

Universal laws for hierarchical systems

Mark A. Changizi*

Department of Psychological and Brain Sciences

Duke University

Durham, NC 27708-0086

changizi@changizi.com

www.changizi.com

Abstract

Organisms are hierarchically organized, both structurally and behaviorally, where one level is defined to be above another level if the objects from the latter are combined together to make objects of the former level. I say that such a pair of levels is a *combinatorial system*. To understand organisms will require appreciating the relationships between all pairs of adjacent hierarchical levels in the towers of hierarchical levels found in organisms. In particular, for any pair of adjacent hierarchical levels, a key question is, (1) What rules govern how objects from the lower level (the components) combine to make objects at the higher level (the expressions)? For example, what are the rules dictating how bird syllables are combined to make songs? Or, what are the rules governing how cells combine to instantiate higher level organism functions? Another question is, (2) Are there universal laws governing hierarchical complexity in systems of the same type, in the sense that two combinatorial systems of the same type will share the same rules? For example, if one bird can sing twice as many songs (expressions) as another bird, what can we say, if anything, about how many syllable types (component types) they have? If there are universal laws of hierarchical complexity for bird vocalization—a “universal grammar”—then it may be possible to say something; if there are no universal laws, it may not. A more general kind of question is, (3) What kinds of such “rules of grammar” for combinatorial systems are found in nature, and why do systems have the kinds of rules they have? This latter question is not concerned with the particular details of the pair of levels—e.g., whether it is structural or behavioral—but on the underlying principles governing adjacent hierarchical levels. The main task of this research is to ask and begin to answer these three questions.

Keywords: Hierarchy, complexity, scaling, language, behavior, organismal

1 Introduction

If you have ever programmed on a computer, you will know that for most programming languages there is a “main” procedure, in which the primary structure of the program is written. Within this

*This paper appeared in *Comments on Theoretical Biology* 6, 25–75, 2001

main procedure, other procedures are called. Perhaps the program's task is to carry out certain manipulations on a list of numbers, and one of the procedures called within the main procedure sorts any list of numbers, another procedure averages any list, and another finds the variance of a list. The main procedure is, then, built out of a combination of "component" procedures. Each of these procedures, in turn, calls other procedures, and they, in turn, call others. At some point, procedures call primitive procedures, which are procedures that are part of the programming language itself, and the programming language knows what these commands mean without needing any programming instructions. This is just to say that programs are written hierarchically. Not only is computer software hierarchically organized, so is computer hardware. At its lowest level, computer hardware consists of primitive components such as wires, resistors, capacitors, inductors and solid state devices. Combinations of these make higher level hardware such as transistors, and combinations of transistors and other components can make logic gates, which in turn can be strung together to carry out complex hardwired programs or general purpose microprocessors. Computers are, then, comprised of "towers" of hierarchical levels. And such hierarchical organization is true of most artifacts.

Hierarchies abound in the brain and biological sciences as well. Consider human language. Sentences are built from words, words built from phonemes, and phonemes from muscle contractions in the face, tongue and throat. In organisms, base pairs combine to make codons, codons combine to build exons, which combine to make genes, which combine to determine, say, cell types, and cells are put together to implement higher level organism functions, and so on. Therefore, organisms are also comprised of towers of hierarchies, and to understand organisms requires understanding what the hierarchical levels are, and how they relate to one another.

For artificial systems we happen to be more confident of where the hierarchical levels are and of what the principles governing the relationship between levels are; we are confident because we (as a culture) built them. But for organisms we must figure out what the levels are and what the underlying principles are. While we may intuitively feel we can with some confidence just "see" where they are—as we just did above for language and genetics—what we need are clear methodological guidelines for such a determination. Furthermore, we need to define what we really mean by a hierarchical level. And we would also like a framework within which we may think about the relationships between hierarchical levels.

First, what is a hierarchical level? There are many definitions one may have^[1–3] but let us try to capture whatever it is we were informally referring to above. The key idea behind hierarchy above was that objects of the lower level are put together to make objects at the higher level, and, in particular, are put together in a *combinatorial* fashion. That is, objects at the lower level are used as if they are symbols in a language, and they may be combined to obtain a "combinatorial explosion" of objects at the higher level. Therefore, adjacent hierarchical levels are, under this notion of hierarchy, *combinatorial systems*, or systems where "words" or *components* of some kind—namely the lower-level objects—combine to make "sentences" or *expressions* of some kind—namely the higher-level objects. By 'expressions', I will mean expression types, unless otherwise stated.

Let us consider a simple combinatorial system. Suppose the system allows just two types of component (i.e., the lower-level object): '0' and '1'. We will say that $C = 2$. We will also suppose that the system's expressions are as follows:

'000', '001', '010', '100', '011', '101', '110' and '111'.

I.e., there are $E = 8$ expressions, and the length L of these expressions is three. How do we know whether a system is combinatorial? To start, is this simple system above combinatorial? It sure *seems* like it is, since with only two component types the system is able to get eight expressions. However, this is not so straightforward. To really know whether a system is combinatorial actually

requires some understanding of the *rules*, or *grammar*, governing the system. If we knew the grammar, we could immediately tell whether the system is combinatorial, and thus whether the two levels count as distinct levels. In describing to you the simple system above, I was careful not to tell you what grammar governs it. All you know is that there are those eight expressions built from the two component types. If we were to learn that the system follows the rule, “Combine any components into sequences of length three,” then we would know that the system is combinatorial. We would know this because the number of expressions E would then be related to the number of component types C and expression length L by $E = C^L$, and, so long as $L > 1$, increasing the number of component types would disproportionately increase the number of expressions. For example, if just two new component types were added, namely symbols ‘2’ and ‘3’, then the number of expressions in the system would increase from $2^3 = 8$ to $4^3 = 64$. Alternatively, we might instead learn that the grammatical rule is, “For every component type there is *just one* other component type with which it may be combined, and combined in any order up to length three.” For example, suppose that two new component types are added, making ‘0’, ‘1’, ‘2’ and ‘3’ the repertoire of component types. But suppose also that ‘0’ and ‘1’ may be combined only with one another, and similarly for ‘2’ and ‘3’. The expressions emanating from ‘0’ and ‘1’ are the same eight we saw earlier, and from ‘2’ and ‘3’ the eight new expressions

‘222’, ‘223’, ‘232’, ‘322’, ‘233’, ‘323’, ‘332’, and ‘333’

can be formed. Doubling the number of component types under this second possible grammar would only double the number of expressions from 8 to 16, and generally, the equation is $E = (C/2)2^L$. For $L = 3$, we have $E = (C/2)8 = 4C$, and thus E and C are directly proportional to one another, or $E \sim C$. Thus, if this second possible grammar applies, the system is, despite first appearances, not combinatorial at all. The components and expressions would accordingly not be at distinct hierarchical levels; rather, they would comprise the structure within just one level.

Getting access to the grammar of a system is rarely straightforward; they are not written on a system’s sleeve. If we do not have access to the grammar, is there some other way to determine if the system is combinatorial? There is, *if* it is the case that systems of the same type follow the same grammar; that is, if there are universal laws governing the relationships between levels for all systems of the same kind. *If* this is the case, then we can look at other combinatorial systems of the same kind to inform us on the nature of the grammar they share. In particular, for a combinatorial system, doubling the number of components should *more* than double the number of expressions. If a system is not combinatorial, however, then doubling the component types should *only* double the number of expressions. Said differently, a system is combinatorial when the number of component types scales disproportionately slowly compared to the number of expressions, and a system is not combinatorial when the number of component types scales up proportionally with the number of expressions. To determine whether a system is combinatorial, then, it suffices to know about what other systems of the same type but of different number of expressions would look like. That is, it suffices to look at many systems of the same type in order to see how the number of component types C scales up as the number of expressions E increases. If there exists a universally-conformed-to grammar by systems of the same kind, then the nature of the scale-up of C should be law-like. However, another possibility is that systems of the same kind do not conform to similar underlying grammars; when systems of that kind have many more expressions, they also tend to have completely different grammars in operation. If this were so, then looking at the scaling behavior of C versus E should not appear law-like, and will be uninformative in regards to whether or not the particular system of interest is or is not combinatorial.

We can in this way—i.e., by looking for law-like scaling behavior for C as a function of E —judge whether two levels are truly at distinct hierarchical levels, and not just part of the structure

found at a single level. We do not need to know the grammar. However, once we have the scaling relationship between the number of components and the number of expressions in a kind of system, the particular nature of the scaling relationship will give us, if it is law-like, a strong indicator as to what the underlying grammar is. The scaling relationship between C and E is, then, not only useful for answering the question of whether there are distinct hierarchical levels; the relationship also can enable us to recover the grammar governing the system itself. The grammar encompasses the fundamental principles of the hierarchical organization of the system in question, and is thus something we would like to discover for hierarchical systems in organisms and behavior.

2 Possible hierarchical laws

We saw in the previous section that by looking across multiple systems of the same kind and seeing how the number of component types C scales up as the number of expressions E increases, we can (i) determine whether the scaling relationship is law-like and accordingly whether there is therefore a single grammar universally applying to systems of that kind, (ii) decide whether the system is or is not combinatorial on the basis of whether C grows disproportionately slowly than E , and (iii) from the particular details of the law-like relationship acquire a strong hint as to what the grammar for that kind of system is. In this section we consider some of the possible relationships between the number of component types and the number of expressions. We will use this taxonomy when we look at a number of actual systems in the following section.

Let us suppose that a given kind of system does have some grammar that universally applies to all systems of that kind. That is, the grammar possesses the rules governing the possible ways components may be put together to make expressions. Recall that if there are C component types and the length of an expression is L , then there are $E = C^L$ expressions. Actually, though, this is how many expressions there would be *if* (i) the expressions are linear sequences, and (ii) every possible sequence is grammatical. More generally, the structure of an expression might not be sequential (e.g., consider muscle-firings combining into a behavioral expression, where many muscles fire simultaneously), and not every possible combination may be allowed. In general, then, we can only expect that the number of expressions will be *proportional* to C^L . That is, we can generally only expect $E = aC^L$, for some constant proportional factor a capturing the fraction of all possible sequences that are actually expressions. Since we do not much care about the constant a , we usually write this as $E \sim C^L$. This is still not quite right: This only holds *if* each possible “slot” for a component in an expression may be filled without regard for the components in the other slots. In actual systems it may be the case that if you put symbol A in slot 1, then B must go in slot 2. There will very often be such dependencies, and this reduces the “effective” length of an expression, the intuition being that such dependencies reduce one’s freedom to build an expression from L down to something lower. Thus, generally, the effective length of expressions in a system may actually be lower than the number of components in the expressions. This “effective length” I call the *combinatorial degree*, d ; more effective length means a greater degree of “combinatorial room” for building expressions. So, for systems governed by the same grammar, we expect that $E \sim C^d$. That is, we expect that there is a fixed proportionality factor relating E and C^d . Accordingly, if we find that in a kind of system $E \sim C^d$, it will be reasonable to presume that the proportionality factor is constant because the grammar is constant across the systems. We should recognize that it is possible, however, unlikely, that the grammatical rules could be different for every system of the same type so long as the grammars maintain the identical proportionality constant. I will also suppose that combinatorial degree and expression length are proportional to one another, even though the former may be less than the latter.

To sum up the last few thoughts, if a kind of system is governed by a universal grammar, then we expect that when we look at systems of that type, we will find conformance to the proportionality $E \sim C^d$, where d is the combinatorial degree (or the “effective length”). Suppose that we have access to the number of component types and the number of expressions for a variety of systems of the same type (as we do in the systems of the next section). If we plot C versus E , what relationships would fit within this universal-grammar (or invariant grammar) possibility? It turns out that there are a few kinds of quantitative relationships one could see, any of which would suggest that there is an invariant grammar. To see this, just imagine how one might increase the number of expressions in a combinatorial system with two component types that may be combined in any fashion into sequences of length three. It has eight expressions now. The most obvious possibility is to increase the length of the allowable expressions: an increase from 3 to 4 ups the number of expressions from 8 to 16. The other obvious way is to just increase the number of component types: an increase from 2 to 3 increases the number of expressions from 8 to 27. A third possibility is to do both: increase the length and the number of component types. These options provide the basis for the following list of possible relationships between C and E that we might expect if there is an invariant grammar (i.e., if $E \sim C^d$).

Possibility (a): C related to E via a power law

The first possibility I will describe is that, as the number of expressions increases, the combinatorial degree stays constant, and, instead, only the number of component types increases. For example, consider bird vocalization, where syllables combine together to produce songs. Possibility (a) is that birds with more songs in their repertoire have more syllable types with which they can compose songs, not longer songs. Since d is constant, the relationship $E \sim C^d$ is something we call a *power law* (which holds for any equation of the form $y = ax^b$, where a and b are constants). If one plots C versus E , one expects the data to fit the power law $C \sim E^{1/d}$. If one expects data to conform to a power law, one plots the data on a log-log plot, i.e., one plots the logarithms of the data rather than the data itself. The reason this is useful can be seen by taking the logarithms of the power law equation, in which case we get

$$\log(C) = \log(\alpha E^{1/d}),$$

where α is a proportionality constant. This can be manipulated into

$$\log(C) = \log(\alpha) + \log(E^{1/d}),$$

which can be changed further into

$$\log(C) = (1/d)\log(E) + \log(\alpha).$$

If we plot $\log(C)$ on the y axis against $\log(E)$ on the x axis, notice that the equation is actually of the form

$$y = mx + b,$$

i.e., an equation for the line, with slope $m = 1/d$ and y -intercept of $\log(\alpha)$. Thus, if a kind of system has a universal grammar, and if a greater number of expressions is achieved by increasing only the number of component types, then we expect a log-log plot of C versus E to be fitted by a straight line. Furthermore, since the combinatorial degree d must be greater than or equal to one, $1/d$ must be less than or equal to one and greater than or equal to 0; i.e., the slope of the best-fit line of the plot should be in the interval $[0,1]$. Slopes nearer to zero imply greater combinatorial degree;

greater combinatorial degree, thus, allows the number of expressions to increase without having to increase the number of component types very quickly. If the combinatorial degree is high enough, say 1000, then $C \sim E^{1/1000}$, in which case C will be effectively constant as E increases, since $1/1000$ is so close to zero. As the slope increases toward one, the combinatorial degree falls toward one, and C must increase more and more quickly as E increases. When the combinatorial degree equals one, the system can no longer be said to be combinatorial at all, since now $C \sim E$; in this case there is no distinct hierarchical level.

Possibility (b): C is constant

The second possibility to consider is that the number of component types C stays constant, and only the combinatorial degree (think: length) d increases as the number of expressions E increases. In the bird vocalization example, greater song repertoires would be made possible not by increasing the number of syllable types, but by increasing the combinatorial degree, or the effective length, of each song. For a plot of C versus E , we simply expect C stay constant. To understand how fast the combinatorial degree must increase, consider that the relationship between the variables under the assumption of an invariant grammar is $E = aC^d$, but now C is a constant and d is variable. Taking the logarithms of both sides, we have

$$\log(E) = \log(a) + d \log(C).$$

Since a and C are constants, we can simplify this to

$$d \sim \log(E)$$

(supposing E is sufficiently large). That is, within this possibility for an invariant grammar, C is invariant, and the combinatorial degree grows logarithmically with E . As mentioned in Possibility (a) above, when Possibility (a) is true and the combinatorial degree is very high, then C will grow so slowly that it will seem constant. Thus, finding that a kind of system has C that does not increase with E could mean that either Possibility (a) or (b) applies. To distinguish between them one would need to acquire data on how the combinatorial degree changes with increasing expression complexity, e.g., by using the average number of components per expression as a measure.

Possibility (c): C related to E via a logarithmic law

Possibilities (a) and (b) are the two extreme possibilities; in (a) C increases and d is invariant, and in (b) d increases and C is invariant. The third possibility is a compromise between these two, namely that both C and d increase as E increases. For bird vocalization this would mean that birds with a greater number of songs have both more syllable types and effectively longer songs. What relationship would we expect if systems of a kind increased their number of expressions in this way? First, note that we expect the number of component types C to increase more slowly than any power law, because if C increased as a power law, Possibility (a) would apply and the combinatorial degree would not need to increase. Similarly, we expect the combinatorial degree to increase more slowly than logarithmically with E , since if it increases logarithmically, Possibility (b) would apply and the number of component types would not need to increase. One natural possibility is that C scales up logarithmically with E —i.e., $C \sim \log(E)$ —which is slower than any power law relationship with a positive exponent. In this case, how fast must d increase? Considering again our scaling relationship for the case of an invariant grammar, we have

$$E = aC^d,$$

for some constant proportionality a , and now where neither C nor d is constant. Under the assumption that $C = k \log(E)$ (with k a constant), we can solve for d in terms of E as follows.

$$\begin{aligned}\log(E) &= \log(a) + d \log(C), \\ \log(E) &= \log(a) + d \log(k \log(E)), \\ \log(E) &= \log(a) + d[\log(k) + \log(\log(E))], \\ d[\log(k) + \log(\log(E))] &= \log(E) - \log(a), \\ d &= \frac{\log(E) - \log(a)}{\log(\log(E)) + \log(k)},\end{aligned}$$

which for sufficiently large E scales as

$$d \sim \frac{\log(E)}{\log(\log(E))}.$$

If a kind of system falls within this possibility, then C should scale proportionally to $\log(E)$, and thus a plot of C versus $\log(E)$ is expected to be a straight line. On a log-log plot, the data are expected not to be straight, but, instead, to increase but with ever-decreasing slope as E increases. The instantaneous slope is a measure of the inverse of the combinatorial degree of the system at that stage.

Possibility (d): C and E are not related in a simple law-like way

I had stated earlier that if there is a universal grammar applying to all the systems of the same kind, then we should expect law-like behavior. We have seen the three principal kinds of law-like behavior that are possible for a universal grammar-bound kind of system. However, it turns out that it *is* possible, in principle, for the relationship between C and E to nevertheless be non-law-like. This would happen if increasing numbers of expressions were sometimes achieved via increasing the number of component types (Possibility (a)), sometimes via increasing the combinatorial degree (Possibility (b)), and sometimes via increasing both (Possibility (c)), with no universal trend in which of these possibilities is followed by any system. For bird vocalizations, this would occur if some birds obtained more songs via increasing the number of syllable types, some by increasing the effective song length, and some by doing both, and which of these is used by a bird is not a function of E . The behavior of a plot of C versus E would, under this possibility, not be law-like. However, there are still things we should expect. First, we expect that as E increases, so should the variance in the number of component types C . This is because some systems will have low C , but some of them very high values for C . (Actually, in particular we expect the variance of the logarithm of C to increase; the variance of C is expected to increase in the Possibility (a) and (c) as well, merely because it is getting larger.) Second, if we can acquire combinatorial degree or expression length information from individual systems, then we can simultaneously plot C versus d versus E in a three dimensional plot, and since they are constrained by the invariant-grammar equation, $E \sim C^d$, we expect that points in this space should fall on single two-dimensional surface. The systems I have studied thus far appear to not be described by this possibility, so we will not discuss this again.

In sum, under the assumption that there is a universal grammar applying to all systems of the same kind, the four possible relationships between the number of component types and the number of expressions are these:

- Possibility (a) is that the number of component types C increases only, in which case C is related to E by a power law with exponent in the interval $[0, 1]$.

- Possibility (b) is that the combinatorial degree d increases only, in which case C does not increase with E .
- Possibility (c) is that both the number of component types and the combinatorial degree increase, in which case C increases logarithmically with E .
- Possibility (d) is “all of the above,” depending on the particular system, meaning that there is no nice law-like relationship between C and E .

The grammar not only determines what the invariant proportionality constant is, but it can also determine which of these Possibilities obtains. For example, it could be that the grammar explicitly allows only so many components per expression, in which case there would be a maximum length expression, and a fixed, maximum combinatorial degree; this would lead to Possibility (a). Alternatively, the grammar could be specific to only certain kinds of component, but allow arbitrarily long expressions composed of them; this would lead to Possibility (b). Or, it could be that the grammar explicitly requires a certain relationship between the component types and the combinatorial degree, so that if one increases so must the other, in which this might lead to Possibility (c).

While it is possible that the grammar could itself determine which Possibility holds, it is also possible that the grammar is weak—not making any strong requirements on C , d or their relationship—but it is nevertheless the case that one of Possibilities (a), (b) or (c) applies. Why the behavior is one of these possibilities rather than another would be due not the grammar, but to “extra-grammatical constraints.” For example, perhaps the grammatical rules allow arbitrarily long expressions, but there are constraints on how long expressions can be because of the nature of kind of system in which the grammar is embedded. For example, in English or any natural language words are combined to make sentences, and English grammar allows arbitrarily long sentences, in principle; but because sentences are uttered by people with limits on the ability to produce and understand sentences, we can expect an upper bound to the combinatorial degree, and thus expect Possibility (a) to apply. Similarly, extra-grammatical constraints could lead to Possibility (b) if there were some limit to the number of component types, or (c) if there were some optimal balance between the number of component types and the combinatorial degree. If there are *no* grammatical requirements on C , d or their relationship, *and* there are *no* extra-grammatical constraints either, then we would expect Possibility (d) to apply: different systems of the same type could choose their own way to increase expressive complexity.

If there is not a universal grammar applying to the kind of system, then these possibilities are irrelevant, and many things are possible. It turns out that the data I have so far accumulated can be accommodated within Possibility (a), (b) or (c), and mainly (a) and (c); thus, the kinds of systems studied thus far do appear have invariant grammars. It is to the studied kinds of system that we turn to in the next section.

3 Universal laws are the rule

Let us now look at some actual systems, systems that appear to be hierarchical, and analyze them in light of the earlier possibilities and discussion. Each system appears to be combinatorial in that components of some kind appear to combine together to make expressions of some kind; each system thus appears to consist of two distinct hierarchical levels. The kinds of system I have acquired data for are listed in Table 1, which shows what the components and expressions are for each kind of system, along with other details we will discuss in turn.

In the following subsections we cover these kinds of system, beginning with artificial systems like universities and electronic devices, and moving to structural hierarchical complexity in or-

Kind of combinatorial system	Component	Expression	Relationship between C and E (and hypothesis from introduction)	R ²	Combinatorial degree
Bird vocalization	Syllable	Song	Power law (a)	0.702	1.23
Human language					
- historical	Word	Sentence	Power law (a)	0.795	5.02
- ontogeny 1	Word	Sentence	Logarithmic (c)	0.984	1 up to 2.5
- ontogeny 2	Phoneme	Morpheme	Logarithmic (c)	0.959	2 up to 4
Behavior	Muscle	Behavior	Logarithmic (c) ?	0.772	3 up to 9
Organism					
- phylogeny	Cell	Expression	Power law (a)	0.438	12.42
- ontogeny	Cell	Expression	Power law (a)	0.988	1.02
University	Concentration	Degree	Power law (a)	0.687	1.65
Electronic device					
- CD player	Button-press	Action	Power law or logarithmic (a or c)	0.489	2.07
- TV	Button-press	Action	Power law or logarithmic (a or c)	0.842	1.58
- VCR	Button-press	Action	Power law or logarithmic (a or c)	0.508	3.95
- calculator	Button-press	Action	Power law or logarithmic (a or c)	0.875	8.77

Table 1: Summary of the kinds of system studied, the components, expressions, which possibility from earlier (for kinds of systems with universal grammars) the data confirm ((a) through (d)), the correlation, and the combinatorial degree. When it is unknown whether (a) or (c) applies, correlation and combinatorial degree are under the assumption of Possibility (a). Correlation is highly significant ($p < 0.01$) in each case. When Possibility (c) applies, the approximate range is shown over which combinatorial degree values increase.

ganisms, and spending the remainder of the section on varieties of behavioral hierarchies, from language to bird song to muscles.^[4–5]

3.1 Artificial systems

Before moving to hierarchical complexity in brain and biological systems, I think it is useful to see examples among the systems we humans built ourselves. We will see that in some kinds of artificial systems, systems of the same kind appear to “self-organize” into conformance with universal laws, there being no central decision-maker dictating this conformance. Although we understand the workings of any one of these systems, these universal laws governing the kind of system have yet to be noticed (at least, not until Changizi^[4]), much less explained. However, because we built these artificial systems ourselves, it will probably be much simpler to understand the universal laws than it will in brain and biological systems.

3.1.1 Electronic device user interface

The first kind of artificial system we will look at is electronic devices. Actions on electronic devices are carried out by the user pressing a sequence of buttons. Button-presses are the components, and each button is a different type of button-press. Button-presses combine to make device actions, the expressions. The question is, how do electronic devices achieve more device actions? Does there appear to be a universal grammar applying to them, and does there appear to be one of the expected simple law-like curves (or the less simple behavior of Possibility (d))? Is greater expressive complexity (i.e., more expressions) handled by having more button-press types (Possibility (a)), or with longer button-press sequences per action (Possibility (b)), or both (Possibility (c))? And if we find law-like scaling behavior of one of these types, can we learn anything about the grammar by the specific nature of the plot?

I obtained measurements for four kinds of electronic device: Compact disk (CD) players ($n=20$), televisions (TVs) ($n=8$), video cassette recorders (VCRs) ($n=36$, averaged from 78) and calculators ($n=17$). I measured the number of button-press types as the number of buttons on the device; for CD players, TVs and VCRs, the number of buttons was taken from remote controls, and for calculators I limited cases to those without full a-to-z keyboards. The number of expressions here is the number of actions the device is capable of, and to measure this I assumed that if a device can do twice as many things as another device, its user’s manual will tend to be roughly double in length. I then used the number of pages in the user’s manual as a proxy for expression complexity.

Figure 1 shows the data plotted on log-log plots for each of the four kinds of electronic device.

The first thing to notice is that, for each type of electronic device, the trend appears law-like, implying that there may be some underlying universal grammar—an invariant proportionality constant—driving the scaling relationship between the number of component types C and the number of expressions E . The grammar for any kind of electronic device is whatever are the rules we have learned when we say we know how to use a device. The fact that these plots are law-like suggests that devices of the same kind use the same grammatical rules; this is why once we learn how to work one VCR we can work any VCR.

Notice, furthermore, in the plots that in each case the number of button-press (component) types increases as the number of device actions (expression) increases. This means that either Possibility (a) or Possibility (c) apply to each of the four kinds of electronic device. Unfortunately, because of the restricted range of values for the number of button-press types, it is not possible to distinguish between these possibilities: the plots are practically identical whether plotted on log-log (which

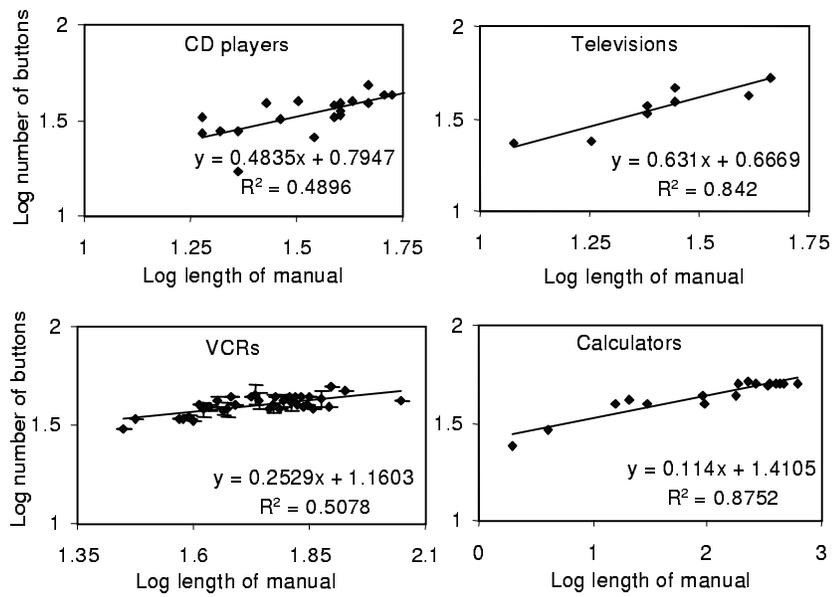


Figure 1: Logarithm (base 10) of the number of button-press types versus the logarithm of the number of device actions (the latter measured by the number of pages in the user's manual), for CD players, VCRs, TVs and calculators.

should be linear if Possibility (a) holds) or plotted with just the x axis logged (which should be linear if Possibility (b) holds).

In either case, we *can* get an estimate of the combinatorial degree from the log-log plot. If Possibility (a) holds, then the kind of system has an invariant combinatorial degree. If, however, Possibility (c) holds, then the kind of system has a combinatorial degree that is increasing, and thus the slope of the log-log plot should progressively decrease. (Recall that the inverse of the slope in a log-log plot is an estimate of the combinatorial degree.) The slopes of the best-fit lines (via linear regression) over the entirety of each of the four log-log plots implies combinatorial degrees of 2.07 for CD players, 1.58 for TVs, 3.95 for VCRs, and 8.77 for calculators. I should reiterate that if Possibility (c) applies, then it is strictly incorrect to give a single combinatorial degree value for the kind of device; instead, the combinatorial degree must actually begin lower than the value mentioned above, and rise above that value. The combinatorial degree estimates from the whole plot nevertheless give us an idea of the size regime of the combinatorial degree, whether Possibility (a) or (c) applies. Each of the four kinds of system has combinatorial degree above one, and is thus truly a combinatorial system; electronic device user-interface languages have (at least) two distinct hierarchical levels. But we knew this, since we all know, at least implicitly, the grammars of these electronic devices (i.e., we know how to use them), and know that the user-interface rules are combinatorial.

Can we make any sense of these combinatorial degree differences for the different kinds of electronic device? The simplest hypothesis for what explains the combinatorial degree in electronic devices is the average or typical number of buttons required to carry out a device action. The combinatorial degree values above, then, may be interpreted as requiring, on average, around two button presses to carry out a function on a CD player, around one and a half on a TV, around four for a VCR, and around nine for calculators. At first glance, the difference between calculators and the other types is expected, since calculators require many more button presses to carry out an action than do the other kinds of device. I have not attempted to make sense of the differences between CD players and TVs on the one hand, and VCRs on the other, the latter whose combinatorial degree is significantly higher, much less have I tried to explain the particular combinatorial degree values for any of these kinds of system. Where the combinatorial degrees come from, though, seems relatively clear: they are due to the average number of buttons required to carry out a function on that kind of device.

3.1.2 Universities

The other artificial system for which I have acquired data concerns universities. In particular, departmental concentrations, the components, are put together by students to make up their academic degrees, the expressions. Many students just choose one major, but many addend minors to their degree, and many others have two majors, or two majors and a minor. Because of the students who choose degrees with more than one departmental concentration, academic degrees can be built in a combinatorial fashion from departmental concentrations. Since pretty much any combination of departmental concentrations is typically allowed by any university, the grammar should be the same across the universities. We thus expect one of Possibilities (a) through (d) to apply.

To see which of these possibilities actually applies, I sought to measure the number of departmental concentration types (e.g., “physics”) and the number of distinct academic degrees (e.g., “physics and math”) in a number of colleges and universities. I used the number of departments at a university as the measure of the number of departmental concentration types. Measurement of the number of distinct academic degree types is difficult to obtain. As a simplifying assumption, I supposed that if a university has twice as many students, then there are twice as many distinct kinds

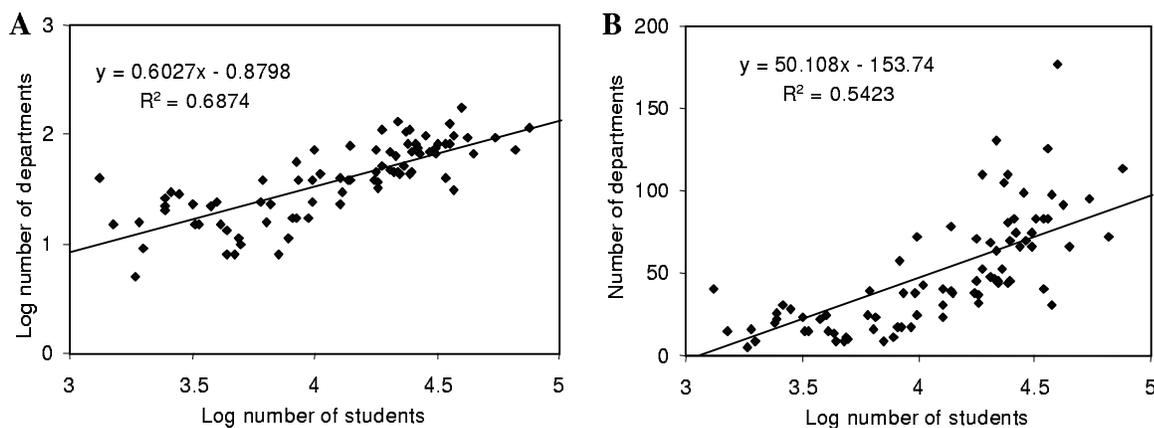


Figure 2: (A) Logarithm (base 10) of the number of departmental concentration types (measured as the number of departments) versus the logarithm of the number of academic degree types (measured as the number of students). (B) Same plot but under the assumption that the number of component types (i.e., departmental concentration types) scales logarithmically with the number of expressions (i.e., academic degrees). The data on the log-log plot conform much better to a line than on the semi-log plot, suggesting that it is Possibility (a) that applies.

of degrees sought after. The intuitive idea behind this assumption is that students are unique individuals, up to a constant proportion; this is why you always find, no matter how many departments there are, some student double majoring in two of the most disparate departments, or dissatisfied with the choices so much that they wish to create their own personal major. With this assumption, the number of students at the university, being proportional to the number of academic degree types, can be used as a proxy for the number of academic degree types.

In this way I obtained measures of the number of departmental concentration types (measured as the number of departments) and the number of distinct academic degree types (measured as the number of students) for 89 U.S. and Canadian colleges and universities.^[6] Figure 2 shows the plots under the assumptions of Possibility (a) (Figure 2, (A)) and Possibility (c) (Figure 2, (B)).

The first thing to observe is that there is a striking law-like relationship followed by these diverse schools. We expected this because we suspected that different universities have the same grammar for combining concentrations into degrees: namely, anything goes. And recall that if there is a universal grammar, then, unless Possibility (d) holds, there is a nice law-like relationship that will

apply to the systems, although which relationship will depend.

The second observation is that the number of departmental concentration types is clearly increasing, not remaining level. This means that either Possibility (a) or Possibility (c) apply to these systems. The inverse of the best-fit (by linear regression) slope in the log-log plot (under the power law assumption) is 1.66 (95% confidence interval [1.45, 1.94]), and is much more linear than the plot under the logarithmic assumption. Thus, Possibility (a) appears to apply: greater academic degree types is accommodated not by increasing the combinatorial degree, but by increasing the number of departmental concentration types from which students may choose. In particular, the relationship between them is given by $C \sim E^{0.6027}$. This is not too surprising, since students can only handle so many concentrations during their four or so years in school, the combinatorial degree cannot grow much past whatever this upper limit may be; this is an extra-grammatical constraint. Perhaps, then the combinatorial degree of 1.66 should be close to the average number of departmental concentrations per student. To test this, I arbitrarily wrote to the registrars at Duke University (because that is where I was), University of Virginia (my undergraduate institution) and Williams College (where my summer student at the time was from), and asked for data on the average number of concentrations per student. Counting only majors and minors as concentrations, the average numbers of majors or minors per person at Duke, UVA and Williams College are, respectively, 1.75, 1.17 and 1.39, in the ballpark of the combinatorial degree (which, recall, is the combinatorial degree measured by examination of the scaling behavior across 89 universities, not just these three universities). [Averages for Duke and UVA are for Spring semester, 2000. The value for Williams College is averaged over the average for each year from 1991 to 2000, and has standard deviation for those ten years of 0.0344.]

It is fascinating to find such universal laws underlying diverse universities over two orders of magnitude in size of student body. The explanation for this order is, as in electronic devices, a “self-organization” explanation, as there would appear to be no central authority or university guidebook dictating how many departments there should be given so and so many students.

3.2 Cells and organisms

Organisms are hierarchically organized in structure. What principles, if any, govern these hierarchies? In this subsection I focus primarily on how cells combine to implement higher level functions. We will be interested in looking at trends at the largest phylogenetic scales. So, for example, if I say that the combinatorial degree is invariant as organisms become more complex, I mean that there is no largest scale trend in the combinatorial degree.

Organisms are complex systems, and they carry out lots of higher level functions, such as inhaling, chewing, digestion and walking. I will generically call such higher level functions “organism-expressions,” and for now I wish to remain elusive about just what exactly an organism-expression is. It will turn out that, so long as whatever organism-expressions are, they follow a certain assumption (“assumption (*)”), then we will be able to make progress without having to become precise about what an organism-expression is. Organisms achieve their organism-expressions via combining cells into complex structures. Cells are, then, the components. In actuality, most organisms probably have hierarchical levels in between that of the cell and the organism-expression, but will attempt to look at only these two levels. Our question becomes, “What principles govern the hierarchical relationship between cells and organism-expressions?” For example, as organisms become more complex—i.e., a greater number of organism-expressions—does the number of cell types increase? And if so, in what fashion, Possibility (a) or (c)? And if Possibility (a) holds, say, then what does the combinatorial degree tell us about the grammar, or about the extra-grammatical

constraints on the system?

Measurement of the number of cell types was possible due to the work of Bell and Mooers,^[7] who acquired estimates from 134 organisms over 31 phyla (namely, plant phyla, animal phyla, fungi phyla, along with Chlorophyta, Phaeophyta, Rhodophyta, Ciliata, Acrasiomycota and Myxomycota). Measuring the number of organism-expressions is more difficult. Not to mention the problem that I would like to not commit myself to any rigid, defined notion of an organism-expression. There are nevertheless two promising strategies toward acquiring a proxy for the number of organism-expressions, the first concerning genomic complexity, and the second concerning organism mass.

3.2.1 Measuring expressive-complexity using coding genome size

The *coding genome size* is the amount of an organism's DNA that codes for proteins. Ultimately, it is this code which possesses the bulk of the information needed for organisms to develop into their final form, with all the highest-level organism-expressions. Coding genome size should thus correlate with the number of organism-expressions. Note that they cannot be expected to be proportional to one another since DNA is a *code*, and so coding genome size should scale up disproportionately slowly as a function of the number of organism-expressions. Note also that, in principle, it need not be the case that genome coding size correlate with the number of cell types. For example, it is *a priori* possible that the number of cell types could be kept invariant as organisms become more expressively complex (i.e., the number of cell types does not change as a function of expressive complexity). Organisms would, in this case, achieve greater expressive complexity via increasing the combinatorial degree, i.e., by increasing the effective number of cells involved in an organism-expression. This would correspond to Possibility (b). So, although it *could* be the case that genome coding size correlates with number of cell types—if, say, Possibilities (a) or (c) apply—it need not. The only aspect of organisms with which genome coding size would *have* to correlate is the number of highest level function expressions, or organism-expressions; *something* has to code for this information, and it is surely primarily DNA that is responsible, even though we do not yet understand all that is involved in the connection between these two hierarchical levels.

With this in mind, our first attempt at analyzing the relationship between the number of cell types and organism expressive complexity is to use coding genome size^[8] as the measure of number of organism-expressions, where we expect this measure to merely correlate. If the number of cell types remains invariant as coding genome size increases, then Possibility (b) probably holds: greater expressive complexity is obtained via increasing the combinatorial degree, not the number of cell types. If, on the other hand, the number of cell types increases, then this implies that either Possibility (a) or (c) applies. Figure 3 shows a log-log plot of the number of cell types against the coding genome size for 8 organisms where both were available in Cavalier-Smith^[8] and Bell and Mooers.^[7]

The number of cell types clearly increases as genome coding size increases, and thus increases as the number of organism-expressions increases. Possibility (a) or (c) applies, then, to this system. We can first ask, Is this system combinatorial? To answer this, first recognize that genome coding size, being a code, must scale disproportionately slowly against the number of organism-expressions. Since the slope in the log-log plot in Figure 3 is less than one, number of cell types scales up disproportionately slowly with coding genome size, and since the latter scales disproportionately slowly with the number of organism-expressions, so must the former. Thus, cells implement organism-expressions in a combinatorial fashion; they do comprise two distinct hierarchical levels. Note that we cannot conclude that they are *adjacent* hierarchical levels; there may well be, and probably are, intermediate hierarchical levels in between.

Can we decide which of Possibilities (a) or (c) applies from Figure 3? We are not able to because

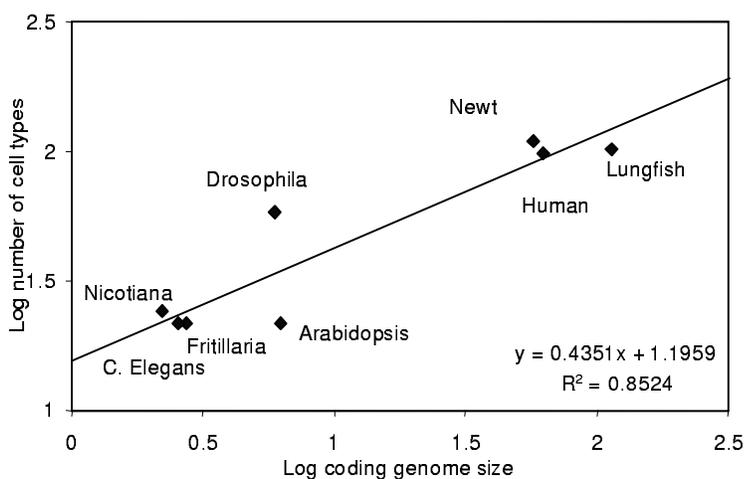


Figure 3: Logarithm (base 10) of the number of cell types versus the logarithm of the coding genome size (picograms of genes coding for proteins) for some plants and animals ($n=8$). Coding genome size data are taken from Cavalier-Smith.^[8] Cell type data are obtained from Bell and Mooers.^[7] A similar plot, but for total DNA rather than coding DNA, appears in Kauffman (Ref. 55, Fig. 13; Ref. 54, Fig. 12.7), motivated by different concerns.

we do not know how genome coding size relates to the number of organism-expressions; all we can be reasonably confident about is that the former grows disproportionately slowly compared to the latter, but we don't know what kind of relationship governs them.

3.2.2 Measuring expressive-complexity using mass

Genome coding size had two downsides. The first we just mentioned, and is that we do not know how it relates to expressive complexity, other than that they are probably correlated. The second difficulty is that it is difficult to acquire genome coding size for organisms, and thus the paucity of data points in Figure 3. It would be convenient if we had some other proxy for the number of organism-expressions for which (i) their relationship is better understood, and (ii) measurement is easy. Mass is our answer; more precisely, total number of cells, but since cells are roughly invariant in size, mass can be used instead of the total number of cells. It is certainly easy to measure mass, and accordingly there exists plenty of mass measurements in the literature for organisms. Mass is also advantageous because, under a simplifying assumption, it is possible to relate it to the expected number of organism-expressions.

Allow me to gloss over the simplifying assumption first, and say how mass relates to the number of organism-expressions. An organism carries out many organism-expressions. Suppose there are E of them. Also suppose that organism-expressions consist of around L cells; this is the expression length. Supposing that every cell is part of some organism-expression, it follows that E times L must equal the total number of cells in the organism. Total number of cells scales proportionally with mass, and thus we can say that $M \sim EL$. Recall also that there is good reason to expect that

the combinatorial degree d is proportional to L , and so

$$M \sim E \cdot d.$$

Now we have a simple scaling equation relating mass to expressive complexity and combinatorial degree. How mass relates to expressive complexity now depends on the combinatorial degree. Under the hypothesis that Possibility (a) holds, the combinatorial degree is invariant, and thus $M \sim E$. That is, if Possibility (a) holds, then we expect that we may use mass as a proxy for expressive complexity, and we expect the number of cell types to scale up with mass as a power law. Alternatively, if Possibility (c) holds, then $d \sim \log(E)/\log(\log(E))$, and so $M \sim E[\log(E)/\log(\log(E))]$. Because the combinatorial degree term scales so slowly, it is approximately the case that $M \sim E$. Thus, mass can be used as a proxy for the number of organism-expressions whether Possibility (a) or (c) applies. Possibility (b) is not an option here because we already know from Figure 3 that the number of cell types increases with expressive complexity.

Before using mass as a proxy to ask which of Possibilities (a) and (c) describe organisms, we should examine the above argument for mass as a proxy in more detail. I have implicitly made a simplifying assumption. Before stating it, let me first provide an intuition pump. Consider calculators for a moment. They are, for the most part, made to be as small as possible subject to the constraints that they can do all the functions they are made for, and that humans can actually punch the buttons. Calculators with more buttons are, in order to fit all the buttons, typically physically larger. Now here is my intuition pump question? If a calculator has more buttons than another calculator, and has some novel type of button, then how many buttons of this novel type does it have? Just one. That is, when new functional buttons are added to calculators, just one is added, not multiple copies. [For the purposes of this discussion, it is not helpful to remember our earlier study of calculators. Our task then concerned how buttons combine to implement higher-level functions. For our purposes now I am only interested in just the level of buttons.] Calculators become more “button-complex” by adding a new button type, not by adding multiple buttons of the same type. More importantly, it is not the case that as calculators acquire more and more buttons, they add a greater and greater number of copies of the same button to the device. For example, the following is not the case:

Calculator A_1 has two buttons, each of a different type. Calculator A_2 adds a new button type, but puts in two copies, making four buttons, three types. Calculator A_3 adds a new type, but puts in three copies of it, resulting in seven buttons, and four types. And so on.

The reason this does not occur in calculators is presumably because of the market pressure to minimize the overall size of the calculator subject to the constraints; there should accordingly be no more buttons than needed. The moral of the intuition is that, if a functional device is under pressure to minimize its overall size, then as it acquires more functions (or in the case of calculators, more buttons) we expect it not to add unnecessary copies of the same function; and, in particular, we expect it not to add an increasing number of copies of the same function.

Let us now consider an analogous idea for organisms. Organisms have organism-expressions in them that carry out functions of some sort, and they are composed of cells. An organism with more organism-expression types is more expressively complex. [I have been using “number of organism-expressions” implicitly to mean number of organism-expressions types.] Suppose an organism with few organism-expressions gains a new type of organism-expression, and in doing so, gains k copies of it (i.e., devotes enough cells to implement k copies of the expressions). Now consider a much more expressively complex organism that gains a new type of expression. How many copies of the new kind of expression does it add? The intuition pump above tried to make it plausible that only k copies will be added, i.e., just as many copies will be added as in the less expressively complex

organism. An alternative is that the more complex organism adds more than k copies of its new expression type, and that still more complex organisms would add even more copies of their new expressions; it is this alternative that our hypothesis is assuming is *not* an appropriate assumption. The simplifying assumption I will call the *Copy-Invariance Assumption*, and is stated as

As organisms get more expressively complex, the number of times any given organism-expression type is instantiated in the organism does not, itself, tend to change.

Organisms conforming to this would thereby achieve their functionality without having to devote any more cells than is necessary.

With this Copy-Invariance Assumption, it follows (as discussed earlier) that mass can be used as a proportional proxy for the number of organism-expressions, and we can proceed and test between Possibilities (a) and (c). Bell and Mooers^[7] have acquired estimates of the number of cell types from many organisms, as we mentioned earlier, but they also estimated the mass of the organisms. Figure 4 shows averages of the number of cell types versus averages for mass for the 31 phyla from Bell and Mooers.^[7] The first observation to be made is that, as portended by the genome coding size plot earlier, the number of cell types certainly increases with the current proxy for organism expressive complexity, namely mass.

Next we need to determine which of Possibility (a) or (c) the data favor. If Possibility (a) holds for organisms, then we expect the data to be approximately linear on a log-log plot of number of cell types versus mass, a plot which is shown in Figure 4A. If, instead, Possibility (c) holds, then we expect the data to be approximately linear when we plot (unlogged) number of cell types versus the logarithm of the mass, the plot which is shown in Figure 4B. One may see from the figures that the data fall more linearly under the hypothesis that Possibility (a) holds than under the hypothesis that Possibility (b) holds. The data therefore appear to confirm Possibility (a), preliminarily suggesting that at the largest phylogenetic scale organism expressive-complexity is primarily achieved via increasing the number of cell types, not via increasing the combinatorial degree. Also, the exponent of the power law is 0.0805, which is lower than one, and thus the system appears to certainly be combinatorial.

3.2.3 The combinatorial degree

Figure 4A has a slope of 0.0805, which leads to an estimated combinatorial degree around 12. There are two questions we might ask. (i) Why might Possibility (a) apply rather than Possibility (c)? That is, is there some reason why the combinatorial degree might have remained invariant? And (ii) why is the combinatorial degree in the rough range of 12? I should point out that we can, at this point, have little confidence that the combinatorial degree for organisms actually *is* invariant, nor much confidence in anything more than the order of magnitude of the combinatorial degree range. Nevertheless, let us now wonder if there might be answers to these questions, recognizing that we are engaging in speculation mostly for its own enjoyment.

First let us ask if there might be any reason for the cell/organism-expression system to have invariant combinatorial degree (supposing it is indeed invariant). Recall that from the Copy-Invariance Assumption we can conclude that $M \sim E \cdot d$, i.e., mass is proportional to the number of organism-expressions times the combinatorial degree. If d is invariant as in Possibility (a), then

$$M \sim E.$$

If, on the other hand, d increases sublogarithmically with E , as in Possibility (c), then

$$M \sim E \cdot \frac{\log E}{\log \log E},$$

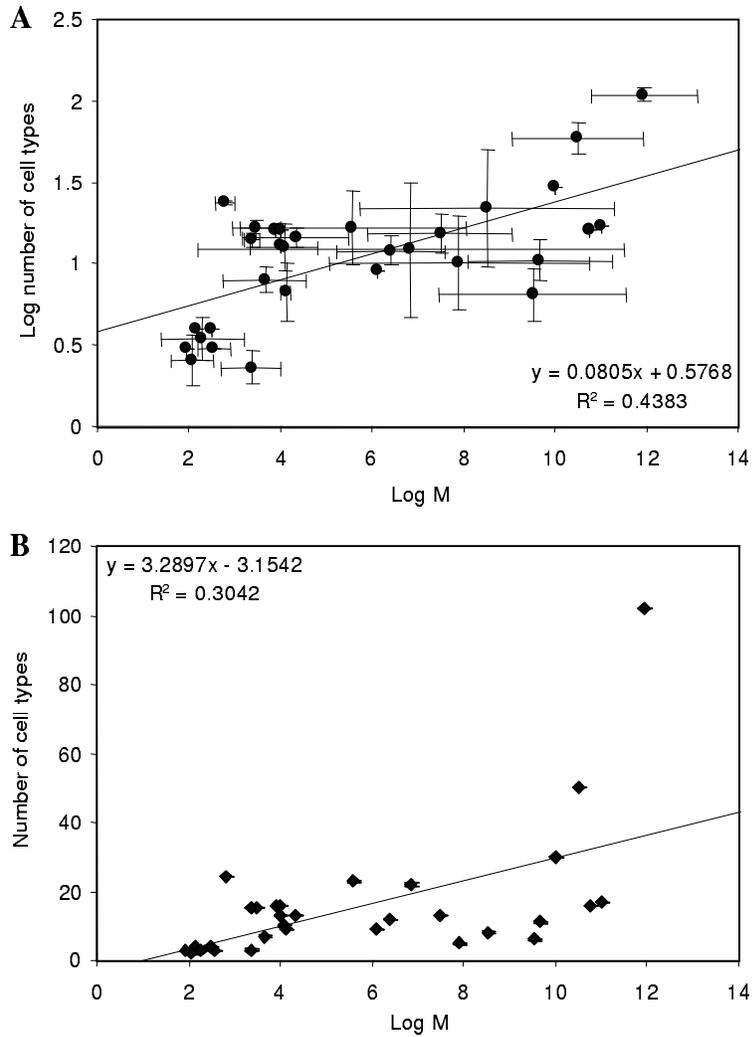


Figure 4: (A) Logarithm (base 10) of the number of cell types versus logarithm of the total number of cells for plant, animal, fungi, Chlorophyta, Phaeophyta, Rhodophyta, Ciliata, Acrasiomycota and Myxomycota phyla (n=31). Error bars show standard deviation. Best-fit line via linear regression. (B) Number of cell types versus logarithm of the total number of cells for same data as in (A).

and thus mass of organisms must scale up more quickly as expressive complexity increases. Finally, if d increases logarithmically with E , as in Possibility (b), then

$$M \sim E \log E,$$

and mass must scale even more quickly. Thus, of these three possibilities, organism mass scales up the least quickly when the combinatorial degree is invariant. Keeping the combinatorial degree invariant, and increasing the number of cell types instead, means that organisms can achieve greater expressive complexity while keeping their overall size low; they can pack in their functional expressions into an overall smaller body.

Now, let us ask why the combinatorial degree might be on the order of 10 or so. Organism-expressions, whatever they may be, are almost surely built out of tremendously more than 10 cells; there are few cases where around 10 cells make up some functional higher level structure. The combinatorial degree of around 10 is probably, then, not due simply to the number of cells in an organism-expression. Let us consider organs as a possible candidate example of an organism-expression, although I do not wish to commit myself to organs being the definition of an organism-expression. Organs are obviously built from cells, but they are more usefully considered to be combinations of *tissues*; that is the way organs are, in fact, more typically described in histology textbooks. Assuming that the number of tissue types scales proportionally with the number of cell types, our plot in Figure 4A can be treated as showing the number of tissue types along the y axis, and the combinatorial degree of around 12 would hold for the hierarchical system of tissues and organism-expressions. While organism-expressions do not have around ten cells, they *may* tend to have around ten tissues. Perhaps, then, organism-expressions are built from tissues as components, and, on average, around ten or so tissues tends to be involved in the construction of an organism-expression.

I eventually plan on testing this by counting up the number of tissues in higher-level functional structures—e.g., organs—in organisms across many phyla, and seeing whether there tends to be around ten or so. In the meantime, as a start I have acquired the number of tissues involved in 63 organs (see legend of Figure 5 for vertebrates from a standard vertebrate histology textbook.^[9] The average number of tissues is 10.52 (± 4.17)), indicated roughly by the arrow in the histogram in Figure 5. This is within the 95% confidence interval of the measured combinatorial degree of 12.42, but I would not make much of this at this point.

Although we have provided some meager evidence toward the conjecture that perhaps the combinatorial degree of around ten is due to tissues combining in groups of around ten into organism-expressions, we have not given any reason for why organism-expressions would have this many tissue components? Why around ten? One general kind of reason is simply that there may be an upper limit to how many tissues may be physically packed near one another to make an expression. For example, imagine that tissues were all spheres of the same size. In that case, one can pack about 12 spheres around one sphere. Tissues are much more convoluted in shape than spheres, but perhaps this packing limit is the driving factor in the combinatorial degree. To overcome this kind of limit for organism-expressions requiring physical contact of their constituent tissues, tissues would have to become convoluted and branched, like neurons. [Neurons combine together to implement higher level neural structures, and since neurons are so branchy they may form expressions built from thousands of neurons simultaneously contributing. We might therefore expect the combinatorial degree for nervous systems to be very high, and thus the number of neuron types would have to scale up very slowly as the number of neural expressions increases.]

I have severe doubts that the combinatorial degree for organisms really has remained invariant over the history of life. The reason is that there may be, for more complex organisms but

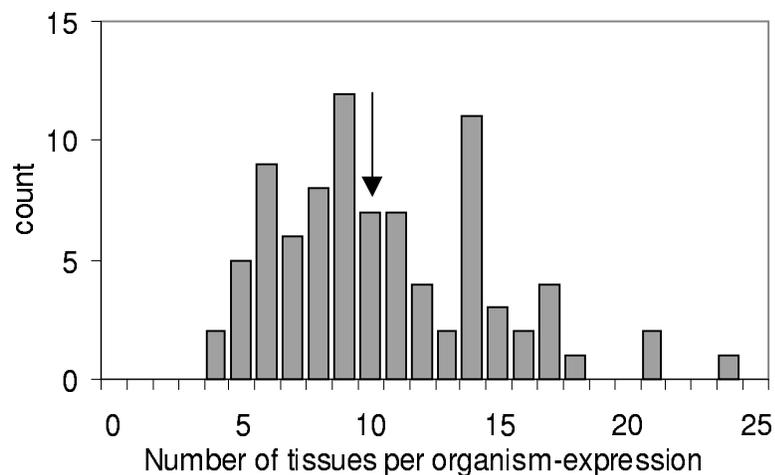


Figure 5: Distribution of numbers of tissues per organ in vertebrates. Arrow indicates the mean. Organs used are: heart, aorta, tonsil, lymph node, spleen, thymus, skin, endocrine and apocrine sweat glands, sweat and sebaceous glands, hair follicle and nail, tongue, submandibular gland, parotid gland, sublingual gland, soft palate, teeth, lip, esophagus, esophagogastric junction, stomach, gastroduodenal junction, duodenum, jejunum, ileum, colon, appendix, anorectal junction, liver, gallbladder, pancreas, olfactory mucosa, larynx, trachea and bronchus, bronchiole and respiratory passages, alveoli, kidney, ureter, urinary bladder, pituitary gland, pineal gland, parathyroid and thyroid glands, thyroid follicle cells, adrenal gland, testis, ductuli efferentes and epididymis, spermatic cord and ductus deferens, prostate gland, seminal vesicle, ovary, corpus luteum, oviduct-uterine tube, uterus, cervix, placenta, vagina, mammary glands, eye, ear, organ of corti.

not less complex organisms, hierarchical levels between the cell level and the highest, “organism-expression,” level. That is, vertebrates may have hierarchical levels between the cell level and the top level, but some ancestor may not have. As we will discuss Section 4, when new hierarchical levels are added in between two existing levels, the consequence is an increased combinatorial degree between the original two levels. Thus, since we might reasonably expect that hierarchical levels have been added over the history of life, we would accordingly expect the combinatorial degree to increase. More data and analysis will be needed to search for signs of hierarchy additions.

3.2.4 Ontogeny of cell types versus mass

We have discussed cell/organism-expression hierarchical complexity at the largest phylogenetic scale. We may naturally wish to know how these hierarchical levels behave during ontogeny. As an organism develops, its number of cell types increases as does its number of organism-expressions. How do they scale relative to one another during ontogeny? If we continue to use mass as a proxy for the number of organism-expressions, then it is relatively simple to obtain measures for expressive complexity as a function of developmental time. Acquiring the number of cell types in an organism as a function of time, however, is very difficult. The nematode worm *Caenorhabditis elegans* provides us with a good starting point: it has only around a thousand cells in all, we know what they are, and how many there are throughout the worm’s development (Sulston and White in Appendix 1 of Wood Ref. [10]).

Figure 6 shows a log-log plot of the number of cell types in *C. elegans* as a function of its mass throughout its gastrulation, which is its primary growth phase; the legend details the 27 kinds of cell types I distinguished.

The plot appears to strongly follow Possibility (a), but the slope is essentially 1, and thus the cell/organism-expression system is not acting combinatorially over the developmental stage. If they were found to act combinatorially, then this would suggest that, at each stage of development, the worm utilizes its current cell types to, in a language-like fashion, build combinatorially many organism-expressions. Instead, the plot suggests that this is not the case: cell types are created only with their ultimate “adult use” in mind. Alternatively, it may be that during ontogeny the Copy-Invariance Assumption is no longer justified, in which case we cannot use mass to measure the number of organism-expressions.

3.2.5 Subcellular parts

As a final topic concerning hierarchical complexity in organisms, we will briefly discuss how subcellular parts implement cellular functions, or cell-expressions (as opposed to organism-expressions). Over all the kinds of hierarchical system I have thus far studied, there is a tendency for the number of component types to increase—meaning either Possibility (a) or Possibility (c) applies. Let us suppose for the moment that this also holds for subcellular components combining to implement cell types.

With this supposition we may make a prediction concerning the difference in internal complexity between free-living eukaryotic cells (protists) and cells in certain multicellular organisms (metazoans and land plants). It is plausible that, on average, free-living cells must carry out more cell functions than cells in multicellular organisms, the latter which may be more specialized and incapable of certain functions (e.g., reproduction). In other words, one reason cells in multicellular organisms may have fewer cell-level functions is because the functionality has been pushed up to a still greater hierarchical level.^[11] Another reason is that higher level functionality requires coordination among the lower-level components, which in turn requires constraint; thus, cells in multicellular

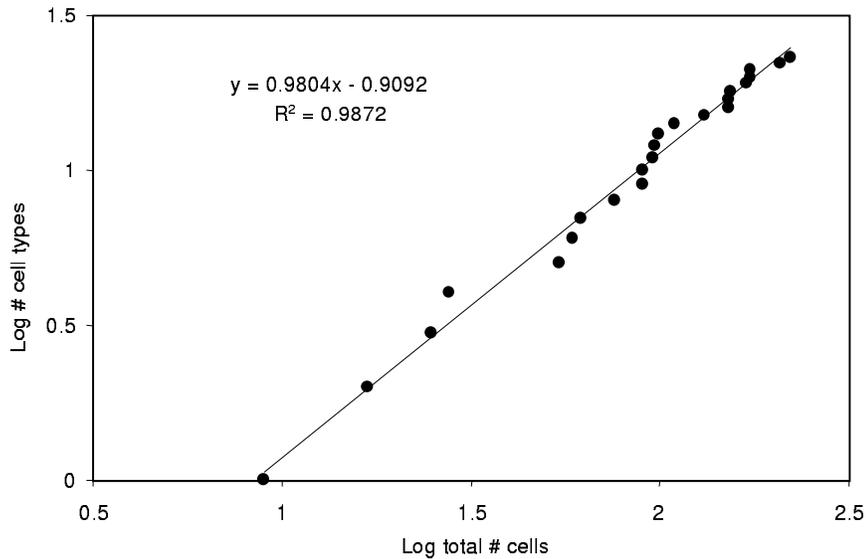


Figure 6: Logarithm (base 10) of number of cell types versus logarithm of total number of cells in *C. elegans* during gastrulation. The scaling exponent is 0.98 ($R^2 = 0.9872$, $n = 22$), or nearly 1. Here I list the cell type distinctions made; in square brackets next to each type I have put (a) the label, if there is one, for the type of cell from Sulston and White in Appendix 1 of Wood (1988), and (b) the number of cells of that type. Six kinds of epithelial cell types: (1) main hypodermis [hyp7, 83], (2) rectal hypodermis [rect, 4], (3) head hypodermis [27], (4) tail hypodermis [hyp8–12, 6], (5) interfacial [arc, 9], (6) seam [35]. Three kinds of nervous tissue: (7) neuron [302], (8) socket [23], (9) sheath [23]. Ten kinds of mesoderm: (10) head [hmc, 1], (11) anal depressor [mu anal, 1], (12) body [mu body, 79], (13) intestinal [mu int, 2], (14) pharynx [m, 46], (15) sphincter [mu sph, 1], (16) uterine [mu ut, 8], (17) vulval [mu vul, 8], (18) coelomocyte [cc, 6], (19) pharyngeal marginal [mc, 9]. Two kinds of intestinal tissue; (20) tube [int, 20], (21) valve [v, 8]. Two kinds of gland: (22) g1 [g1, 3], (23) g2 [g2, 2]. Finally, four kinds of excretory cell: (24) exc cell [1], (25) duct [1], (26) gland [2], (27) socket [1]. Data are for hermaphrodite only, and the founder and blast cells were excluded from the analysis.

organisms should have fewer cell-level functions.^[11] If free-living cells have more cell-level functions, or expressions, at their disposal, and if cells as combinatorial systems increase the number of component types to obtain greater expressive complexity, then free-living cells would be predicted to have a greater number of subcellular part types than cells in multicellular organisms. That is, as cells become members of colonies, their internal complexity should tend to decrease. Analogous arguments hold for the difference between any kind of free-living unit versus units within colonies. Preliminary research^[11–13] (McShea, 2000, 2001; McShea and Anderson, 2001) provides evidence confirming this prediction. Subcellular part type counts for metazoans and land plants tend to be significantly lower than that for protists, choanoflagellates and green algae.

However, organisms with a greater number of cell types may, over all the cell types, have a greater number cell-level expressions than any free-living cell. That is, the total number of cell-level expressions in a multicellular organism is greater the more cell types there are, and probably greater than the number of cell-level expressions a free-living cell is capable of, even though the free-living cell may have more cell-level expressions than any one cell type in a multicellular organism. If one organism has, over all its cell types, more cell-level expressions than does another organism, we expect that there must be, over all its cell types, more subcellular part types than does the other organism. So, for example, we would expect that, over all cell types in human, the total number of subcellular part types would be greater than the union of all subcellular part types found in *C. Elegans*. (This is true despite the fact that each cell type in each organism may have, on average, the same number of subcellular part types.) This is a prediction I would like to test in the future.

3.3 Hierarchies in behavior

We have now studied hierarchical complexity in artificial systems, and for some of the structure of organisms. We have not, however, touched on hierarchies concerned with the brain. Although I would eventually like to study neuroanatomical hierarchical levels, at this point I have concentrated only on hierarchical complexity in behavioral organization: namely, bird vocalization, human language and mammalian behaviors.

3.3.1 Bird vocalization

Bird vocalization is a convenient hierarchical system within behavior to study because the behavioral components and expressions are relatively easy to distinguish (compared to many non-linguistic behaviors). In this case, the components are called syllables, and the expressions are songs. Do birds that have more songs in their repertoire have more syllable types, or do they have longer songs (and greater combinatorial degree)? Is there a law-like relationship at all?

Figure 7 shows the data for a log-log plot and a semilog plot, and the data clearly conform better to the log-log plot. Birds appear to increase their song repertoire size by adding new syllable types, not by increasing their combinatorial degree. In particular, the number of syllable types scales against number of songs with exponent 0.813, i.e., $C \sim E^{0.813}$; the 95% confidence interval is [0.599, 1.027]. The inverse of this exponent gives the estimated combinatorial degree, which is 1.23, with 95% confidence interval [0.97, 1.67]. The combinatorial degree is thus not significantly different from one, and we should not conclude that bird songs are combinatorial at all. Birds with twice as many songs tend to have roughly twice as many syllables. This is despite birds having, on average, around 3 or 4 syllables per song. [Inverse log of the mean of the log number of syllables per song is 3.72, and median is 3, for 109 birds catalogued by Read and Weary.^[14]] Although syllables appear to act in a language-like manner to achieve songs, this may be illusory, and syllables and songs may not be at different hierarchical levels at all.

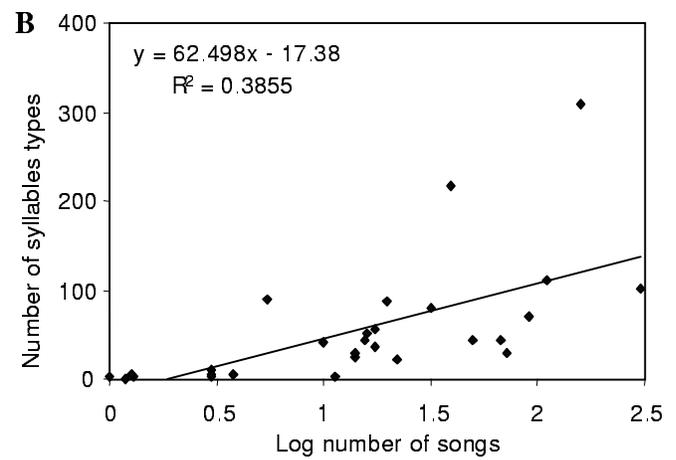
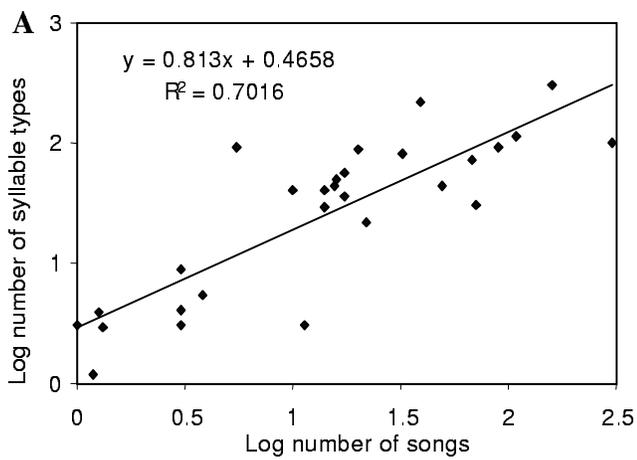


Figure 7: (A) Logarithm (base 10) of number of syllable types versus logarithm of number of songs in 28 birds. (B) Number of syllable types versus logarithm of number of songs for the same data. The birds are: 10 thrush species,^[56] 7 wren species,^[57] 5 male Bewick Wrens,^[57] 1 magpie,^[58] Tufted and Bridled Titmouses,^[59] 1 canary,^[60,61] and Alder and Willow Flycatchers.^[62]

3.3.2 English throughout history

Just as vocal behaviors are more amenable to measurement in birds than many other kinds of bird behavior, language in humans is more amenable to study than other kinds of human behavior. In particular, one of the most natural kinds of hierarchical structure, and a behavioral one at that, is the word/sentence hierarchical system. Our system will be the English language, and not for one individual, but for the entire community of English speakers. We will treat the English-speaking community as an entity that has a word type repertoire and a sentence repertoire. How many word types—i.e., entries in the dictionary, such as ‘dog’, ‘race’ and ‘the’—does this entity have at its disposal? And, how many things does entity express using those words? If we could answer those questions for the current English-speaking community, we would then have *one* data point. How may we acquire other data points where the number of sentences in the entity’s repertoire would be very different, in order to get scaling information? The English-speaking community has grown in population over time, and accordingly this entity has had more and more things to say. Thus, we may look at this same English-speaking community over time.

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; 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.

What relationship should we expect between the number of word types and the number of sentences? First, we know the grammar of English well enough to conclude that it is combinatorial. The number of word types should, then, scale disproportionately slowly against the number of expressions. However, the grammar of natural language does not constrain the number of word types, nor does it constrain how long sentences may be. Thus, if there are scaling laws relating the number of word types to the number of sentences uttered by the English-speaking community, they are extra-grammatical constraints. Which of Possibilities (a) through (d) applies?

Figure 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 rates for each are 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 $du/dt \sim e^{rt}$. And if you recall your calculus, it follows that the quantity itself scales exponentially with time, and, in fact, it scales proportionally

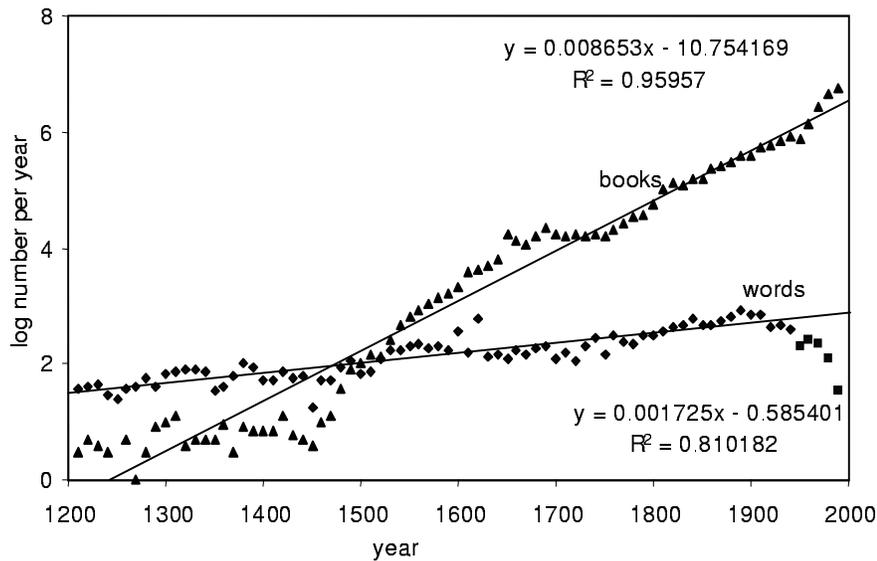


Figure 8: Growth rates in the decades from the years 1200 to 1990 for the number of new English word types and the number of new English books. Regression equations and correlation coefficients are shown for each (79 data points each). Unsure etymological dates tend to cluster at century and half century 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.

with the growth rate: i.e., $u \sim du/dt$. Thus, Figure 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 combinatorial degree is less than one and thus English is combinatorial. Thus, via some kind of extra-grammatical, or de facto, constraint, greater expressive complexity was achieved over the last 800 years not by increasing the combinatorial degree (or average sentence length), 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 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.^[15–16]

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.^[17] 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.^[17] 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,

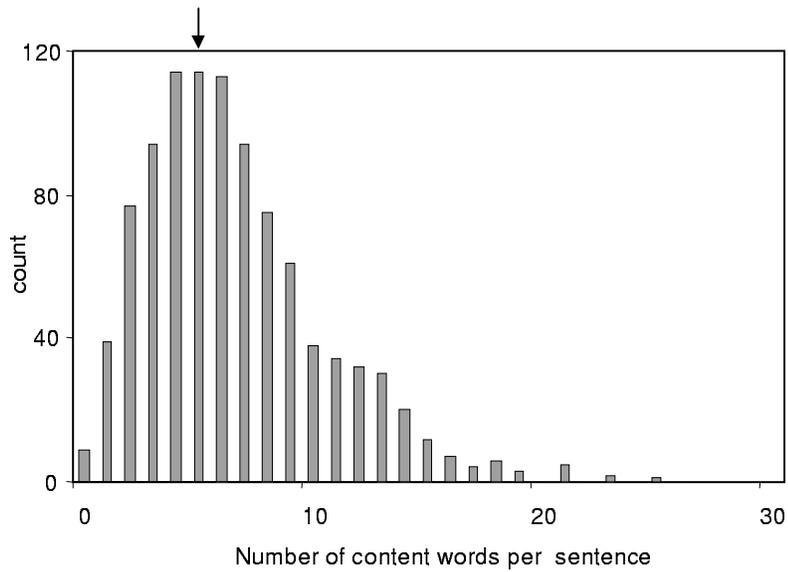


Figure 9: Distribution of numbers of content words per sentence in English. Arrow indicates the log-transformed mean. 984 sentences from 155 authors were measured from texts in philosophy, fiction, science, politics and history. I chose the second sentence on each odd numbered page. A word was deemed 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.

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 9). The distribution is log-normal (Figure 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^[18]; working memory is the extra-grammatical constraint for the combinatorial degree of around 5. 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.

3.3.3 Ontogeny of language

In addition to studying human language at the level of the entire English-speaking community, we may study it at the level of an individual developing child. Children combine phonemes into words, and words into sentences, and they eventually begin to do so combinatorially. We know that developing children learn new phoneme types and word types through time, and thus either Possibility (a) or (c) apply. Developing children also have an increasing ability to string words together.^[19–24] It is therefore possible that their combinatorial degree increases through time. However, recall that the combinatorial degree is operationally measured by determining how the number of component types scales with the *actual* number of sentences expressed (namely, the inverse of the scaling exponent), not the potential number of sentences. For all we know at the moment, developing children could indeed increase their ability to string components together, but they may not actually use this ability to express all that is within their power, in which case their combinatorial degree may not increase as fast as their (latent) ability to string words together. To test whether children's combinatorial degree keeps pace with their ability to combine, we may compare their combinatorial degree and their expression lengths through development. If combinatorial degree “keeps up” with expression length, then this implies that children are using their combinatorial powers to the fullest. If the combinatorial degree lags behind, growing disproportionately slowly, then children may be *able* to combine components into longer expressions, as evidenced by their longer sentences, but they do not actually much use this ability. I present data—for just two children—below for the phoneme/word hierarchical system and for the word/sentence hierarchical system, each for individual developing children, and we will see that in each case children do appear to have a combinatorial degree that is increasing as sentence length increases: at any point in development, they use their productive power to its fullest.

Ontogeny of words and sentences

To study the developing word/sentence hierarchical system in children, I compiled data for the number of word types and the number of distinct sentences produced by a child named Damon for 41 weeks from 12 to 22 months of age.^[25] The numbers of word types and sentences do not appear to be related by a power law, as can be seen in Figure 10 by how the plot flattens out, decreasing its slope. A logarithmic plot (not shown)—i.e., C versus $\log E$ —appears comparatively linear, providing support for Possibility (c), which is what we expected.

The plot is probably best interpreted as consisting of two power law regimes. In the first regime the estimated combinatorial degree is about 1, meaning that the child has not yet begun to treat

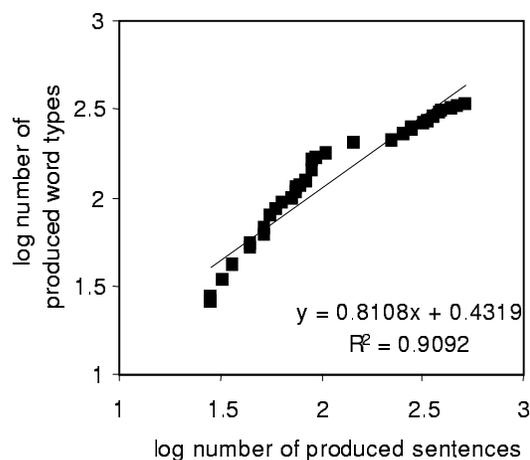


Figure 10: Logarithm (base 10) of the number of word types versus logarithm of the number of sentences, as produced by one child named Damon from 12 to 22 months.^[25] Plot is confined to multiword utterance ages, which began at about 14 months.

words and sentences as distinct hierarchical levels; it is as if his sentences are all built with just one word. In the second regime, after the bend, the estimated combinatorial degree increases to around 2.5, meaning that the child has begun using words in a combinatorial fashion to build sentences. Sentence length data do not exist for Damon, and it is not possible to directly compare his combinatorial degree increase with his sentence length increase. However, this is consistent with the typical increases in the mean length utterances of children during this period.^[23]

Ontogeny of phonemes and words

To study the developing phoneme/word hierarchical system, I compiled from Velten^[26] the number of phoneme types and the number of morphemes, as produced by a child named Jean from 11 to 30 months of age. [A *morpheme* is the smallest meaningful linguistic unit.] Figure 11 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 development, meaning the combinatorial degree is increasing. The plot of (unlogged) number of phoneme types versus the logarithm of the number of morphemes (not shown here) is comparatively linear, again suggesting Possibility (c), as we expected. Does this increasing combinatorial degree keep up with the child's apparent ability to combine phonemes into words? The combinatorial degree increases from around 2 to around 4, and scales up well with the maximum number of phonemes per morpheme over this period (Figure 11). This, again, suggests that this child utilizes his combinatorial potential to its fullest. It is as if the child has so many expressions it wishes to say that it says all that is possible with its abilities at any time.

3.3.4 Muscles and behavior

In our discussion of hierarchies in behavior thus far, the behavior has been linguistic or vocal: bird song, English throughout history, and the ontogeny of human language. We have not yet considered

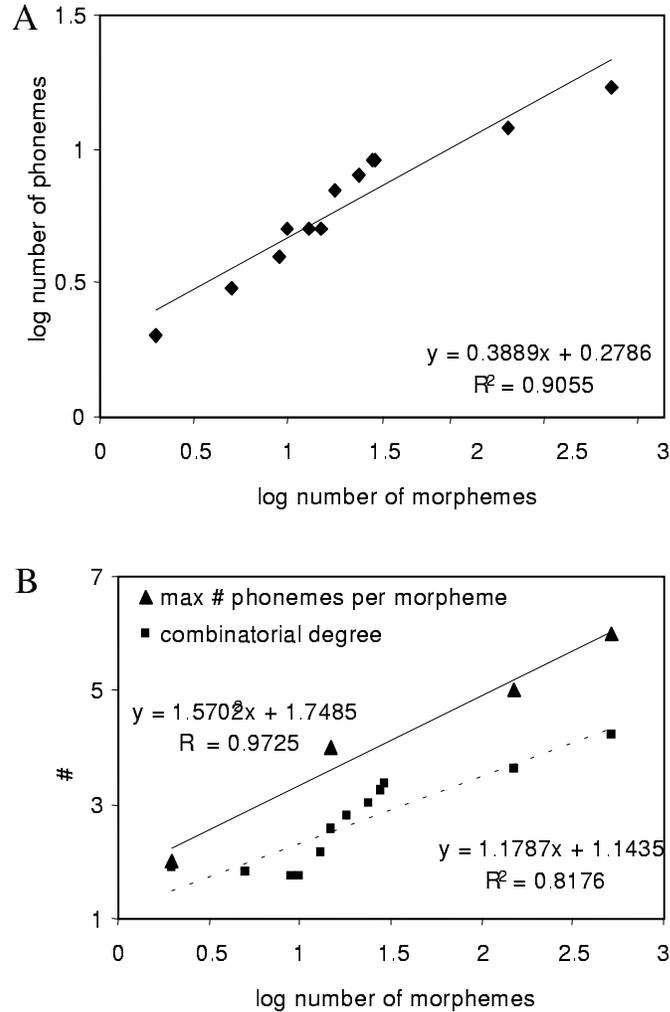


Figure 11: (A) Logarithm (base 10) of the number of phoneme types versus logarithm of the number of morphemes, as produced by one child named Jean from 11 to 30 months.^[26] Morphemes are the smallest meaningful linguistic unit, and are mostly words in this case. (B) Maximum number of phonemes per morpheme (a measure of expression length) and combinatorial degree versus logarithm of number of morphemes. Combinatorial degree is measured from the inverse of the instantaneous slope (measured via linear regression for a moving window twelve data points wide) from the log-log plot. One can see that both the maximum number of phonemes per morpheme and the combinatorial degree increase together, and are well correlated with one another.

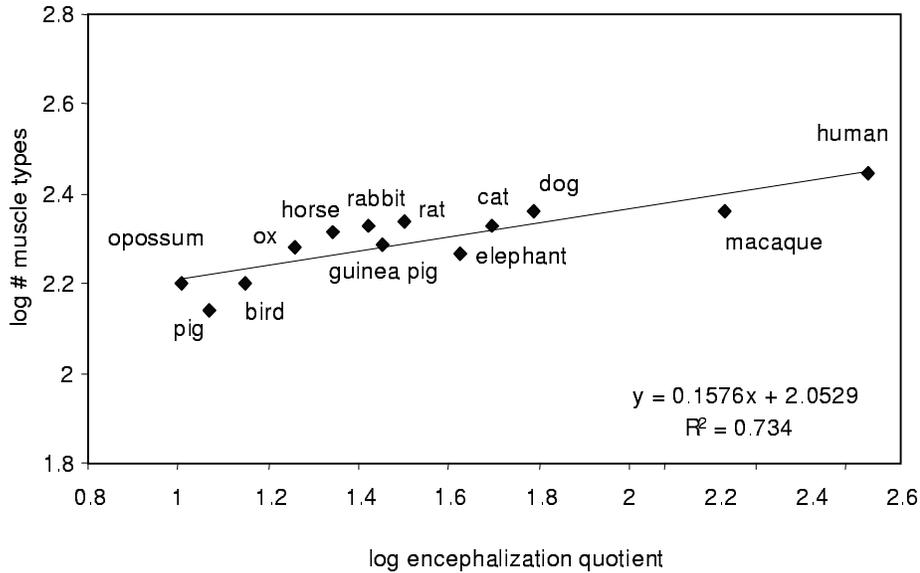


Figure 12: Logarithm (base 10) of the number of muscle types versus logarithm of the encephalization quotient, for 12 land mammals and for birds. One concern is that humans may have a greater muscle type count purely because they are much more well studied, and that the correlation may be being driven by this. Removing human from the plot leads to the equation $y = 0.164x + 2.044$ with a correlation only slightly reduced to $R^2 = 0.621$. Numbers of muscle types were taken as the maximum estimate counted from the following sources for each animal: human,^[63–66] macaque,^[67] cat,^[68–71] rat,^[72–74] rabbit,^[75–78] guinea pig,^[79] dog,^[80,68] horse,^[81,82] ox,^[83–84] pig,^[85] elephant,^[86] bird,^[87–89] opossum,^[90] Encephalization quotients were computed using brain volumes from Ref. 91-98; and body masses from Ref. 98-99.

the more run-of-the-mill non-linguistic behaviors. Behaviors, generally, are implemented via complex scores of muscle contractions^[27–31]; muscle-contractions combine to carry out behaviors. The question, then, is whether or not there is an invariant grammar governing how muscle-contractions make up behavioral expressions, and if so, which of Possibilities (a) through (d) apply. To attempt to answer this I concentrated on 12 land mammals (and one bird species), and counted up the number of muscles types in each (see legend of Figure 12).

The difficulty was in measuring the number of behaviors the animal is capable of. Ethograms provide one kind of measure, and I have begun to explore the possibility of using them, but in the research I describe here I use a notion of brain size as a proxy, the idea being that bigger brain should correlate well with behavioral complexity. Using simply brain volume, however, is not a good correlated proxy for behavioral complexity because animals with larger bodies tend to have larger brains even if they do not appear to be smarter. Thus, any interesting notion of brain bigness that is hoped to correlate with intelligence will have to factor out body mass. The *encephalization quotient* does just this, measuring how large the brain is relative to what we would expect given the animal's body size.^[32] Since, across the higher mammalian taxa, brain volume scales roughly as the 3/4 power of body mass^[33,4] the encephalization quotient is $V_{brain}/M^{3/4}$, where V_{brain} is the

brain volume and M the body mass. The encephalization quotient does, indeed, highly correlate with the apparent intelligence and behavioral complexity of mammals and birds; however, we do not know what exactly the relationship is between encephalization quotient and size of behavioral repertoire. I will make a working, and undefended, assumption that encephalization quotient is related to behavioral complexity by a power law with some unknown exponent. The encephalization quotient was computed for each of the 12 mammals and one bird species (see legend of Figure 12).

Figure 12 shows the log-log plot of the number of muscle types versus the encephalization quotient for the 13 species, and one may observe a law-like trend, with the number of muscle types clearly increasing. This suggests Possibility (a) or Possibility (c). A plot of the data under the assumption that Possibility (c) is true (not shown) actually leads to a slightly better fit; also, analysis of the instantaneous slope of the log-log plot shows significantly decreasing slope as behavioral complexity increases. These observations allow the tentative favoring of the hypothesis that the combinatorial degree actually increases as behavioral complexity increases. This is not surprising since greater encephalization might be expected to allow animals to string together longer and more complex combinations of muscle contractions (as well as coding for a greater number of behavioral expressions). I do not have any conjectures at this time for *why* mammals appear to follow Possibility (c) in this regard. Is there an optimal tradeoff between increasing the number of muscle types and increasing the combinatorial degree leading to the scaling relationship?

3.3.5 Further behavioral directions

There are a number of interesting directions I would like to take in the future.

I have, as discussed already, acquired data for how muscles combine together to implement behaviors, but the proxy for number of behaviors was encephalization quotient, which, at best, is known to correlate reasonably well with behavioral complexity. Furthermore, encephalization quotient probably best correlates with the highest hierarchical levels of behavior, and there are doubtless many hierarchical levels of behavior between muscles and the highest level. I would like to acquire behavioral repertoire size data from published ethograms for many kinds of animal, which exist for a wide variety of animals including, but in no way confined to, snails and slugs^[35,36], squid,^[37] beetles,^[38,39] true bugs,^[40] birds,^[41,42] fruit bats,^[43] horse,^[44] and monkey.^[45,46] The expectation is that the behaviors recorded in ethograms are at a hierarchical level above muscles but below the highest level behaviors. For example, ethograms have 120 behaviors for rhesus monkey,^[47] 111 for human children (McGrew in Ref. 48), 67 for cat,^[49] and 27 for *Leptothorax curvispinosus* ants.^[50] Common sense tells us that rhesus monkeys have orders of magnitude more behaviors than 120, which means ethograms catalog behaviors at a lower hierarchical level, a level nearer to the lowest level of muscles. I therefore expect that the number of muscle types should scale up more slowly than the number of ethogram-counted behaviors, which, in turn, should scale up more slowly than the number of higher level behaviors, as proxied by the encephalization quotient, and I hope in this way to gain more insight into the rules governing behavioral hierarchies.

It would also be interesting to concentrate on certain behavioral subsystems for a single species, namely human. The number of human muscle types and number of human behaviors (as estimated by encephalization quotient) gives us only one data point, and no consequent understanding of humans as behavioral combinatorial systems. But by partitioning human behavior into certain subsystems, it may be possible to acquire scaling data for how human brains combine muscles into behaviors. One behavioral subsystem would be the vocal system, wherein there are a certain number of vocal muscles and a certain number of things we do with those muscles. This would provide one data point. Another behavioral subsystem might be the human arm (not including the hand): how many muscle types are in the human arm, and roughly how many different things do humans

do with their arms (e.g., throw, lift, turn, etc.)? As a third example behavioral subsystem, facial muscles combine to make facial expressions. Other systems include hand, finger and eye. In each case there is a relevant literature from which it should be possible to get estimates for the number of expressions—e.g., phoneme counts for vocal muscles from the linguistics literature; arm or hand expression counts from motor disorder literature and the American Sign Language literature; facial expression counts from the psychology of facial expression literature. Scaling information obtained from such data could illuminate whether or not the human brain has some characteristic combinatorial degree for muscle combinations.

Animal colonies is another place ripe for study. Insect and other kinds of colonies are convenient because the notion of a behavioral component type is relatively easy to define: the set of behavioral component types just is the set of highest level behaviors of the individual within the colony, i.e., what an individual insect can do. Other measures for component types might be the number of castes, or the number of worker sizes within the colony, the idea being that greater size variability tends to correlate with greater range of lower level functionality. Data for the latter have been collected in Jaffe^[51] (and first plotted in Anderson and McShea, Ref. 52), and show a significant increase in size variability with colony size. An undergraduate student of mine named Michael A. McDannald has put together two exciting plots, each measuring the number of ant size types versus the colony size.^[53] The first is shown in Figure 13 (A), and plots the logarithm of the number of distinct worker size types, or physical castes, as a function of the logarithm of colony size. The number of physical castes clearly correlates well with colony size, and increases disproportionately slowly, as we would expect of a hierarchical system. The second plot is shown in Figure 13 (B), and it uses the ratio of maximum to minimum head width for workers in ant colonies as a proxy for the number of physical ant types within the colony: if the “maxmin head ratio” is doubled, there is, informally, twice as much room for physical size differences in ants, and so we might expect this measurement to be proportional to the number physical size types. The log-log plot of the maxmin head ratio versus colony size (Figure 13 (B)) leads, again, to a strong correlation; this serves as a replication of the first figure. Furthermore, the slopes of each plot are very similar, hovering around 0.1, suggesting a combinatorial degree of around 10. Might this be due to there being, on average, around 10 ants involved in the implementation of a colony-level expression? And if so, why 10? As for measuring the total number of expressions in a colony, if an assumption akin to the Copy-Invariance Assumption from earlier (see Subsection 3.2) holds for colonies—i.e., colonies gain in expressive complexity by having more kinds of expressions, not more expressions in absolute number (where, here, each expression is a combination of individual insect actions)—then colony size can be used as a proxy for the colony’s expressive complexity. Thus, preliminary evidence supports the conclusion that colonies as combinatorial systems increase the number of component types, and thus follow either hypothesis (a) or (c) from the introduction.

As a final future direction, one may look at ourselves as an example colony, where human workers (the components) combine together in an economy to create products (the expressions). The economy may be studied as a combinatorial system longitudinally, by studying the growth in the number of occupation types through time and also the number of different product types as a function of time.

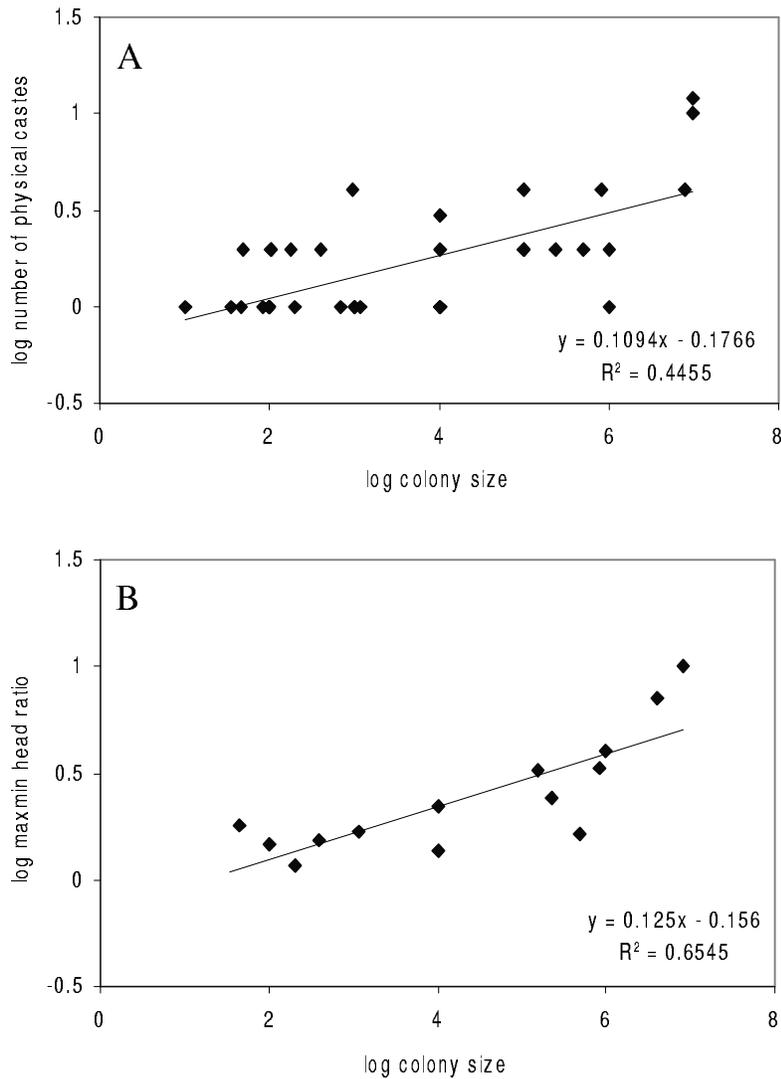


Figure 13: (A) Logarithm (base 10) of the number of physical castes for workers in ant colonies versus the logarithm (base 10) of the colony size. Data are for 24 species from Jaffe^[51] and 16 species from other places in the literature compiled by my student Michael McDannald. [Note that Jaffe mistakenly says in the legend of his Table 1 that the caste values are the logarithm of variation, but in personal communication with him we learned that the quantities refer to the number of physical castes, as measured by the number of bumps in the frequency distribution of sampled physical sizes in the colony.] (B) Logarithm (base 10) of the ratio of the maximum to minimum head width for workers in ant colonies versus the logarithm (base 10) of the colony size. We expect that this ratio should scale proportionally with the number of distinct physical size types, and note that its slope is similar to that in (A). Also, for each plot, when colony size is 1 we expect the number of castes and the head ratio to be 1, and we see that the y -intercepts are indeed each close to zero.

4 Hierarchical laws and new levels

4.1 Why an invariant combinatorial degree?

In the previous section we saw that a number of radically different kinds of hierarchical system appear to conform to universal laws (see Table 1), suggesting there are invariant grammars underlying each system of a given kind. Furthermore, each kind of system studied thus far increases its expressive complexity, at least in part, by increasing the number of component types, something that is not, in principle, a necessity (i.e., Possibility (b)). In most cases the combinatorial degree appeared to be invariant (Possibility (a)), and in some other cases it appeared to increase (Possibility (c); see Figure 1). Although I put forth ideas for some of the kinds of system for why the scaling behavior may be what it is, and why the combinatorial degree may be what it is, I have not discussed what kinds of general principles might be found across these various kinds of system.

There are at least three (extra-grammatical) reasons why a hierarchical system might conform to Possibility (a), i.e., to an invariant combinatorial degree. The first is simply that there is some upper limit constraining it. [This is also the easiest way to make sense of those cases where the combinatorial degree is increasing as well as the number of component types: the upper limit happens to be increasing. However, this does not help us to understand the principles underlying the “decision” for a kind of system to follow Possibility (c).] The second is that there may sometimes be a pressure to minimize the growth in the total number of components—not component types. The total number of components in a system is the number of expressions times the number of components per expression, i.e., EL . Supposing that the expression length is proportional to the combinatorial degree, to minimize EL requires minimizing the growth rate of the combinatorial degree, and keeping it invariant is thus optimal. This possibility only seemed relevant in our discussion the cell/organism-expression hierarchical system. The last reason I have thought of for why a kind of system would have invariant combinatorial degree is that when systems of that kind have more expressions, they have expressions that have, intuitively, completely novel meanings or functions. The meanings are so novel that no combination of component types from a less expressive system could possibly capture the meaning. Instead, a new component type had to be invented, with its own novel constituent meaning, and only with it could the new needed expressions be built. This scenario requires that the kind of system have a *compositional semantics*, which means that component types have fixed meanings, and the meaning of an expression is a function of the meanings of the constituent component types, and of the expression’s syntax. A kind of system that increased expressive complexity only in order to obtain such “completely novel” expressions I will call a *rich* kind of system. Rich kinds of systems with compositional semantics will have invariant combinatorial degree not because of an upper bound on the invariant degree, but because it never serves any expressive purpose to increase the combinatorial degree, as the only way to get the new, needed expressions is by the addition of new component types instead.

4.2 Why hierarchical levels at all?

One issue we have not yet touched upon is why there should be hierarchical levels at all. We have just taken them for granted, and sought to study the laws governing them, but given no explanation for why there are any.

The most obvious reason concerns the combinatorial power the system achieves: a small number of component types suffices for the construction of an exorbitantly larger number of expressions, and a small increase in the number of component types has a disproportionately large influence on the number of expressions. This is, for example, why there is not an extraordinary differ-

ence in the number of cell types between *C. Elegans* and mammals—about 25 and 100 cell types, respectively—and yet the difference in expressive complexity is probably many orders of magnitudes. This is also why an adult individual with a vocabulary of around 50,000 word types can utter an effectively infinite number of sentences. Bird vocalization, on the other hand, appears that it may not be combinatorial, or at least not very; doubling the number of songs therefore requires doubling the number of syllable types. There is, however, a cost to being combinatorial, i.e., a cost to having hierarchy: the kind of system must have some apparatus by which to impose a grammar, so that components can be combined appropriately to obtain expressions. Components must be guided somehow, or guide themselves, into only certain allowable arrangements. Such guidance requires information, or programming, from somewhere. For electronic devices the grammar for the user interface is in the heads of the users. For English the grammar is, again, in the head. For cell types the cells possess the information, and may utilize natural principles of self-organization to do much of the work in determining the grammar.^[54] Bird vocalization may avoid this need for a grammar: these systems may, instead, just effectively add a new syllable type every time they wish to add a new song. In this sense, syllable types just *are* the songs.

4.3 Why new hierarchical levels?

In all the discussion thus far, only two hierarchical levels were discussed. In some of the studied kinds of system, the levels were probably not adjacent; there were surely levels in between. For example, there may well be levels in between that of cells and organism-expressions. And there are certainly hierarchical levels between muscle contractions and the total behavioral repertoire.

At some point in the history of a kind of hierarchical system that has more than two levels, there may have been fewer levels, and at some point just two levels. What reasons are there for adding a new hierarchical level? What use does it serve the system? To understand this, consider a kind of system with an invariant grammar, and having an invariant combinatorial degree d_{max} due to an upper limit. While the system has only two hierarchical levels, the scaling relationship is given by

$$C \sim E^{1/d_{max}}.$$

Now suppose that a new hierarchical level is added between them, and that the number of entity types at this level is D . I will call the lowest level objects C -objects, the middle level D -objects, and the highest level E -objects; the number of object types at each level is C , D and E , respectively. While it was the case that C -objects combined directly to make E -objects, now C -objects combine to make D -objects, which, in turn, combine to make E -objects. Let us suppose that just as many E -types must be expressed as before, it is just that now their components are D -objects. And, let us suppose that each E -type requires just as many components as before; that is, each E -type, or expression, now has L D -objects in it instead of L C -objects; more weakly, I am supposing that the combinatorial degree between D and E is whatever it was between C and E , namely d_{max} . So,

$$D \sim E^{1/d_{max}}.$$

The lowest level's task has now shifted: it must now just worry about making D -objects, not E -objects. Suppose now that the combinatorial degree for C -objects is the same as it used to be, d_{max} . That is, C -objects have gotten no better (or worse) at combining into higher level objects. Then it follows that

$$C \sim D^{1/d_{max}}.$$

Now suppose we only have access to C and E , and wish to see how they scale relative to one another. By combining the last two equations we may solve for C in terms of E .

$$C \sim D^{\frac{1}{d_{max}}},$$

$$C \sim [E^{\frac{1}{d_{max}}}]^{\frac{1}{d_{max}}},$$

$$C \sim E^{\frac{1}{d_{max}} \cdot \frac{1}{d_{max}}},$$

$$C \sim E^{\frac{1}{d_{max}^2}}.$$

That is, C now scales against E with exponent of not $1/d_{max}$ as before, but with exponent $1/(d_{max}^2)$. Thus, the combinatorial degree for the C/E pair of levels is no longer just d_{max} , but d_{max}^2 . By adding just one hierarchical level, the combinatorial degree for the C/E system has been squared. If it was 5, it will now be 25. The growth in the number of lowest level component types— C types—will thus be dramatically slowed, without losing expressive power. This is an extraordinary gain in combinatorial degree. More generally, adding n new levels would modify the combinatorial degree from d_{max} to d_{max}^{n+1} . Adding hierarchical levels, then, allows systems to exponentially increase their combinatorial degree, and accordingly to keep their number of lowest-level component types very low! Miniscule changes in the the number of C types results in an explosion of new expressions at the highest level.

This is why organisms vary so astronomically in complexity, yet vary so little in number of cell types; there are hierarchical levels in between. This is why such a small number of phonemes—typically around 40—can produce effectively infinitely many complex ideas; there are many hierarchical levels in between. This is why the variability in the number of muscle types among mammals is low, and yet the highest level behavioral repertoires would appear, intuitively, to vary wildly between the apparently less behaviorally complex mammals and ourselves. Generally, if one looks at the towers of hierarchical levels in organisms, and focuses at whatever is the rock bottom lowest level, we expect to find that the number of component types at that level has stayed invariant across organisms, the reason being that the combinatorial degree for it in building whatever are the objects at the highest level should be so large that the growth in the number of component types is effectively zero. For example, a combinatorial degree of 25 would imply that $C \sim E^{1/25}$. Supposing the proportionality is equality here, when $C = 2$, $E = 3.36 \cdot 10^7$. Suppose that a system wants a bit more expressions than this. Then C will increase to 3, in which case the system will suddenly be capable of $8.47 \cdot 10^{11}$ expressions, even though it only initially wanted a little more than $3.36 \cdot 10^7$ expressions. This is more than 10,000 times more than the system initially wanted. As systems of that kind become more expressive, they will not need to increase C again until their expressive complexity increases by more than 10,000 fold, and they will thus appear to have an invariant number of component types, when, in fact, they just have a large combinatorial degree due to the many intervening hierarchical levels. Combinatorial degrees as high as 100 would be possible with just four levels with combinatorial degrees between adjacent levels of, respectively, 5, 5 and 4; in such a case, the bottom level number of component types may *never* have to increase over the course of the universe to obtain the highest level expressive complexity. It is for observations of this kind that, for example, four base pairs may forever suffice as the bottom-level component types in organisms; or alternatively, why less than two dozen amino acids may forever be sufficient building blocks for proteins.

These observations give us another possible interpretation for hierarchical systems that appear to follow Possibility (c), i.e., where the number of component types and the combinatorial degree increases. The earlier interpretation was that the expression length was lengthening, but it is also

possible, instead, that a hierarchical level is being added between the two. A new hierarchical level will not appear all at once (as it did in my simple example above). Instead, it will appear incrementally, and thus we would expect to see the combinatorial degree increasing incrementally. In a certain sense, adding a hierarchical level does increase the expression length, since there are more C objects involved in each E -object, so increasing expression length is just a more general interpretation, consistent with the possibility of the addition of a new hierarchical level.

Bibliography

1. Simon HA (1962) The architecture of complexity. *Proc. Am. Phil. Soc.* 106: 467–482.
2. Valentine JW and May CL (1996) Hierarchies in biology and paleontology. *Paleobiology* 22: 23–33.
3. McShea DW (1996) Metazoan complexity and evolution: Is there a trend? *Evolution* 50: 477–492.
4. Changizi MA (2001b) Universal scaling laws for hierarchical complexity in languages, organisms, behaviors and other combinatorial systems. *J. Theor. Biol.* 211, in press.
5. Changizi MA (2002) *The Brain from 25,000 Feet: High Level Explorations of Perception, Anatomy, Hierarchical Organization, Learning, and Vagueness*. Kluwer, Dordrecht, in press.
6. (2000) *World of Learning*. Europa, London.
7. Bell G and Mooers AO (1997) Size and complexity among multicellular organisms. *Biol. J. Linnean Soc.* 60: 345–363.
8. Cavalier-Smith T (1985) Eukaryote gene numbers, non-coding DNA and genome size. In Cavalier-Smith T (ed), *The Evolution of Genome Size*. John Wiley and Sons, New York, pp. 69–103.
9. Ross MH, Romrell LJ and Kaye GI (1995) *Histology: A Text and Atlas*. Williams and Wilkins, Baltimore.
10. Wood WB (ed) (1988) *The Nematode Caenorhabditis elegans*. Cold Spring Harbor: Cold Spring Harbor Press, Cold Spring Harbor.
11. McShea DW (2001) A complexity drain on cells in the evolution of multicellularity. Under review.
12. McShea DW (2000) Parts and integration: the consequences of hierarchy. In Jackson JBC, Lidgard S and McKinney FK (eds) *Evolutionary Patterns: Growth, Form, and Tempo in the Fossil Record*. University of Chicago Press, Chicago.
13. McShea DW and Anderson C (2001) The remodularization of the organism. Under review.
14. Read AF and Weary DM (1992) The evolution of bird song: Comparative analyses. *Phil. Trans. R. Soc. Lond. B* 338: 165–187.
15. Scudder HH (1923) Sentence length. *English J.* 12: 617–620.
16. Hunt KW (1965) A synopsis of clause-to-sentence length factors. *English J.* 54: 300–309.
17. Clark HH and Wasow T (1998) Repeating words in spontaneous speech. *Cog. Psychol.* 37: 201–242.
18. Miller GA (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* 63: 81–97.

19. Pascual-Leone J (1970) A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica* 32: 301–345.
20. Case R, Kurland DM and Goldberg J (1982) Operational efficiency and the growth of short-term memory span. *J. Exp. Child Psychol.* 33: 386–404.
21. Siegel LS and Ryan EB (1989) The development of working memory in normally achieving and subtypes of learning disabled children. *Child Develop.* 60: 973–980.
22. Adams AM and Gathercole SE (2000) Limitations in working memory: Implications for language development. *Int. J. Lang. Comm. Dis.* 35: 95–116.
23. Robinson BF and Mervis CB (1998) Disentangling early language development: Modeling lexical and grammatical acquisition using an extension of case-study methodology. *Devel. Psychol.* 34: 363–375.
24. Corrigan R (1983) The development of representational skills. In Fischer K (ed) *Levels and Transitions in Children's Development*. Jossey-Bass, San Francisco, pp. 51–64.
25. Clark EV (1993) *The Lexicon in Acquisition*. Cambridge University Press, Cambridge.
26. Velten HV (1943) The growth of phonemic and lexical patterns in infant language. *Language* 19: 281–292.
27. Tinbergen N (1950) The hierarchical organization of nervous mechanisms underlying instinctive behaviour. *Symp. Soc. Exp. Biol.* 4: 305–312.
28. Gallistel CR (1980) *The Organization of Action: A New Synthesis*. Lawrence Erlbaum, Hillsdale.
29. Fentress JC and Stilwell FP (1973) Grammar of a movement sequence in inbred mice. *Nature* 244: 52–53.
30. Baerends GP (1976) The functional organization of behaviour. *Anim. Behav.* 24: 726–738.
31. Fentress JC (1983) Ethological models of hierarchy and patterning in species-specific behavior. In Satinoff E and Teitelbaum P (eds) *Handbook of Behavioral Neurobiology*. Plenum Press, New York, pp. 185–234.
32. Jerison H (1973) *The Evolution of the Brain and Intelligence*. Academic Press, New York.
33. Allman JM (1999) *Evolving Brains*. Scientific American Library, New York.
34. Changizi MA (2001a) Principles underlying mammalian neocortical scaling. *Biol. Cybern.* 84: 207–215.
35. Everett RA, Ostfeld RS and Davis WJ (1982) The behavioral hierarchy of the garden snail *Helix aspersa*. *J. Compar. Ethology* 59: 109–126.
36. Leonard JL and Lukowiak K (1986) The behavior of *Aplysia californica cooper* (Gastropoda; Opisthobranchia): I. Ethogram. *Behaviour* 98: 320–360.
37. Hanlon RT, Maxwell MR, Shashar N, Loew ER and Boyle K-L (1999) An ethogram of body patterning behavior in the biomedically and commercially valuable squid *Loligo pealei* off Cape Cod, Massachusetts. *Biol. Bull.* 197: 49–62.
38. Danoff-Burg JA (1996) An ethogram of the ant-guest beetle tribe sceptobiini (Coleoptera: Staphylinidae; Formicidae). *Sociobiol.* 27: 287–328.
39. Betz O (1999) A behavioural inventory of adult *Stenus* species (Coleoptera: Staphylinidae). *J. Natural Hist.* 33: 1691–1712.

40. Brewer DW and Sites RW (1994) Behavioral inventory of *Pelocoris femoratus* (Hemiptera: Naucoridae). *J. Kansas Entomol. Soc.* 67: 193–198.
41. Schleidt WM, Yakalis G, Donnelly M and McGarry J (1984) A proposal for a standard ethogram, exemplified by an ethogram of the bluebreasted quail (*Coturnix chinensis*). *J. Compar. Ethol.* 64: 193–220.
42. Figueredo AJ and Ross DM (1992) The quantitative ethology of the Zebra finch: A study in comparative psychometrics. *Multivariate Behavioral Research* 27: 435–458.
43. Courts SE (1996) An ethogram of captive Livingstone's fruit bats *Pteropus livingstonii* in a new enclosure at Jersey wildlife preservation trust. *Dodo J. Wildl. Preserv. Trusts* 32: 15–37.
44. McDonnell SM and Haviland JCS (1995) Agonistic ethogram of the equid bachelor band. *Appl. Anim. Behav. Sci.* 43: 147–188.
45. Kaufman C and Rosenblum LA (1966) A behavioral taxonomy for *Macaca nemestrina* and *Macaca radiata*: Based on longitudinal observation of family groups in the laboratory. *Primates* 7: 205–258.
46. Oswald M and Lockard JS (1980) Ethogram of the De Brazza's guenon (*Cercopithecus neglectus*) in captivity. *Appl. Anim. Ethol.* 6: 285–296.
47. Altmann S (1965) Sociobiology of rhesus monkeys. II: Stochastics of social communication. *J. Theor. Biol.* 8: 490–522.
48. Hutt SJ and Hutt C (1971) *Direct Observation and Measurement of Behaviour*. C. C. Thomas, Springfield, Ill.
49. Fagen RM and Goldman RN (1977) Behavioural catalogue analysis methods. *Anim. Behav.* 25: 261–274.
50. Wilson EO and Fagen R (1974) On the estimation of total behavioral repertoires in ants. *J. NY Entomol. Soc.* 82: 106–112.
51. Jaffe K (1987) Evolution of territoriality and nestmate recognition systems in ants. In Pasteels JM and Deneubourg JL (eds), *From Individual to Collective Behavior in Social Insects*. Birkhäuser Verlag, Basel, pp. 295–311.
52. Anderson C and McShea DW (2001) Individual versus social complexity, with particular reference to ant colonies. *Biological Reviews* 76: 211–237.
53. Changizi MA, McDannald MA and McShea DW (2002) Hierarchical complexity in colonies. In progress.
54. Kauffman SA (1993) *The Origins of Order*. Oxford University Press, New York.
55. Kauffman SA (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theoret. Biol.* 22: 437–467.
56. Ince SA and Slater PJB (1985) Versatility and continuity in the songs of thrushes *Turdus* spp. *Ibis* 127: 355–364.
57. Kroodsma DE (1977) Correlates of song organization among North American wrens. *Am. Naturalist* 111: 995–1008.
58. Brown ED, Farabaugh SM and Veltman CJ (1988) Song sharing in a group-living songbird, the Australian Magpie, *Gymnorhina tibicen*. Part I. Vocal sharing within and among social groups. *Behavior* 104: 1–28.

59. Hailman JP (1989) The organization of major vocalizations in the paradae. *Wilson Bull.* 101: 305–343.
60. Mundinger PC (1999) Genetics of canary song learning: Innate mechanisms and other neurobiological considerations. In Hauser MD and Konishi M (eds) *The Design of Animal Communication* MIT Press, Cambridge, pp. 369–390.
61. Devoogd TJ, Krebs JR, Healy SD and Purvis A (1993) Relations between song repertoire size and the volume of brain nuclei related to song: comparative evolutionary analyses amongst ocine birds. *Proc. R. Soc. Lond. B* 254: 75–82.
62. Kroodsma DE (1984) Songs of the Alder Flycatcher (*Empidonax alnorum*) and Willow Flycatcher (*Empidonax Traillii*) are innate. *Auk* 101: 13–24.
63. Agur AMR and Lee MJ (1991) *Grant's Atlas of Anatomy*. Williams and Wilkins, Baltimore.
64. Netter FH (1997) *Atlas of Human Anatomy*. East Hanover, New Jersey.
65. Rohen JW and Yokochi C (1993) *Color Atlas of Anatomy*. Igaku-Shoin, New York.
66. Anson BJ (1966) *Morris' Human Anatomy*. McGraw-Hill, New York.
67. Bast TH, Christensen K, Cummins H, Geist FD, Hartman CG, Hines M, Howell AB, Huber E, Kuntz A, Leonard SL, Lineback P, Marshall JA, Miller GS Jr, Miller RA, Schultz AH, Stewart TD, Straus WL Jr, Sullivan WE and Wislocki GB (1933) *The Anatomy of the Rhesus Monkey*. Williams and Wilkins, Baltimore.
68. Boyd JS, Paterson C and May AH (1991) *Clinical Anatomy of the Dog and Cat*. Mosby, St. Louis.
69. McClure RC, Dallman MJ and Garrett PD (1973) *Cat Anatomy: An Atlas, Text and Dissection Guide*. Lea and Febiger, Philadelphia.
70. Reighard J and Jennings HS (1929) *Anatomy of the Cat*. Henry Holt and Company, New York.
71. Hudson LC (1993) *Atlas of Feline Anatomy for Veterinarians*. W. B. Saunders, Philadelphia.
72. Greene EC (1935) *Anatomy of the Rat*. The American Philosophical Society, Philadelphia.
73. Hebel R and Stromberg MW (1976) *Anatomy of the Laboratory Rat*. Williams and Wilkins, Baltimore.
74. Howell AB (1926) *Anatomy of the Wood Rat*. Williams and Wilkins, Baltimore.
75. Wingerd BD (1985) *Rabbit Dissection Manual*. The Johns Hopkins University Press, Baltimore.
76. Craigie EH (1966) *A Laboratory Guide to the Anatomy of the Rabbit*. University of Toronto Press, Toronto.
77. McLaughlin CA and Chiasson RB (1990) *Laboratory Anatomy of the Rabbit*. Wm. C. Brown Publishers, Dubuque.
78. Popesko P, Rajtov V and Horák J (1990) *A Colour Atlas of the Anatomy of Small Laboratory Animals*. Wolfe Publishing, Bratislava.
79. Cooper G and Schiller AL (1975) *Anatomy of the Guinea Pig*. Harvard University, Cambridge.
80. Adams DR (1986) *Canine Anatomy*. The Iowa State University Press, Ames.

81. Budras K-D and Sack WO (1994) *Anatomy of the Horse: An Illustrated Text*. Mosby-Wolfe, London.
82. Way RF and Lee DG (1965) *The Anatomy of the Horse*. J. B. Lippincott, Philadelphia.
83. Ashdown RR and Done S (1984) *Color Atlas of Veterinary Anatomy: The Ruminants*. University Park Press, Baltimore.
84. Singh H and Roy KS (1997) *Atlas of the Buffalo Anatomy*. Indian Council of Agricultural Research, Pusa, New Delhi.
85. Sisson S and Grossman JD (1953) *The Anatomy of the Domestic Animals*. W. B. Saunders, Philadelphia.
86. Mariappa D (1986) *Anatomy and Histology of the Indian Elephant*. Indira Publishing House, Oak Park, Michigan.
87. Chamberlain FW (1943) *Atlas of Avian Anatomy*. Hallenbeck, East Lansing.
88. Kaupp BF (1918) *The Anatomy of the Domestic Fowl*. W. B. Saunders, Philadelphia.
89. Nickel R, Schummer A, Seiferle E, Siller WG and Wight PAL (1977) *Anatomy of the Domestic Birds*. Springer-Verlag, New York.
90. Ellsworth AF (1976) *The North American Opossum: An Anatomical Atlas*. Robert E. Krieger Publishing, Huntington.
91. Frahm HD, Stephan H, Stephan M (1982) Comparison of brain structure volumes in Insectivora and Primates. I. Neocortex. *J. Hirnforschung* 23: 375–389.
92. Hofman MA (1982a) Encephalization in mammals in relation to the size of the cerebral cortex. *Brain Behav. Evol.* 20: 84–96.
93. Hofman MA (1982b) A two-component theory of encephalization in mammals. *J. Theor. Biol.* 99: 571–584.
94. Hofman MA (1983) Evolution of brain size in neonatal and adult placental mammals: A theoretical approach. *J. Theor. Biol.* 105: 317–332.
95. Hofman MA (1985) Size and shape of the cerebral cortex in mammals. I. The cortical surface. *Brain Behav. Evol.* 27: 28–40.
96. Stephan H, Frahm H and Baron G (1981) New and revised data on volumes of brain structures in insectivores and primates. *Folia primatol.* 35: 1–29.
97. Haug H (1987) Brain sizes, surfaces and neuronal sizes of the cortex cerebri: A stereological investigation of man and his variability and a comparison with some mammals (primates, whales, marsupials, insectivores, and one elephant). *Am. J. Anatomy* 180: 126–142.
98. Hrdlicka A (1907) Brain weight in vertebrates. Washington, D. C.: Smithsonian Miscellaneous Collections, pp 89–112.
99. Nowak RM (1999) *Walker's Mammals of the World*. The Johns Hopkins University Press, Baltimore.