universal language case. The invariant length approach is optimal in the sense that network size grows minimally. Note that it also entails that the number of types must be increased, and we will see in Section 1.2 that this is indeed true for neocortex, and for networks generally.*

1.2 Complexity in brain and behavior

I took up brain scaling in the previous section, and we saw that many of the ways in which larger brains are "more complex" are consequences of brains maintaining an invariant degree of economical well-connectedness. That is, bigger brains *seem* more complex since they are more highly convoluted, they have more synapses per neuron, and they have a greater number of cortical areas; but these greater "complexities" are not due to the brains themselves being "smarter" in any fashion. Rather, these "complexities" are purely due to the brains being bigger. These seemingly complex qualities of larger brains are thus epiphenomenal, where by that I mean that their increase does not signify any increasing functional complexity of the brain.

In this section I concentrate on brain complexity, both in the nervous network itself, and in the behaviors exhibited by brains. I will be interested in understanding how greater complexity is achieved. First, however, we will need to become clear concerning what I mean by complexity.

The central intuitive notion of 'complexity' I rely upon here is that *an entity, or system, is more complex if it can do more kinds of things.* For example, if my radio has one more type of function than does yours—say, scanning—mine is more complex. Note that under this idea of 'complexity,' doing more of the *same* kinds of thing does *not* make something more complex. For example, a book is complex in some sense, but stapling together two copies of the same book does not create a more complex book; in each of these two cases all the same sentences are uttered. Similarly, if two birds have the same song repertoires, they have the same complexity even if one sings twice as often. Complexity, then, concerns the diversity of things done by an entity. Rather than referring to these things that are done by an entity as "things," I will call them *expressions*, and a system or entity of some kind is more complex than another of that kind if it has, or does, more expression types. The number of expression types, E, is thus the complexity of the entity, and I will sometimes refer to this number, E, as the *expressive complexity* of the entity.

The question we are interested in asking in this section is, How does a system, or entity, of a given kind accommodate greater expressive complexity? For example, how is greater song repertoire size handled in birds? And, how is greater brain complexity achieved? The first-pass answer to these questions is that expressions are always built out of lower-level components that come in different types. For example, bird songs are built out of bird syllables of distinct types. And functional expressions of the brain are built out of combinations of

neurons of distinct types. Let L be the average number of components in an expression (for some given kind of entity); L is the *expression length*. For example, for bird song L is the number of syllables per song. Also, let C be the total number of component types from which the E expression types of the system are buildable. For bird song, C is the total number of syllable types in the repertoire of a bird (from which that bird's E different songs are built).

If expressions are of length L, and each spot in the expression can be filled by one of C component types, then there are a maximum of $E = C^L$ many expression types buildable. For example, if there are $C = 2$ component types—labeled A and B—and expression length $L = 4$, then there are $E = 2^4 = 16$ expression types, namely $AAAA$, $AAAB$, $AABA$,..., $BBBB$. However, this is insufficiently general for two reasons. First, only some constant fraction α of these expression types will generally be grammatical, or allowable, where this proportionality constant will depend on the particular kind of complex system. The relationship is, then, $E \sim C^L$. Second, the exponent, L, assumes that all L degrees of freedom in the construction of expressions are available, when only some fixed fraction β of the L degrees of freedom may generally be available due to inter-component constraints. Let $d = \beta \cdot L$ (where, again, what β is will depend on the particular kind of system). Call this variable d the *combinatorial degree*. The relationship is, then

$$E \sim C^d,$$

where C and d may each possibly be functions of E. Using the same example as above, let us suppose now that As always occur in pairs, and that Bs also always occur in pairs. The "effective components" in the construction of expressions are now just AA and BB, and the expression types are $AAAA$, $AABB$, $BBAA$, and $BBBB$. The number of degrees of freedom for an expression is just 2, not 4, and thus $E = 2^2 = 4$. d is a measure of how combinatorial the system is. The lowest d can be is 1, and in this case there is effectively just one component per expression, and thus the system is not combinatorial at all. Higher values of d mean the system is more combinatorial.

Given this above relationship between expressive complexity E, the number of component types C, and combinatorial degree d, let us consider a few of the ways that a system might increase its expressive complexity.

The first way is the *universal language approach*. The idea here is that there exists a fixed number of component types from which any expression type is constructable. For example, for computable functions there exists such a language: from a small number of basic computable functions it is possible

to build, with ever longer programs, any computable function at all. If this universal language approach were taken, then the number of component types, C, would not change as a function of the expressive complexity, E. Something *would* change, however, and that is that the average length, L, of an expression increases as a function of E. In particular, since C is invariant in the equation $E \sim C^d$, it follows that $d \sim \log E$. This approach may be advantageous for systems where new component types are costly, or where there is little cost to increasing the expression length; it is generally difficult to achieve however, since the invention of a universal language is required.

The second way to increase expressive complexity is via the *specialized component types approach*. In this case, for each new expression type, a new specialized set of component types is invented just for that expression type. Here the expression length L may or may not be > 1, but the combinatorial degree $d = 1$. Thus, $E \sim C$. Note that if this possibility holds for a complex system, then a log-log plot of C versus E should have a slope of 1. The advantage of this approach is that expressions are short, and no complex grammatical rules need to be invented. The disadvantage is that the number of new component types must be scaled up very quickly (namely, proportionally with expressive complexity).

The third way to raise E is via the *invariant-length approach*. This is like the previous approach in that the combinatorial degree d (and expression length L) is invariant, except that now it is > 1. Thus, it is combinatorial ($d > 1$), and its "degree of combinatorialness" remains invariant. The expected relationship is the power law $E \sim C^d$, with d invariant and > 1. On a log-log plot of C versus E, we expect a straight line with slope of $1/d$. A log-log straight line with fractional slope means that a small increase in the number of component types gives a disproportionately large number of new expression types; and this is characteristic of combinatorial systems. An advantage to this approach is that the rate at which new component types must be invented is very slow ($C \sim E^{1/d}$, where d is constant and > 1). The disadvantage is that expressions tend to be longer, and that there must exist a relatively sophisticated set of rules, or grammar, allowing the expressions to be built in a combinatorial fashion.

The final way to increase expressive complexity is via the *increasing-C-and-d approach*. This is similar to the previous case, except that now expressive complexity increase is accommodated by increasing C *and* increasing d. If d increases logarithmically with E, then this is the universal language approach from above, where C does not increase. Thus, d must increase sublogarithmi-

cally, such as $d \sim [\log E]/[\log \log E]$. In this case, it follows that $C \sim \log E$; C is increasing here less quickly than a power law. As in the previous possibility, a small increase in C gives a disproportionately large increase in E, except that now the size of the combinatorial explosion itself increases as a function of E (since d is increasing). This is a kind of middle ground between the universal language approach and the invariant-length approach.

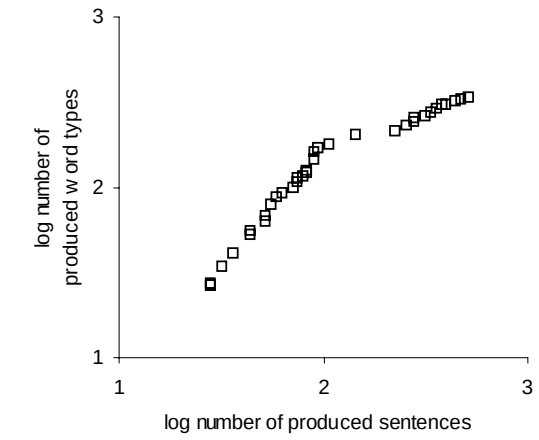
These are the four key manners in which expressive complexity may be increased in complex systems, and our central question may be stated more rigorously now: In which of these ways do complex systems related to brain and behavior increase expressive complexity? And why? In the next two subsections we discuss behavioral and brain complexity in light of the above ideas. (See Changizi (2001e) for the connections between these above ideas and the notion of 'hierarchy'.)

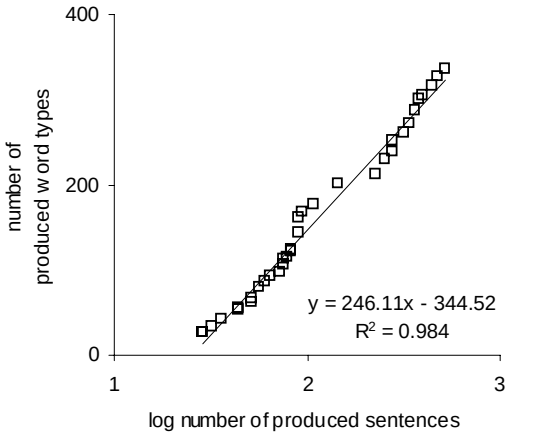
1.2.1 Complexity of languages and behaviors

Behavior is a complex system, consisting of behavioral expressions, which are built out of multiple behavioral components of some kind. The main question concerning behavior here is, In what manner is behavioral repertoire size—i.e., expressive complexity—increased? That is, which of the earlier approaches to increasing complexity holds for behaviors? Do animals have a "universal language" of component behaviors from which any complex behavior may be built, or do animals with more behaviors have more behavioral component types? We examine this question in three distinct kinds of behavior: human linguistic behavior, bird vocalization behavior, and traditional non-vocal animal behaviors.

Ontogeny of language

Human natural language is a complex system, where components of various kinds are combined into expressions. For example, phonemes are combined into words, and words into sentences. We begin by studying expressive complexity increase during the development of language in children. That is, in light of the earlier approaches to increasing complexity, how do children increase their expressive complexity? We already know a few things about the development of language in children. First, we know that children increase the number of component types (e.g., their phoneme repertoire and word repertoire) as they age. That is, C increases. Second, we know that their ability to string together components increases with age (Pascual-Leone, 1970; Case et





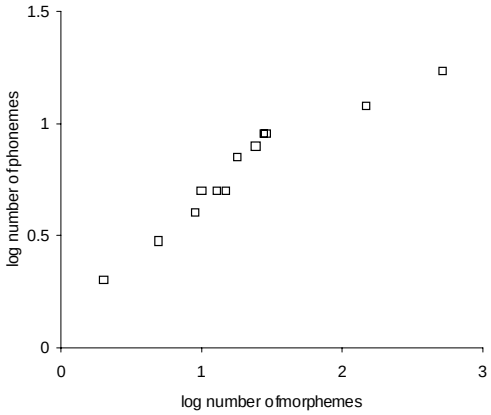
age (wks)	# words	# sentences
61	8	5.49
62	9	5.49
63	10	5.49
64	11	9.42
65	12	13.34
66	13	21.19
67	14	21.19
68	15	29.04
69	16	29.04
70	17	32.97
71	18	36.89
72	19	40.82
73	20	48.67
74	21	52.59
75	22	52.59
76	23	56.51
77	24	60.44
78	25	60.44
79	26	68.29
80	27	68.29
81	28	72.21
82	29	83.99
83	30	123.23
84	31	201.73
85	32	233.12
86	33	256.67
87	34	256.67
88	35	295.92
89	36	315.54
90	37	339.09
91	38	358.71
92	39	374.41
93	40	421.51
94	41	452.90
95	42	495.00

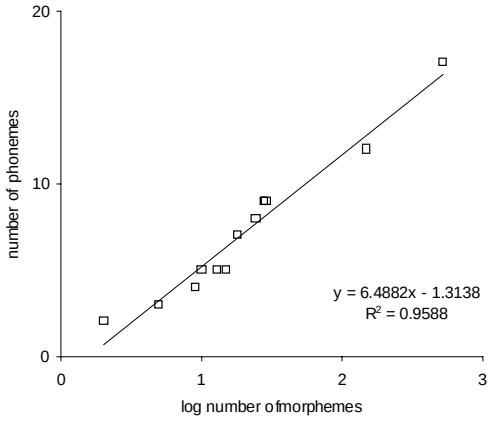
Figure 1.6: Top. *Logarithm (base 10) of the number of word types versus logarithm of the number of sentences, as produced by one child named Damon from 12 to 22 months (Clark, 1993).* Bottom. *Semi-log plot of the same data. Plot is confined to multiword utterance ages, which began at about 14 months. On the right are shown the data; sentence data is fractional because I obtained it via measuring from Clark's plots. [Note that Damon grew up and went on to do some research on aspects of how the brain scales up.]*

al., 1982; Siegel and Ryan, 1989; Adams and Gathercole, 2000, Robinson and Mervis, 1998; Corrigan, 1983). However, from this increasing combinatorial ability we cannot straightforwardly conclude that the child's combinatorial degree, d, will keep pace. Recall that d is measured by the relative rates of growth of the number of component types C and the number of expression types E. A child's combinatorial *ability* could increase, and yet the child could simply choose not to speak much, in which case the combinatorial ability growth would not be reflected in the combinatorial degree measured from the C versus E plot (since E would not increase much). Nevertheless, learning new component types is costly, and efficiency considerations would lead one to expect that new component types are added no more quickly than needed for the given level of expressive complexity. If this were true, we would expect the combinatorial degree to increase as a function of E, and thus the log-log slope of C versus E to fall as E increases. That is, the increasing-C-and-d approach would be used, and $C \sim \log E$. We examine this prediction for the development of words and sentences, and also for the development of phonemes and words.

The number of word types and number of distinct sentences uttered by a single child named Damon for 41 weeks from 12 to 22 months of age (Clark, 1993) are shown in Figure 1.6. One can see that, as expected, the combinatorial degree (i.e., the inverse of the slope in the log-log plot) falls as a function of E. At the start, the combinatorial degree is 1, which means that the child is not yet using words in a combinatorial fashion. Note that the data here are only for the multi-word utterance stage of the child; thus, although the child may *seem* to be using words in a combinatorial manner since his sentences have more than one word, he is not. By the end of the recorded data, the combinatorial degree has increased to about 2.5. This combinatorial degree range is similar to the sentence length range for children of this period (Robinson and Mervis, 1998). Since the combinatorial degree and number of component types are increasing, the increasing-C-and-d length approach is being employed for increasing expressive complexity, and thus we expect $C \sim \log E$. Indeed, a plot of C versus $\log E$ appears linear.

The growth in the number of phonemes and distinct morphemes was compiled from Velten (1943), as produced by a child named Jean from 11 to 30 months of age. [A *morpheme* is the smallest meaningful linguistic unit; or, a word that is not decomposable into meaningful parts.] Figure 1.7 shows the log-log plot of the number of phoneme types versus the number of morpheme types, and one can see that the slope tends to decrease somewhat through





age (mos.)	# phonemes	# morphemes	max # phonemes per morpheme
11	2	2	2
12	3	5	4
13	4	9	4
14	5	10	4
15	5	13	4
16	5	15	4
17	7	18	4
20	8	24	4
21	9	28	4
22	9	29	4
23	12	150	5
30	17	524	6

Figure 1.7: Top*: Logarithm (base 10) of the number of phoneme types versus logarithm of the number of morphemes, as produced by one child named Jean from 11 to 30 months (Velten, 1943).* Bottom*: Semi-log plot of the dame data. Morphemes are the smallest meaningful linguistic unit, and are mostly words in this case. On the right are shown the data.*

development, meaning the combinatorial degree is increasing. The plot of (unlogged) number of phoneme types versus the logarithm of the number of morphemes is comparatively linear, again implicating the increasing-C-and-d approach, as predicted above. The combinatorial degree begins at around 2 ("ma"), and increases to around 4 ("baby"). The number of phonemes per morpheme—i.e., expression length—increases over a similar range during this period (Changizi, 2001d, 2001e).

In each of these language development cases, we see that the expressive complexity is increased via the increasing-C-and-d invariant approach, and that the combinatorial degree appears to increase in proportion to the child's ability to combine components. This accords with the efficiency hypothesis mentioned above, that children will learn component types no more quickly than needed to express themselves.

English throughout history

Here I consider complexity in the English language. Not the complexity of the language of one English speaker as above, but, instead, the complexity of the entire English language. Our "individual" here is the entire English-speaking community. This individual has, over history, said more and more new expression types. Namely, new distinct sentences are being uttered throughout history. How has this increase in expressive complexity been accommodated? Via an increase in the average length of a sentence, or via the addition of new vocabulary words—word types—with which sentences may be built, or via a combination of the two? That is, which of the earlier approaches to complexity increase is employed in the evolution of the English language?

I estimated the growth in the number of English word types by using the Oxford English Dictionary (OED), Second Edition. It is possible to search for years within only the etymological information for all entries in the OED. In this way it was possible to estimate the number of new word types per decade over the last 800 years. To obtain an estimate of the growth rate for the number of sentences the English-speaking entity expresses, I used the number of books published in any given year as an estimate of the number of new sentences in that year. This would be a problematic measure if different books tended to highly overlap in their sentences, but since nearly every written sentence is novel, never having been uttered before, there is essentially no overlap of sentences between books. This would also be a problematic measure if the length of books, in terms of the number of sentences, has been changing through time;

Table 1.3: The data for the history of English from 1200 to the present. Decades covering century or half-century years do not include those years (since they tend to be overcounted). The new words were measured from the Oxford English Dictionary, and the number of new books from WorldCat.

decade	# new words	# new books		decade	# new words	# new books
1210	36	3		1610	161	3705
1220	40	5		1620	606	4174
1230	43	4		1630	130	4736
1240	29	3		1640	140	6321
1250	24	0		1650	125	17891
1260	36	5		1660	169	13976
1270	42	1		1670	144	11274
1280	59	3		1680	181	16548
1290	41	8		1690	202	21868
1300	65	10		1700	117	16962
1310	72	12		1710	156	15513
1320	77	4		1720	111	17398
1330	78	5		1730	207	15685
1340	73	5		1740	273	17717
1350	33	5		1750	146	16113
1360	42	9		1760	296	21161
1370	61	3		1770	231	25254
1380	104	8		1780	214	33542
1390	83	7		1790	295	38186
1400	54	7		1800	315	54326
1410	54	7		1810	372	99069
1420	72	12		1820	440	129734
1430	56	6		1830	481	114817
1440	63	5		1840	602	148800
1450	18	4		1850	454	158774
1460	51	10		1860	452	234706
1470	51	13		1870	554	257161
1480	83	38		1880	661	290921
1490	107	79		1890	809	391372
1500	65	103		1900	686	402769
1510	75	139		1910	685	541294
1520	126	134		1920	425	587809
1530	171	265		1930	484	680614
1540	167	468		1940	397	808427
1550	201	626		1950	206	730849
1560	223	814		1960	249	1354217
1570	178	1038		1970	216	2600020
1580	195	1363		1980	120	4594985
1590	163	1688		1990	35	5618350
1600	372	2055				

I have no data in this regard, but it seems plausible to assume that any such trend is not particularly dramatic. The number of new books published per year was obtained by searching for publication dates within the year for literature listed in WorldCat, an online catalog of more than 40 million records found in thousands of OCLC (Online Computer Library Center) member libraries around the world. In this way I was able to estimate the number of new books per decade over the last 800 years, the same time period for which I obtained word type data. These data are shown in Table 1.3.

Figure 1.8 shows the logarithm of the number of new word types and books per decade over the last 800 years, measured as described above. Note that the plot shows estimates for the number of *new* word types per decade, and the number of *new* sentences per decade; i.e., it measures dC/dt and dE/dt versus time. The plot does not, therefore, show the growth in the actual magnitude of the number of word types or the number of sentences. But it is the scaling relationship between the actual magnitudes of C and E we care about, so what can we do with a plot of growth rates over time? Note first that the growth rate for each is exponential (this is because the plots fall along straight lines when the y axis is logarithmic and the x axis not). If a growth rate for some quantity u increases exponentially with time, then this means $du/dt \sim e^t$. And if you recall your calculus, it follows that the quantity itself scales exponentially with time, and, in fact, it scales proportionally with the growth rate: i.e., $u \sim du/dt$. Thus, Figure 1.8 has effectively measured the growth in the number of word types and the number of books. By looking at the growth in the number of word types compared to that for the number of books, we can determine how the first scales against the second.

From the figure we can, then, determine that

$$dC/dt \sim C \sim 10^{0.001725t} \sim e^{0.003972t},$$

and

$$dE/dt \sim E \sim 10^{0.008653t} \sim e^{0.01992t}.$$

We may now solve for C in terms of E, and we obtain

$$C \sim E^{0.003972t/0.01992t} = E^{0.1994}.$$

The number of word types scales as a power law against the number of sentences, and, unsurprisingly, the slope is less than one and thus English is combinatorial. Thus, greater expressive complexity was achieved over the last 800 years not by increasing the combinatorial degree (or average sentence length),

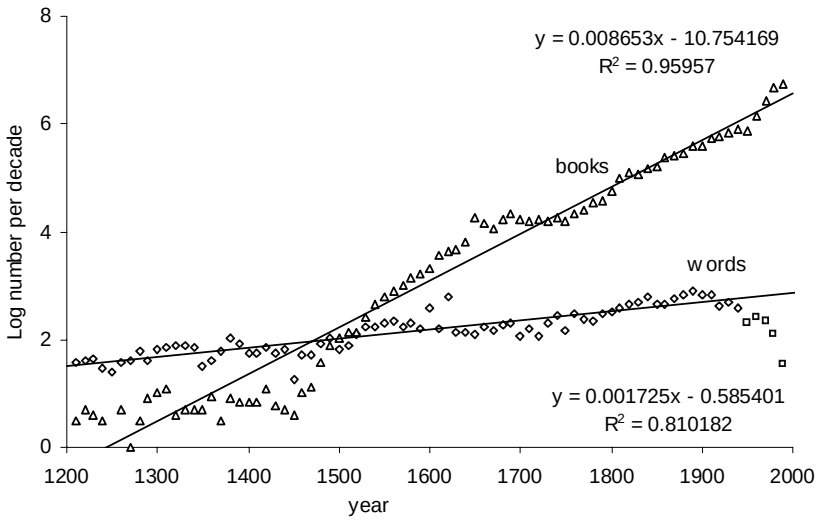


Figure 1.8: Growth rates in the decades from the years 1200 to 1990 for the number of new English word types and the number of new English books. Regression equations and correlation coefficients are shown for each (79 data points each). Unsure etymological dates tend to cluster at century and half century marks and therefore century and half-century marks tend to be overcounted; accordingly, they were not included in the counts. The OED is conservative and undercounts recently coined word types; consequently, the exponential decay region (the last five square data points) was not included when computing linear regression. I do not have any way to similarly measure the number of word type extinctions per year, and so I have not incorporated this; my working assumption is that the extinction rate is small compared to the growth rate, but it should be recognized that the estimated combinatorial degree is therefore an underestimate.

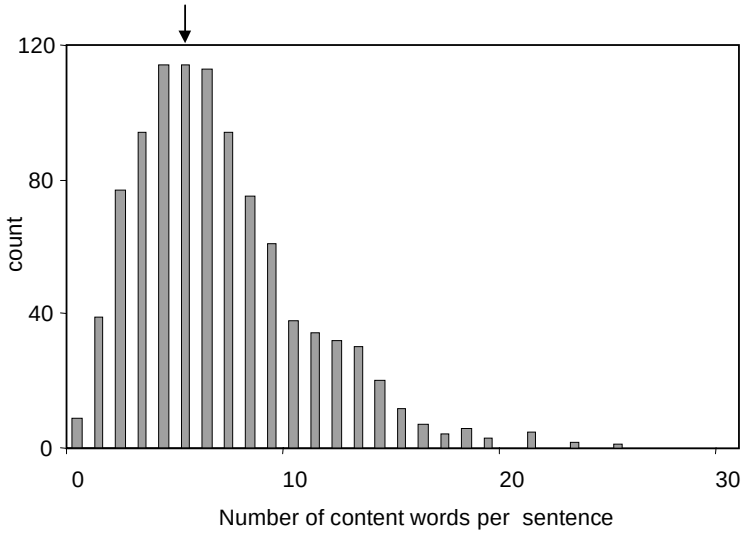


Figure 1.9: Distribution of numbers of content words per sentence in English. Arrow indicates the log-transformed mean. 984 sentences from 155 authors were measured from texts in philosophy, fiction, science, politics and history. I chose the second sentence on each odd numbered page. A word was deemed a function word if it was among a list of 437 such words I generated. A string of words was deemed a sentence if it represented a complete thought or proposition. So, for example, semicolons were treated as sentence delimiters, multiple sentences combined into one long sentence by ", and" were treated as multiple sentences, and extended asides within dashes or parentheses were not treated as part of the sentence.

but, instead, by increasing the number of word types with which to build sentences. The scaling exponent of around 0.2 implies an estimated combinatorial degree of about 5. There appears to be nothing about the English grammar that implies a fixed combinatorial degree (or sentence length), much less any particular value of it. What explains this value of 5? [Or, a little more than 5; see legend of Figure 1.8 concerning word type extinctions.] It cannot simply be due to the typical number of words in an English sentence, since there are typically many more words than that, namely around 10 to 30 words (Scudder, 1923; Hunt, 1965).

To make sense of the combinatorial degree, we must distinguish between two kinds of word in English: *content* and *function*. The set of content words, which refer to entities, events, states, relations and properties in the world, is large (hundreds of thousands) and experiences significant growth (Clark

and Wasow, 1998). The set of function words, on the other hand, which includes prepositions, conjunctions, articles, auxiliary verbs and pronouns, is small (around 500) and relatively stable through time (Clark and Wasow, 1998). The scale-invariant combinatorial degree of English suggests that the average number of words per sentence is invariant. Imagine, for simplicity, that there on average n places for content words in a sentence, and m places for function words, and that these values, too, are invariant. (And thus the average sentence length is $n + m$.) The total number of possible sentences is then

$$E \sim N^n M^m,$$

where N is the total number of content words in English and M the total number of function words. n and m are invariant, as mentioned just above, and so is the total number of function words M. Thus, the equation above simplifies to the power law equation

$$E \sim N^n.$$

Also, note that the number of content words, N, is essentially all the words, since it dwarfs the number of function words; i.e., $C \approx N$. Thus, $E \sim C^n$, and so,

$$C \sim E^{1/n}.$$

That is, the combinatorial degree is expected to be equal to the typical number of *content* words per sentence—not the typical total number of words per sentence—and, up to a constant factor, they may be combined in any order. To test this reasoning, I measured the number of content words in nearly one thousand sentences (see legend of Figure 1.9). The distribution is log-normal (Figure 1.9), and the mean of the logs is 0.7325 (± 0.2987); the log-transformed mean is thus 5.401, and one standard deviation around this corresponds to the interval [2.715, 10.745]. This provides confirmation of the hypothesis that the combinatorial degree is due to there being five content words per sentence.

But *why* are there typically five content words per sentence? One obvious hypothesis is that sentences can convey only so much information before they overload the utterer's or listener's ability to understand or absorb it. In this light, five content words per sentence is probably due to our neurobiological limits on working memory, which is a bit above five (Miller, 1956). The fingerprint of our working memory may, then, be found in the relative rate at which new words are coined compared to the number of sentences uttered by the English-speaking community.

Table 1.4: Number of syllable types and song types for a variety of species of bird.

Species	Number of syllable types	Number of song types	Citation
Turdus nudigenis (Bare-eyed Thrush)	1.20	1.19	Ince and Slater (1985)
Turdus tephronotus (African Bare-eyed Thrush)	3.83	1.26	Ince and Slater (1985)
Turdus iliacus (Redwings)	2.93	1.31	Ince and Slater (1985)
Turdus torquatus (Ring Ouzels)	5.49	3.79	Ince and Slater (1985)
Turdus viscivorus (Song and Mistle Thrush)	43.08	15.74	Ince and Slater (1985)
Turdus pilaris (Fieldfare)	80.65	32.05	Ince and Slater (1985)
Turdus merula (Blackbird)	216.09	38.97	Ince and Slater (1985)
Turdus abyssinicus (Olive Thrush)	43.08	49.99	Ince and Slater (1985)
Turdus migratorius (American Robin)	30.10	71.33	Ince and Slater (1985)
Turdus philomelos (Song Thrush)	309.22	158.77	Ince and Slater (1985)
Catherpes mexicanus (Canyon Wren)	9	3	Kroodsma (1977)
Cistothorus palustris (Long-billed Marsh Wren)	44—118	40—114	Kroodsma (1977)
Cistothorus platensis (Short-billed Marsh Wren)	112	110	Kroodsma (1977)
Salpinctes obsoletus (Rock Wren)	69—119	69—119	Kroodsma (1977)
Thryomanes bewickii (Bewick's wren)	25—65	9—22	Kroodsma (1977)
Thryomanes bewickii (Bewick's wren)	87.5	20	Kroodsma (1977)
Thryomanes bewickii (Bewick's wren)	56	17.5	Kroodsma (1977)
Thryomanes bewickii (Bewick's wren)	50	16	Kroodsma (1977)
Thryomanes bewickii (Bewick's wren)	40.5	10	Kroodsma (1977)
Thryomanes bewickii (Bewick's wren)	35.5	17.5	Kroodsma (1977)
Thryothorus ludovicianus (Carolina Wren)	22	22	Kroodsma (1977)
Troglodytes troglodytes (Winter wren)	89—95	3—10	Kroodsma (1977)
Gymnorhina Tibicen (Australian Magpie)	29	14	Brown et al. (1988)
Paridae bicolor (Tufted titmouse)	3	11.25	Hailman (1989)
Parus wollweberi (Bridled titmouse)	3	3	Hailman (1989)
Serinus canaria (Canary)	101	303	Mundinger (1999)
Empidonax alnorum (Alder Flycatcher)	3	1	Kroodsma (1984)
Empidonax traillii (Willow Flycatcher)	4	3	Kroodsma (1984)

Bird vocalization

Bird songs are built out of bird syllables, and the question we ask is, Do birds with more songs in their repertoire have longer songs, or more syllables, or both? In particular, which of the four earlier approaches to expressive complexity increase is used in bird vocalization?

To answer this I surveyed the bird vocalization literature and compiled all cases where the authors recorded the number of syllable types and the number of song types in the repertoire of the bird. Although song repertoire size counts are common, syllable type counts are much rarer, especially when one is looking for papers recording both. Table 1.4 shows the data and the sources from which I obtained them.

Plotting the number of syllable types, C, versus the number of song types, E, on a log-log plot (Figure 1.10) reveals that (i) they are related by a power law (i.e., the data are much more linear on a log-log plot than on a semi-log plot), and (ii) the exponent is approximately 0.8. That is, $C \sim E^{0.8}$. Since the relationship is a power law, the combinatorial degree is an invariant; i.e., there appears to be no tendency for the combinatorial degree (or expression length) to increase in birds with greater numbers of songs. Instead, greater expressive complexity is achieved entirely through increasing the number of bird syllable types. Since the exponent is about 0.8, the combinatorial degree is its inverse, and is thus about 1.25, which is not much above 1. In fact, it is not significantly different from 1 (Changizi, 2001d). Birds with twice as many songs therefore tend to have roughly twice as many syllable types, and thus bird vocalization may not be combinatorial at all, and, at most, it is not very combinatorial. Birds therefore appear to conform to the specialized component types approach to complexity increase. Using bird vocalization as a model for language is thus inappropriate. Note that the combinatorial degree for bird vocalization is near 1 despite the fact that birds have, on average, around 3 or 4 syllables per song (Read and Weary, 1992; Changizi, 2001d).

Traditional animal behavior

Thus far, the cases of behavior we have discussed have been vocalizations, whether bird or human. Now we wish to consider run-of-the-mill behaviors, and ask how greater behavioral repertoire size is accommodated in animals. Do animals with more distinct behaviors (i.e., more expression types) have more muscles with which the behaviors are implemented? Or do they have the same number of muscles, and behaviorally more complex animals have longer, and

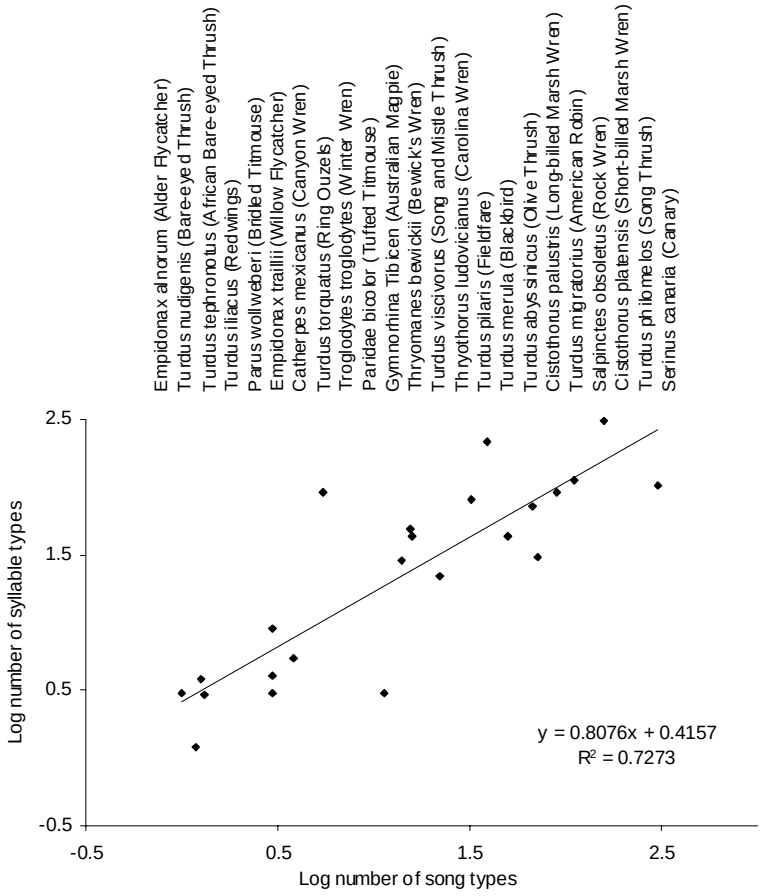


Figure 1.10: Logarithm (base 10) of the number of bird syllable types versus the logarithm of the number of song types. When a min and a max are given in Table 1.4, 10 to the power of the average of the logged values is used. (The multiple measurements for Bewick's wren are averaged and plotted as one data point.) The slope is not significantly different from 1 (see Changizi, 2001d), suggesting that bird vocalization may not be combinatorial, and thus not language-like.

more complex, behaviors? (I am assuming that each distinct muscle is its own component type.) What we desire now are data showing how the number of muscles varies as a function of the number of distinct behaviors.

By exhaustively reviewing the animal behavior and ethology literature over the last century, I was able to compile estimates of the behavioral repertoire size in 51 species across seven classes within three phyla. Such behavior counts are recorded in what are called *ethograms*, and I only used ethograms where the aim was to record all the animal's behaviors, not just, say, mating behavior. Behaviors recorded in ethograms tend to be relatively low level behaviors, and probably serve as components themselves in higher level behaviors. I refer to behaviors listed in ethograms as *ethobehaviors*. Table 1.5 shows these data, and Figure 1.11 displays them. There are not enough data in each of these classes to make any strong conclusions concerning the relative ethobehavioral repertoire sizes, other than perhaps (i) that the range of ethobehavioral repertoire sizes for mammals is great, and greater than that for the other classes, and (ii) the number of ethobehavior types for vertebrates tends to be higher than the number for invertebrates.

Recall that our main purpose is to examine how the number of muscles scales with the number of ethobehavior types. There are two reasons to focus on only one class of animals at a time. First, it seems reasonable to expect that if there are universal laws governing the relationship between number of muscles and number of ethobehavior types, the general form of the relationship may be similar across the classes, but the particular constants in the mathematical relationships may depend on the class of animal. For example, perhaps fish with E ethobehavior types tend to have half the number of muscles as a mammal with E ethobehavior types, but within each class the scaling relationships are identical. Second, the community standards for delineating behaviors are more likely to be similar within a class than across classes. For example, it may be that ethologists tend to make twice the number of behavioral delineations for insects than for mammals. Here I examine behavioral complexity within mammals only. One reason to choose this class is because there exists more ethobehavior data here (from 23 species), and it covers a wider range, than the data for the other classes (see Figure 1.11). The other reason is that we also require estimates of the number of muscle types, and I have been able to acquire muscle counts for only a few non-mammals.

Table 1.6 shows the behavioral repertoire sizes for just the mammals, along with estimates of the number of muscles and of index of encephalization (which is a measure of brain mass that corrects for how big it is due merely to the mass

Table 1.5: Number of ethobehavior types (i.e., the number of behaviors listed in the authors' ethogram) for 51 species.

Phylum	Class	Latin name	Name	# etho-behaviors	citation
Arthropoda	Insecta	*Apis mellifera*	Worker honey bee	30	Kolmes, 1985
		Ropalidia marginata	Social wasps	37	Gadagkar & Joshi, 1983
		Camponotus colobopsis	Mangrove ants	36	Cole, 1980
		Automeris aurantinca Weym	Butterfly	15	Bastock & Blest, 1958
		Pelocoris femoratus	Water bug	22	Brewer & Sites, 1994
		Sceptobiini	Ant-guest beetle	42	Danoff-Burg, 1996
		Stenus	Stenus beetle	73	Betz, 1999
Mollusca	Gastropod	*Aplysia californical Cooper*	Sea slug	45	Leonard & Lukowiak, 1986
		Navanax inermis	Sea slug	28	Leonard & Lukowiak, 1984
		Strombidae	Sea snail	7	Berg, 1974
	Cephalopo	*Callianassa subterranea*	Burrowing shrimp	12	Stamhui et al., 1996
		Eledone moschata	Cuttlefish	14	Mather, 1985
Chordata	Osteichthy	*Haplochromis buroni*	Mouth-brooding african cichlid fish	19	Fernald & Hirata, 1977
		Lepomis gibbosus, Linneaus	Pumpkinseed sunfish	26	Miller, 1963
		Parablennius sanguinolentus parvicornis	Blennies	40	Santos & Barreiros, 1993
		Pleuronectes platessa L.	Juvenile plaice fish	8	Gibson, 1980
		Colisa	Colisa fish	23	Miller & Jearld, 1983
		Gasterosteus aculeatus	Three-spined stickleback	19	Wooton, 1972
Chordata	Reptilia	*Gopherus agassizii*	Desert tortoise	80	Ruby & Niblick, 1994
		Caiman sclerops	Caimen	188	Lewis, 1985
		Lampropholis guichenoti	Scincid lizard	45	Torr & Shine, 1994
Chordata	Aves	*Ara ararauna and A. macao*	Parrot	23	Uribe, 1982
		Melopsittacus undulatus	Budgerigar parakeet	60	Brockway, 1964a, 1964b
		Hydrophasianus chirurgus	Pheasant-tailed and bronzewinged jacana, duck	19	Ramachandran, 1998
		Phalacrocorax atriceps bransfieldensis	Blue-eyed shag (a cormorant)	21	Bernstein and Maxson, 1982
		Coturnix chinensis	Bluebreasted quail	60	Schleidt et al., 1984
		Gallus bankvia	White leghorn-type hen	13	Webster & Hurnik, 1990
		Poephila guttata	Zebra finch	52	Figueredo et al., 1992
Chordata	Mammalia	*Alces alces andersoni*	North Am. Moose	22	Geist, 1963
		Meriones unguiculatus	Mongolian gerbil	24	Roper & Polioudakis,1977
		Peromyscus maniculatus gambelii	Deer mouse	29	Eisenberg, 1962
		Dolichotis patagonum	Mara	30	Ganglosser & Wehnelt, 1997
		Rattus rattus	Albino lab rat	43	Bolles and Woods, 1964
		Marmota monax	Woodchuck	43	Ferron & Ouellet, 1990
		Castor canadensis	Beaver	51	Patenaude, 1984
		Sciuridae (four species)	Squirrel	55	Ferron, 1981
		Rattus norvegicus	White rat	33	Draper, 1967
		Spermophilus beecheyi	California Ground squirrel	34	Owings et al., 1977
		Leporidae (family)	White rabbit	30	Gunn & Morton, 1995
		Pteropus livingstonii	Fruit bat	93	Courts, 1996
		Blarina brevicaudo	Short-tailed shrew	54	Martin, 1980
		Mustela nigripes	Black-footed ferret	74	Miller, 1988
		Felis catus	Cat	69	Fagen & Goldman, 1977
		Tursiops truncatus	Bottlenose dolphin	123	Muller et al., 1998
		Calithrix jacchus jacchus	Common marmoset	101	Stevenson & Poole, 1976
		Nycticebus coucang	Malaysian slow loris	80	Ehrlich & Musicant, 1977
		Galago crassicaudatus	Great Galagos	97	Ehrlich, 1977
		Cercopithecus neglectus	De Brazza monkey	44	Oswald & Lockard, 1980
		Macaca nemestrina	Macaque monkey	184	Kaufman & Rosenblum, 1966
		Papio cynocephalus	Baboon	129	Coehlo & Bramblett, 1981
		Homo sapiens	Human child	111	Hutt & Hutt, 1971

of the animal's body). Muscle counts were estimated from atlases of anatomy, and I used the maximum estimate I could find, since lower estimates in an atlas are due to a lack of detail. Here I have listed all the muscle estimates for each mammalian order, only the maximum which was used in the analysis.

- Artiodactyla: 89 (Walker, 1988), 116 (Sisson and Grossman, 1953, ox), 138 (Sisson and Grossman, 1953, pig), 186 (Singh and Roy, 1997), 191 (Ashdown and Done, 1984), 203 (Raghavan, 1964).
- Carnivora: 160 (Sisson and Grossman, 1953), 169 (Bradley and Grahame, 1959), 197 (Reighard and Jennings, 1929), 204 (Boyd et al., 1991, cat), 204 (Boyd et al., 1991, dog), 208 (McClure et al., 1973), 212 (Hudson and Hamilton, 1993), 229 (Adams, 1986), 322 (Evans, 1993).
- Didelphimorphia: 159 (Ellsworth, 1976).
- Lagomorpha: 67 (Busam, 1937), 85 (Chin, 1957), 112 (Wingerd, 1985), 126 (McLaughlin and Chiasson, 1990), 128 (Craigie, 1966), 214 (Popesko et al., 1990).
- Perissodactyla: 146 (Sisson and Grossman, 1953), 172 (Way and Lee, 1965), 194 (Budras and Sack, 1994), 245 (Pasquini et al., 1983).
- Primates: 160 (Schlossberg and Zuidema, 1997), 190 (Stone and Stone, 1997), 228 (Rohen and Yokochi, 1993), 230 (Bast et al., 1933), 255 (Anson, 1966), 267 (Agur and Lee, 1991), 278 Netter, 1997), 316 (Williams et al., 1989).
- Proboscidea: 184 (Mariappa, 1986).
- Rodentia: 104 (Popesko et al., 1990, mouse), 113 (Popesko et al., 1990, hamster), 134 (Popesko et al., 1990, rat), 143 (Popesko et al., 1990, guinea pig), 183 (Howell, 1926), 190 (Hebel and Stromberg, 1976), 206 (Cooper and Schiller, 1975), 218 (Greene, 1935).

Index of encephalization, P, was computed as body mass, M, divided by brain mass, B, to the power of $3/4$; i.e., $P = M/B^{3/4}$. This is appropriate since brain mass scales as body mass to the $3/4$ power (Allman, 1999; Changizi, 2001b). Body and brain masses were acquired from Hrdlicka (1907), Bonin (1937), Crile and Quiring (1940), and Hofman (1982a, 1982b). [These data were first presented in Changizi, 2002.]

Figure 1.12 is reminiscent of Figure 1.8 in that the number of component types (respectively, muscles and words) increases disproportionately slowly compared to the number of expression types (respectively, ethobehaviors and sentences) as a function of some third parameter (respectively, encephalization and time). From the relative rate at which the number of muscles and number of ethobehavior types scale as a function of encephalization, we can compute

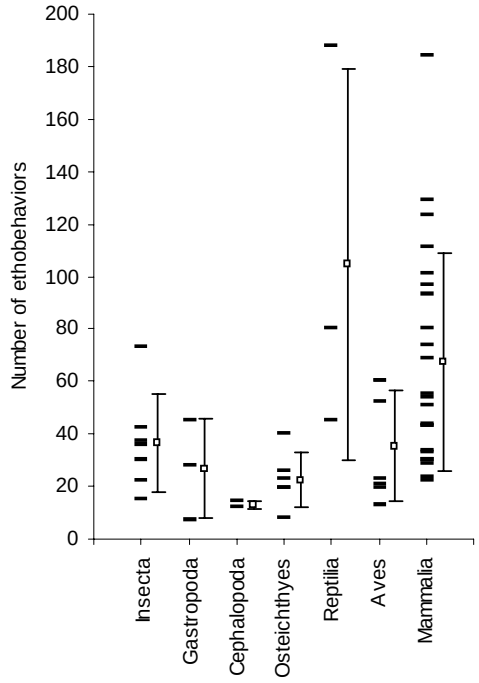


Figure 1.11: Number of ethobehavior types for each species, and average over all species (and standard deviation) within the seven classes shown.

the combinatorial degree. In particular, Figure 1.12 shows that $E \sim P^{0.8}$ and that $C \sim P^{0.27}$. From this we may conclude that $E \sim C^3$, and thus the combinatorial degree is roughly 3. However, the data are insufficient to statistically distinguish between whether the combinatorial degree is invariant (as a function of E), or whether the combinatorial degree may be slowly increasing. Greater behavioral complexity is achieved, at least in part, by increasing the number of behavioral component types, or the number of muscles. Muscles therefore do not serve as a universal behavioral language from which any behavior may be built.

The combinatorial degree for mammalian behavior is roughly 3 (possibly not invariant), and there are several interesting implications. (1) Since it is greater than one, it means that behavior is, indeed, language-like. There are many who already believe that behavior is language-like in this sense (Fentress and Stilwell, 1973; Slater, 1973; Dawkins and Dawkins, 1976; Douglas and Tweed, 1979; Rodger and Rosebrugh, 1979; Gallistel, 1980; Lefebvre, 1981; Fentress, 1983; Schleidt et al., 1984; Berkinblit et al., 1986; Greenfield, 1991; Allott, 1992; Bizzi and Mussa-Ivaldi, 1998), but the mere fact that multiple muscles are involved in each behavior is not an argument that behavior is language-like, as we saw in bird vocalization. The results here provide a rigorous test of the language-likeness of behavior. (2) A combinatorial degree of around 3 is surprisingly low, given that behaviors have dozens or more muscles involved. The actual number of degrees of freedom is well below the actual number of muscles involved, and this is due to the stereotyped mutual dependencies between muscles. (3) This value for the combinatorial degree is not too much lower than the combinatorial degree of 5 for human natural language. Since the combinatorial degree for mammalian behavior is effectively an average over many mammals, it is possible that the behavioral combinatorial degree for humans is actually nearer to 5, and that perhaps there are similar neurobiological constraints underlying these values. Preliminary data in my own experiments (Changizi, 2002) show that the combinatorial degree is also around three for the ontogeny of behavior in rats (Figure 1.13), where low-level components were the total number of degrees of freedom exhibited by the joints of the pups (i.e., the behavior of the pup parts), and the high-level behaviors were ethobehaviors. (I expect that my estimates scale in proportion to the true counts, but I do not expect that my counts reflect the actual magnitudes of the repertoire sizes, especially for the low-level components where I suspect severe undercounting.)

Another interesting conclusion we may draw from Figure 1.12 is that ethobe-

Table 1.6: Number of ethobehavior types, number of muscles, and index of encephalization (i.e., brain size corrected for body size) for mammals.

Order and *species latin name*	Species common name	# behavior types	behavior citation	index of enceph.	# muscle types	muscle citation
Artiodactyla		27.0		0.0297	203	Raghavan
Alces alces	North Am. Moose	22	Geist	0.0342		
Cephalophus monticola	Duikers	32	Dubost	0.0252		
Carnivora		71.5		0.0862	322	Evans
Felis catus	Cat	69	Fagen & Goldman	0.0888		
Mustela nigripes	Black-footed ferret	74	Miller	0.0837		
Cetacea		123.0		0.1721		
Tursiops truncatus	Bottlenose dolphin	123	Muller et al.	0.1721		
Chiroptera		93.0		0.0679		
Pteropus livingstonii	Fruit bat	93	Courts	0.0679		
Didelphimorphia				0.0185	159	Ellsworth
Insectivora		54.0		0.0490		
Blarina brevicaudo	Short-tailed shrew	54	Martin	0.0490		
Lagomorpha		30.0		0.0345	214	Popesko et al.
Leporidae (family)	White rabbit	30	Gunn & Morton	0.0345		
Perissodactyla				0.0388	245	Pasquini et al.
Primates		106.6		0.1789	316	Williams et al.
Cercopithecus neglectus	De Brazza monkey	44	Oswald & Lockard	0.1454		
Nycticebus coucang	Malaysian slow loris	80	Ehrlich & Musicant	0.1231		
Galago crassicaudatus	Great Galagos	97	Ehrlich	0.0977		
Calithrix jacchus	Common marmoset	101	Stevenson & Poole	0.1445		
Homo sapiens	Human child	111	Hutt & Hutt	0.3502		
Papio cynocephalus	Baboon	129	Coehlo & Bramblett	0.1793		
Macaca nemestrina	Macaque monkey	184	Kaufman & Rosenblum	0.2116		
Proboscidea				0.0731	184	Mariappa
Rodentia		38.0		0.0555	218	Greene
Meriones unguiculatus	Mongolian gerbil	24	Roper & Polioudakis	0.0569		
Peromyscus maniculatus	Deer mouse	29	Eisenberg	0.0569		
Dolichotis patagonum	Mara	30	Ganglosser & Wehnelt	0.0394		
Rattus norvegicus	White rat	33	Draper	0.0337		
Spermophilus beecheyi	Ground squirrel	34	Owings et al.	0.0803		
Rattus rattus	Albino lab rat	43	Bolles and Woods	0.0337		
Marmota monax	Woodchuck	43	Ferron & Ouellet	0.0803		
Castor canadensis	Beaver	51	Patenaude	0.0383		
Sciuridae (four species)	Squirrel	55	Ferron	0.0803		

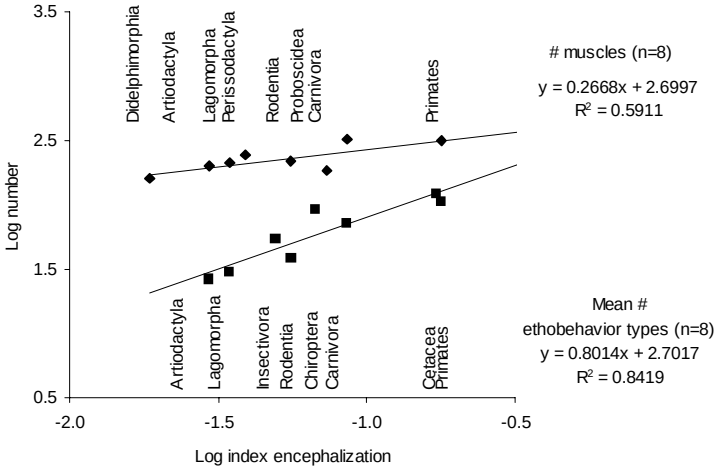


Figure 1.12: A log-log plot of the number of ethobehavior types and number of muscles in mammals, each as a function of index of encephalization.

havioral repertoire size is strongly correlated with index of encephalization. In fact, they are roughly proportional to one another. This provides a kind of justification for the use of encephalization as a measure of brain complexity.

Summing up scaling in languages

Table 1.7 summarizes the results for the behavioral systems we have covered above. One of the first generalizations we may make is that in no case do we find the universal language approach employed. For behavioral complexity across adults—i.e., not the cases of the ontogeny of behavior—the combinatorial degree is, in every case, consistent with its being invariant, implicating the length-invariant approach to complexity increase. We cannot, however, reject the possibility that the combinatorial degree is increasing in mammalian behavior. For the ontogeny of human language, the combinatorial degree clearly, and expectedly, increases as expressive complexity increases, and the relationship thus conforms to a logarithmic law; the increasing-C-and-d length approach is followed. For the ontogeny of rat behavior we are unable to say whether the relationship is a power law or logarithmic, but can conclude that the combinatorial degree is of the same order of magnitude as that for the phylogeny of mammalian behavior.

1.2.2 Scaling of differentiation in the brain

We have looked at the manner in which behavioral complexity increases, and now we consider how the brain itself increases in complexity. When a brain is built to do more things, does it do these "more things" via using the same basic building blocks—the same set of neuron types—but by stringing them together into longer functional expressions, or does it achieve greater complexity via the invention of new kinds of basic building blocks—i.e., new neuron types? Consider digital circuits as an example kind of network. Digital circuits consist of logic gates like AND and OR and NOT. For example, AND gates have two inputs and one output, and output a '1' if and only if both inputs are '1'. OR gates output a '1' if and only if at least one of the inputs is a '1'. And a NOT gate has just one input, and outputs the opposite number as the input. The set of all possible digital circuits is an infinite set of circuits, carrying out infinitely many different digital circuit functions. It turns out that, no matter how complex a digital circuit function is, it can be implemented using just these three logic gate types. (In fact, there exists a single logic gate that suffices to build any digital circuit.) No increase in the number of gate types—i.e., no

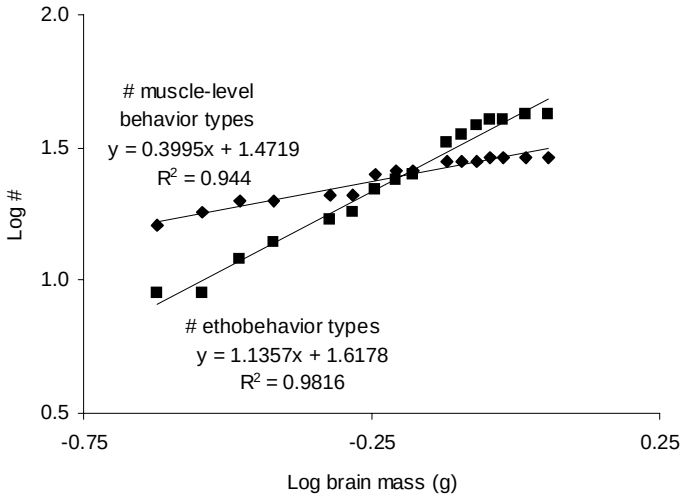


Figure 1.13: Top*: Logarithm (base 10) of the number of muscle-level behavior types versus the logarithm of brain mass (g) for the first 20 days of rat development.* Bottom*: Logarithm (base 10) of the number of ethobehavior types versus the logarithm of brain mass (g) for the first 20 days of rat development. Brain masses taken from Markus and Petit (1987). Ethobehavior types recorded from rat pups during the first 20 days are here recorded, followed by the day of first appearance in at least one pup:* back up, 8; bite cage, 14; bite sib, 15; break self from falling forward, 14; burrow into pile of pups, 1; clean face, 3; clean head, 12; climb wall, 8; dig chips with hands, 13; dig with hind feet, 18; eat chow or poop, 9; fight, 13; free self from pile or mother, 1; grasp bar, 18; grasp feet, 12; head search for nipple, 1; head shake, 4; jump, 15; lick body, 12; lick feet, 8; lick hands, 6; lick sib, 6; lie on back (to lick self), 12; manipulate object, 12; mouth floor, 3; push off pup, 8; righting, 1; run, 12; scratch body with hind leg, 4; scratch ears with front leg, 6; scratch ears with hind legs, 8; seeking nipple, 19; shoveling chips with head, 12; sit on haunches, 12; sleep, 1; sniff air, 10; stand, 14; suckle, 1; turn, 1; walk, 3; walk away from pile, 7; yawn, 1. *Muscle-level behavior types recorded from rat pups during the first 20 days are here recorded, followed by the day of first appearance in at least one pup:* arm lateral push, 2; arm push at elbow, 1; arm push at shoulder, 1; arm push body back, 8; arm stretch, 1; body bend left-right, 1; body bend sit-up, 1; body stretch, 1; body twist, 1; chew, 12; eye open-close, 12; hand grasp, 9; hand to face, 3; head left-right, 1; head twist, 1; head up-down, 1; head rotate, 3; leg burst, 15; leg lateral push, 8; leg push at ankle, 1; leg push at knee, 1; leg stretch, 2; leg to body, 9; leg to face, 8; lick, 6; mouth open-close, 1; suck, 1; tail left-right, 1; tail up-down, 1.

increase in differentiation—needs to occur. For digital circuits there exists a universal language. Perhaps nervous systems are like digital circuits, then: a handful of neuron types are sufficient to carry out any function, and thus brain differentiation remains invariant in more complex brains. The alternative is that there is no universal language employed, and more complex brains have new neuron types.

We address this question first by examining networks generally, rather than just nervous networks. That is, I will present a theory that applies to any kind of network under economic or selective pressure, and then show that many networks, including nervous systems, appear to conform to the theory.

Hypothesis

Nodes in networks combine together to carry out functional expressions, and let L be the average number of nodes involved in an expression. For example, employees in businesses group together to carry out tasks for the business, and neurons in a brain work together to implement brain functions. Let C be, as earlier, the number of component, or node, types; C is a measure of the degree of *differentiation* of the network. If the network can accommodate E distinct expression types, then there must be nodes in the network doing the work. Supposing that, on average, each node can participate in s expressions (where s is a constant depending on the kind of network), the number of nodes in the network, N, must satisfy the inequality

$$N \geq EL/s.$$

For example, if there are $E = 3$ expression types in the network, each of length $L = 4$, and each node can participate in $s = 2$ expressions, then there must be at least $N = 3 \cdot 4/2 = 6$ nodes in the network to realize these expressions.

I am interested here only in networks that are under selective or economic pressure of some kind, and for such networks the following optimality hypothesis plausibly applies (Changizi, 2001d, 2001e; Changizi et al, 2002a): *Network size scales up no more quickly than "needed" to obtain the E expression types.* The motivation for this is that nodes in a network are costly to build and maintain, and network size should accordingly be optimized subject to the functional requirements of the network. Note that networks *not* under selective pressure would not be expected to conform to this hypothesis. For example, a salt crystal is a network with nodes of different types, and the nodes interact with other nodes to carry out functional connective, lattice-related expressions.

Table 1.7: Summary of the kinds of behavior studied. When it was not possible to distinguish between a power law ($C \sim E^a$) and a logarithmic law ($C \sim \log E$), "$\sim$" is written before the rough value of the combinatorial degree to indicate that it might be increasing.

Kind of behavior	Combinatorial degree
Across adults	
Human language over history	Invariant and 5.02
Bird vocalization across phylogeny	Invariant and 1.23
Mammalian behavior across phylogeny	~3.00
Ontogeny	
Ontogeny of language	
- phoneme-morpheme	Increasing from 2 to 4
- word-sentence	Increasing from 1 to 2.5
Ontogeny of behavior	~3

However, salt crystals are not under selective pressure, and the optimality hypothesis does not apply, for a salt crystal that is twice as large will tend to have no more expression types (i.e., no new kinds of interactions among the nodes).

We derived just above that $N \geq EL/s$, and from the optimality hypothesis we may thus conclude that $N \sim EL$. Furthermore, if L were to increase as a function of E, then network size would scale up more quickly than needed, and thus L must be invariant. It follows that

$$N \sim E.$$

For networks under selective pressure, then, we expect network complexity, E, to be directly proportional to network size, N.

How does the network's differentiation, C, relate to network size? Recall from earlier in this chapter that $E \sim C^d$, where d is the combinatorial degree. We may immediately conclude that

$$N \sim C^d.$$

Since we just saw that L must be invariant, d will also be invariant. Therefore, for networks under selective or economic pressure, we predict that network

differentiation and size are related by a power law. Do networks under selective pressure—*selected networks*—conform to this prediction? And, in particular, do nervous networks conform to it? We will see that a wide variety of selected network conform to the predicted relationship, and by measuring the inverse of the log-log slope of C versus N we can, as in the earlier cases of behaviors, compute the combinatorial degree, d.

Example networks, and nervous networks

Changizi et al. (2002a) presented data on the scaling of differentiation in a wide variety of networks, and some of the key plots are shown in Figure 1.14; a summary of the studied networks are shown in Table 1.8. The plots on the left in Figure 1.14 are for human-invented networks, and those on the right are for biological networks. Pairs on the same row are analogous to one another. In particular, (i) Legos are akin to organisms in that in each case geographically nearby nodes interact to carry out functions, (ii) universities are akin to ant colonies in that in each case there are individual animals interacting with one another, and (iii) electronic circuits are akin to nervous systems in that each are electrical, with interconnecting "wires."

The data sources are discussed in Changizi et al. (2002a), and I will only mention the neocortex plot here in detail. The data are obtained from Hof and colleagues, who have used immunoreactive staining and morphological criteria to compare the neuron types in mammals from 9 orders (Hof et al., 1999), and in great ape (Nimchinsky et al., 1999). For each mammalian order, indices of encephalization P (i.e., the brain mass after normalizing for body size) were computed from brain and body weights (grams) for all species in that order found in the following references: Hrdlicka (1907), Bonin (1937), Crile and Quiring (1940), Hofman (1982a, 1982b). Since brain mass scales as body mass to the 3/4 power (Allman, 1999; Changizi, 2001a), P is defined as brain mass divided by body mass to the 3/4 power. Averages were then taken within families, and the family averages, in turn, averaged to obtain the average for an order. Index of neuron encephalization Q (i.e., the number of neurons after normalizing for body size) was computed as $Q = P^{2/3}$, since the number of neurons in neocortex scales as brain volume to the 2/3 power (see previous section). Number of neuron types and index of neuron encephalizations are as follows: Monotremata (7, 0.0699), Artiodactyla (8, 0.0860), Dasyuromorphia (7, 0.1291), Insectivora (8, 0.1339), Rodentia (8, 0.1522), Chiroptera (6, 0.1664), Carnivora (9, 0.1830), Cetacea (9, 0.3094), Primate (not great apes)

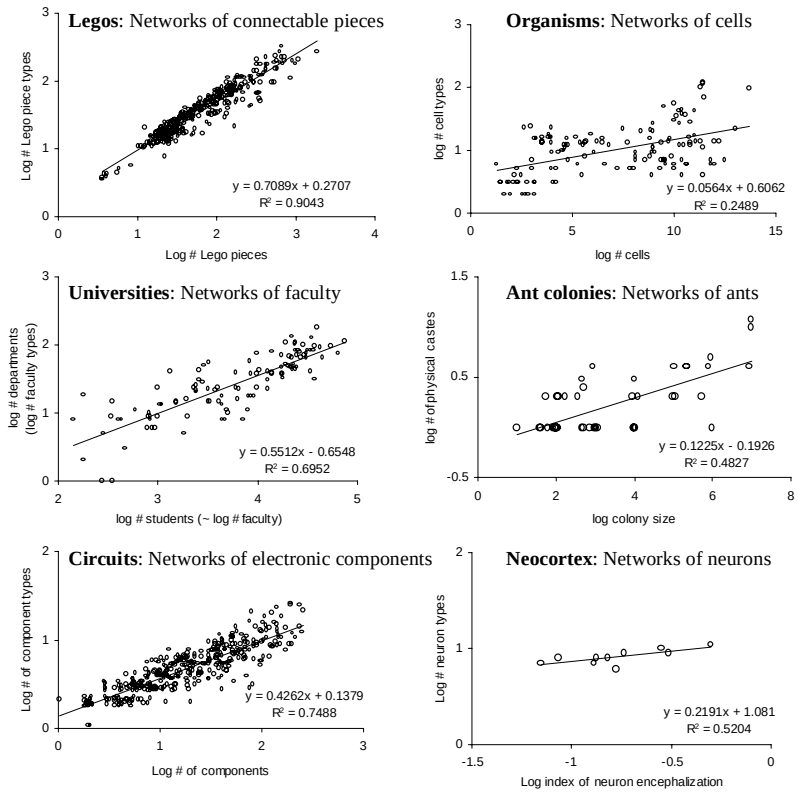


Figure 1.14: Log-log plots of the number of node types versus network size for six kinds of network.

(10, 0.2826), Great Ape (11, 0.4968).

Each of the networks shown in Figure 1.14 and mentioned in Table 1.8 have differentiation increasing as a function of network size. They therefore do not take the universal language approach. Also, the data are consistent with a power law in every case studied thus far, and the logarithmic relationship can be ruled out in the majority of the kinds of network (Changizi et al., 2002a). For neocortex in particular, a logarithmic relationship cannot be excluded (Changizi et al., 2002a) due to the insufficient range. Because of the tendency for selected networks to follow power laws, it seems reasonable to expect that the neocortex does as well, and that with more data a logarithmic relationship could be excluded. In fact, recalling our discussion in the previous section concerning invariant-length minicolumns in neocortex, we have reason to believe that expressions are length-invariant in neocortex, and thus we expect differentiation to scale as a power law of neocortical network size. In sum, then, it appears that, as predicted earlier for network optimality reasons, networks increase in complexity by scaling differentiation as a power law with network size. It also means that in all these networks there are invariant-length expressions; neocortex is hardly, then, unique in this regard.

The combinatorial degree for neocortex is approximately 5—i.e., $N \sim C^5$—and what might this signify? It means that whatever expressions are, there are around five degrees of freedom in their construction. Presumably, most functional expressions in neocortex are carried out by many more neurons than five. That is, it seems plausible that whatever expressions might be, their length L is significantly larger than five. The number of degrees of freedom in an expression may nevertheless be lower than the expression length, as we have seen for human language over history, bird vocalization, and mammalian behavior. What, then, might expressions be given that they have on the order of five degrees of freedom? Consider electronic circuits as an example, where the combinatorial degree is around 2.5. The basic functional expressions here are simple circuits, such as voltage dividers, Zener regulators, and diode limiters (Changizi et al., 2002a), where there are around 2 to 3 electronic components, and this gives the roughly 2 to 3 degrees of freedom, which, in turn, determines the rate at which differentiation scales as a function of network size. For neurons, we must ask what are the functional groupings of neurons in the neocortex? There is no known answer for neocortex here, but one plausible conjecture is the minicolumn, which is a functional grouping of neurons extending along a line through the thickness of the neocortex (Mountcastle, 1957; Tommerdahl et al., 1993; Peters, 1994; Mountcastle, 1997). Minicolumns are invariant in size

Table 1.8: The seven general categories of network for which I have compiled data for scaling of differentiation. The second column says what the nodes in the network are, and the third column gives the estimated combinatorial degree (the inverse of the log-log best-fit slope for differentiation C versus network size N).

Network	Node	Comb. degree
Electronic circuits	component	2.29
Legos™	piece	1.41
Businesses		
- military vessels	employee	1.60
- military offices	employee	1.13
- universities	employee	1.37
- insurance co.	employee	3.04
Universities		
- across schools	faculty	1.81
- history of Duke	faculty	2.07
Ant colonies		
- caste = type	ant	8.16
- size range = type	ant	8.00
Organisms	cell	17.73
Neocortex	neuron	4.56

(see previous section), which is what we expect since the combinatorial degree is invariant. Minicolumns also typically have roughly five layers to them, corresponding to the five cell-rich layers of the neocortex. Perhaps each layer contributes a degree of freedom?

1.3 The shape of limbed animals

Why are limbed animals shaped like they are? Why do animals have as many limbs (digits, parapodia, etc.) as they do? Questions like this are sometimes never asked, it being considered silly, or unscientific, or impossible to answer, or so likely to depend on the intricate ecological details of each individual species that there will be a different answer for each species. Or, if the question is asked using those words, the question will really concern the mechanisms underlying why animals have as many limbs as they do (e.g., certain gene complexes shared by all limbed animals). But the question I asked concerns whether there may be universal principles governing limb number, principles that cut across all the diverse niches and that apply independently of the kinds of developmental mechanisms animals employ.

I began this research (Changizi, 2001a) with the hypothesis that the large-scale shapes of limbed animals would be economically organized. Three reasons motivating this hypothesis were, as mentioned more generally earlier, (1) that animal tissue is expensive and so, all things equal, it is better to use less, (2) that any tissue savings can be used to buy other functional structures, and (3) that economical animal shape can tend to lower information delays between parts of the animal. It is this last motivation that makes this limb problem also a nervous system problem: even if tissue is inexpensive for some species, as long as (i) the animal has a nervous system, and (ii) the animal is under selective pressure to respond to the world relatively quickly, there will be pressure to have a large-scale morphology with low transmission delays.

To make any sense of a hypothesis about optimality, one needs to be precise about what is being optimized. Also, when one says that some shape is optimal, it is implicitly meant that that shape is more economical than all the other shapes in some large class of shapes; so, we must also be clear about what this class of shapes is.

1.3.1 Body-limb networks

To characterize the class of possible (but not necessarily actual) limbed animal shapes, I have developed the notion of a *body-limb network*. The basic idea of a body-limb network is to treat the body and limb tips of an animal as nodes in a network, and the limbs as edges connecting the body to the limb tips. The limbs are required to emanate from the body at points that lie along a single plane—this is the *limb plane*, and the cross-section of the body lying in this plane is what we will represent with our node for the body. More precisely, a body-limb network is any planar network with a central *body node*, and any number of *limb tip nodes* uniformly distributed at some distance X from the body node. Edges are all the same cost per unit length, and may connect any pair of nodes. When an edge connects the body node to a limb tip node, the edge is called a *limb edge*, or a *limb*. Figure 1.15 shows some example body-limb networks. In every network I have ever studied, nodes are points. For the purpose of characterizing animal bodies, this will not do: animal bodies are often not point sized compared to limb length. To accommodate this, body nodes are allowed to have a size and shape. For example, they are circles in Figure 1.15, and a *stretched circle* is shown in Figure 1.16. [Stretched circles are circles that have been cut in two equal halves and pulled apart a stretched-circle length L.] Body-limb networks are general enough to cover both many animal-like networks—e.g., a starfish—and many non-animal-like networks. Body-limb networks with body nodes having stretched-circle shapes have the following important parameters (see Figure 1.16):

- The body radius, R. I.e., the distance from the body's center to the body's edge. This parameter accommodates all those animals for which the body is not negligible compared to limb length.
- The stretched-circle length, L. I.e., the "length" of the body, but where $L = 0$ implies that the body node is a circle. This parameter accommodates long animals, like millipedes.
- The distance from the body node's edge to a limb tip, X. When there are edges from the body to a limb tip, these edges are limb edges, and X is then the limb length. More generally, though, X is the separation between the body-node and the limb-tip nodes. Since a connected body-limb network will always have at least one limb, this distance is always the length of this one limb, at least; accordingly, I will typically refer to it as the limb length.
- The number of limbs, N. I.e., the number of edges actually connecting the body node to a limb tip. To emphasize, N is not the number of limb tip nodes, but the number of limb edges; thus, the number of limb tip nodes must be $\geq N$.

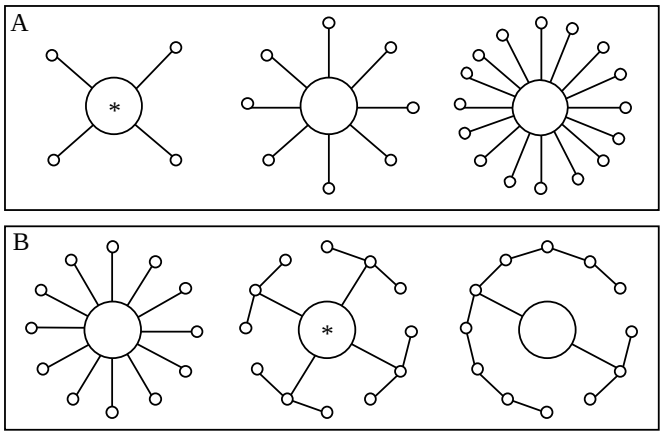


Figure 1.15: Some example body-limb networks with the same body radius and limb length [as well as the same stretched-circle length (namely 0), see Figure 1.16], and limb length. Body nodes are not required to be points; here they are circles. (A) Three body-limb networks where all the edges are limbs, but where some networks have more limb tip nodes than others. (B) Three body-limb networks where there are just as many nodes as one another, but where some have more limb edges than others. The networks with asterisks on the body-limb node have different numbers of limb tip nodes, but have the same number of limbs.

To the extent that real limbed animals can be treated as body-limb networks, the treatment is extremely crude. The limb tips in a body-limb network are all equidistant from the body node, whereas real limbed animals often have limb length variation. They are also required to be uniformly distributed around the body, but real animals often violate this. The edges in body-limb networks must have equal costs per unit length, but real animals sometimes have limbs with different cross-sectional areas (and thus different costs per unit length). The positions of the nodes in a body-limb network are all fixed in place, whereas limbed animals move their limbs. Furthermore, although the limbs of an animal might emanate from the animal along a single plane—the limb plane—and although limbs of many animals can, if the animal so wishes, lie roughly flat in that plane, animals rarely keep their limbs within this limb plane. For example, the limbs of an octopus emanate from along the same planar cross-section of the animal, and the limbs *can* lie flat in the plane; but they rarely if ever do. With regard to reaching out into the world, there *is* something special about the plane, special enough that it justifies modeling the shape of

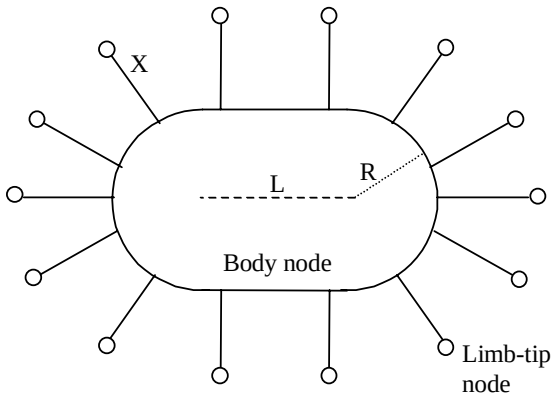


Figure 1.16: An example body-limb network with a stretched circle body node. The limb ratio *is* $k = X/(R + X)$*; the* stretched-circle ratio *is* $s = L/X$.

animals like an octopus as if the limbs are always lying in the plane. Imagine that all the limbs have the same angle relative to the plane; e.g., they are all pointing down and out of the plane, with an angle of 30° with the plane, as is depicted in Figure 1.17. For each such "way of pointing the limbs," let us calculate the total perimeter made by drawing lines connecting the limb tip nodes. Now ask ourselves, At which angle relative to the plane is this perimeter the greatest? Well, it is least when all the limbs are pointing either straight down or straight up; it is greatest when the limbs are lying in the limb plane. Animals have limbs in order to reach out, and since there is more reaching out to do when the limbs are in the limb plane, we might expect that it is the geometry when in the limb plane that is the principal driving force in determining the nature of the network. If an animal's limbs cannot lie in the limb plane—as is, for example, the case for most mammals, who have ventrally projected limbs—then they cannot be treated via body-limb networks as I have defined them. Despite all these idealizations, body-limb networks allow us to capture the central features of the highest level descriptions of limbed animals, and these networks are simple enough that we can easily think about them, as well as answer questions about optimality.

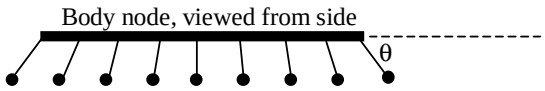


Figure 1.17: Real limbed animals often project their limbs out of the limb plane. [The limb plane *is the plane defined by the points where the limbs intersect the body. It is also the plane in which the body node lies.] This figure shows an example "animal" viewed from the side, where all the limbs are pointing below the body at an angle θ relative to the limb plane. The perimeter made by the limb tips is greatest when θ is zero, i.e., when the limbs lie in the plane. There is accordingly the greatest need for limbs in the limb plane, and this is my justification for treating limbed animals as if their limbs lie in the limb plane.*

1.3.2 The optimization hypothesis

We now know what body-limb networks are, and how they may be used, to a first approximation at least, to characterize the large-scale morphology of many kinds of limbed animals. They are also sufficiently general that there are many body-limb networks that do not describe real limbed animals. The question now is, If limbed animals *are* economically arranged, then what body-limb networks would we expect to describe them? Or, said another way, which body-limb networks are optimal? To make this question more precise, suppose that an animal has body radius R, stretched-circle length L, and limb length X. Now we consider the class of all body-limb networks having these three values—the class of "R-L-X body-limb networks"—and ask, Which ones are optimal? For example, all the example networks in Figure 1.15 have the same body radius, same stretched-circle length (namely zero), and same limb length; they are therefore all in the same class of body-limb networks from which we would like to find the optimal one. However, rather than asking which such body-limb network is optimal, I will ask a weaker question: How many limbs does an optimal R-L-X body-limb network have? The reason I want to ask this question is that, ultimately, it is the number of limbs that I am interested in. From our point of view, two body-limb networks that differ in their number of limb tip nodes but have the same number of limb edges are the same. For example, the networks with asterisks in Figure 1.15 have the same number of limbs, and so we do not wish to distinguish them.

The answer to the question "How many limbs does an optimal R-L-X body-limb network have?" is roughly that these networks cannot have too

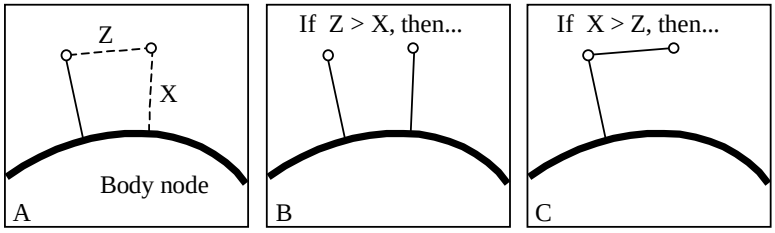


Figure 1.18: The basic idea behind the argument for why there cannot be too many limbs in an optimal body-limb network. (A) Part of the body node is shown at the bottom, two limb tip nodes on top. One limb edge is presumed to already exist. To connect the network, the other limb tip node must either have an edge straight to the body node, which is of length X, or have an edge to the tip of the existing limb, which is of length Z. (B) When $Z > X$ it is less costly to have a limb go to the limb tip node straight from the body node. (C) But when $X > Z$ it is cheaper to have an edge go to the tip of the existing limb.

many limbs, where "too many" depends on the parameter values of R, L and X. Figure 1.18 illustrates the argument. The basic idea is that if two limb tip nodes are close enough to one another, then it is cheaper to send an edge directly from one to the other, and to have only one of the limb tips connect to the body node. This occurs when the distance, Z, between the limb tips is smaller than the limb length; i.e., when $Z < X$. However, when limb length is smaller than the distance between the limb tips—i.e., when $X < Z$—it is cheaper to connect the limb tip nodes directly to the body node. That is, it is then cheaper to have a limb for each limb tip node. With this observation in hand, we can say that *an optimal R-L-X body-limb network must have its limbs sufficiently far apart that no limb tip nodes at the end of a limb are closer than X.*

Because our body node shapes are confined to stretched circles, it is not difficult to calculate what this means in regards to the maximum number of limbs allowed for an optimal, or wire-minimal, R-L-X body-limb network. Let us consider the stretched circle's two qualitatively distinct parts separately. First consider the straight sides of the body node. These sides are of length L, and the limbs are all parallel to one another here. It is only possible to fit L/X many limbs along one of these edges. Actually, L/X can be a fraction, and so we must round it down; however, for simplicity I will ignore the truncation from now on, and compute just the "fractional number of limbs." So, along the two sides of a stretched circle there are a maximum of $2L/X$ limbs; letting $s =$

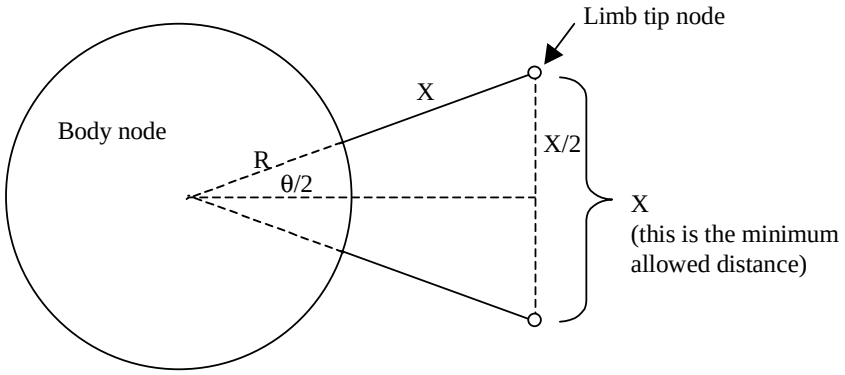


Figure 1.19: The simple trigonometry involved in computing the minimum allowed angle between two limbs for a circular node. The two nodes cannot be closer than X. We can compute $\theta/2$ as the $\arcsin[(X/2)/(R+X)]$. Since the limb ratio—i.e., a measure of how long the limbs are compared to the body—is $k = X/(R+X)$, we can rewrite this as $\theta = 2\arcsin(k/2)$. (An alternative derivation leads to the equivalent $\theta = \arccos(1 - k^2/2)$.)

L/X be the *stretched-circle ratio*, the maximum number of limbs is $2s$. The remaining parts of the body node are two semicircles, which we will imagine pushing together. Limbs on a circular body node poke out radially. Consider the angle between the lines reaching from the body node to two limb tips. What must this angle be in order to make the distance between the two limb tip nodes greater than the limb length X? Figure 1.19 illustrates the simple trigonometry involved. The conclusion is that, for circular body nodes, the angle, θ, between adjacent limbs must satisfy the inequality

$$\theta \geq 2\arcsin(k/2),$$

(or equivalently $\theta \geq \arccos(1 - k^2/2)$), where $k = X/(R + X)$ is the *limb ratio*. The maximum number of limbs that can be placed around a circular body node is therefore

$$\frac{2\pi}{2\arcsin(k/2)} = \frac{\pi}{\arcsin(k/2)}.$$

In total, then, for an R-L-X body-limb network to be optimally wired the number of limbs N must satisfy the inequality,

$$N \leq N_{max} = 2s + \frac{\pi}{\arcsin(k/2)},$$

where $s = L/X$ and $k = X/(R + X)$. Note that this inequality no longer refers to the body radius R, the stretched-circle length L or the limb length X. Instead, it refers only to the stretched-circle ratio s and the limb ratio k. The absolute size of the network therefore does not matter; all that matters are the relative proportions of an animal's body and limbs. It should be noted that this treatment of stretched circle nodes engages in a simplification since I have made the argument for the sides separately from that for the circular ends; a more precise mathematical treatment would determine the maximum number of limbs for the stretched-circle node shape as it is. For our purposes this approximation suffices. The notion of optimality we have employed here is something called a *minimal spanning tree*, or *MST*. *Spanning trees* are networks that connect up all the nodes, but where there are no loops. *Minimal* spanning trees are spanning trees that use the least amount of wire. What we have found thus far is that *if* an R-L-X body-limb network is a minimal spanning tree, *then* it must have fewer than N_{max} limbs.

That encompasses the volume-optimality part of the hypothesis. All it concludes, though, is that there must not be more than N_{max} many limbs; it does not predict how many limbs an animal will actually have. This is where I made a second hypothesis, which is that animals are typically selected to maximize their number of limbs subject to the volume-optimality constraint. The simple intuition is that limbed animals have limbs in order to reach out (for many different reasons), and need to "cover" their entire perimeter.

These two hypotheses lead to the prediction that, for those limbed animals describable as R-L-X body-limb networks, the number of limbs N satisfies the equation,

$$N = 2s + \frac{\pi}{\arcsin(k/2)}.$$

(I.e., that $N = N_{max}$.) Because the first hypothesis concerns minimal spanning trees and the second concerns maximizing the number of limbs, I have labeled this composite hypothesis the *max-MST hypothesis*. Notice that the max-MST hypothesis says nothing about life as we know it; it is a general hypothesis, so general that one might expect it to apply to any limbed animals anywhere, so long as they are describable by body-limb networks.

Let us ask what this equation means for the relationship between predicted limb number and the body and limb parameters s (the stretched-circle ratio $s = L/X$) and k (the limb ratio $k = X/(R+X)$). First consider what happens as s is manipulated. When $s = 0$ it means that the stretched-circle length is very small compared to the limb length. The consequence is that the stretched-

circle term in the equation for the number of limbs drops out, which means that the network can be treated as having a circular body node. As s increases, and keeping R and X constant, the equation is of the form $N = 2s + N_c(k)$, where $N_c(k)$ is a constant referring to the number of limbs for a circle node with limb ratio k. Thus, N increases proportionally with s. For this reason, the stretched-circle length parameter is rather uninteresting; that is, it just leads to the obvious prediction that, for sufficiently large values of s, animals with bodies twice as long have twice the number of limbs. Now consider what happens as the limb ratio is manipulated. When $k = 1$ it means the limbs are very long compared to the body radius, and the number of limbs becomes $N = 2s + 6$. When the body node is circular $s = 0$ and $N = 6$; that is, when the limbs are so long that the body node may be treated as a point, the predicted number of limbs falls to its minimum of 6. As k approaches zero the limbs become very short compared to the body radius. Using the approximation $x \approx \sin x$ for x near 0 radians, it follows that $\sin(k/2) \approx k/2$, and so $\arcsin(k/2) \approx k/2$, and the predicted number of limbs becomes

$$N \approx 2s + \pi/(k/2) = 2s + 2\pi/k.$$

In fact, even when k is at its maximum of 1, $\arcsin(k/2) \approx k/2$; e.g., $\arcsin(0.5) = 0.52 \approx 0.5$. The error at this maximum is only about 4%, and the error gets lower and lower as k drops toward zero. Therefore, the approximation above is *always* a reasonable one. When the body node is either a circle or the limb length is very large compared to the stretched-circle length (but still much smaller than the body radius), the equation becomes $N \approx 2\pi/k$. That is, the number of limbs becomes inversely proportional to the limb ratio. In short, when $s = 0$, the number of limbs falls to six for very long limbs compared to the body, but increases toward infinity in a particular quantitative fashion as the limbs become shorter relative to the body. The reader may examine the kinds of body-limb networks that conform to the hypothesis by playing with a little program built by Eric Bolz at www.changizi.com/limb.html.

Before moving to data, it is important to recognize that the hypothesis does not apply to animals without limbs. The hypothesis states that there is a relationship between an animal's number of limbs and its body-to-limb proportion (i.e., limb ratio). Without limbs, the model can say nothing. Alternatively, if having no limbs is treated as having zero limb ratio, then the model predicts infinitely many non-existent limbs. Snakes and other limbless organisms are therefore not counterexamples to the max-MST hypothesis.

1.3.3 Comparing prediction to reality

At this point I have introduced the prediction made by the max-MST hypothesis. With this prediction in hand, I sought to discover the extent to which real limbed animals conform to the prediction. To obtain data for actual body-limb networks, I acquired estimates of the stretched-circle ratio s and the limb ratio k from published sources for 190 limbed animal species over 15 classes in 7 phyla (Agur, 1991; Barnes, 1963; Bishop, 1943; Brusca and Brusca, 1990; Buchsbaum, 1956; Buchsbaum et. al., 1987; Burnie, 1998; Downey, 1973; Hegner, 1933; Netter, 1997; Parker, 1982; Pearse et. al., 1987; Pickwell, 1947; Stebbins, 1954). The studied phyla (classes) were annelids (Polychaeta), arthropods (Myriapoda, Insecta, Pycnogonida, Chelicerata, Malacostraca), cnidarians (Hydrozoa, Scyphozoa), echinoderms (Holothuroidea, Asteroidea), molluscs (Cephalopoda), tardigrades and vertebrates (Mammalia (digits only), Reptilia (digits only), Amphibia). An appendix subsection at the end of this section shows these values. Measurements were made on the photographs and illustrations via a ruler with half millimeter precision. The classes were included in this study if six or more data points from within it had been obtained. Species within each class were selected on the basis of whether usable data could be acquired from the sources above (i.e., whether the limb ratio and stretched-circle ratio were measurable); the number of limbs in the measured animals ranged from 4 to 426. What counts as a limb? I am using 'limb' in a general sense, applying to any "appendage that reaches out." This covers, e.g., legs, digits, tentacles, oral arms, antennae and parapodia. Although for any given organism it is usually obvious what appendages should count as limbs, a general rule for deciding which appendages to count as limbs is not straightforward. Some ad hoc decisions were required. For vertebrate legs only the those of Amphibia were studied, as their legs are the least ventrally projected of the vertebrates. For amphibians, the head and tail were included in the limb count because there is an informal sense in which the head and tail also "reach out". (Thus, amphibians have six "limbs" in this study.) For insects (and other invertebrates with antennae studied), antennae appear to be similar in "limb-likeness" to the legs, and so were counted as limbs unless they were very small (around $< 1/3$) compared to the legs. The head and abdomen of insects were not counted as limbs because, in most cases studied, they are well inside the perimeter of the legs and antennae, and thus do not much contribute to "reaching out" (the head was treated as part of the body). Since I obtained the data for the purposes of learning how body-limb networks scale up when there are

more limbs, and since scaling laws are robust to small perturbations in measurement (being plotted on log-log plots), these where-to-draw-the-line issues are not likely to much disturb the overall scaling behavior. Digits are treated in the same manner as other types of limbs, the only difference being that only a fraction of the body (i.e., hand) perimeter has limbs (i.e., digits). Cases of digits were studied only in cases where the "hand" is a stretched circle with digits on roughly one half of the stretched circle. For these cases hands may be treated as if the digits emanate from only one "side" of the node. Digits like those on a human foot are, for example, not a case studied because the foot is not a stretched circle for which the toes are distributed along one half of it. In 65 of the cases presented here the stretched-circle ratio $s \neq 0$, and to observe in a single plot how well the data conform to the max-MST hypothesis, the dependence on the stretched-circle length can be eliminated by "unstretching" the actual number of limbs as follows: (i) given the limb ratio k and the stretched-circle ratio s, the percent error E between the predicted and actual number of limbs is computed, (ii) the predicted number of limbs for a circular body is computed by setting $s = 0$ (and keeping k the same), and (iii) the "unstretched actual number of limbs" is computed as having percent error E from the predicted number of limbs for a circular body. This rids of the dependence on s while retaining the percent error.

After unstretching, each measured limbed animal had two remaining key values of interest: limb ratio k and number of limbs N. The question is now, How do N and k relate in actual organisms, and how does this compare to the predicted relationship? Recall that, for $s = 0$ as in these unstretched animals, the predicted relationship between N and k for limbed animals is

$$N \approx 2\pi/k.$$

If we take the logarithm of both sides, we get

$$\log_{10} N \approx \log_{10}(2\pi/k),$$

$$\log_{10} N \approx -\log_{10} k + \log_{10}(2\pi) = -\log_{10} k + 0.798.$$

Therefore, if we plot $\log_{10} N$ versus $-\log_{10} k$, the predicted equation will have the form of a straight line, namely with equation $y = x + 0.798$. This is shown in the dotted lines in Figure 1.20.

Figure 1.20 shows a plot of the logarithm (base 10) of the number of limbs versus the negative of the logarithm (base 10) of the limb ratio for the data I acquired. If the max-MST hypothesis is true, then the data should closely

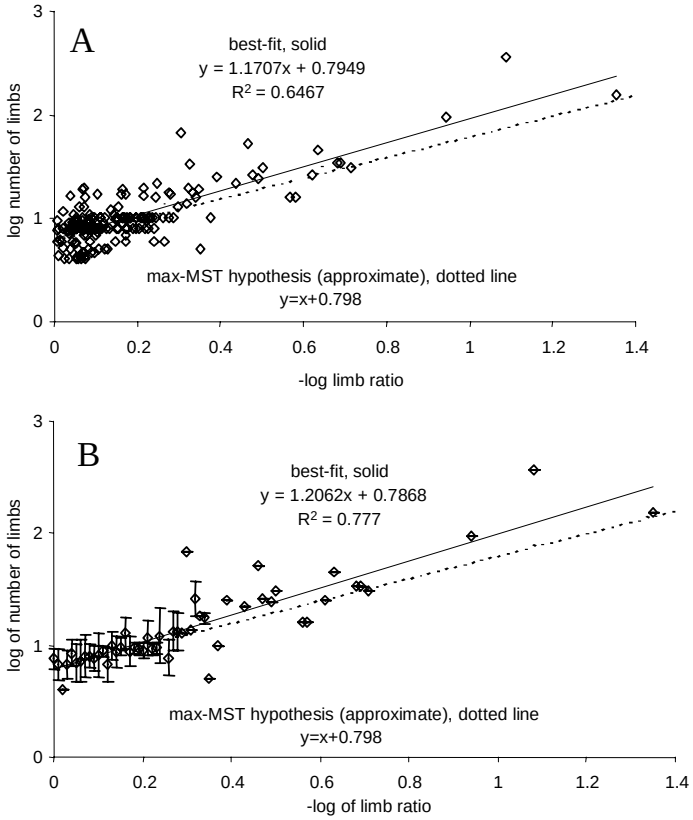


Figure 1.20: (A) The logarithm (base 10) of the unstretched number of limbs versus the negative of the logarithm of the limb ratio, for all 190 limbed animals. The best fit equation via linear regression is $y = 1.171x + 0.795$ *(solid line) (*$R^2 = 0.647$, $n = 190$, $p < 0.001$*), and predicted line* $y = x + 0.798$ *(dotted line). The* 95% *confidence interval for this slope is* $[1.047, 1.294]$*. The three rightmost data points exert a disproportionate influence on the best-fit line, and removing them leads to the best fit equation* $y = 1.089x + 0.8055$ *(*$R^2 = 0.487$, $n = 187$, $p < 0.001$*), with a* 95% *confidence interval for the slope of* $[0.900, 1.279]$*. (B) The average of* $\log_{10} N$ *values versus* $-\log_{10} k$*, where the* $-\log_{10} k$ *values are binned with width* 0.01*. Error bars indicate standard deviation (for points obtained from bins with 2 or more cases). The best fit equation is now* $y = 1.206x + 0.787$ *(solid line) (*$R^2 = 0.777$, $n = 52$, $p < 0.001$*), again very close to the predicted line (dotted line). points still exert a disproportionate influence on the best-fit line, and removing them results in the equation* $y = 1.112x + 0.807$ *(*$R^2 = 0.631$, $n = 49$, $p < 0.001$*).*

follow the equation $y = x + 0.798$ in the plot, shown as dotted lines. Examination of the plots show that the data closely follow the predicted lines. When $-\log(k) = 0$, $k = 1$, meaning that the body radius R is extremely small compared to the limb length X; and when this is true, the number of limbs falls to a minimum of around six (see legend of Figure 1.20). As $-\log(k)$ increases, the limb ratio decreases toward 0, meaning that the limbs are getting smaller relative to the body radius; and when this is true, the number of limbs increases higher and higher. Not only does limb number clearly increase as limb ratio decreases (and the x axis increases), it appears to be well described by the linear regression equation $\log(N) = 1.171[-\log(k)] + 0.795$ (and, without the three points on the far right, $\log(N) = 1.089[-\log(k)] + 0.8055$). Manipulation of this equation leads to $N = 6.24k^{-1.171}$ (and without the three stray points, $N = 6.39k^{-1.089}$): the number of limbs appears to be roughly inversely proportional to the limb-ratio, with a proportionality constant around 6. This is extraordinarily similar to the predicted relationship which, recall, is $N = 6.28k^{-1}$.

In summary, many limbed animals across at least seven phyla conform well to the max-MST hypothesis, which suggests that their large-scale morphologies are arranged to minimize the amount of tissue needed to reach out in the world; they also appear to have the maximum number of limbs subject to the constraint that they are still optimal trees. And this is despite the complete lack of any details in the hypothesis concerning the ecological niches of the animals, and despite the extreme level of crudeness in the notion of body-limb networks. It is worth emphasizing that, even without the max-MST hypothesis to explain the data, these empirical results are interesting because they reveal that limbed animals follow universal laws relating their body-to-limb ratio to their number of limbs. It happens that this universal law is just what one might *a priori* suspect of limbed animals—as I *a priori* suspected—if they are driven by volume-optimization considerations. It is also worth mentioning that this limb problem is a kind of network scaling problem: the issue is, what changes do body-limb networks undergo as they acquire more limbs? That is, how do animals change as their number of limbs is scaled up? The answer is that limbed animals scale up in such a way as to keep the value $N \cdot k$ invariant; and, in particular, limbed animals satisfy the constraint that $N \cdot k \approx 2\pi$.

Appendix for section: Raw limb data

In this appendix I have included my raw limb data. It appears on the following three consecutive pages, with the phylum, class (with type of "limb" in parentheses), name of animal (species name, or whatever information was available from the source), limb ratio ($X/(R+X)$), stretch ratio (L/X), and the number of limbs. I mention in passing that it may be interesting to look at conformance to this model in two new ways. One, to look at spherical nodes, where the limbs point radially outward in all directions; mathematical research from Coxeter (1962) can be used to determine roughly how many limbs are optimal. [The predicted relationship is $N \approx 4\pi/k^2$, where k is again the limb ratio.] Second, one may look at non-animals, and perhaps even viruses: e.g., the T4 bacteriophage conforms well to the model, having six "lunar-lander-like" limbs attached to a very small "body" (the shaft).

Phylum	Class (limb type)	Name	limb ratio X/(R+X)	stretch ratio (L/X)	# limbs
Annelida	Polychaeta (parapodia)	Glycera americana	0.3658	158.0550	426
		Tomopteris	0.1929	11.3734	52
	(parapodia, long)	Halosydna	0.3214	5.2222	37
	(parapodia, short)	Halosydna	0.2400	7.8333	41
		Syllis cornuta	0.9102	2.6316	30
		Nereis virens	0.4545	23.6000	72
		unnamed	0.3333	69.7500	218
		Nereis diversicolor	0.6087	10.4286	50
Arthropoda	Myriapoda (legs)	Lithobius	0.7865	3.4611	32
		Scolopendra gigantea	0.7037	7.7895	36
		a California centipede	0.4795	11.5429	38
		Scutigera coleoptrata	0.8474	2.5210	34
		Scolopendra cingulata	0.5233	8.5506	42
		Scutigerella	0.6061	9.5333	28
		a millipede	0.6842	8.6923	56
	Insecta (legs and antennae)	Thraulodes salinus	0.7929	0.5732	8
		Pediculus humanus	0.7108	0.0407	6
		Phthirus pubis	0.5725	0.0000	6
		a cockroach	0.9696	0.8784	8
		Microcoema camposi	0.9555	0.5276	6
		Lonchodes brevipes	0.9764	1.1774	6
		Velinus malayus	0.8919	0.1212	6
		an ant	0.8810	0.1622	6
	Pycnogonida (legs)	Nymphopsis spinosossima	0.9189	0.1961	8
		Achelia echinata	0.8882	0.2517	10
		Dodecolopoda mawsoni	0.9556	0.1395	12
		Decolopoda australis	0.9808	0.2157	10
		Tanystylum anthomasti	0.9218	0.3392	8
		Nymphon rubrum	0.9853	0.1660	8
	Chelicerata (legs)	spider larva	0.8123	0.1320	8
		spider nymph	0.8364	0.0372	8
		spider	0.8467	0.0849	8
		Argiope	0.9015	0.0000	8
		Scytodes	0.8551	0.0000	8
		Pardosa amentata	0.8952	0.1064	8
		a generalized spider	0.8571	0.1667	8
		a spider (in amber)	0.8815	0.0640	8
		a crab spider	0.8545	0.0000	8
		Tegenaria gigantea	0.8956	0.0245	8
		Brachypelma emilia	0.7947	0.0397	8
		Buthus martensi	0.8556	0.2412	10
		Ricinoides crassipalpe	0.8464	0.0000	8
		unnamed	0.9256	0.2009	8
		daddy long legs	0.9735	0.0181	8
		Mastigoproctus	0.8773	0.2721	10
		Heterophrynus longicornis	0.9167	0.0121	8
		Stegophrynus dammermani	0.8802	0.0000	8
		Koenenia	0.8477	0.5689	10
		Galeodes arabs	0.8991	0.2801	10
		Chelifer cancroides	0.8985	0.3349	10
		Eurypterus	0.5145	0.0000	10
		Pterygotus buffaloensis	0.7316	0.0000	12
		Limulus	0.9231	0.3750	10
	Malacostraca (legs)	Pachygrapsus crassipes	0.6785	0.0000	10
		Chionoecetes tanneri	0.7745	0.0000	10
		Gecarcoidea natalis	0.6537	0.0000	10
		Carcinus maenas	0.8261	0.0702	10
		Maja squinado	0.7078	0.0183	10
		Callianassa	0.7625	0.4177	10
		Pleuroncodes planipes	0.7874	0.3681	8
		Petrolisthes	0.6676	0.1770	10
		Cryptolithodes	0.8229	0.3472	10
		a crab	0.8012	0.0025	10
		Loxorhynchus	0.7035	0.1533	10
		Pugettia	0.7111	0.1060	10
		Stenorhynchus	0.9554	0.0800	8

Phylum	Class (limb type)	Name	limb ratio X/(R+X)	stretch ratio (L/X)	# limbs
Cnidaria	Hydrozoa (tentacles)	Hydra A	0.9815	0.0000	6
		Hydra B	0.8041	0.0000	8
		Hydra C	0.9620	0.0000	6
		Polyorchis	0.4719	0.0000	33
		Tubularia hydroid adult polyp A	0.6111	0.0000	20
		Tubularia hydroid adult polyp B	0.7536	0.0000	10
		Tubularia hydroid actinula larva	0.8264	0.0000	8
		Tubularia hydroid new polyp	0.9048	0.0000	9
		Tubularia indivisa hydroid	0.8387	0.0000	16
		Niobia medusa	0.5426	0.0000	6
		Sarsia medusa	0.8784	0.0000	4
		Rathkea medusa	0.3137	0.0000	31
		"typical" medusa	0.7957	0.0000	9
		Proboscidactyla	0.2071	0.0000	34
		Obelia medusa	0.3404	0.0000	52
		"typical" medusa	0.4950	0.0000	68
		a hydranth A	0.6747	0.0000	6
		a hydranth B	0.7188	0.0000	11
		a hydranth C	0.6897	0.0000	17
		Linmocnida medusa	0.4750	0.0000	20
		Aglaura medusa	0.2308	0.0000	45
	Scyphozoa (tentacles)	Stomolophus meleagris scyphistoma	0.8511	0.0000	13
		Stomolophus meleagris stobila	0.7391	0.0000	8
		Stomolophus meleagris late strobila	0.8458	0.0000	7
		Stomolophus meleagris ephyra	0.2712	0.0000	16
		Cassiopea andromeda	0.5946	0.0000	8
		Mastigias medusa	0.8485	0.0000	8
		Haliclystis	0.8333	0.0000	8
		Pelagia adult scyphomedusa	0.7748	0.0000	8
	(oral arms)	Pelagia adult scyphomedusa	0.9412	0.0000	4
		Aurelia adult medusa	0.0444	0.0000	154
		Aurelia ephyra	0.2632	0.0000	16
		Aurelia scyphistoma A	0.5652	0.0000	22
		Aurelia scyphistoma B	0.6735	0.0000	17
		Aurelia scyphistoma C	0.8317	0.0000	8
		"typical" medusa A	0.1136	0.0000	96
	(oral arms)	"typical" medusa A	0.8861	0.0000	4
		"typical" medusa B	0.0816	0.0000	368
	(oral arms)	"typical" medusa B	0.9231	0.0000	4
Echinodermata	Holothuroidea (arms)	Cucumaria crocea	0.8864	0.0000	10
		Cucumaria planci	0.8571	0.0000	10
		Enypniastes	0.4643	0.0000	18
		Pelagothuria	0.8958	0.0000	11
		Holothuria grisea	0.5313	0.0000	18
		Stichopus	0.5926	0.0000	10
		Euapta	0.7059	0.0000	8
	Asteroidea (arms)	Luidia phragma	0.7547	0.0000	5
		Luidia ciliaris	0.8590	0.0000	8
		Luidia sengalensis	0.8235	0.0000	9
		Luidia clathrata	0.8281	0.0000	5
		Ctenodiscus	0.7451	0.0000	5
		Astropecten irregularis	0.7800	0.0000	5
		Heliaster microbranchius A	0.2040	0.0000	34
		Heliaster microbranchius B	0.4051	0.0000	25
		Solaster	0.7733	0.0000	10
		Acanthaster planci	0.4500	0.0000	19
		Pteraster tesselatus	0.4455	0.0000	5
		Solaster notophrynus	0.6741	0.0000	7
		Linckia guildingii	0.8752	0.0000	5
		Linckia bouvieri	0.9148	0.0000	5
		Ampheraster alaminos	0.9091	0.0000	6
		Odinia	0.8553	0.0000	19
		a starfish A	0.5476	0.0000	10
		a starfish B	0.8395	0.0000	8
		Freyella	0.8717	0.0000	13
		Crossaster papposus	0.5050	0.0000	13
		Coscinasterias tenuispina	0.8333	0.0000	7
		Coronaster briorcus	0.8182	0.0000	11

Phylum	Class (limb type)	Name	limb ratio X/(R+X)	stretch ratio (L/X)	# limbs
Mollusca	Cephalopoda (arms)	Sepia A	0.7463	0.0000	8
		Sepia B	0.7755	0.0000	8
		Architeuthis	0.8889	0.0000	8
		Octopus	0.8826	0.0000	8
		Octopus dofleini	0.9111	0.0000	8
		Octopus vulgaris	0.9068	0.0000	8
		Loligo	0.8163	0.0000	8
		Loligo pealeii	0.7987	0.0000	8
		Histioteuthis	0.8224	0.0000	8
		a juvenile	0.7000	0.0000	8
Vertebrata	Amphibia (limbs, tail and head)	a salamander A	0.8458	1.0666	6
		a salamander B	0.8128	1.1219	6
		a salamander C	0.8653	1.6132	6
		a salamander D	0.8636	1.2982	6
		a salamander E	0.8854	1.4243	6
		a salamander F	0.8452	1.5385	6
		a salamander G	0.8582	1.3125	6
		a salamander H	0.8582	1.0593	6
		a salamander I	0.8438	1.7284	6
		a salamander J	0.7971	1.0727	6
Tardigrada digits	(digits)	Echiniscus	0.6667	0.0000	8
		Halobiotus crispae	0.6667	0.0000	8
		Echiniscoides sigismundi	0.5714	0.0000	16
		Wingstrandarctus corallinus	0.8384	0.0000	8
		Styraconyx qivitoq	0.7857	0.0000	8
		Halechiniscus	0.8530	0.0000	8
		Orzeliscus	0.8000	0.0000	8
		Batillipes	0.8225	0.0000	6
Vertebrata digits	Mammalia (digits)	homo sapien A	0.5879	0.0000	10
		homo sapien B	0.6135	0.0000	10
		homo sapien C	0.5889	0.0000	10
		homo sapien D	0.5782	0.0000	10
		chimpanzee	0.6358	0.0000	10
		Tarsius bancanus	0.7059	0.0000	10
	Reptilia (digits)	Triturus cristatus (rear limb of a newt)	0.6691	0.0000	10
		Triturus cristatus (front limb of a newt)	0.8023	0.0000	8
		Sceloporus occidentalis biseriatus	0.6861	0.0000	10
		Lacerta lepid (a lizard)	0.6067	0.0000	10
		Cnemidophorus tessalatus tessellatus	0.6833	0.0000	10
		Eumeces skiltonianus (a skink)	0.7738	0.0000	10
		Dasia (a skink)	0.8108	0.0000	10
	Amphibia (digits)	a salamander front limb 1	0.7699	0.0000	8
		a salamander rear limb 2	0.4202	0.0000	10
		a salamander front limb 3	0.6329	0.0000	8
		a salamander rear limb 4	0.5868	0.0000	10
		a salamander front limb 5	0.6237	0.0000	8
		a salamander rear limb 6	0.6209	0.0000	10
		Plethodon vandyke (front limb)	0.6053	0.0000	8
		Plethodon vandyke (rear limb)	0.5704	0.0000	10
		Anneides lugubris (front limb)	0.6464	0.0000	8
		Anneides lugubris (rear limb)	0.6578	0.0000	10
		Laeurognathus marmorata (a front limb)	0.6732	0.0000	8
		Laeurognathus marmorata (a rear limb)	0.6552	0.0000	10
		Pseudotriton ruber ruber (a front limb)	0.5769	0.0000	8
		Pseudotriton ruber ruber (a rear limb)	0.5303	0.0000	10
		Plethodon vehiculum (a rear limb)	0.6358	0.0000	10