

Human Synaptic Development is Near-Optimal

Robert Sneddon
Division of Biology
California Institute of Technology
Email: rsneddon@caltech.edu
Phone: (626) 395-6818

William Rodman Shankle
Depts. of Cognitive Sciences and Pharmacology
University of California, Irvine
Email: rshankle@uci.edu

Junko Hara
Division of Information and Computer Science
University of California, Irvine
Email: jhara@ics.uci.edu

James H. Fallon
Dept. of Anatomy and Neurobiology
University of California, Irvine
Email: jfallon@uci.edu

Abstract

Information processing in the brain is a complex and sophisticated process that would work poorly if synaptic connections were randomly organized. It would be interesting if a developmental organizing principle exists which describes synaptic development. We postulate that the optimal development of synaptic connectivity is this organizing principle and is a likely product of evolution.

From a Kolmogorov structure measure of synaptic development and an *a priori* assumption of maximizing connective structure (optimal synaptic development quantified), we derive an equation for the growth and pruning of synapses. We compare the values predicted from this equation to those of 18 sets of human cortical data. The predicted values account for 98% of the variance of the observed values. This finding indicates that optimal synaptic development may be a driving principle for human synaptic development.

Introduction

Discovering the rules that govern the development of connections within the cerebral cortex is an important step towards understanding how brains process information. The complexity of at least 50 cytoarchitecturally different human cortical areas developing both separately and jointly in a highly reproducible fashion is a remarkable bioengineering feat suggesting strong underlying rules. The existence of evolutionary rules involved in establishing brain connectivity is suggested by data on the relatively constant lateral extent of basal dendrites and on the homeobox-like development of new cortical areas (Stevens, 1993; Krubitzer, 1995).

Since this development of connections creates a mechanism which processes information in a highly sophisticated manner, it is extremely unlikely that neural connectivity is random. Purely random connectivity has the highest level of disorganization as measured by the Kolmogorov complexity measure (Cover & Thomas, 1991). Instead, we expect that synaptic connectivity will be highly complex, but nonrandom. That is, highly structured. In terms of brain development, this means that synaptic connectivity will develop in an optimal fashion (defined further on). The assumptions behind this hypothesis are listed below.

An Optimal Development Hypothesis

Assumptions:

1. a. A given biological structure, e.g., the brain, has a particular function(s).
b. This function is reflected in the structure's morphology.
2. a. Certain portions of this morphology are more important to function than others.
b. The basic function of the brain is information processing/computation. Thus synaptic connectivity is the most important aspect of brain morphology.
3. a. Evolution pushes the brain towards optimal function. However, evolution is not perfect and every biological organ has a metabolic cost.
b. Thus, aspects of brain structure which are more important to brain function (such as synaptic connectivity) will be pushed harder (by evolution) towards optimality than other portions whose specific structure are less important to computation and information processing, e.g., glial cells.
c. This means that components of the brain which are most relevant to brain function will develop to a (near-) optimal end-point.
d. However, metabolic costs limit this optimality. For example, mammalian cortex appears to have a fixed number of synapses which it can support. [1] demonstrated that this value appears to be fixed at about 10^9 neurons per mm^3 . Within this limit, this synaptic connectivity should be organized in a manner which optimizes information processing and computation .

4. a. There are a potentially infinite number of paths to the final (optimal) state of synaptic connectivity. However, this final state is not known beforehand, as the developing brain does not “know” the future. Therefore, synaptic development always moves in the most optimal direction. This insures (given a reasonable amount of time) that the end-point of development will be near-optimal¹.

b. Since this near-optimal path of development is a product of evolution, synaptic development should be mostly genetic. However, genetics must allow the environment to influence the brain’s behavior so that behavior is optimal for the specific environment. It appears to be that for a variety of levels in the hierarchy from synapse to behavior, the amount of variability induced by the environment is about 25%. (Moore, et. al, 1999, 2000; Haug 1987; Johnson 1996)

From these considerations, we make the following qualitative hypothesis.

Qualitative Hypothesis:

Human synaptic development will proceed in a near-optimal manner.

In order to address this hypothesis directly, a meaningful definition of optimal development must be given.

Defining Optimal Synaptic Development

At an intuitive level, an optimal biological structure is a structure which performs its function with the least amount of noise/chaos and the most amount of efficiency. In the case of the brain, (near-) optimal information processing/computation would be reflected in a high level of synaptic connection complexity. Therefore, the synaptic structure of the brain should be highly complex without being chaotic.

An everyday analogy to this concept is the television. This mechanism is highly complex and requires a very long description to fully describe it. On the other hand, random behavior, such as the results of many dice rolls will also take a great deal of description to accurately describe them. This is because there is no descriptive redundancy in the results of these dice rolls. And that is due to the fact that the outcomes of the dice rolls can’t be predicted. Thus, to make our qualitative hypothesis quantitative, we need a quantitative measure which demonstrates a high level of organization in a mechanism that is complex and non-random, without assigning a high level of organization to a complex, but randomly organized mechanism. To answer this question, we will first define complexity in a formal manner.

What is complexity? One of the best approaches to defining complexity is the information-theoretic approach. Specifically, some physical structures are more difficult to describe than others. This difficulty is reflected in the length of

¹This is similar to the “hill-climbing” maximization algorithm. This algorithm does not know where the maximum of a function is. However, it blindly seeks it by maximizing the gradient.

the description. Or more specifically, the length of a computer program needed to create this description.

We adopt the formal version of this way of defining complexity. This formal method is the Kolmogorov complexity [3]. The Kolmogorov complexity of a piece of information is defined as the length of the shortest (universal) computer program (partially recursive function) necessary to produce this information. In principle, this information can be anything, including a complete description of the brain's synaptic connectivity.

Definition: Complexity of an Object. The complexity of an object is determined by the length of the shortest (universal) computer program which will print a description of the object.

Simply put, the longer the program, the more complex the object. The outcomes of purely random behavior are the most complex, because they can not be predicted. (Note that this description can be limited to a certain aspect of the object, e.g., the synaptic connectivity of the human brain.)

Given this measure of complexity, we need to relate it to a measure which demonstrates how much of this complexity is random. Fortunately, Kolmogorov defined such a measure. This measure is called the Kolmogorov structure function [3]. The Kolmogorov structure function furthers Kolmogorov's description of complexity. It takes the measurement given by the Kolmogorov complexity function and divides it into two parts. These parts are essentially the random and non-random components of the complexity measure.

Definition: Structure of an Object. The structure of an object is determined by the portion of its complexity which is non-random.

If an object is high in complexity and mostly non-random, it has a high level of structure. On the other hand, if it is high in complexity, but mostly random, it will be low in structure. For example, the outcomes of roulette wheel spins can not be predicted. Thus, a computer program which prints out these outcomes can have almost no redundancy. Thus, this program will be about the same length as the actual list of roulette outcomes, i.e., there is almost no compression of information by the program. Therefore, a program which describes the outcomes of a roulette wheel will have a high level of complexity. However, these outcomes are random, so it will have a low level of structure. (It should be noted that this is a cursory description of Kolmogorov's sophisticated complexity theory. A more detailed explanation and discussion of these structure/information measures is beyond the scope of this paper. The interested reader is referred to Cover & Thomas, 1991, for an introduction to the theory of Kolmogorov complexity.)

This knowledge of object complexity and structure allows us to define optimal synaptic development.

Definition: Optimal Synaptic Development. Optimal synaptic development is that development which maximizes the structure of synaptic connectivity.

This development is subject to metabolic constraints such as an initial rate of synaptic development. We are now in a position to state a quantitative version of our hypothesis.

Quantitative Hypothesis:

Human synaptic development maximizes the increase in Kolmogorov structure at each point in time.

This structure is measured by the Kolmogorov structure function and it is bounded by the initial conditions of development². In lay terms, this hypothesis is stated as “life fights chaos” and has an intuitive appeal. This appeal is based on Darwin’s notion of survival of the fittest which suggests that the more organized brain connectivity is, the better the brain processes information. In this study, we take this intuitive statement, define it in a rigorous, quantitative, manner and apply it to actual human neurodevelopmental data. We do this by first creating a zeroth order random variable model of synaptic development, i.e., a model with maximum complexity, but little structure. This zeroth order model allows for a very large number of possible developmental paths from birth to 72 months (the endpoint of our data). We take this set of developmental paths and constrain it to the path that maximizes the increase in synaptic structure at each point in time. This is our first order model which quantifies the statement, “life fights chaos.”

Thus, the thrust of the paper is to model the development over time of synapses on an individual neuron according to the principle that the system has evolved so as to maximize, at every time period, the increase in synaptic connection complexity, subject to the condition of minimizing the amount of randomness in this increase. This model is worked out in detail leading to a damped cosine expression for synaptic growth. We examine two versions of this model, one with 3 free parameters and one with 5 free parameters. The model is tested against data on human synaptic growth between birth and 72 months. This quantity is estimated from existing data according to the formula: synapses per neuron = synaptic density divided by the neuron packing density. The model is compared with several *ad hoc* ones functions, as well as three standard growth models.

By testing this axiomatically derived hypothesis for brain development of synaptic connections, we avoid the problem of applying ad hoc tests, e.g., finding the best curve to fit the data. Instead, we use a predictive approach. As brains develop, there are in principle, many possible types of synaptic connectivity. We hypothesize that the brain’s actual wiring maximizes the structure of synaptic connectivity and derive a quantitative prediction of synaptic development for each layer of three different cortical areas (2 primary, 1 association) of the postnatal human brain from birth to 72 months.

A Model of Synaptic Development Based on Maximizing Connective Structure

²It is important to note that we are not saying that the developing brain computes the Kolmogorov structure at each point in time. This would be like hypothesizing “force equals mass times acceleration” and then assuming that a flying ball solves the necessary differential equations to follow its trajectory. Instead, we assume that evolution has genetically programmed the brain to develop – in conjunction with environmental influences, in this optimal manner.

To apply the hypothesis that brain development maximizes connective structure, we must apply the Kolmogorov measure of structure to synaptic development. However, we must first make some pragmatic considerations. Most important is the fact that synaptic connectivity is not the only pertinent fact in understanding human information processing and computations. Other important considerations are “wiring length,” type of neuron, strength of synapse, type of synapse etc.

At this point, we reassert our hypothesis that synaptic connectivity is the most important aspect of brain computation. To control for these other important aspects, we average over a large number of neurons with the assumption that these other aspects of brain computation will be averaged out.

With this assumption, we examine the average change in Kolmogorov structure for a group of n neurons. These changes are denoted by the difference in Kolmogorov structure, ΔK_k , where $\overline{\Delta K_k}$ is the average change from a time, t_m to a time t_{m+1} . This change is:

$$\begin{aligned} \overline{\Delta K_k}(s_{m+1}^n|n) &= \frac{K_k(s_{m+1}^n|n) - K_k(s_m^n|n)}{n} \\ &= \frac{K(s_{m+1}^n|n) - k_{m+1} - K(s_m^n|n) + k_m}{n} \end{aligned} \quad (1)$$

Here, s_{m+1}^n represents the changes in synaptic connectivity for n neurons from t_m to t_{m+1} , K is the Kolmogorov complexity function. This function can be divided into two parts. They are, K_k , the Kolmogorov structure function and k_m , the least sufficient structure statistic at time t_m . The term “structure function” is something of a misnomer. K_k is essentially a measurement of the amount of randomness in a physical system or piece of information. Thus, the smaller the value of K_k , the greater the Kolmogorov structure. See the Appendix, Part I for a detailed explanation of this formulation.

This change in the average value of the Kolmogorov structure is (approximately) equal to the change in entropy from time t_m to time t_{m+1} :

$$\overline{\Delta K_k}(s_{m+1}^n|n) \simeq H(\mathbf{S}_{m+1}) - H(\mathbf{S}_m) \quad (2)$$

See the Appendix, Part I for a detailed explanation and proof of Eq.(2). Here, \mathbf{S} is a nonhomogeneous random variable, i.e., its distribution is dependent on time. For example, \mathbf{S}_m represents the random variable at time t_m . with outcomes s . These outcomes represent possible changes in synaptic connectivity. Eq.(2). can be rewritten in terms of the rate of Kolmogorov structure change:

$$\frac{\overline{\Delta K_k}(s_{m+1}^n|n)}{t_{m+1} - t_m} \simeq \frac{H(\mathbf{S}_{m+1}) - H(\mathbf{S}_m)}{t_{m+1} - t_m} \simeq \frac{dH(\mathbf{S})}{dt} \quad (3)$$

The larger the value of K_k , the less Kolmogorov structure there is in the synaptic connectivity. Thus, maximizing the increase in the structure of synaptic connectivity is equivalent to minimizing $\overline{\Delta K_k}$ and hence, minimizing the rate of entropy, $\frac{dH(\mathbf{S})}{dt}$ (Eq. 3).

Thus, to maximize the structure of synaptic connectivity, we must first model synaptic development in terms of random variables. To do so, requires the formulation of a probability distribution for synaptic development. This distribution must be nonhomogeneous, because it is unrealistic to assume that the probabilities of synaptic connection events are constant over time. We must also account for a basic metabolic constraint: the average maximum number of supportable synapses for a cortical neuron. On the other hand, there are important synaptic possibilities, e.g., type of synapse, strength of synapse etc. which we will ignore. Again, this is due to the fact that we are addressing average behavior over all neurons and synapses.

The Measurement of Possible Synaptic Configurations is Equivalent to an Entropy Measurement

The data set we are dealing with has indistinguishable neurons and synapses. Thus, consider this data set with S_T total number of synapses and N_T total number of neurons. The distribution of synapses per neuron needs to be described in the best manner possible.

First, define, S^i to be the number of synapses for the i th neuron. Then, $\sum_{i=1}^{N_T} S^i = S_T$. Furthermore, let S_M be the maximum number of synapses that can be supported by this neuron. Let C_i denote the number of different configurations that a neuron N_i can have. In other words this neuron has S^i synapses to be distributed between a total of S_M possible synaptic connection sites. Then the number of possible different configurations of S^i for this neuron is the number of binomial combinations:

$$C_i = \binom{S_M}{S^i} = \frac{S_M!}{S^i! (S_M - S^i)!} \quad (4)$$

Applying Stirling's approximation for large numbers, $S! \approx \sqrt{2\pi S} S^S e^{-S}$ yields:

$$C_i \approx \left(\sqrt{\frac{S_M}{2\pi S^i (S_M - S^i)}} \right) \frac{S_M^{S_M} e^{-S_M}}{S^i e^{-S^i} (S_M - S^i)^{(S_M - S^i)} e^{-(S_M - S^i)}} \quad (5)$$

Since we are dealing with large numbers, we'll take the logarithm of C_i :

$$\begin{aligned} \ln [C_i] &\approx \ln \left[\sqrt{\frac{S_M}{2\pi S^i (S_M - S^i)}} \right] + \ln \left[\frac{S_M^{S_M} e^{-S_M}}{S^i e^{-S^i} (S_M - S^i)^{(S_M - S^i)} e^{-(S_M - S^i)}} \right] \quad (6) \\ &= \frac{1}{2} \ln \left[\frac{S_M}{2\pi S^i (S_M - S^i)} \right] - (S_M - S^i) \ln \left[\frac{(S_M - S^i)}{S_M} \right] - S^i \ln \left[\frac{S^i}{S_M} \right] \\ &= -\frac{1}{2} \ln [2\pi] - \frac{1}{2} \ln [S^i] - \left(S_M - S^i + \frac{1}{2} \right) \ln \left[\frac{(S_M - S^i)}{S_M} \right] - S^i \ln \left[\frac{S^i}{S_M} \right] \end{aligned}$$

Now, the values of S_M and S^i are on the order of 10^4 and 10^3 respectively. Furthermore, $-\frac{1}{2} \ln [2\pi] \approx -0.92$. Thus, $S_M - S^i \gg \frac{1}{2}$, $-S^i \ln \left[\frac{S^i}{S_M} \right] \gg -\frac{1}{2} \ln [S^i]$

as well as $-S^i \ln \left[\frac{S^i}{S_M} \right] \gg -\frac{1}{2} \ln [2\pi]$. Thus, these terms can be dropped from the approximation, yielding:

$$\ln [C_i] \approx - (S_M - S^i) \ln \left[\frac{(S_M - S^i)}{S_M} \right] - S^i \ln \left[\frac{S^i}{S_M} \right] \quad (7)$$

Define the ratio $\frac{S^i}{S_M} \equiv \mathbf{S}^i$ ($0 \leq \mathbf{S}^i \leq 1$) then rewriting:

$$\ln [C_i] \approx S_M \{ -\mathbf{S}^i \ln [\mathbf{S}^i] - (1 - \mathbf{S}^i) \ln [1 - \mathbf{S}^i] \} \quad (8)$$

Further, defining $H(\mathbf{S}^i) \equiv -\mathbf{S}^i \ln [\mathbf{S}^i] - (1 - \mathbf{S}^i) \ln [1 - \mathbf{S}^i]$, yields the following identity:

$$\ln [C_i] = S_M H(\mathbf{S}^i) \quad (9)$$

Thus, the total number of possible synaptic configurations, C is:

$$C = \sum_{i=1}^{N_T} C_i \approx \sum_{i=1}^{N_T} \text{Exp} [S_M H(\mathbf{S}^i)] \quad (10)$$

Minimizing "Wiring"

The wiring minimization hypothesis is well established in the neuroscience literature. We note that the larger the number of different possible synaptic combinations, the greater the chance there is to find a minimum total wiring length. Likewise, this result also occurs from maximizing metabolic efficiency. Thus, we seek the maximum of Eq.(10). Since the maximum number of synapses is the same for each neuron (by assumption) the value of Eq.(10) is maximized when the number of synapses for each neuron, \mathbf{S}^i is distributed evenly. In other words, for all \mathbf{S}^i :

$$\mathbf{S} = \frac{S_T}{N_T} \quad (11)$$

Defining $\mathbf{S} \equiv \frac{S_T}{N_T}$ and $\mathbf{S} \equiv \frac{S}{S_M}$ yields:

$$C \approx \sum_{i=1}^{N_T} \text{Exp} [S_M H(\mathbf{S})] = N_T \text{Exp} [S_M H(\mathbf{S})] \quad (12)$$

In (logarithmic) terms of the number of configurations per number of neurons:

$$\ln \left[\frac{C}{N_T} \right] \approx S_M H(\mathbf{S}) = S_M \{ -\mathbf{S} \ln [\mathbf{S}] - (1 - \mathbf{S}) \ln [1 - \mathbf{S}] \} \quad (13)$$

Thus, Eq.(10) represents the optimal (most metabolically efficient) state of configurations for data with indistinguishable neurons and synapses. Note that this equation takes the same form as Shannon's measure of information as well as the entropy given by the canonical ensemble in statistical mechanics.

The Maximum Structure Model

We maximize the Kolmogorov structure by using Eq.(13) and minimizing its increase, i.e., $\frac{d^2[S_M H(\mathbf{S})]}{dt^2} = 0$. This results in the following differential equation:

$$f'' + f'^2 \left(1 - \frac{1}{[f - \ln[1 - e^f]](1 - e^f)} \right) = 0 \quad (14)$$

Here, and $f \equiv \ln \left[\frac{S(t)}{S_M} \right]$. An approximate solution to this equation is:

$$\ln[S(t)] = F e^{-\frac{t}{t_c}} \cos(\omega t + \phi) + G \quad (15)$$

Here, ω , ϕ , t_c , F and G are constants. The constants come from \bar{S}_M as well as the initial conditions for the differential equation. See the Appendix, parts II-VII for a derivation of this equation, the definition of the constants, as well as a demonstration that this approximation is a good one. Also, see the Appendix, part VIII, for a demonstration that Eq. (15) yields a minimum.

Empirical Data Analyzed

In this study, we used previously reported data on synaptic and neuronal packing densities in the postnatal human cerebral cortex from birth to 72 months to estimate the number of synapses per neuron in each cortical layer of the three cortical areas for which human data are available. The neuronal packing density data were originally gathered by [8] and subsequently corrected by [9]. The synaptic density data were originally collected by Huttenlocher et al. (1979, 1982, 1987, 1997). The final synaptic and neuronal density data sets we analyzed were corrected for stereologic counting error and tissue shrinkage before deriving estimates of the number of synapses per neuron in each layer of the three cortical areas at each of eight age points from birth to 72 months. Corrections for shrinkage and stereological counting may introduce error, so we also applied our theory to uncorrected data. Together, these data sets allow us to test an entropy-based theory of synaptic development.

Conel's Neuronal Packing Density Data

The age-specific laminar neuron packing density data reported by Conel came from 5 to 9 grossly and histologically normal postmortem human brains for each of eight postnatal age points (0, 1, 3, 6, 15, 24, 48 and 72 months). Clinically, each case showed no evidence of neurologic disease and did not have diseases known to affect the central nervous system. All cases had normal brain weights for their age. Conel reported uncorrected neuron packing density values per cortical layer per cytoarchitectonic area per age point studied. Shankle et al. corrected Conel's neuron packing density data for stereologic counting error and tissue shrinkage, and found that the corrected values agree closely with those values reported by contemporary studies [9]. Conel's data were also analyzed using correspondence analysis [13]. Confidence ellipses from the first two factors demonstrate that within cortical layer variance is generally much smaller than between cortical layer variance [13, Figure 3, P. 4025].

Buttenlocher’s Synaptic Density Data

The only age-specific data on laminar synaptic densities in postnatal human cerebral cortex comparable to the Conel neuron packing density data come from Huttenlocher et al. (1979, 1982, 1987, 1997). Huttenlocher examined one brain per age point to measure age-specific laminar synaptic densities in primary visual cortex – Brodmann area 17 (BA 17), primary auditory cortex (BA 41) and midfrontal association cortex (BA 10). While Huttenlocher corrected these data for stereologic counting errors, he reported that only BA 17 was corrected for tissue shrinkage [14]. We therefore applied his reported age-specific tissue shrinkage factors to BA 10 and 41.

Computing the number of synapses per neuron

Using the tissue-shrinkage and stereological corrections of the neuron packing and synaptic density data of Conel and Huttenlocher, we computed the number of synapses per neuron for each layer of each cortical area at each of the eight postnatal age-points, 0 (birth), 1, 3, 6, 15, 24, 48 and 72 months. The formula was simply: synapses per neuron = synaptic density divided by neuron packing density.

To estimate laminar synaptic densities at the same age points as Conel’s neuron packing density data, we used the synaptic densities at the closest age points on both sides of Conel’s age points and performed linear interpolations. This interpolation appears to be accurate as synaptic development is relatively smooth (Bourgeois, Rakic & 1993; Bourgeois, Goldman-Rakic & Rakic, 1994).

Statistical analysis of the data

For each layer of each cortical area, the goodness-of-fits of the analytical solution for number of synapses per neuron to these data were then computed as a least squares fit:

$$\chi^2(\hat{\theta}) = \sum_{i=1}^8 (\ln[\mathbf{S}_i] - \ln[S_i])^2 \tag{16}$$

Here, \mathbf{S}_i denotes the number of synapses observed at the i th point in time, S_i is the number of synapses predicted by Eq. (15) and $\chi^2(\hat{\theta})$ represents the sum of least-squares, where $\hat{\theta}$ is a vector of the five parameters from Eq. (15)³. It is not clear that the value given by Eq. (16) is actually distributed as a chi-squared random variable. This is especially true for least-squares fits computed from nonlinear regressions. Therefore, to compute a proper test statistic, we used an F-statistic given by [19]:

$$F_{q,8-p} = \frac{\chi^2(\tilde{\theta}) - \chi^2(\hat{\theta})}{\chi^2(\hat{\theta})} \left(\frac{8-p}{q} \right) \tag{17}$$

³The parameter values which minimized Eq. (16) were computed using the conjugate gradient method. The basic method is detailed in [17]. Useful modifications are described in [18].

Here, $\tilde{\theta}$ is a vector of parameters from a null hypothesis and $q = (n-p_0)-(n-p)$, where p_0 is the number of parameters in the null hypothesis. Furthermore, p is the number of parameters of the alternative hypothesis. Thus, this statistic accounts for the number of parameters in both the null and alternative hypotheses. Also, it should be noted that Eq. (16) uses logarithms of the synaptic numbers instead of the original values. This was done for pragmatic (numerical analysis) reasons⁴. However, the resulting values computed from Eq. (17) are essentially the same as those from the original synaptic values. See the Appendix, part IX for a proof.

We first used Eq. (17) to compare our predicted values to the null hypothesis – a horizontal line (one parameter). This gave us the values of five parameters (in Eq. 15) for each cortical layer in each of the three cortical areas. Since five parameters is a large number of parameters relative to eight data points, we fixed two of the parameters for each of the three areas. These fixed parameters are the frequency ω and the decay constant t_c . These parameters were fixed to their average value for each of the three areas. This allowed us to test a three parameter model.

A number of additional models were compared to the our five parameter model (Eq. 15). They are: first, a basic exponential growth model: $S(t) = Ae^{Bt}$ (two parameters) second, a logistic growth model: $S(t) = \frac{Ae^{Bt}}{1+Ce^{Bt}}$ (three parameters) and third, the Lotka-Volterra competition equations (five parameters): $S'(t) = S(t)r_1(1 - \frac{S(t)+\alpha_1 S_2(t)}{K_1})$ and $S_2'(t) = S_2(t)r_2(1 - \frac{S_2(t)+\alpha_2 S(t)}{K_2})$. Here, $A, B, C, D, E, r_1, r_2, \alpha_1, \alpha_2, K_1,$ and K_2 are parameters. S_2 represents a “competitor,” e.g., a pruning mechanism. Note that a solution to the Lotka-Volterra differential equations yields eight parameters. Our model has five, so some of the parameters of the Lotka-Volterra equations were equated to reduce it to a five parameter model. (For an introduction to the Lotka-Volterra theory, see [20].)

To further test our theory, we compared the fits obtained using Eq. (16) to four ad hoc functions. Each of these functions has five parameters. They are, a modified (five parameter) logistic: $S(t) = \frac{Ae^{Bt}}{e^{Ct}+De^{Bt}}$, polynomial: $S(t) = A + Bx + Cx^2 + Dx^3 + Ex^4$, Fourier series: $A + B \cos(Ct) + D \cos(2Ct) + E \cos(3Ct)$, and sinusoidal: $S(t) = A + B \cos(Ct) + D \sin(Et)$.

Main findings

The goodness of fits of our theory are given in Table 1 (3 parameter model) and Table 2 (5 parameter model). We used both the raw (uncorrected) data and the corrected data to test our theory. The 3 parameter model yielded significant fits for the raw data at all 18 data points. All but two points are significant for the corrected data. Similarly for the 5 parameter model, 16 out of 18 data points have a significant fit for the raw data. 15 out of 18 are significant for the corrected data. Figure 1 shows the observed and expected synapses per neuron values for the cortical layer and area with the best fitting (5 parameter,

⁴The data varies over a number of orders of magnitude, making it difficult to find a minimum. However, taking the logarithm of the data makes minimization much easier, as these data are all at the same order of magnitude.

corrected data) solution. Figure 2 shows the worst fitting case (5 parameter, corrected data). Note that in some cases, the 3 parameter model fits better than the 5 parameter model. This may seem counterintuitive. However, it is due to the fact that goodness-of-fit statistic used (Eq. 17) accounts for the number of parameters.

A comparison of our 5 parameter model to three standard growth models is given in the Table 3. These models are the exponential, logistic and the Lotka-Volterra. Here, we used the corrected data for our comparisons. Statistical significance was computed using a sign test. In all three cases, our theory explains these data better in a highly significant manner.

Similarly, Table 3 also shows the results of four ad hoc functional fits. These functions are the modified logistic, polynomial, Fourier and sinusoidal. Again, we used a sign test to compare our theory to these other functions. And again, our theory fits significantly better in all four cases. In three of these cases, the fit is highly significant.

Discussion

The common solution of the maximum Kolmogorov structure equation (Eq. 14) to these different cortical areas and layers is a damped sinusoidal function of the *logarithm* of the number of synaptic connections. This logarithmic dependency is not unusual. Other neurobiological systems, the sensory processes in particular, function on a logarithmic scale [21]. Thus this finding is congruent with the nature of other brain systems. However, a deeper examination of this solution is necessary to see how it correlates to synaptic development. This is done by an examination and interpretation of the parameters of Eq. (15)

Interpreting the parameters

There are five parameters in Eq. (15). Two of these parameters \bar{S}_M and b , were defined earlier. The other three arise from (mathematical) initial conditions. However, it is more informative to examine the parameters in the context of biology. In this context, the growth constant, t_c , is a good candidate for significance beyond being just a function of initial values.

To examine its meaning, one must look at the nature of synaptogenesis in different brain areas. In many theories of synaptic development, there are two basic forms of synaptogenesis in cerebral cortex [22]. The first is “selective,” the other “non-selective.” Selective synaptogenesis occurs when developing axons grow directly to their targets. There is little overgrowth or pruning. On the other hand, non-selective cortical areas show significant axonal growth beyond their final target. They also exhibit significant pruning after this overgrowth. These sorts of theories assert that selective and non-selective cortical areas are fundamentally different.

This is one view of synaptic development. On the other hand, our theory postulates a single equation for both selective and non-selective cortical areas. This means that these two types of cortices form the extremes on a spectrum of developmental possibilities. Thus, we postulate that synaptic development is basically the same in different parts of the cortex, but the degree of “developmental control” (via synaptotrophins etc.) varies from area to area. To test this

hypothesis we chose two cortical areas, one selective area and one non-selective area. We postulated that a single parameter, the growth constant t_c , could distinguish between these two areas.

A good cortical candidate for selective synaptogenesis is Brodmann area 10. Synaptogenesis in area 10 of rhesus monkeys has been carefully examined [16]. Commenting on synapses on dendritic shafts (p. 87), “synaptic contacts ... do not appear either to be overproduced or subject to pruning during development or adulthood.” Similarly, when discussing synaptic development on dendritic spines (also p. 87), they note, “Their density begins to decline mainly after the onset of puberty...” This means that there is little synaptic pruning prior to adolescence. This finding for prefrontal cortex is similar to Huttenlocher’s results and suggests that area 10 engages in selective synaptic development.

Similarly, a good candidate for non-selective synaptogenesis is area 17. Bourgeois and Rakic carefully documented synaptic development in the primary visual cortex of the macaque monkey [15]. On page 2807, Fig. 5, top, they quantified synaptic contacts per neural probe. Their data show a peak of about 5000 synapses at around 3 months. On the other hand, at 18 months (equal to about 72 human months in development) this number is halved to about 2500 synapses. This factor of two synaptic pruning is in agreement with the area 17 data of Huttenlocher. Thus, it appears that primary visual cortex engages in non-selective synaptic development.

To examine the hypothesis that the growth constant could differentiate between selective and non-selective synaptogenesis, a one-way ANOVA as well as a Wilcoxon rank order test were carried out. These tests were used to compare growth constants computed for area 10 with those of area 17. The result of the ANOVA was a significant difference between the two sets at a $p < 0.01$ level of significance. Similarly, the Wilcoxon test demonstrated a significant difference at $p < 0.01$. These results suggest that the growth constant parameter differs between selective and non-selective cortical areas.

Conclusion

In summary, this application of optimal synaptic development to human neurobiology provides useful knowledge which suggests that there may be a general set of rules governing the laminar development of the number of synapses per neuron in different areas of the cerebral cortex in humans, and perhaps other mammals. Direct examination of synaptic development data in various primate and nonprimate species is an important next step in the testing of this new hypothesis.

References

- [1] Stevens, C. F. *Nature* **361**, 500–501 (1993).
- [2] Krubitzer, L. *Trends in Neuroscience* **18**, 408–417 (1995).
- [3] Cover, T. M. and Thomas, J. A. *Elements of information theory*. John Wiley & Sons, New York, (1991).
- [4] Moore, C. C., Romney, A. K., and Hsia, T. *Proceedings of the National Academy of Sciences* **97**, 5007–5010 (2000).
- [5] Moore, C. M., Romney, A. K., Hsia, T., and Rusch, C. D. *Journal of Comparative Neurology* **406**(3), 375–383 (1999).
- [6] Haug, H. *American Journal of Anatomy* **180**, 126–142 (1987).
- [7] Johnson, M., Perry, R. H., Piggot, M. A., Court, J. A., Spurden, D., Lloyd, S., Ince, P. G., and Perry, E. K. *Neurobiology of aging* **17**(4), 639–651 (1996).
- [8] Conel, J. *The postnatal development of the human cerebral cortex*. Harvard University Press, Cambridge, MA, (1939).
- [9] Shankle, W. R., Landing, B. H., H., R. B., A., S., M., C. J., and Hara, J. *Journal of Theoretical Biology* **191**, 115–140 (1998).
- [10] Huttenlocher, P. R. *UCLA Forum on Medicine* **163**, 195–205 (1979).
- [11] Huttenlocher, P. R., De Courten, C., Garey, L., and Van der Loos, H. *International Journal of Neurology* **16**, 144–154 (1982).
- [12] Huttenlocher, P. R. and De Courten, C. *Human Neurobiology* **6**, 1–9 (1987).
- [13] Shankle, W. R., Romney, A. K., Landing, B. H., and Hara, J. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 4023–8 (1998).
- [14] Bloom, F. E. and Aghajanian, G. K. *Journal of Ultrastructure Research* **22**, 361–375 (1968).
- [15] Bourgeois, J. P. and Rakic, P. *The Journal of Neuroscience* **13**, 2801–2820 (1993).
- [16] Bourgeois, J. P., Goldman-Rakic, P. S., and Rakic, P. *Cerebral Cortex* **4**, 78–96 (1994).
- [17] Powell, M. J. D. *Computer Journal* **7**, 155–162 (1964).
- [18] Brent, R. P. *Algorithms for maximization without derivatives*. Prentice Hall, Englewood Cliffs, (1973).

- [19] Seber, G. A. F. and Wild, C. J. *Nonlinear Regression*. Wiley, New York, (1989). Chapter 5.
- [20] Goel, N. S. G., Maitra, S. C., and Montroll, E. W. *On the Volterra and other nonlinear models of interacting populations [by] N. S. Goel, S. C. Maitra [and] E. W. Montroll*. Academic Press, New York, (1971).
- [21] Luce, R. D., Bush, R. R., and Galante, E., editors. *Handbook of Mathematical Psychology*, volume 13. Wiley, New York, (1965).
- [22] Haydon, G. H. and Drapeau, P. *Trends in Neuroscience* **18**, 196–201 (1995).

Acknowledgement 1 *We wish to dedicate this work to the memory of Benjamin Landing, mentor, colleague and friend. We thank R. D. Luce for his helpful criticism and feedback. We also thank the Sloan foundation for their generous support.*

Appendix

Appendix

Part I, Maximizing the Increase in Kolmogorov Structure of an Ensemble of Developing Neurons is Equal to Minimizing its Entropy Rate.

Consider a very large ($\sim 10,000$) group of developing neurons. Assume that for a small, fixed period of time, Δt_i , each neuron can do one of three things: 1) Add a new synapse, 2) Prune a synapse or 3) Neither prune nor add a new synapse. Note that the exact length of this time period, Δt_i may be different at different times of development. However, it is chosen in such a way that the probabilities of other synaptic development events (than the three listed above) are negligible. Then in terms of the number of synaptic connections for a given neuron, (denoted as S), this number is changed by either -1, 0, or 1. For the i th neuron, that change is denoted by s^i .

Thus, for a group of n neurons, we can describe the change in synaptic connectivity as a Bernoulli sequence of length n , e.g., 1,0,-1..., and denoted as s^n . (It is important to note that this is a zeroth order description based on the assumption that the outcomes are i.i.d. as well as the zeroth order assumption of synaptogenesis being uniformly random for a fixed time period, Δt_i .)

These assumptions allow us to perform a Kolmogorov structure analysis of this group of neurons. Much of the analysis described in this section of the Appendix is taken from Cover and Thomas (1991, Ch. 7). The reader is referred to this excellent text for important details. [3] defines the Kolmogorov structure function as⁵:

$$K_k(s^n|n) = \min_{\substack{p: l(p) \leq k \\ \mathcal{U}(p,n) = C \\ x^n \in S \subseteq \{-1,0,1\}^n}} \log|C| \quad (18)$$

Here, \mathcal{U} is a universal Turing machine, p is a program, $l(p)$ is its length, x^n is a Bernoulli sequence of length n drawn from $\{-1, 0, 1\}^n$. C is a set containing s^n , $|C|$ is its cardinality and k is a constant. The conditionality on n refers to the computer having prior knowledge of the value of n . The minimum sufficient statistic for the Kolmogorov structure is defined as the minimum k such that:

$$K_k(s^n|n) + k = K(s^n|n) \quad (19)$$

Here, $K(x^n|n)$ is the Kolmogorov complexity function which is defined as $K = \min_{p: \mathcal{U}(p)=x} l(p)$. For the rest of the Appendix, ' k ' will refer to this minimum value.

Let s_m^n refer to the value of the Bernoulli sequence at time $t_m = \sum_{i=1}^m \Delta t_i$. Also, let k_m denote the value of k at time t_m . Then the difference in Kolmogorov structure between the times t_m and $t_{\bar{m}+1}$ is:

$$K_k(s_{m+1}^n|n) - K_k(s_m^n|n) = K(s_{m+1}^n|n) - k_{m+1} - K(s_m^n|n) + k_m \quad (20)$$

⁵This structure function, K_k should not be confused with the Kolmogorov third order structure function used in atmosphere and turbulence research. Also, note that this function is not the same as the Kolmogorov complexity function, though the two are closely related.

Thus, the mean value of this difference is:

$$\lim_{n \rightarrow \infty} E \left[\frac{K_k(s_{m+1}^n | n) - K_k(s_m^n | n)}{n} \right] \quad (21)$$

$$= \lim_{n \rightarrow \infty} E \left[\frac{K(s_{m+1}^n | n) - k_{m+1} - K(s_m^n | n) + k_m}{n} \right] \quad (22)$$

$$= \lim_{n \rightarrow \infty} E \left[\frac{K(s_{m+1}^n | n)}{n} \right] - \lim_{n \rightarrow \infty} E \left[\frac{K(s_m^n | n)}{n} \right] + \lim_{n \rightarrow \infty} E \left[\frac{k_m - k_{m+1}}{n} \right] \quad (23)$$

Here, E is the expected value. Since the actual number of neurons is very large, the infinite limit provides a good approximation. There is a basic relationship between the mean value of the Kolmogorov complexity and the entropy H . Mathematically, this relationship is: $\lim_{n \rightarrow \infty} E \left[\frac{K(s^n | n)}{n} \right] = H(\mathbf{S})$ (Theorem 7.3.1, Cover & Thomas, 1991). Here, \mathbf{S} is the random variable with outcomes s_i . Applying this theorem to Eq.(23) yields:

$$= H(\mathbf{S}_{m+1}) - H(\mathbf{S}_m) + \lim_{n \rightarrow \infty} E \left[\frac{k_m - k_{m+1}}{n} \right] \quad (24)$$

Note that k_m and k_{m+1} are the lengths of the programs used to write the indexed sets, s_m and s_{m+1} . Members of these sets have the exact same number of elements, n . Furthermore, the members of both sets are taken from the same (or very similar) Bernoulli sequence. Thus, the difference $k_m - k_{m+1}$ will be a relatively small number compared to n . Therefore, $\lim_{n \rightarrow \infty} E \left[\frac{k_m - k_{m+1}}{n} \right] = 0$. Applying this to Eq.(24) yields:

$$= H(\mathbf{S}_{m+1}) - H(\mathbf{S}_m) \quad (25)$$

Thus, the average change in Kolmogorov structure is equal to the difference in entropies. This fact can be rewritten in terms of rates by dividing by the time step value, Δt_i :

$$\lim_{n \rightarrow \infty} \frac{1}{\Delta t} E \left[\frac{K_{k_{m+1}}(s^n | n) - K_{k_m}(s^n | n)}{n} \right] = \frac{H(\mathbf{S}_{m+1}) - H(\mathbf{S}_m)}{\Delta t_i} \quad (26)$$

$$\approx \frac{dH \left(\mathbf{S}(t = \sum_{i=1}^{m+1} \Delta t_i) \right)}{dt} \quad (27)$$

Somewhat counterintuitively, the greater the value of the Kolmogorov structure function, the less structure there is. Therefore, to maximize the increase in synaptic structure over time, we wish to minimize Eq.(26). This is done in the standard way, i.e., by taking its derivative and setting it equal to zero:

$$\frac{d}{dt} \left[\frac{dH \left(\mathbf{S}(t = \sum_{i=1}^{m+1} \Delta t_i) \right)}{dt} \right] = \frac{d^2 H \left(\mathbf{S}(t = \sum_{i=1}^{m+1} \Delta t_i) \right)}{dt^2} \quad (28)$$

$$= 0$$

Thus, maximizing the increase in average synaptic structure over time is equivalent to minimizing its entropy rate.

Part II, Differential Equation for Synaptic Development

The entropy for a discrete probability distribution is defined as $H(\mathbf{S}) = \sum_{i=1}^N -p(s_i) \ln[p(s_i)]$ [3]. Here, $p(s_i)$ is the probability of a particular outcome, s_i . Inserting Eqs.(??-??) into $H(\mathbf{S})$ yields:

$$\begin{aligned} H(t) = & - \left(1 - \frac{S(t)}{\bar{S}_M}\right) \ln\left[1 - \frac{S(t)}{\bar{S}_M}\right] \\ & - \left(b \frac{S(t)}{\bar{S}_M}\right) \ln\left[b \frac{S(t)}{\bar{S}_M}\right] \\ & - \left((1-b) \frac{S(t)}{\bar{S}_M}\right) \ln\left[(1-b) \frac{S(t)}{\bar{S}_M}\right] \end{aligned} \quad (29)$$

Note that the random variable, $\mathbf{S}(t = \sum_{i=1}^{m+1} \Delta t_i)$ is time-dependent, that is, nonhomogeneous. To solve for a continuous function of synaptic development, we will approximate $\mathbf{S}(t)$ as a continuous function of time, i.e., $S(t)$. This allows us to find a differential equation for the maximum Kolmogorov structure increase by inserting Eq.(20) into Eq.(19). To solve this equation, we define $f(t) \equiv \ln[b \frac{S(t)}{\bar{S}_M}]$. This yields the differential equation:

$$f'' + f'^2 \left(1 - \frac{1}{\left[f + k - \ln\left[1 - \frac{ef}{b}\right]\right] (1 - ef)}\right) = 0 \quad (30)$$

Here, $k \equiv b \ln[b] + (1-b) \ln[1-b]$.

Part III, The linear approximation to Eq. (30)

The linear approximation to Eq. (30). can be computed using Taylor series expansions. Dividing both sides of Eq.(30) by f' yields:

$$\frac{f''}{f'} + f' \left(1 - \frac{1}{\left[f + k - \ln\left[1 - \frac{ef}{b}\right]\right] (1 - ef)}\right) = 0. \quad (31)$$

Then a Taylor series expansion of the $\frac{1}{f'}$ portion of the first term of Eq. (31) ($\frac{f''}{f'}$), yields:

$$\frac{f''}{f'} \approx \frac{f''}{f'_c} - \frac{f''}{f'^2_c} (f' - f'_c) \quad (32)$$

Here, f'_0 is shorthand notation for $f'(t_0)$, i.e., the function evaluated at a fixed time. This term is further approximated by inserting the first-order value for f'' into the second term of Eq. (32), yielding:

$$\frac{f''}{f'} \approx f'' \frac{2}{f'_0} - \frac{f''_0 f'}{f'^2_0} \quad (33)$$

To approximate the second term, define K as $\frac{1}{(f_0+k-\ln[1-e^{f_0}])(1-e^{f_0})}$ and

$$\begin{aligned} & \left(\frac{1}{(\ln[1-e^f]-f-k)(1-e^f)} \right)'_0 \\ &= \frac{-1}{[(f_0+k-\ln[1-e^{f_0}])(1-e^{f_0})]^2} \left[(1-e^{f_0}) \left(1 + \frac{e^{f_0}}{1-e^{f_0}} \right) - e^{f_0} (f_0+k-\ln(1-e^{f_0})) \right] \\ & \text{as } -H. \text{ Then, Eq. (30) can be written as:} \end{aligned}$$

$$A f'' + B f' + C f - D = 0 \quad (34)$$

where,

$$\begin{aligned} A &\equiv \frac{2}{f'_0}, \\ B &\equiv 1 - \frac{f''_0}{f'^2_0} - H \\ C &\equiv f'_0 H \\ D &\equiv f_0 K \end{aligned}$$

This equation has the well-known damped harmonic oscillator solution:

$$f = G e^{-\frac{t}{t_c}} \cos(\omega t + \phi) + \frac{D}{C} \quad (35)$$

Here, $t_c \equiv \frac{2A}{B}$ and $\omega \equiv \sqrt{\frac{C}{A} - \frac{B}{4A}}$. G and ϕ are parameters.

Rewriting in terms of $\ln[s(t)]$ this equation becomes:

$$\ln[s(t)] = G e^{-\frac{t}{t_c}} \cos(\omega t + \phi) + F \quad (36)$$

Here, F is defined as $F \equiv \frac{D}{C} + \ln\left[\frac{\bar{S}_M}{b}\right]$.

Part IV, Asymptotic solution for $0 \lesssim s(t) \ll \bar{S}_M$

At this asymptote, $e^f \approx \frac{e}{\bar{S}_M} (f + \ln[\bar{S}_M])$ and $\ln[1-e^f] \approx \frac{-e}{\bar{S}_M} (f + \ln[\bar{S}_M])$. Inserting these values into Eq. (30) and dividing by f' yields:

$$\ln[f']' + f' \left[1 - \frac{1}{\left[f + k + \frac{e}{\bar{S}_M} (f + \ln[\bar{S}_M]) \right] \left[1 - \frac{e}{\bar{S}_M} (f + \ln[\bar{S}_M]) \right]} \right] = 0 \quad (37)$$

This equation can be rewritten as:

$$\ln[f']' + f' \left[1 - \frac{1}{A f^2 + B f + C} \right] = 0 \quad (38)$$

Here, $A \equiv (1 + \frac{e}{\bar{S}_M})\frac{e}{\bar{S}_M}$, $B \equiv (1 + \frac{e}{\bar{S}_M})\frac{e \ln[\bar{S}_M]}{\bar{S}_M} + \frac{ke}{\bar{S}_M} + \frac{e^2 \ln[\bar{S}_M]}{\bar{S}_M}$ and $C \equiv \frac{ke \ln[\bar{S}_M]}{\bar{S}_M} + (\frac{e \ln[\bar{S}_M]}{\bar{S}_M})^2$. Integrating Eq. (38) and exponentiating yields:

$$f' = De^{\frac{1}{\sqrt{A}} \left[\frac{1}{\sqrt{F}} \tan^{-1} \left(\frac{f\sqrt{\frac{A}{F}} + \frac{B}{2\sqrt{AF}}}{\sqrt{F}} \right) - \sqrt{A}f + \frac{B}{2\sqrt{A}} \right]} \quad (39)$$

Here $F \equiv C - \frac{B^2}{4A}$, D is an integration constant. A good approximation is $\tan^{-1} \left(\frac{f\sqrt{\frac{A}{F}} + \frac{B}{2\sqrt{AF}}}{\sqrt{F}} \right) = \frac{\pi}{2}$. Using this approximation and integrating a second time yields:

$$f = L \ln[Mt + N] \quad (40)$$

Here, L , M and N are arbitrary constants.

Part V, Comparison of the oscillator solution to the asymptotic solution for $0 \lesssim t$

The second order Taylor series expansion of Eq. (40) is:

$$L \ln[Mt + N] \approx L \ln[N] + \frac{LM}{N}t - L \frac{M^2}{N^2} \frac{t^2}{2} \quad (41)$$

This can be compared to the Taylor series expansion of Eq. (35) for $t \approx 0$:

$$\begin{aligned} f(t) \approx & -L \cos(\phi) \\ & + L \left[\frac{\cos(\phi)}{t_c} + \omega \sin(\phi) \right] t \\ & - L \left[\frac{\cos(\phi)}{t_c^2} + \frac{2\omega \sin(\phi)}{t_c} - \omega^2 \cos(\phi) \right] \frac{t^2}{2} \end{aligned} \quad (42)$$

Choosing ϕ such that:

$$\phi = \cos^{-1}(\ln[N]) \quad (43)$$

as well as ω such that:

$$\omega = \frac{\frac{M}{N} - \frac{\ln[N]}{k}}{\sin(\cos^{-1}(\ln[N]))} \quad (44)$$

equates the first two terms of the series expansions. The third terms are approximately equal if one of the following two conditions hold:

1. $\phi \approx 0$ and $\frac{1}{t_c} \gg \omega$
2. $\phi \approx \frac{\pi}{2}$ and $\frac{2}{t_c} \approx \omega$

Part VI, Asymptotic solution for $0 \ll s(t) \lesssim \bar{S}_M$

At this asymptote, $e^f \approx 1 + f$ and $\ln[1 - e^f] \approx -\ln[\bar{S}_M]$. Inserting these values into Eq. (30) and dividing by f' yields:

$$\ln[f']' + f' \left[1 - \frac{1}{(f + k + \ln[\bar{S}_M])(-f)} \right] = 0 \quad (45)$$

Using the approximation $f^2 \approx 0$, integrating and exponentiating yields:

$$f' = Le^{-f} \left(2 + \frac{k + \ln[\overline{S}_M]}{f} \right)^{\frac{-1}{(k + \ln[\overline{S}_M])}} \quad (46)$$

Here, L is an integration constant. In general, this equation cannot be integrated. However, it is clear that it has a vertical asymptote to $-\infty$ at $f = 0$. To illustrate this, set $k + \ln[\overline{S}_M] = 1$. In this case, an implicit solution for f is:

$$2e^f + \int_f^\infty \frac{e^x}{x} dx = Lt + M \quad (47)$$

Here, M is a second constant. Fig. 3 shows a graph of this function. Note the vertical asymptote at zero ($s = \overline{S}_M$, $M = 0$). This corresponds to a horizontal asymptote for $f(t)$. A zeroth-order approximation demonstrates this directly. Approximate $e^f \approx 1$. Then Eq. (30) becomes:

$$f'^2 = 0 \quad (48)$$

This equation solves to:

$$f = L \quad (49)$$

i.e., a horizontal line. Note that Eq. (35) also has a horizontal asymptote.

Part VII, Critical points of the function f , as given by Eq. (30).

An examination of Eq. (30) demonstrates that the solution, f has an critical point ($f' = 0$) when:

$$\left[f + k - \ln\left[1 - \frac{e^f}{b}\right] \right] (1 - e^f) = 0 \quad (50)$$

The two factors yield two possible solutions. The first factor becomes a zero when:

$$f + k = \ln\left[1 - \frac{e^f}{b}\right] \quad (51)$$

Where f is defined as:

$$f(t) \equiv \ln\left[b \frac{S(t)}{\overline{S}_M}\right] \quad (52)$$

Solving for $\frac{S(t_1)}{\overline{S}_M}$ yields:

$$\frac{1}{be^k + 1} = \frac{S(t_1)}{\overline{S}_M} \quad (53)$$

Here, t_1 denotes a point in time where the first factor in Eq. (50) yields an critical point. Numerically speaking, this inversion point occurs when $S(t_1)$ is between $\frac{1}{2}$ and 1 of \overline{S}_M (depending on the value of b), i.e., intermediate values of f . Examining the second factor becomes a zero when:

$$1 = e^f \quad (54)$$

The definition of f (Eq. 52) yields the zero condition:

$$S(t_2) = \bar{S}_M \quad (55)$$

This zero point can only be reached when $b = 1$.

Thus, for $b \neq 1$, there is at most, one possible critical point. On the other hand, for $b = 1$, Eq. (52) yields an exact solution to Eq. (53):

$$S(t_1) = \frac{\bar{S}_M}{2} \quad (56)$$

Thus, for $b = 1$ there are critical points at $S(t_2) = \bar{S}_M$ and at $S(t_1) = \frac{\bar{S}_M}{2}$.

Critical points as constraints to Eq. (30)

There appears to be no clear way to assign the critical points the status of maximum, minimum or inflection without assuming a solution to the Eq. (30). However, if the linear approximation solution, as given in Eq. (15) is assumed, then it is accurate, with added constraints. These constraints yield five possible solution forms. These are:

1. An over-damped harmonic oscillator, i.e., no local maxima or minima.
2. An over-damped harmonic oscillator, i.e., no local maxima or minima, but an inflection point at $S(t) = \frac{\bar{S}_M}{2}$ ($b = 1$).
3. A critically damped harmonic oscillator, i.e., one local maximum.
4. A critically damped harmonic oscillator, i.e., one local maximum and an initial inflection point at $S(t) = \frac{\bar{S}_M}{2}$ ($b = 1$).
5. A damped harmonic oscillator with an inflection point at the first time that $S(t) = \frac{\bar{S}_M}{2}$, a maximum at $S(t) = \bar{S}_M$ and a minimum at the second point where $S(t) = \frac{\bar{S}_M}{2}$ ($b = 1$). This is followed by a horizontal asymptote.

It should be noted that the initial critical point – a saddle point is smoothed over by a growing exponential. That is, the saddle point behavior does not actually occur in the linear approximation solution. There appears to be no good way to get around this by-product of the approximation.

Part VIII, Eq. (15) yields a minimum

In order to show that Eq. (15) is a minimum, it is necessary to show that the third derivative of the entropy is positive, i.e.,

$$H'''[S(t)] > 0 \quad (57)$$

This differential inequality does not have an analytical solution. However, use of the linear solution (Eq. 36) and numerical analysis yields H''' as a function of $\frac{s}{\bar{S}_M}$, i.e., a ratio between zero and one. As an example, area 17, layer 6 is shown in Fig. 4

The parameter values used to compute this curve are the same as those used to fit area 17, layer 6 data, with the additional assumption of the parameter b being $b = 0.5$. In general, the third derivative of the entropy was positive. However, an important exception is when the parameter b is below about 0.08. See Fig. 5 below, for an example. This finding is discussed in the main text.

Part IX, An F statistic is equal to an F statistic with logarithmic values

Let \tilde{S}_i denote the i th synaptic number computed using hypothesis $\tilde{\theta}$, \hat{S}_i be the i th synaptic number computed using hypothesis $\hat{\theta}$ and \mathbf{S}_i be the i th value of the measured synaptic number. Then, Eq.(17) becomes:

$$\begin{aligned} \mathbf{LR} &\approx \frac{(8-p) \sum_{i=1}^8 (\ln[\mathbf{S}_i] - \ln[\tilde{S}_i])^2 - \sum_{i=1}^8 (\ln[\mathbf{S}_i] - \ln[\hat{S}_i])^2}{q \sum_{i=1}^8 (\ln[\mathbf{S}_i] - \ln[\tilde{S}_i])^2} \quad (58) \\ &= \frac{(8-p) \sum_{i=1}^8 (\ln[\frac{\tilde{S}_i}{\mathbf{S}_i}])^2 - \sum_{i=1}^8 (\ln[\frac{\hat{S}_i}{\mathbf{S}_i}])^2}{q \sum_{i=1}^8 (\ln[\frac{\tilde{S}_i}{\mathbf{S}_i}])^2} \\ &= \frac{(8-p) \sum_{i=1}^8 (\ln[1 + \frac{\tilde{S}_i - \mathbf{S}_i}{\mathbf{S}_i}])^2 - \sum_{i=1}^8 (\ln[1 + \frac{\hat{S}_i - \mathbf{S}_i}{\mathbf{S}_i}])^2}{q \sum_{i=1}^8 (\ln[1 + \frac{\tilde{S}_i - \mathbf{S}_i}{\mathbf{S}_i}])^2} \end{aligned}$$

Using two approximations, first, $\ln(1+x) \approx x$ and then $\frac{1}{\mathbf{S}_i^2} \approx E\left[\frac{1}{\mathbf{S}_i^2}\right]$, where E is the expected value, yields:

$$\begin{aligned} &\approx \frac{(8-p) \sum_{i=1}^8 (\frac{\tilde{S}_i - \mathbf{S}_i}{\mathbf{S}_i})^2 - \sum_{i=1}^8 (\frac{\hat{S}_i - \mathbf{S}_i}{\mathbf{S}_i})^2}{q \sum_{i=1}^8 (\frac{\tilde{S}_i - \mathbf{S}_i}{\mathbf{S}_i})^2} \quad (59) \\ &\approx \frac{(8-p) \sum_{i=1}^8 (\tilde{S}_i - \mathbf{S}_i)^2 (E\left[\frac{1}{\mathbf{S}_i^2}\right]) - \sum_{i=1}^8 (\hat{S}_i - \mathbf{S}_i)^2 (E\left[\frac{1}{\mathbf{S}_i^2}\right])}{q \sum_{i=1}^8 (\tilde{S}_i - \mathbf{S}_i)^2 (E\left[\frac{1}{\mathbf{S}_i^2}\right])} \\ &= \frac{(8-p) E\left[\frac{1}{\mathbf{S}_i^2}\right] \left[\sum_{i=1}^8 (\tilde{S}_i - \mathbf{S}_i)^2 - \sum_{i=1}^8 (\hat{S}_i - \mathbf{S}_i)^2 \right]}{q E\left[\frac{1}{\mathbf{S}_i^2}\right] \left[\sum_{i=1}^8 (\tilde{S}_i - \mathbf{S}_i)^2 \right]} \\ &= \frac{(8-p) \sum_{i=1}^8 (\tilde{S}_i - \mathbf{S}_i)^2 - \sum_{i=1}^8 (\hat{S}_i - \mathbf{S}_i)^2}{q \sum_{i=1}^8 (\tilde{S}_i - \mathbf{S}_i)^2} \end{aligned}$$

which completes the proof.

Table 1. Goodness of fit for 3 parameter model by area, layer and data type.

Layer	Data:	Area 10		Area 17		Area 41	
		Raw	Corrected	Raw	Corrected	Raw	Corrected
		Fit	Fit	Fit	Fit	Fit	Fit
1		99%	99%	99%	99%	99%	99%
2		99%	99%	99%	99%	99%	99%
3		99%	99%	96%	99%	98%	82%
4		99%	99%	99%	99%	99%	98%
5		99%	99%	99%	99%	99%	93%
6		99%	99%	99%	99%	99%	99%

Table 2. Goodness of fit for 5 parameter model by area, layer and data type.

Layer	Data:	Area 10		Area 17		Area 41	
		Raw	Corrected	Raw	Corrected	Raw	Corrected
		Fit	Fit	Fit	Fit	Fit	Fit
1		98%	99%	99%	99%	98%	97%
2		99%	99%	99%	99%	99%	97%
3		98%	97%	99%	99%	70%	53%
4		99%	99%	99%	99%	97%	92%
5		97%	98%	99%	96%	99%	95%
6		99%	97%	97%	99%	92%	94%

Table 3. Comparison of Maximum Structure Model to Other Models.

Comparison Model	Number of Parameters	# of better fits for our theory (out of 18)	Significance
Exponential Growth	2	18	$p < 4 \times 10^{-6}$
Logistic Growth	3	18	$p < 4 \times 10^{-6}$
Lotka-Volterra	5	17	$p < 7 \times 10^{-5}$
Modified Logistic	5	14	$p < 0.02$
Polynomial	5	16	$p < 0.0007$
Fourier	5	18	$p < 4 \times 10^{-6}$
Sinusoidal	5	18	$p < 4 \times 10^{-6}$

Area 17
Layer 2

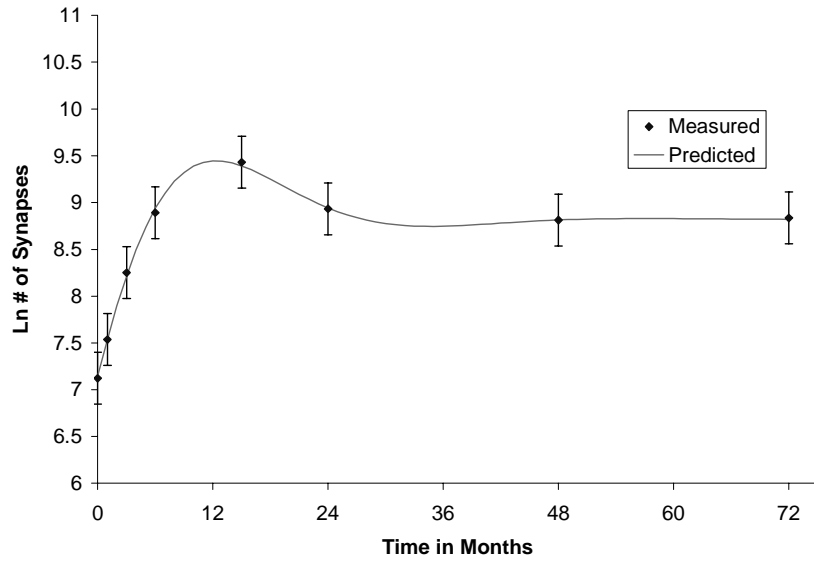


Figure 1:

Fig. 1 Best Predicted Area and Layer

**Area 41
Layer 3**

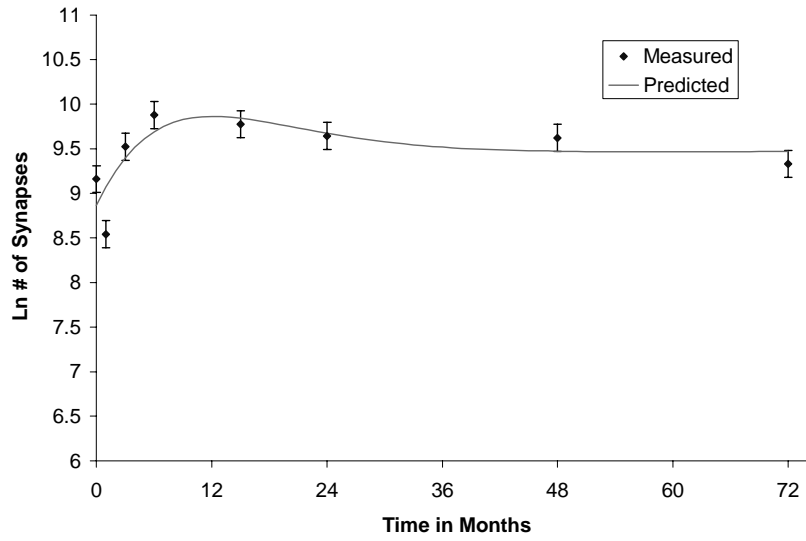


Figure 2:

Fig. 2 Worst Predicted Area and Layer

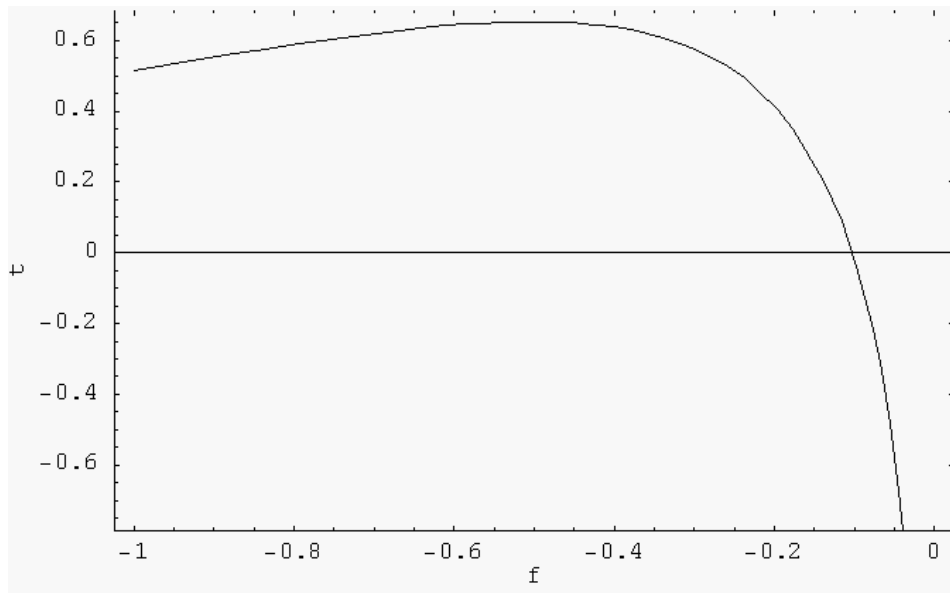


Figure 3:

Fig. 3 An example of Eq. (47)

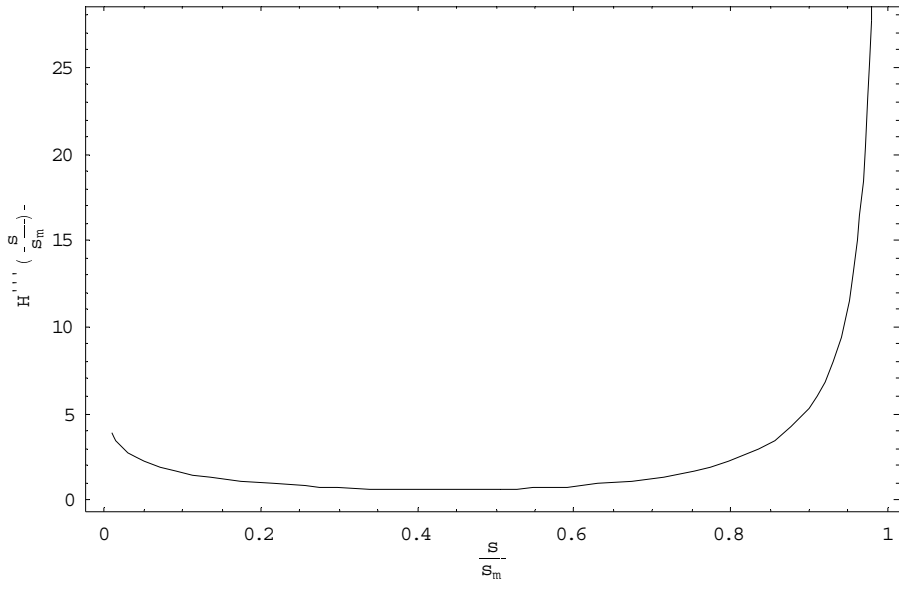


Figure 4:

Fig 4 An example of H'''

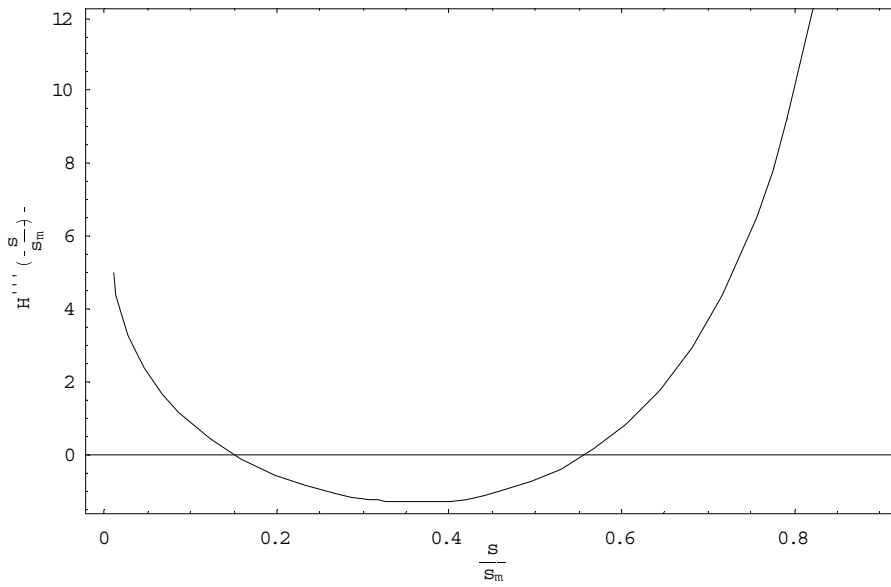


Figure 5:

Fig. 5 An example of $H''' < 0$ for $b = 0.05$