Biol Cybern (2006)
DOI 10.1007/s00422-006-0056-x

ORIGINAL PAPER

Biological Cybernetics

Mark A. Changizi

The optimal human ventral stream from estimates of the complexity of visual objects

Received: 6 July 2005 / Accepted: 16 January 2006

Abstract The part of the primate visual cortex responsible for the recognition of objects is parcelled into about a dozen areas organized somewhat hierarchically (the region is called the *ventral stream*). Why are there approximately this many hierarchical levels? Here I put forth a generic information-processing hierarchical model, and show how the total number of neurons required depends on the number of hierarchical levels and on the complexity of visual objects that must be recognized. Because the recognition of written words appears to occur in a similar part of inferotemporal cortex as other visual objects, the complexity of written words may be similar to that of other visual objects for humans; for this reason, I measure the complexity of written words, and use it as an approximate estimate of the complexity more generally of visual objects. I then show that the information-processing hierarchy that accommodates visual objects of that complexity possesses the minimum number of neurons when the number of hierarchical levels is approximately 15.

Keywords Visual hierarchy · Number of visual areas · Sizes of areas · Optimization · Information theory · Object recognition · Intermediate-level features · Visual object complexity

1 Introduction

The mammalian visual cortex involved in object-recognition is partitioned into multiple areas, and appears to be hierarchically organized (Rockland and Pandya 1979; Van Essen and Maunsell 1983; Felleman and Van Essen 1991; Coogan and Burkhalter 1993; Scannell et al. 1995), where "lower" areas possess finer spatial feature specificity, and computations made there are subsequently used by "higher" areas possessing coarser spatial feature specificity, but finer "object" specificity. One striking feature about visual cortical organization in animals like macaque and cat is that there are so many hierarchical levels. Although current anatomical information does not determine a unique hierarchy for these animals, candidate hierarchies typically consist of about 10–20 levels (Hilgetag et al. 1996, 2000). Here I ask, Why are there approximately this many hierarchical levels? I put forth a generic model of an information processing hierarchy for vision, and derive a formula relating the total number of neurons in the hierarchy to the complexity of visual objects it must process, and to the number of hierarchical levels. I estimate the complexity of visual objects for humans, and then show that the total number of neurons in a hierarchy capable of processing such objects is minimized when there are approximately 15 hierarchical levels.

2 Generic information-processing hierarchy for vision

I begin by considering a simple generic information-processing hierarchy for vision, capturing just the bare essentials of a (bottom-up) visual hierarchy (see Appendix A), and making no specific assumptions about the computational mechanisms. Each level is partitioned into modules (e.g., columns, barrels, blobs), each module which consists of neurons having the same receptive field which does not overlap that of other modules, and the union of all the receptive fields amounts to the entire visual field (or the entire retina). (The main results we present here hold if there is receptive field overlap, so long as the percentage of overlap does not vary as a function of level.) Modules in higher levels have larger receptive fields, and receive input converging from multiple modules from the level below it; call this group of lower-level modules a *convergence zone*. See Fig. 1a.

The activation pattern of an $i+1$-level module depends on the state of the modules in its convergence zone, and there must exist neurons encoding instructions, or commands, telling the $i+1$-level module how to activate depending on the state of the convergence zone. This generic model assumes that these "command neurons" are within the convergence

M.A. Changizi
Sloan-Swartz Center for Theoretical Neurobiology,
MC 139-74, Caltech, Pasadena,
CA 91125, USA
E-mail: changizi@caltech.edu,
URL: http://www.changizi.com

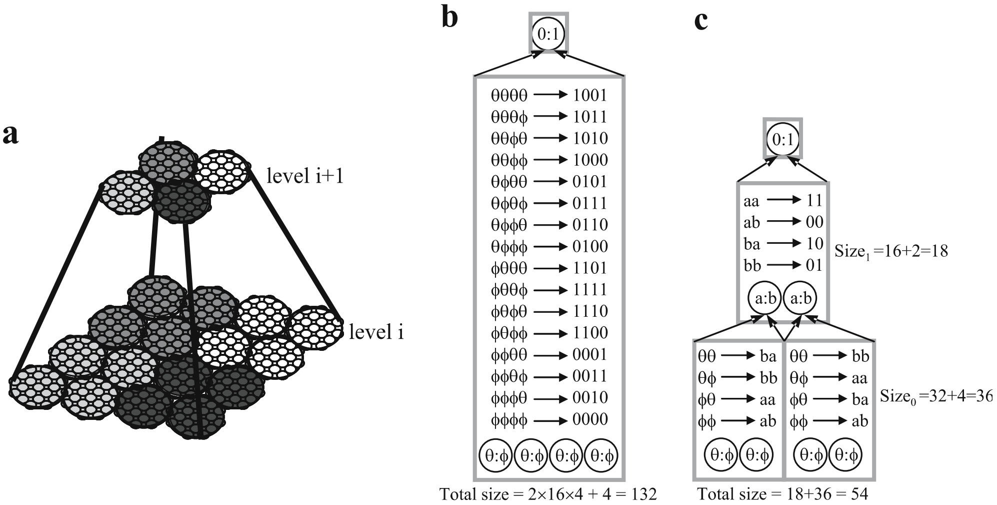

Fig. 1 **a** Illustration of the model for visual representations. Two levels are shown, and each level consists of modules (*the larger circles*) built from neurons (*the tiny circles*), each module representing a portion of the visual field. Multiple modules in level i (shown here sharing a *gray-level*) converge into a single module in level $i+1$, where here the level–level convergence factor is $\mu = 4$. This only illustrates the neurons responsible for representing visual features, and does *not* show the neurons responsible for instructing the next level on how to activate. **b** Illustration of the instructions required if only two levels – the top and bottom – are allowed. Because there is no intermediate level, there is just one convergence zone in the bottom level, and so the level–level convergence $\mu = 4$. For illustration-sake, I presume that each module (shown as *circles*) is capable of only two states; I indicate the two module states possible for the bottom level an θ and φ (analogous to 0 and 1, but I want different binary symbols for the different levels). The number of states possible for the zone is 2^4=16. For each of these 16 different convergence zone states, the zone must cause a different sequential activation pattern in the top level module to which it converges, where such an activation pattern is a sequence of four 0s and 1s. This is achieved with neurons coding up these 16 instructions, as shown, where for the illustration we assume that one (binary) neuron is required per symbol shown. For example, the first instruction says, "If the convergence zone modules are all θs, then the top-level module must sequentially activate in the pattern 1,0,0,1." The total number of symbol tokens (or binary neurons) required for these instructions, along with the four binary modules (each requiring just one binary neuron here), is 132. **c** Illustration of how adding an intermediate hierarchical level can reduce the overall number of neurons required in the hierarchy. Suppose that one intermediate level is placed between the bottom and top; it possesses two modules, and each module is capable of two states, a and b . There are, then, two convergence zones in level 0, and each is capable of $2^2 = 4$ states. For each of these four different convergence zone states, the zone must cause a different sequential activation pattern in the intermediate level module to which it converges, where such an activation pattern is now a sequence of two as and bs. This is achieved with neurons coding up these four instructions, as shown. This requires 16 symbol tokens (or binary neurons) per zone, for each of the two bottom level zones; for 32 symbol tokens in the bottom level, plus 4 tokens for the four binary modules. The same occurs for the single zone in the intermediate level instructing the top level, where there are 16 symbol tokens for instructions, plus two new modules (each here with just one binary neuron). In all, then, with three levels the number of symbol tokens required for the hierarchy is 54, which is 41% the size required when there was no intermediate level, despite the two hierarchies being computationally identical. (Note that these hierarchies have no redundancies, for the sake of the example, but real visual hierarchies certainly do, and the model accommodates this.)

zone of the ith level (but it makes no difference to my predictions if the commands are placed in the $i+1$st level). Consider, for example, Fig. 1b, which shows four modules (the four circles) in the single bottom level zone. Because each module is capable of two possible states, θ or ϕ, the bottom level zone is capable of $2^4 = 16$ possible states. If the single top level module (the single circle at the top) to which this zone converges is (for example) to activate differently for each of these 16 possible states, there must be instructions in the hierarchy somewhere telling the top level module how to respond for each of these 16 different possible zone states. These instructions are shown as the list of 16 instructions in the bottom zone as to how the top module should sequentially activate conditional on the 16 different possible zone states. This example assumes that the top module must discriminate among all the different possible lower level zone states – i.e., it assumes no redundancy, or no loss of information – but more generally the model allows redundancy, as discussed later.

From this generic information-processing hierarchy for vision it is possible to derive an equation for the total number of neurons in the hierarchy (Eq. (6) in the Appendix):

$$N_{\mathrm{reg}}\,(n,\sigma,\mu_{\mathrm{tot}},d_{\mathrm{tot}}) \approx \log(\sigma)\times\left[1+(d_{\mathrm{tot}}/\mu_{\mathrm{tot}})^{1/n}\sigma^{d_{\mathrm{tot}}^{1/n}}\right] \times\frac{\left[\mu_{\mathrm{tot}}^{(n+1)/n}-1\right]}{\left[\mu_{\mathrm{tot}}^{1/n}-1\right]} \qquad (1)$$

Actually, the equation shows the number of *regularized* neurons, $\mathrm{N}_{\mathrm{reg}}$, which is the total number of neurons in the entire hierarchy below a single module in the top level; i.e., it is the sum of all the neurons in the hierarchy that are involved in the information that eventually converges to one top-level module. The (regularized) number of neurons is a function of four parameters: n (the number of hierarchical levels above the bottom level), σ (the number of visual representational states a single module is capable of), μ_{tot} (the total convergence factor over the entire hierarchy), and d_{tot} (the total combinatorial degree over the entire hierarchy). In the following four sections (Sect. 3–6) I discuss the meaning of these four parameters and how they may be determined. This will involve the introduction of principles of optimization and efficient coding, as well as empirical estimation.

3 Number of levels, $n + 1$: set to minimize the total number of neurons in hierarchy

The first parameter in Eq. (1) is n, the number of hierarchical levels above the bottom (so that $n + 1$ is the total number of hierarchical levels). Recall the illustrative hierarchy in Fig. 1b, where four modules in the bottom level converge to one module at the top level. Consider what happens if an intermediate level, with two modules, is allowed, as shown in Fig. 1c. First, notice that the number of modules in a convergence zone drops from 4 to 2. Athough two new modules must be added, the total number of neurons required for the commands is greatly reduced, achieving a significant savings in the number of required neurons to implement the hierarchy. More generally, as we will see later in Fig. 2, the (regularized) number of neurons required from Eq. (1) tends to be enormous when only two levels are allowed, and increasing the number of levels above two tends to drastically reduce the total number of neurons required for the hierarchy, often by many orders of magnitude. Increasing the number of intermediate levels still further leads to further reductions in the number of neurons, up to a point after which the number of neurons increases fairly slowly. [A similar phenomenon may drive the large-scale organization of the natural language lexicon (Changizi MA, in review).]

Evidence of volume-optimization in the brain has been found in a variety of ways (Cajal 1995; Kaas 2000; Cowey 1979; Mead 1989; Durbin and Mitchison 1990; Mitchison 1991, 1992; Ringo 1991; Cherniak 1992, 1994, 1995; Cherniak et al. 1999; Jacobs and Jordan 1992; Traverso et al. 1992; Ruppin et al. 1993; Van Essen 1997; Chklovskii and Koulakov 2000; Changizi 2001a,b, 2005; Changizi and Shimojo 2005b), and central to my hypothesis is the following principle of parsimony:

> The number of hierarchical levels above the bottom, n, has been selected by evolution so as to minimize the total number of neurons, N, required in the hierarchy.

I will denote this optimal value of n as n_{opt}. In other words, given values for the other three parameters (σ, μ_{tot} and d_{tot}), determine the value of n that minimizes $N_{\text{reg}}(n)$ from Eq. (1). That value of n is called n_{opt}, and is the predicted number of hierarchical levels above the bottom. The number of hierarchical levels above the bottom, n, is, then, no longer a free parameter, but is set to n_{opt}. This leaves three parameters which we discuss in the three following sections.

4 The total combinatorial degree, d_{tot}: set to efficiently code visual objects

To understand a second parameter in Eq. (1), the total combinatorial degree, d_{tot}, it is helpful to first understand the *total convergence*, μ_{tot}, which is the total number of modules from the bottom level that eventually converge, over the entire hierarchy, to a single top-level module. The total convergence can be interpreted as the maximum possible number of degrees of freedom a top level module might have to accommodate. (We will discuss the total convergence more in Sect. 5.) But actual visual objects that people recognize may well possess redundancies (e.g., due to statistical regularities in the ecology), so that the true number of degrees of freedom required of a system is well below μ_{tot}. Let d_{tot} be the total number of degrees of freedom a top-level module is capable of, where $d_{\text{tot}} = \beta_{\text{tot}}\mu_{\text{tot}}$, and β_{tot} is the *total redundancy fraction*. I will refer to d_{tot} as the *total combinatorial degree* of the system (Changizi 2001c, 2003b; Changizi et al. 2002); it is the entropy in base-σ of a top-level module. Total combinatorial degree values can be as low as $d_{\text{tot}} = 1$, intuitively meaning that the μ_{tot} modules in the bottom level that ultimately converge to a top level module do not interact combinatorially, and as high as $d_{\text{tot}} = \mu_{\text{tot}}$, meaning that all the μ_{tot} potential degrees of freedom are utilized. For example, although the average English sentence may possess about 20 words – and therefore the maximum possible number of degrees of freedom for a sentence is 20 – there are redundancies (e.g., due to some words being highly predictive of adjacent words), and so the total number of degrees of freedom is nearer to 5 (Changizi 2001c).

An efficient visual system (Attneave 1954; Barlow 1961; Simoncelli and Olshausen 2001) is expected to be only as complex as needed to accommodate the ecologically typical visual objects, and one therefore expects that the total combinatorial degree, d_{tot}, of the hierarchy has been selected by evolution to roughly match the number of degrees of freedom found in visual objects. For this reason, it would suffice to acquire an estimate of the number of degrees of freedom in visual objects, and then, on the basis of the "efficient visual system" assumption, to infer that the ventral stream hierarchy has a total combinatorial degree, d_{tot}, roughly matched to this.

In an effort to estimate the number of degrees of freedom for visual objects for humans, I will provide such an estimate for human visual signs that are plausibly analogous to visual objects: *the written word*. Written words are similar to other visual objects in at least two respects. First, written words are structurally somewhat analogous to other visual objects in the sense that written words are composed of letters which are analogous to object-junctions (Changizi et al. 2006), and letters, in turn, are built from strokes which are analogous to contours (Changizi and Shimojo 2005b). Second, written words activate similar regions of the inferotemporal visual cortex as other visual objects (Hasson et al. 2002). This latter phenomenon is presumably related to the fact that words tend to be the largest written linguistic entities that are recognizable;

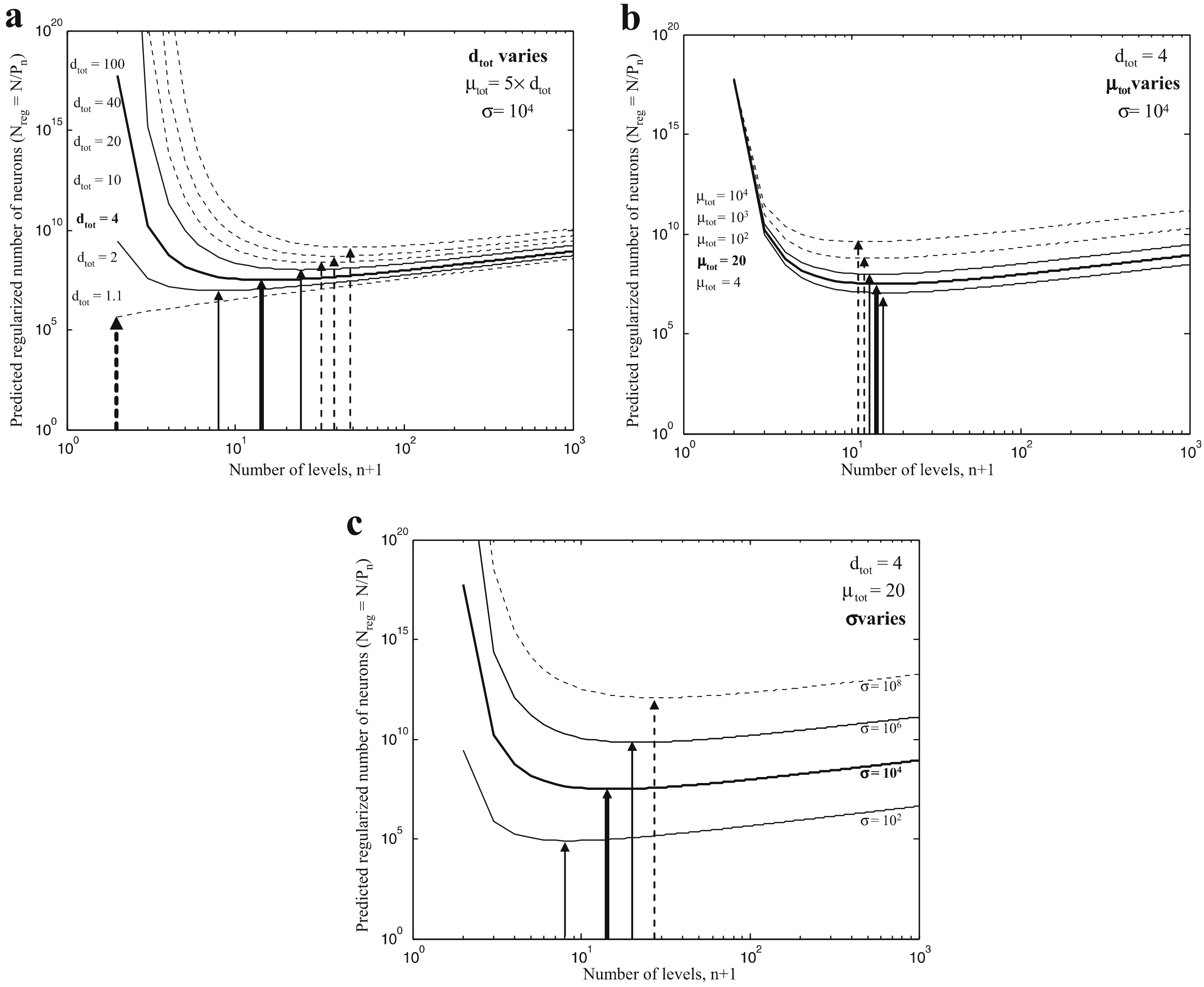

Fig. 2 The predicted number of hierarchical levels for the individual perturbation of each of the three parameters within its plausible range. The plots are of the model's regularized total number of neurons, N_{reg} (Eq. (1)), versus the number of hierarchical levels, $n + 1$. *Solid curves* are within the empirically plausible range for the parameter being varied, and *solid arrows* indicate the optimal number of hierarchical levels, $N_{\mathrm{opt}} + 1$; the *bold curve* and *arrow* indicate the best estimate of the parameter. *Dotted curves* and *dotted arrows* are for values of the varying parameter outside of its empirically plausible range. **a** Plots of N_{reg} versus n for seven values of the total combinatorial degree, $d_{\mathrm{tot.}}$ The optimal number of levels for each curve are 2, 8, 15, 25, 32, 39 and 49, respectively, and the relationship between n_{opt} and d_{tot} is given by $n_{\mathrm{opt}} = 24.0 \log(\mathrm{d_{tot}}) - 0.2$. For the range of empirically plausible values of d_{tot} (2–10), the number of hierarchical levels ranges from 8 to 25, with a "best" estimate of 15 levels when $d_{\mathrm{tot}} = 4$. **b** Plots of N_{reg} versus n for five values of the total convergence, μ_{tot}. The optimal number of levels for each curve are 16, 15, 14, 13 and 12, respectively, and the relationship between n_{opt} and μ_{tot} is given by $n_{opt} = -1.17 \log(\mu_{\mathrm{tot}}) + 15.5$. For the range of empirically plausible values of $\mu_{\mathrm{tot}}(4$ to $100)$, the number of hierarchical levels ranges from 14 to 16, again with a "best" estimate of 15 levels when $\mu_{\mathrm{tot}} = 20$. **c** Plots of N_{reg} versus n for four values of the number of module states, σ The optimal number of levels for each curve are 9, 15, 21 and 28, respectively, and the relationship between n_{opt} and σ is given by $n_{\mathrm{opt}} = 3.15 \log(\sigma) + 1.5$. For the range of empirically plausible values of $\sigma(10^2–10^6)$, the number of hierarchical levels ranges form 9 to 21, again with a "best" estimate of 15 levels when $\sigma = 10^4$. Note that we are assuming neurons are binary for the computation here, which affects only the overall height of the plots

i.e., other than a few highly stereotyped phrases, we do not recognize written sentences, but, instead, must construct sentences out of the recognized words.

It is possible that the complexity of written words has, in fact, been culturally selected to match our visual processing constraints. To understand this, it is important to realize that the visual complexity of written words is not determined by spoken language. This is for two reasons. *First*, there is the choice of how or whether to visually represent the phonemic constituents of a word; i.e., whether to use a writing system that is logographic, a syllabary, an abugida, an alphabet, or an abjad. Spoken language does not determine this choice, and the outcome of this choice affects the visual complexity of the written word. *Second*, even once this decision is made, there is the choice of how to visually represent the characters themselves; e.g., whether to use color modulations or contour modulations to code characters, how many strokes per character to use, how much redundancy, etc. (see Changizi and Shimojo 2005a). This also is not a consequence of spoken language, and strongly affects the visual complexity of written words. Because the visual complexity of written words is not determined by spoken language, it is possible that writing has been culturally selected in such a way that the visual complexity of written words is approximately that of visual objects. While this cultural selection hypothesis for written words is a possibility, and is not contradicted by spoken language, it is not my intent to defend such a hypothesis here; I only mean to argue that it is a live possibility. My arguments that the visual complexity of written words may be representative of visual objects is based on the two arguments in the previous paragraph.

Because of the linguistic nature of this special kind of visual object – written words – it is possible, as I show below, to compute information theoretic quantities, in particular the combinatorial degree (or the number of degrees of freedom), something not at all easily done for visual objects generally.

Changizi and Shimojo (2005b) measured how strokes combine to make characters in 115 writing systems over humans history, and found an average combinatorial degree for strokes combining into characters of approximately $d_{\text{stroke}-\text{char}} \approx 1.5$ (compared to 3 strokes per character on average, meaning a redundancy of approximately 50%). [Strokes are defined via visual discontinuities, so that "U" possesses one stroke but "V" possesses two strokes (Changizi and Shimojo 2005b; Changizi et al. 2006).] Shannon (1951) found that the entropy for English words is approximately 11.82. Because there are 26 letters, the average combinatorial degree for characters combining into words is the base-26 entropy, which is $d_{\text{char}-\text{word}} \approx 2.5$ (compared to 4.5 characters per word on average, meaning a redundancy of approximately 50%). The total combinatorial degree for strokes combining into words is the product of these two combinatorial degree values, and is $d_{\text{stroke}-\text{word}} \approx d_{\text{stroke}-\text{char}} \times d_{\text{char}-\text{word}} \approx 1.5 \times 2.5 = 3.75$.

In the empirical estimation of d_{tot} thus far, I have been treating strokes as the bottom-level symbols. Strokes themselves require recognition, however, and this may occur above V1. How many degrees of freedom are there in possible strokes? Here I examined the strokes found across the 115 writing systems of Changizi and Shimojo (2005b), and determined the number of distinct concavities for each stroke type in each writing system; for example, a straight line has zero distinct concavities, a "C" has one, and an "S" has two. I also determined the frequency distribution of the stroke types as they occur within the character types of the writing system. For each writing system, I computed the frequency-weighted average number of distinct concavities per stroke type. I then averaged these values across the writing systems, obtaining an average of 0.497 distinct concavities per stroke. Considering a straight stroke as having primarily just one degree of freedom, namely orientation, and each distinct concavity as adding another potential degree of freedom, the average potential number of degrees of freedom is then $1.497 \approx 1.5$. But just as there are redundancies in building words from characters, and characters from strokes, there are probably redundancies in building strokes from "distinct concavities." I do not have the ability to compute the redundancy for the latter, however, because this would require some notion as to what the set of symbol types are from which strokes are built, something I do not have. The combinatorial degree for the construction of strokes is, then, $d_{\text{bottom}-\text{stroke}} = 1$–$1.5$ (recall combinatorial degree values are ≥ 1). The total combinatorial degree for a written English word would then be $d_{\text{bottom}-\text{word}} \approx d_{\text{bottom}-\text{stroke}} \times d_{\text{stroke}-\text{char}} \times d_{\text{char}-\text{word}} \approx$ 3.75–5.6.

Using this combinatorial degree measurement for written words as an estimate of the combinatorial degree for visual objects more generally (as discussed earlier), we expect that an efficient visual system will have a total combinatorial degree, d_{tot}, that approximately matches the combinatorial degree of visual objects, and so $d_{\text{tot}} \approx 3.75$–$5.6$. I will somewhat arbitrarily choose $d_{\text{tot}} \approx 4$ as the "best" estimate, and I will consider as empirically plausible values of d_{tot} from about 2 to 10.

5 The total convergence, μ_{tot}: empirical estimates

The *total convergence*, μ_{tot}, is the total number of modules from the bottom level that eventually converge, over the entire hierarchy, to a single top-level module. The total convergence is the maximum possible number of degrees of freedom a top level module might have to accommodate. Here I present two very approximate estimates for μ_{tot}.

We saw above that although the number of degrees of freedom from strokes to words is $d_{\text{stroke}-\text{word}} \approx 4$, the average number of strokes per word is $L_{\text{stroke}-\text{word}} \approx 3 \times 4.5 = 13.5$, and thus $d_{\text{stroke}-\text{word}} \approx 0.28 \times L_{\text{stroke}-\text{word}}$, and the redundancy fraction is $\beta_{\text{stroke}-\text{word}} \approx 0.28$. Supposing for simplicity that, at some hierarchical level, each stroke is recognized in a separate module, then around 13.5 modules would be needed, despite there being only about 4 degrees of freedom. Therefore, this suggests a total redundancy fraction for the visual system of $\beta_{\text{tot}} \approx 0.28$, and if we set $d_{\text{tot}} \approx 4$ (see above), then the total convergence $\mu_{\text{tot}} = d_{\text{tot}}/\beta_{\text{tot}} \approx 15$. Supposing that there are further redundancies in the construction of strokes themselves, say 50% (as is the case for strokes-to-letters and letters-to-words), the total convergence may be double this, or 30.

We may also estimate μ_{tot} by considering the relative receptive field size from V1 to a top level area in inferotemporal cortex. Receptive field sizes for inferotemporal modules are on the order of 10° for natural stimuli (e.g., Rolls et al. 2003); meaning there are $P_n \sim (180°)^2/(10°)^2$ modules in an inferotemporal area. The receptive field sizes for the set of modules from V1 that eventually converge to an inferotemporal module will depend on the range of eccentricities of those modules (because receptive field size in V1 grows with eccentricity). Receptive field sizes in V1 vary eccentrically from about a quarter of a degree to 10° (Van Essen et al. 1984; Rosa 1997), with a (log-transformed) average of approximately 1.5°. Using this average, the number of modules in V1 is $P_0 \sim (180°)^2/(1.5°)^2$. Recalling that $\mu_{\text{tot}} = P_0/P_n$, we have $\mu_{\text{tot}} \approx 10^2/1.5^2 = 44.4$.

I have just discussed two separate approaches to estimating the total convergence, μ_{tot}. The first used redundancy estimated from visual objects and concluded $\mu_{\text{tot}} \approx 15$–$30$. The second approach estimated the total convergence by utilizing the relative receptive field size from top to bottom, and concluded that $\mu_{\text{tot}} \approx 50$. I will accordingly set the total convergence to be approximately within this range, and more generally to consider as empirically plausible values of μ_{tot} from 4 to 10^2 (the former being the "best" estimate for d_{tot} discussed above, which would correspond to zero redundancy). The logarithmically mid-way point within this range is 20, which I will somewhat arbitrarily use as the "best" estimate in what follows. Note that Fig. 2b shows that the predicted number of hierarchical levels only negligibly depends on the choice of μ_{tot}, varying by only several levels despite μ_{tot} ranging over nearly four orders of magnitude.

6 The number of states per module, σ: approximate range

Recall that modules are composed of neurons in the same level who share the same receptive field. Let σ be the number of states a module is capable of. Intuitively, it is akin to the "pixel depth" of a computer monitor, such as 16 bit color per pixel versus 32 bit. The number of neurons required to specify σ many states will be logarithmic in σ, and the total number of neurons in the entire hierarchy will depend on σ, as Eq. (1) shows. Empirically estimating σ is difficult, but as we will see later (Fig. 2c), the predicted number of hierarchical levels varies only weakly with σ.

As one attempt at an approximation, consider that at the very top level there may be a relatively small number of modules (compared to V1), each with a much larger receptive field size than that of lower levels, and each module capable of responding to a large repertoire of visual objects. For example, if the top level is deemed to be an object-recognition area in the inferotemporal lobe such as human VOT (Malach et al. 2002; Hasson et al. 2003), then there still exists a large-scale retinotopic map, with, for example, buildings more eccentric than faces, and the total visual object repertoire is the union of the repertoires across all these modules. Now consider that humans have verbal vocabulary sizes of about 5×10^4, and one might reasonably expect a total visual object repertoire size around the same order of magnitude. Because the total visual object repertoire is accommodated by the union of the modules in the inferotemporal area (say, VOT), any one module will have a lower visual object repertoire size. That is, 5×10^4 provides a reasonable upper bound to the number of states for a module in a high-level object-recognition area. In this light, I will suppose that, very approximately, $\sigma \approx 10^4$, but more weakly, I will assume that σ is in the range of 10^2–10^6 . As mentioned above, the predicted number of levels will only weakly depend on the setting of σ, as Fig. 2c shows; for example, the predicted number of hierarchical levels varies only by about a factor of three despite varying σ over six orders of magnitude.

7 Main result: predicted number of hierarchical levels in the ventral stream

In the previous sections I introduced the generic hierarchical information-processing model for vision, which captures essential features of any visual hierarchy. I showed how the total (regularized) number of neurons varies as a function of four parameters, n, d_{tot}, μ_{tot} and σ. The number of hierarchical levels above the bottom, n, was hypothesized to be set by a principle of parsimony, namely set to minimize the total number of neurons. The total combinatorial degree, d_{tot}, for human was presumed to match (for reasons of efficient coding) the number of degrees of freedom found in visual objects, and as an estimate of the latter I measured the number of degrees of freedom found in written words. I concluded that the range of plausible values for d_{tot} are from 2 to 10, with a "best" estimate of $d_{\text{tot}} \approx 4$. The total convergence, μ_{tot}, was estimated by two different techniques, and I concluded that the range of plausible values for it are from 4 to 10^2, with a "best" estimate of $\mu_{\text{tot}} \approx 20$. And the number of states per module, σ, was estimated to be in the range of 10^2–10^6, with a "best" estimate of $\sigma \approx 10^4$.

It is now possible to examine the hypothesis' predicted number of hierarchical levels for the human ventral stream. Figure 2 shows plots of the total (regularized) number of neurons, N_{reg}, versus the number of levels, $n + 1$, (from Eq. (1)). One can see that the total number of neurons precipitously falls when increasing the number of levels above two, in some cases falling by as much as ten orders of magnitude at the minimum. After reaching the minimum, the number of neurons increases relatively slowly. The three parts of Fig. 2 – i.e., (a), (b) and (c) – differ in that in each one, one of these three parameters is being varied around its "best" setting, and plots of N_{reg} versus $n + 1$ are shown for these varied values of that parameter. Solid-line curves are cases where the varying parameter is still within its "plausible" range, as discussed in the previous sections, and the dotted-line curves are cases where the parameter is outside this range. For each curve, an arrow indicates the optimal number of hierarchical levels, $n_{\text{opt}} + 1$.

Using the previous sections' "best" estimates of the three parameters – $d_{\text{tot}} = 4$, $\mu_{\text{tot}} = 20$, $\sigma = 10^4$ – it is possible to determine from Eq. (1) the predicted number of hierarchical levels, $n_{\text{opt}} + 1$. The N_{reg} versus $n_{\text{opt}} + 1$ curve for these "best" parameter settings is shown in bold in each of the three plots in Fig. 2, and the optimal number of hierarchical levels indicated with a bold arrow. In particular, the optimal number of levels is 15. Perturbations of any one of the three parameters – d_{tot}, μ_{tot}, σ – within its "reasonable" range as discussed earlier lead to $n_{\text{opt}} + 1 \in [8, 25]$ for $d_{\text{tot}} \in [2, 10]$, $n_{\text{opt}} + 1 \in [13, 15]$ for $\mu_{\text{tot}} \in [4, 10^2]$, and $n_{\text{opt}} + 1 \in [9, 28]$ for $\sigma \in [10^2, 10^6]$. That is, reasonable empirical settings of these three parameters predicts between approximately 8 to 28 hierarchical levels, and a "best" prediction of approximately 15. (See Fig. 3h in Appendix B for the predicted relative sizes of the areas as a function of hierarchical level in the human ventral stream.)

8 Discussion

In this paper I have provided a framework that allows us to quantitatively determine the optimal number of levels (i.e., the number of hierarchical levels that minimizes the total number of neurons), and that, more generally, connects the recognition demands of the visual system (i.e., the combinatorial degree, d_{tot}) to the organization of the visual system (e.g., the number of hierarchical levels and the number of neurons per level). The main hypothesis was that actual visual hierarchies will possess the number of hierarchical levels that minimizes the total number of neurons required to implement the hierarchy. To make quantitative predictions, empirical estimates had to be made of three properties: the number of states per module, σ; the total convergence from bottom to top, μ_{tot}; and the total combinatorial degree of the system (or its "complexity"), d_{tot}. As seen in Fig. 2, the predicted number of levels (i.e., the optimal number of levels) depends only weakly on the first two (σ and μ_{tot}), but strongly on the last (d_{tot}).

In order to estimate the total combinatorial degree, d_{tot}, I made an "efficient coding" hypothesis that d_{tot} will be roughly matched to the actual number of degrees of freedom found in the visual objects encountered in natural scenes. Because written words appear to possess similarities to other visual objects – similar overall contour-junction-whole structure, and activation in similar parts of inferotemporal cortex – as an estimate of the "complexity" of visual objects for humans, I measured the number of degrees of freedom in written words, a kind of visual object for which, because of its linguistic nature, it is possible to estimate the number of degrees of freedom (for which I utilized, in part, earlier work on the complexity of writing systems (Changizi and Shimojo 2005a)).

With empirical estimates of the three parameters in hand, the hypothesis predicts approximately 10–20 levels in the human ventral stream, with a "best" estimate of approximately 15 levels. (The hypothesis also predicts exponential decay of area size with hierarchical level, and level–level convergence values of $\mu \approx 1.24$, as discussed in Appendix B.)

There are a number of reasons why this prediction cannot be more than very approximate. (1) The empirical settings of the three parameters – the total combinatorial degree, d_{tot}; the total convergence, μ_{tot}; and the number of module states, σ – were only very approximate estimates (especially the latter two). (2) Furthermore, I have assumed – primarily for the sake of simplicity – that the parameters (namely d, μ and σ, see Appendix A) are constant across the levels, something that is certainly an idealization. (3) The N_{reg} versus $n + 1$ plots in Fig. 2 show that the minimum is fairly broad, and increases much more slowly after $n_{\text{opt}} + 1$ than before it; this means that the number of levels could deviate somewhat from $n_{\text{opt}} + 1$ and still be near optimal. It is therefore expected that even if one could confidently make precise predictions about the optimal number of levels, there is "evolutionary wiggle room" for the actual number of levels to deviate somewhat from the optimal, and still be very near optimal. (4) I must reiterate that the model is an extreme idealization of the hierarchy, where for each level there is only one level above it. In reality, areas connect to multiple levels above it (something also found for the hierarchy for the English lexicon (Changizi MA, in review)). (5) Finally, the economical principle stated that the number of hierarchical levels is set to minimize the number of neurons, but in reality the system may, in fact, be selected to minimize the total amount of wire volume as well. And, of course, selection acts at the level of the whole animal, not just the brain (or the ventral stream). For all these reasons, the hypothesis I have presented here can only hope to explain the first order features of the visual hierarchy: the approximate number of levels, and the approximate manner in which the relative sizes of levels decrease with hierarchical level (see Appendix B).

Note that the kind of explanation given here is "epiphenomenal," in that the visual system in the model would work even if there were just two levels; it is only selection pressure to minimize the total number of neurons that leads to parcellating into multiple hierarchical levels. This explanation is similar, then, in kind to past conjectures for why there are so many visual areas (Kaas 1977, 1989, 1995, 1997b, 2000; Cowey 1979, 1981; Barlow 1986), an explanation for why the number of areas over the entire neocortex changes as it does as a function of brain size (Changizi 2001b, 2003a, 2005; Changizi and Shimojo 2005b), and why areas are positioned where they are within the neocortex (Klyachko and Stevens 2003; Cherniak et al. 2004).

The question of why there are as many visual areas as there are is inextricably connected to the question of why there should be areas specialized for intermediate-level complexity visual features. Some have suggested that information maximization explains this (Ullman et al. 2002), and my approach can be interpreted as a complementary suggestion: rather than maximizing information for a given volume of hardware, which may lead to areas specialized for intermediate-level complexity, I have considered minimizing the volume of hardware for a given load of information processing that must be implemented, again concluding that this leads to areas specialized for intermediate-level complexity. My approach is on the "neuroanatomy" side of the coin, and viewing the problem in this fashion allows me to predict the number of hierarchical levels (and also the level–level convergence as discussed in Appendix B). However, it is possible that the causality is reversed. Rather than the organization of the ventral stream being what it is because it has been selected

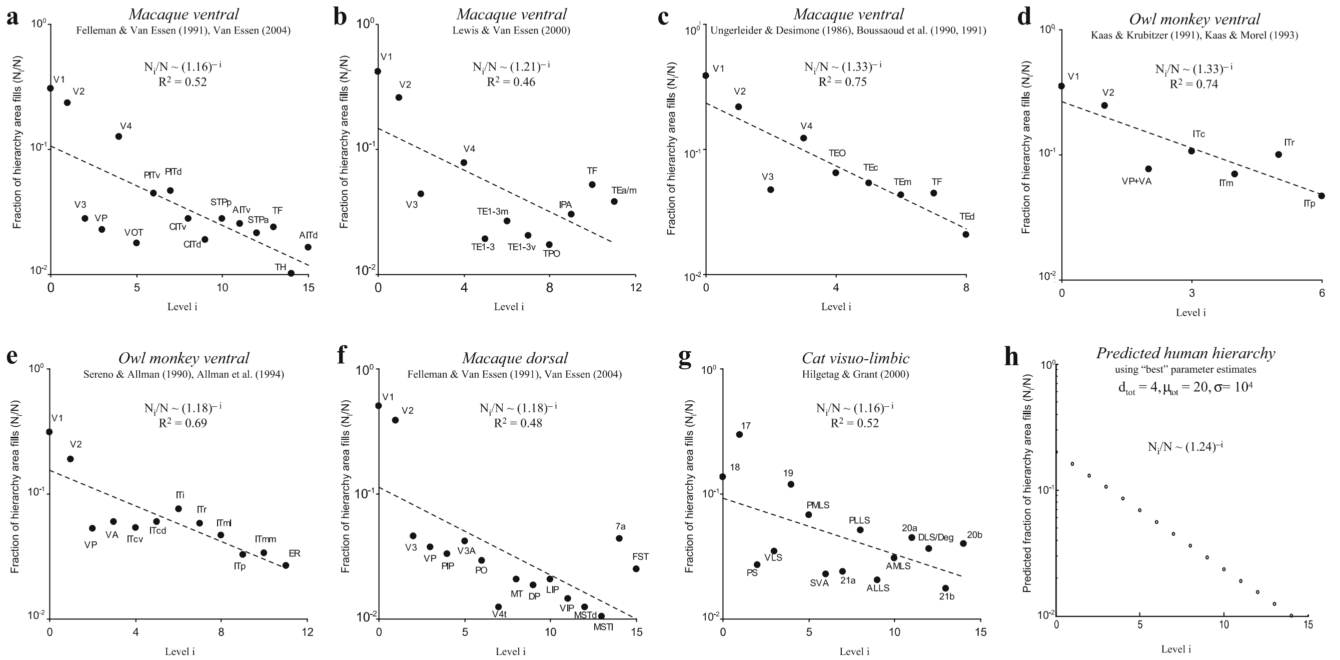

Fig. 3 **a–e** Logarithm (base 10) of area size (measured as the fraction of the surface area within the hierarchy) versus its approximate level in the hierarchy of the ventral stream. Three different parcellation schemes for macaque are shown in (**a**), (**b**) and (**c**), and two different schemes for owl monkey shown in (**d**) and (**e**). The parcellations and area sizes are taken from the citations shown, and also from Van Essen (2004, Fig. 2), for (**b**) and (**c**), and Kaas (1997a, Fig. 6), for (**d**) and (**e**). Some of the unusually low, outlier, areas – such as V3 (in(**a**)), VP, VOT and VA – are "improbable" areas, representing only one quadrant (Lyon and Kaas 2002; Van Essen et al. 1990; Van Essen 2004; Kaas 1997a), and are likely to be approximately twice the size shown when the other quadrant is adjoined. Hierarchical levels for (**a**) are taken from Hilgetag et al. (2000, Fig 5a), where each area is placed on a distinct level. For **b–e**, the hierarchical orders were set as follows: for each area not in the Felleman and Van Essen (1991) parcellation scheme, its relative level in the hierarchy is set to that of the approximately cortically nearest area in the Felleman and Van Essen (1991) scheme. (**f**) Logarithm of area size versus its approximate level in the hierarchy of the macaque *dorsal* stream. Only non-frontal areas are shown here. The sources for the area sizes are shown, and the hierarchical levels are taken from Hilgetag et al. (2000, Fig. 5a), where each area is placed on the distinct level. (**g**) Logarithm (base 10) of area size versus its approximate level in the hierarchy of the cat *visuo-limbic* system. The sources for the area sizes are shown, and the hierarchical levels are taken from Hilgetag et al. (2000, Fig. 11) (see also Scannel et al. 1995), where each area is placed on a distinct level. Because the hierarchial orderings from Hilgetag et al. (2000) are not unique best orderings, and because the method of determining level is not unambiguous, the level orderings shown in these plots are only expected to be *correlated* with whatever the "true" hierarchical order might be. The main morals of the figures are that (i) areas higher in the hierarchy trend to be smaller, (ii) the relationship is consistent with an exponential, i.e., area size falls exponentially as a function of level in the hierarchy, (iii) the number of hierarchical levels is around a dozen, and (iv) the scaling equation is approximately $N_i/N \sim \mu^{-i}$, where level–level convergence μ ranges form about 1.15 to 1.33, with an average among these plots of approximately 1.2. **h** The predicted fraction of the hierarchy filled by an area, as a function of the area's level within the hierarchy, for humans. The parameters are set to be "best" estimates from the main text: the number of module states $\sigma = 10^4$, the total convergence factor μ_{tot}=20, and the total combinatorial degree d_{tot}=4. It predicts that there should be very approximately 15 levels (i.e., 15 dots in the plot), and that the relative sizes of the levels should exponentially decrease as approximately $N_i/N \sim (1.24)^{-i}$

to accommodate visual objects of a certain degree of complexity, it could be that we recognize visual objects of that degree of complexity because the organization of the ventral stream is what it is, due to developmental constraints, and information is maximized for the given hardware.

Finally, it is natural to ask what the model predicts for non-human primates and other animals having smaller brains, and probably fewer areas (Orban et al. 2004; Rosa and Tweedale 2005; Sereno and Tootell 2005; Changizi and Shimojo 2005b; see Appendix Fig. 3 for summary information on non-human ventral stream hierarchies). Although the model presented here is potentially useful for answering this question, we are not currently in a position to know how the three parameters – the total combinatorial degree d_{tot}, the total convergence μ_{tot}, and the number of states per module σ – vary as a function of brain size, or even whether there are scaling laws that describe them. That is, for primates having a smaller ventral stream, we do not know which of these three parameters is modulated. It is useful, however, to note how the predicted (i.e., optimized) total number of neurons in the ventral stream varies as these parameters are varied. (1) Lowering the complexity of visual objects (i.e., lowering the combinatorial degree, d_{tot}) from $d_{\text{tot}} = 100$ to $d_{\text{tot}} = 1.1$ reduces the predicted (regularized) total number of neurons from about 10^9 to 10^6, as one can see in Fig. 2a by looking at how the height of the minimum of the curves (i.e., the optimal total number of neurons) falls as d_{tot} falls. (2) Lowering the total convergence, μ_{tot}, from 10^4 to 4 reduces the predicted total number of neurons from approximately 10^{10} to 10^7, as shown in Fig. 2b. Although such a modulation to the ventral stream would leave the complexity (i.e., combinatorial degree) of visual objects intact, the redundancy would decrease, and the ventral stream would become more error-prone. (3) Lowering the number of states per module, σ, from 10^8 to 10^2 reduces the predicted number of neurons from approximately 10^{13} to 10^5, as shown in Fig. 2c. Again this modulation would not affect the complexity of visual objects the ventral stream can accommodate, but the representational power of each module would decrease, and, intuitively, the "pixel depth" of a percept would deteriorate (akin to moving from 32 bit color displays to 16 bit). (4) In addition to lowering any of these parameters, there is another aspect that can vary, and thereby lower the total number of neurons in the ventral stream. Recall that the total number of neurons discussed in Fig. 2 is actually the *regularized* total number of neurons, which is the total number of neurons in the hierarchy below – or eventually converging to – a top level module. Two different primates having the same settings for the three parameters mentioned above can still differ in the number of modules in the top level. Doubling the number of modules in the top level doubles the number of modules at every level in the hierarchy, increasing the resolution at every level, and doubling the total number of neurons in the ventral stream. So, for example, macaque could have the same settings for the three parameters as human, but have, say, 1/10 the number of modules in the top IT level (corresponding, say, to a repertoire of top-level visual objects that is one-tenth of what we have), and thereby have 1/10 the total number of neurons in the ventral stream. . . . and yet still accommodate visual objects of the same complexity as do we, have the same redundancy as do we, and possess modules with the same representational power (or pixel depth) as do we. Addressing these scaling questions in regards to the ventral stream is the subject of continuing work.

Appendix a: generic information-processing hierarchy

Here I describe a generic information-processing hierarchy for vision, derive a general equation stating how the number of neurons in a level varies as a function of level, and derive an equation stating how the total number of neurons varies as a function of the complexity of visual objects it must recognize, and the total number of hierarchical levels.

A.1 Modules, levels and convergence

Each level of the hierarchy is partitioned into modules (e.g., columns, barrels, blobs), each of which consists of neurons having the same receptive field, and the union of all the receptive fields amounts to the entire visual field (or the entire retina). We assume for simplicity that the receptive fields do not overlap, but the main results will hold so long as the overlap percentage itself does not vary as a function of level in the hierarchy. Each module is capable of σ distinct states. Let P_{i} be the number of modules in level i, where $i = 0$ is the bottom level (i.e., V1), and $i = n$ is the top level (e.g., some IT area). There are therefore $n + 1$ levels in all. Levels above the bottom progressively have fewer modules, but where each module has greater receptive field size. See Fig. 1a in main text. On average, one module in level $i + 1$ receives inputs from μ modules from level i; μ is called the *level-level convergence*. It follows that level i has μ times more modules than level $i + 1$, i.e., $P_{i+1} = P_i/\mu$. Therefore, $P_i = P_n\mu^{n-i}$, and so $P_0 = P_n\mu^n$. The relationship between the level–level convergence factor, μ, and the number of hierarchical levels above the bottom, n, is, then,

$$\mu(n) = (P_0/P_n)^{1/n} = (\mu_{\text{tot}})^{1/n}, \tag{2}$$

where $\mu_{\text{tot}} = P_0/P_n$ is the total convergence over the entire hierarchy. This equation indicates that the level–level convergence, μ, depends on n, the number of hierarchical levels above the bottom; namely, it falls and approaches 1 as the number of hierarchical levels increases.

A.2 Instructions for the activation of the next level

The sequential activation pattern of a module in level $i + 1$ depends upon the activations of the μ modules in level i that converge to it. These μ modules that converge to the same i+1-level module are called a *convergence zone*. At any given time, the zone is potentially capable of σ^μ states, where recall that σ is the number of potential module states; one may think of a zone state at any given time as a "sentence" of length μ, where each of the μ spots in the sentence can be filled by one of σ many different module-words. More generally, however, not all these σ^{μ} states of the zone may need to be treated differently by the animal; there may, for example, be redundancies among the μ modules due to statistical regularities in the ecology. Although there are potentially μ degrees of freedom in each possible zone state, there may in fact only be $d = \beta\mu$ many degrees of freedom relevant for the i+1-level module, where β is a fraction and is called the *level-level redundancy constant*. d is called the *level–level combinatorial degree* (Changizi 2001c, 2003b; Changizi et al. 2002), and is the entropy in base-σ; $d \geq 1$, and if $d = 1$ then the module states in a zone do not interact combinatorially, whereas greater values of d (until it reaches a maximum of μ) imply that the modules in a zone act together more combinatorially in the construction of zone states. The total number of states per convergence zone is, then, $D_{\mathrm{z}} = \sigma^{\beta\mu(n)} = \sigma^{d(n)}$, where I have explicitly noted the dependency of μ and d on the total number of levels above the bottom, n.

Because the i+1-level module must, by assumption, activate differently for each of these $D_{\mathrm{z}} = \sigma^{d}$ many states from the zone that converges to it, and because the module is capable of only σ distinct instantaneous states, it follows that the i+1-level module must encode these D_{z} many different representations via its *pattern of sequential activation*. In particular, I presume for simplicity that it carries this out without redundancy, which means that the i+1-level module is capable of σ^{d} distinct sequential activation patterns, each of length d, in response to the σ^{d} many distinct convergence zone states. How does the i+1-level module know which activation pattern to carry out? There must exist neural tissue encoding the *instructions* that tell the module how to activate in each of the D_{z} many cases. These "neural instructions" for the i+1-level module are placed within the zone in level i. See Fig. 1b. [It would make no substantive difference to the model if we assumed that the neural instructions for the i+1-level module are placed in the level i+1.]

A.3 Number of neurons per level

How many neurons are required in a zone to encode both the visual representations (i.e., the modules) and the instructions? The visual representations are carried out by the P_i many modules in a level, where each module is capable of σ states. Let N_{pix} be the number of neurons needed to accommodate the σ states of a module, which we assume to be logarithmic in σ; i.e., $N_{\mathrm{pix}} \approx \log(\sigma)$. The number of module neurons in a zone is, then, just $\mu \log(\sigma)$. For the instructions, recall there are $D_{\mathrm{z}} = \sigma^{d}$ many instructions, each which must tell the i+1-level module how to activate in a sequence of length d, and where each activation of the module is capable of σ states. The number of neurons needed to specify one of σ many states is logarithmic in σ, and so the number of neurons required for the instructions in a zone is $N_{\mathrm{z,instr}} \approx \sigma^{d} \times d \times \log(\sigma)$. The total number of neurons in a zone is therefore $N_{\mathrm{z}} = \mu N_{\mathrm{pix}} + N_{\mathrm{z,instr}} = \mu \log(\sigma) + d\sigma^{d} \log(\sigma)$. Because there are P_i/μ many zones per level, the total number of required neurons in level i is $N_i = (P_i/\mu)[\mu \log(\sigma) + d\sigma^{d} \log(\sigma)]$. Recalling that $P_i = P_n\mu^{n-i}$ and $\beta = d/\mu$, we may manipulate this into $N_i = P_n\mu^{n-i} \log(\sigma)[1 + \beta\sigma^{d}]$. We may therefore write

$$N_i \approx \left[P_n \log(\sigma)\right] \times \left[1 + \beta\sigma^{\beta\mu(n)}\right] \times \left[\mu(n)^{n-i}\right] \qquad (3)$$

where the explicit dependency of convergence, μ, on the number of levels above the bottom, n, is shown here. One consequence of this equation is that $N_i \sim \mu^{-i}$.

A.4 Total number of neurons

The total number of neurons in the visual hierarchy is the sum of these N_i. Only the last term in Eq. (3) depends on the level, i, and using the geometrical progression identity, we can derive that

$$N = \sum N_i(n) \approx \left[P_n \log(\sigma)\right] \times \left[1 + \beta\sigma^{\beta\mu(n)}\right] \times \frac{\left[\mu(n)^{n+1} - 1\right]}{\left[\mu(n) - 1\right]} \qquad (4)$$

Recall from Eq. (2) that $\mu(n) = (\mu_{\mathrm{tot}})^{1/n}$, where $\mu_{\mathrm{tot}} = P_0/P_n$ is the total convergence over the entire hierarchy. Also let $\beta_{\mathrm{tot}} = \beta^{n}$, which is the *total redundancy constant* over the entire hierarchy, and $d_{\mathrm{tot}} = \beta_{\mathrm{tot}}\mu_{\mathrm{tot}}$, which is the *total combinatorial degree* over the hierarchy. Then we may write

$$N \approx \left[P_n \log(\sigma)\right] \times \left[1 + (d_{\mathrm{tot}}/\mu_{\mathrm{tot}})^{1/n}\sigma^{d_{\mathrm{tot}}^{1/n}}\right] \times \frac{\left[\mu_{\mathrm{tot}}^{(n+1)/n} - 1\right]}{\left[\mu_{\mathrm{tot}}^{1/n} - 1\right]}. \qquad (5)$$

The total combinatorial degree of the system, d_{tot}, measures how many degrees of freedom there are in the construction of top-level representations; it is the base-σ entropy of a high-level representation. Intuitively, it is a measure of how complex a high-level representation is, such as of a visual object. If there were no redundancy – i.e., $\beta_{\mathrm{tot}} = 1$ – then $d_{\mathrm{tot}} = \mu_{\mathrm{tot}}$, and the number of degrees of freedom would just be the total number of bottom-level modules per top-level module to which they ultimately converge.

It will be useful to define the *regularized number of neurons*, $N_{\mathrm{reg}} = N/P_n$, which is the total number of neurons in the entire hierarchy below a single module in level n (the top level); i.e., it is the sum of all the neurons in the hierarchy that are involved in the information that eventually converges to one top-level module.

$$N_{\mathrm{reg}}(n, \sigma, \mu_{\mathrm{tot}}, d_{\mathrm{tot}}) \approx \log(\sigma) \times \left[1 + (d_{\mathrm{tot}}/\mu_{\mathrm{tot}})^{1/n}\sigma^{d_{\mathrm{tot}}^{1/n}}\right] \times \frac{\left[\mu_{\mathrm{tot}}^{(n+1)/\mathrm{n}} - 1\right]}{\left[\mu_{\mathrm{tot}}^{1/n} - 1\right]} \qquad (6)$$

Appendix B: summary of hierarchical levels and sizes for non-humans

The estimates of the visual complexity of objects relied upon written words, and for this reason the predictions in the main text do not directly apply to non-human primates and other mammals. Nevertheless, the model is expected to apply just as well (or just as poorly) to the ventral stream of other animals, supposing we can find estimates for the complexity of visual objects (d_{tot}), the total convergence (μ_{tot}), and the number of states per module (σ). We note in this appendix that, although we must recognize that hierarchical orderings are far from unambiguous, and that there is a tremendous amount of arbitrariness in the determination of borders for higher areas (and thus the relative surface areas are unreliable for higher areas), current estimates from different studies lead to the same conclusion that hierarchically higher areas tend to be smaller than lower areas (Fig. 3). This is true for three studies of the parcellation for macaque ventral stream (Fig. 3a, b, c), and two studies of the parcellation for owl monkey ventral stream (Fig. 3d, e). This also appears to hold for the macaque dorsal stream (Fig. 3f), and even for the cat visuo-limbic system (Fig. 3g). These plots are very likely to change considerably as greater knowledge of the parcellation maps and connectivity are obtained, but because these studies all conform to the higher-areas-are-smaller rule, it seems reasonable to tentatively suppose that this will remain to be true. In the future, the hope is that the generic information-processing, hierarchical model can be brought to bear on these other hierarchies, predicting not just that higher areas should be smaller, but predicting the rate of exponential decrease of the areas as a function of level.

The predicted such plot for human – from the results of this paper – is shown in Fig. 3h, and emanates from Eq. (3). To understand why this is the predicted plot for human, consider first that the predicted *level–level convergence* – i.e., the convergence from one level to the next, or the number of modules in a convergence zone – is given by $\mu_{\text{opt}} = (\mu_{\text{tot}})^{1/n_{\text{opt}}}$ (see Eq. (2)). Given the "best" prediction for n_{opt} in the main text (namely $\text{n}_{\text{opt}} = 14$), and recalling that the "best" estimate of the total convergence was $\mu_{\text{tot}} \approx 20$, it follows that the "best" prediction for the level-level convergence is $\mu_{\text{opt}} = (20)^{1/14} = 1.24$. Perturbations of any one of the parameters within its "reasonable" range (as discussed in Sect. 7 of the main text) lead to $\mu_{\text{opt}} \in [1.13, 1.65]$ for $d_{\text{tot}} \in [2, 10]$, $\mu_{\text{opt}} \in [1.10, 1.43]$ for $\mu_{\text{tot}} \in [4, 10^2]$, and $\mu_{\text{opt}} \in [1.16, 1.45]$ for $\sigma \in [10^2, 10^6]$. Therefore, the predicted convergence, μ_{opt}, may range from about 1.1 to 1.65, with a "best" prediction of 1.24. From Eq. (3) in Appendix A, $N_{i,\text{opt}} \sim \mu_{\text{opt}}^{-i}$, and thus I expect $N_{\text{i,opt}}/N_{\text{opt}} \sim (1.24)^{-i}$, and this prediction for human is shown in Fig. 3. In other words, I expect to find that the sizes of areas for the human ventral stream should fall as approximately $N_i/N \sim \mu^{-i}$, where μ is approximately 1.24, ranging as low as perhaps 1.1 and as high as 1.6.

Acknowledgements I wish to thank Shinsuke Shimojo and Dan Ryder, and one very helpful referee, for comments on the paper. Support for this research was given by NIH grant 5F32EY015370-02.

References

Allman J, Jeo R, Sereno M (1994) The functional organization of visual cortex in owl monkeys. In: Baer JF, Weller RE, Kakoma I (eds) Aotus: the owl monkey. Academic, Orlando, pp 287–320

Attneave F (1954) Some informational aspects of visual perception. Psychol Rev 61:183–193

Barlow HB (1961) Possible principles underlying the transformation of sensory messages. In: Rosenblith WA (ed) Sensory communication. MIT Press, Cambridge, pp 217–34

Barlow HB (1986) Why have multiple cortical areas? Vis Res 26:81–90

Boussaoud D, Ungerleider LC, Desimone R (1990) Pathways for motion analysis: cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque. J Comp Neurol 296:462–495

Boussaoud D, Desimone R, Ungerleider LG (1991) Visual topography of area TEO in the macaque. J Comp Neurol 306:554–575

Cajal SR (1995) Histology of the nervous system, vol. 1. Oxford University Press, New York

Changizi MA (2001a) The economy of the shape of limbed animals. Biol Cybern 84:23–29

Changizi MA (2001b) Principles underlying mammalian neocortical scaling. Biol Cybern 84:207–215

Changizi MA (2001c) Universal scaling laws for hierarchical complexity in languages, organisms, behaviors and other combinatorial systems. J Theor Biol 211: 277–295

Changizi MA (2003a) The brain from 25,000 feet: high level explorations of brain complexity, perception, induction and vagueness, Kluwer, Dordrecht

Changizi MA (2003b) The relationship between number of muscles, behavioral repertoire, and encephalization in mammals. J Theor Biol 220:157–168

Changizi MA (2005) Scaling the brain and its connections. In: Kaas JH (ed) Evolution of nervous systems, Elsevier, Amsterdam

Changizi MA, McDannald MA, Widders D (2002) Scaling of differentiation in networks: nervous systems, organisms, ant colonies, ecosystems, businesses, universities, cities, electronic circuits, and legos. J Theor Biol 218:215–237

Changizi M, Shimojo S (2005a) Character complexity and redundancy in writing systems over human history. Proc R Soc Lond B 272:267–275

Changizi MA, Shimojo S (2005b) Parcellation and area–area connectivity as a function of neocortex size. Brain Behav Evol 66:88–98

Changizi MA, Zhang Q, Ye H, Shimojo S (2006) The structures of letters and symbols throughout human history are selected to match those found in objects in natural scenes. Am Nat (in press)

Cherniak C (1992) Local optimization of neuron arbors. Biol Cybern 66:503–510

Cherniak C (1994) Component placement optimization in the brain. J Neurosci 14:2418–2427

Cherniak C (1995) Neural component placement. Trends Neurosci 18:522–527

Cherniak C, Changizi MA, Kang D (1999) Large-scale optimization of neuron arbors. Phys Rev E 59:6001–6009

Cherniak C, Mokhtarzada Z, R-Esteban R, Changizi B (2004) Global optimization of cerebral cortex layout. Proc Nat Acad Sci 101:1081–1086

Chklovskii DB, Koulakov AA (2000) A wire length minimization approach to ocular dominance patterns in mammalian visual cortex. Physica A 284:318–334

Coogan TA, Burkhalter A (1993) Hierarchical organization of areas in rat visual cortex. J Neurosci 13:3749–3772

Cowey A (1979) Cortical maps and visual perception. The Grindley Memorial Lecture. Q J Exp Psychol 31:1–17

Cowey A (1981) Why are there so many visual areas? In: Schmitt FO, Warden FG, Adelman G, Dennis SG (eds) The organization of the cerebral cortex. MIT Press, Cambridge, pp 395–413

Durbin R, Mitchison G (1990) A dimension reduction framework for understanding cortical maps. Nature 343:644–647

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1–47

Hasson U, Levy I, Behrmann M, Hendler T, Malach R (2002) Eccentricity bias as an organizing principle for human high-order object areas. Neuron 34:479–490

Hasson U, Harel M, Levy I, Malach R (2003) Large-scale mirror-symmetry organization of human occipito-temporal object areas. Neuron 37:1027–1041

Hilgetag CC, Grant S (2000) Uniformity, specificity and variability of corticocortical connectivity. Phil Trans R Soc Lond B 355:7–20

Hilgetag CC, O'Neill MA, Young MP (2000) Hierarchical organization of macaque and cat: cortical sensory systems explored with a novel network processor. Phil Trans R Soc Lond B 355:71–89

Jacobs RA, Jordan MI (1992) Computational consequences of a bias toward short connections. J Cogn Neurosci 4:323–336

Kaas JH (1977) Sensory representations in mammals. In: Stent GS (ed) Function and formation of neural systems. Dahlem Konferenzen, Berlin, pp 65–80

Kaas JH (1989) Why does the brain have so many visual areas? J Cogn Neurosci 1:121–135

Kaas JH (1995) The evolution of isocortex. Brain Behav Evol 46:187–196

Kaas JH (1997a) Theories of visual cortex organization in primates. In: Rockland KS, Kaas JH, Peters A (eds) Cerebral cortex. vol 12. extrastriate cortex in primates. Plenum, New York, pp 91–125

Kaas JH (1997b) Topographic maps are fundamental to sensory processing. Brain Res Bull 44:107–112

Kaas JH (2000) Why is brain size so important: design problems and solutions as neocortex gets bigger or smaller. Brain Mind 1:7–23

Kaas JH, Krubitzer LA (1991) The organization of extrastriate visual cortex. In: Dreher B, Robinson SR (eds) Neuroanatomy of the visual pathways and their development. MacMillan, London, pp 302–323

Kaas JH, Morel A (1993) Connections of visual areas of the upper temporal lobe of owl monkeys: the MT crescent and dorsal and ventral subdivisions of FST. J Neurosci 13:534–546

Klyachko VA, Stevens CF (2003) Connectivity optimization and the positioning of cortical areas. Proc Nat Acad Sci 100:7937–7941

Lewis JW, Van Essen DC (2000) Architectonic parcellation of parieto-occipital cortex and interconnected cortical regions in the macaque monkey. J Comp Neurol 428: 79–111

Lyon DC, Kaas JH (2002) Evidence for a modified V3 with dorsal and ventral halves in macaque monkeys. Neuron 33:453–461

Malach R, Levy I, Hasson U (2002) The topography of high-order human object areas. Trends Cogn Sci 6:176–184

Mead C (1989) Analog VLSI and neural systems. Addison-Wesley, Boston

Mitchison G (1991) Neuronal branching patterns and the economy of cortical wiring. Proc R Soc Lond B 245:151–158

Mitchison G (1992) Axonal trees and cortical architecture. Trends Neurosci 15:122–126

Orban G, Van Essen D, Vanduffel W (2004) Comparative mapping of higher visual areas in monkeys and humans. Trends Cogn Sci 8:315–324

Ringo JL (1991) Neuronal interconnection as a function of brain size. Brain Behav Evol 38:1–6

Rockland KS, Pandya DN (1979) Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. Brain Res 179:3–20

Rolls ET, Aggelopoulos NC, Zheng F (2003) The receptive fields of inferior temporal cortex neurons in natural scenes. J Neurosci 23:339–348

Rosa MGP (1997) Visuotopic organization of primate extrastriate cortex. In: Rockland KS, Kaas JH, Peters A (eds) Cerebral cortex. vol 12. extrastriate cortex in primates. Plenum, New York, pp 127–203

Rosa MGP, Tweedale (2005) Brain maps, great and small: lessons from comparative studies of primate visual cortical organization. Phil Trans R Soc Lond B 360:665–691

Ruppin E, Schwartz EL, Yeshurun Y (1993) Examining the volume efficiency of the cortical architecture in a multi-processor network model. Biol Cybern 70:89–94

Scannell JW, Blakemore C, Young MP (1995) Analysis of connectivity in the cat cerebral cortex. J Neurosci 15:1463–1483

Sereno, Allman JM (1990) Cortical visual areas in mammals. In: Leventhal AG (ed) The neural basis of visual function, vol. 4. Macmillan, London, pp 160–172

Sereno MI, Tootell RBH (2005) From monkeys to humans: what do we now know about brain homologies? Curr Opin Neurobiol 15:135–144

Shannon CE (1951) Prediction and entropy of printed English. Bell Syst Tech J 30:50–64

Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. Annu Rev Neurosci 24:1193–1216

Traverso S, Morchio R, Tamone G (1992) Neuronal growth and the Steiner problem. Riv Biol 85:405–418

Ullman S, Vidal-Naquet M, Sali E (2002) Visual features of intermediate complexity and their use in classification. Nat Neurosci 5:682–687

Ungerleider LG, Desimone R (1986) Cortical projections of visual area MT in the macaque. J Comp Neurol 248:190–222

Van Essen DC (1997) A tension-based theory of morphogenesis and compact wiring in the central nervous system. Nature 385:313–319

Van Essen DC (2004) Organization of visual areas in macaque and human cerebral cortex. In: Chalupa LM, Werner JS (eds) The visual neurosciences. MIT Press, Cambridge, pp 507–521

Zeki SM (2003) Improbable areas in the visual brain. Trends Neurosci 26:23–26

Van Essen DC, Maunsell JHR (1983) Hierarchical organization and the functional streams in the visual cortex. Trends Neurosci 6:370–375

Van Essen DC, Newsome WT, Maunsell JHR (1984) The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. Vis Res 24:429–448

Van Essen DC, Felleman DF, DeYoe EA, Olavarria JF, Knierim JJ (1990) Modular and hierarchical organization of extrastriate visual cortex in the macaque. Cold Spring Harb Symp Quant Biol 55: 679–696