# Dynamic Suppression of Sensory Detail Saves Energy

Terry Bossomaier*, Lionel Barnett†, Michael Harré,‡ and Herbert F. Jelinek*

*Centre for Research in Complex Systems
Charles Sturt University
Bathurst, Australia
Email: {tbossomaier,hjelinek}@csu.edu.au

†Sackler Centre for Consciousness Science
School of Informatics
University of Sussex
Brighton, BN1 9QH, UK
Email: lionelb@sussex.ac.uk

‡Complex Systems Group
Faculty of Engineering and Information Technologies
The University of Sydney
Sydney, Australia
Email: mike.harre@gmail.com

*Abstract*—**High functioning autistic people can exhibit exceptional skills with numbers, eidetic imagery and recall of concrete detail, as brought to popular attention in the film Rain Man. However, it now transpires that these skills are to some extent latent within all of us. We do not have access under normal circumstances to this concrete detail, because it is dynamically inhibited by higher level concepts. Brain stimulation using Trans Cranial Magnetic Stimulation or Direct Current Stimulation, which blocks this inhibition, releases savant-like skills in non-savants. This paper proposes that one of the reasons for this lies in the brain's need to conserve energy. Computer simulations using a spiking neural network support this hypothesis. A spiking neural network was set up with a number of feature detectors feeding an output unit, which in turn generates inhibition of the input neurons. This reduces the spike activity of the input, and thus overall energy usage. We introduce a theoretical analysis for the gains, which might be made. Thus, we demonstrate that energy conservation is a possible cause of inhibition of sensory detail by high level concepts.**

*Keywords*—**energy, spiking neuron, inhibition, simulation, sensory detail.**

## I. INTRODUCTION

The evidence from high functioning autistic individuals shows the overwhelming advantage of concept formation in the human brain. Such individuals tend to have weak concept formation but can have very powerful perception and memory for detail. The evolutionary significance is abundantly clear. What is less clear is *why the raw detail, to which these people have access is not available to everybody else.* The surprising thing, revealed by direct brain stimulation, is that this detail is *not* destroyed on the way to conscious awareness, but is somehow *blocked from access.* This paper provides a novel solution to this conundrum, which was first described in Bossomaier et al. [1].

To demonstrate that such a conjecture is feasible, we ran a series of computer simulations of spiking neurons and found, that, indeed, energy could be saved in this way. Access to raw detail could create a better brain, able to accomplish a greater range of tasks and arguably be more creative [14]. Thus, a powerful argument is needed for why such a brain has not evolved. Energy could be one such factor.

The energy cost of neural computation is split between the generation of spikes and synaptic activity. Thus, the energy used by a neuron communicating with another neuron depends on the rate at which it fires spikes and the excitatory post synaptic potentials associated with each synapse. Thus, the number of other neurons with which a neuron communicates significantly affects the energy requirements. The relative proportion varies across species [45], but the focus in this article is on the spike activity. We show that spiking neural networks, even using the most basic approximation to the established Hodgkin-Huxley spike-dynamic equations [22], can exhibit significant energy savings in these inhibition models.

We consider two cases. The first implements our concept model of the previous paragraph. The second uses a Bayesian or attention approach to reduce energy costs even further. The essential feature of both models is the inhibition of inputs as soon as a concept has been activated.

To begin with, Section II surveys some of the related background work on the suppression of detail and energy constraints in the brain. Section III describes the simulation model. The methodology of the paper and the parameters for the simulation are given in Section IV. Section V provides

theoretical analysis, Section VI gives results from the simulation and Section VII provides a detailed discussion and contextualisation. Section VIII concludes.

## II. RELATED WORK

Early indicators that some of this low-level detail might be accessible came from studies on victims of stroke and brain injury, where, for example, a person might discover the ability to draw realistically. Snyder and Mitchell [52] predicted that such access might be obtained using brain stimulation techniques, in which the conceptual part of the brain was blocked, because concepts inhibit lower level detail [51].

It transpired that this was indeed the case. The direct brain stimulation techniques, *Transcranial Magnetic Stimulation* (TMS) and the more recent technique, *Transcranial Direct-Current Stimulation* (TDCS) can be used to "switch off" part of the brain. By targeting the anterior temporal lobe in the left hemisphere—a brain area highly involved in concept formation and storage—it is possible to block access to concepts and thus release access to lower-level detail. In the first such study, now nearly a decade ago, drawing and proof-reading [53] were found to be enhanced by TMS. So, for example, it is hard for many people to see the word "the" when it is repeated on a following line. The ability to spot the error is enhanced when the meaning of the sentence is blocked by brain stimulation. Likewise numerosity [50] (rapidly estimating the number of objects in the field of view, and inspired by an incident in the film Rain Man), also goes up with TMS to the left anterior temporal lobe, i.e., blocking left temporal lobe activity. Over the subsequent decade, a diverse range of higher-level cognitive phenomena have been shown to be enhanced through dis-inhibition with brain stimulation. False memory, where like objects may get mixed up in memory tests (e.g., chair instead of stool), can be reduced [4]. Even the ability to solve visual puzzles can be enhanced [7].

There are numerous arguments for why this might be the case, such as the possibility of computational overload, discussed further in Section VII. In this era of information overload, such an explanation is at first sight appealing, but is hard to quantify with our existing knowledge of the brain.

Closely linked to computational overload is the energy cost of neural computation. The human brain uses about 20% of the body's energy [41] and various evolutionary changes, such as the appearance of meat in the diet, may have allowed the brain's energy consumption to grow. In fact estimates of brain energy usage per gram is about the same as human leg muscle during a marathon [2]. Navarette et al. [36] show that in over 100 species of mammal, adipose deposits correlate negatively with encephalisation. Thus, fat storage and increased brain size are oppositional strategies for avoiding starvation. But the cost of neural computation appears as a constraint across the animal kingdom. Plaçais and Preat show that in flies the brain disables costly memory mechanisms in the face of starvation [39].

Laughlin and Sejnowski [31] show that the brain's overarching network structure is consistent with preserving energy.

The energy required for the transmission of nerve impulses, or spikes, and synaptic transmission are very tightly optimized, approaching the thermodynamic limits within cellular constraints [30]. The energy cost of transmission of a single bit of information turns out to cost around $10^4$ ATP molecules. A single protein molecule, switching conformational states, was estimated to be able to switch at 1 ATP/bit, but the incorporation of this switch into the rest of the cellular circuitry is very costly.

The fundamental unit of energy across most animal systems is the energy released by conversion of the adenosine triphosphate (ATP) to adenosine diphosphate (ADP). This is about 10kT at human body temperature, where k is Boltzmann's constant and T is absolute temperature. The theoretical limit for transmission of a single bit would be kT, so the cellular cost is about 5 orders of magnitude above the theoretical limit.

Neuronal spikes account for a significant fraction of the neuronal energy usage [2].

The idea that the number of spikes might be kept to a minimum to save energy began with the idea of *sparse coding* in sensory systems [49][38]. More recently, cells have been observed, which fire strongly when the subject is exposed to stimuli corresponding to a particular person, say Bill Clinton, and to very little else [11][40]. They respond to the *concept,* and can be activated by pictures, voice or unique events. Obviously, for most people such a cell would fire very infrequently. The alternative distributed representation might have many cells coding for all US presidents. All of these cells would be active for any president, thus making their average activity much higher.

However, sparse coding is not the only way to reduce energy consumption by neurons using action potentials (APs). Changing the kinetics of the ion channels involved in generating the spike can reduce the energy requirements of the APs. Sengupta et al. [45] show that considerable differences in the relative cost of spike transmission versus the energy of synaptic transmission may be found, depending upon the exact ion channel kinetics, for example between giant squid neurons and those in mouse cortex.

This strong need to conserve energy suggests a possible explanation for why the raw sensory input is not accessible other than through external means such as TMS. *It is turned off to save energy.* Snyder et al. [51] and Bossomaier and Snyder [5] propose a *concept model* for how inhibition mechanisms might generate the observed effects of TMS. The effect is to turn off the inhibitory mechanisms, dis-inhibiting their targets.

Inhibition is of course widespread in the brain, and the prefrontal cortex—the area with most development over other primates—abounds in inhibitory effects. But, evidence is now emerging that even sensory perception in early areas such as primary visual area V1 depends upon top-down modulation, of which a large part is inhibitory [12][44].

From an evolutionary perspective, the human brain is more economical on energy use than other mammals. Lennie [32] calculates the rate of glucose (the brain's only source of metabolic energy) is three times lower than in rat and one and

a half times lower in monkey. Thus, since the human brain is much bigger, this implies that fewer cells are active.

Feedback mechanisms are a common way of modulating input from lower processing areas of the cortex to higher processing areas. Visual processing streams provide a good example where higher order visual areas display an inhibitory top-down activity to lower visual processing areas such as V1 [12][44]. However, these models only consider connectivity patterns in the cortex related to visual processing. Jelinek and Elston [25] have shown that on a cellular level, processing complexity increases from V1 to prefrontal cortex, with layer-III pyramidal cell dendritic branching patterns becoming more complex and larger, thus requiring more energy. Higher visual processing areas deal more with conceptual phenomena by integrating simple information bits from lower processing areas.

Higher processing areas are not always necessary for rapid concept formation suggesting that single spike or a limited number of spikes representing different input to sensory areas can be already sorted prior to higher level processing and therefore enhancing processing speed. This has been shown in the auditory cortex. Higher cognitive processes, such as extracting similar patterns in varying acoustic input, or anticipation of acoustic input, already occur at the level of the sensory system rather than requiring higher cortical input. Thus, less information is passed on to higher levels such as the prefrontal cortex even if attention is not directed to the current sensory input [35].

In addition, reciprocal feedback has been shown between prefrontal areas of the cortex and hippocampus. Two pathways exist, which carry highly specific information from the prefrontal cortex to the hippocampus and the reciprocal connection allow rapid retrieval of information from the hippocampus by the prefrontal cortex [15].

Such top-down effects reduce activity at lower levels. Zhang et al. [56] show that in inferotemporal cortex, activity corresponding to a particular object is vastly different depending upon whether attention is focussed on that object.

### III. THE SIMULATION MODEL

The simplest approximation to the Hodgkin-Huxley equations is the Leaky Integrate and Fire (LIF) model. Izhikevich [23] points out that this neuron is capable of only a few of the 20, or so, behaviors, of which the full Hodgkin-Huxley model is capable. However, it is used here because *if a very simple model can generate the behavior we observe, then so can any of the more complex models.* This assures that the model is reasonably robust to parameter variations. Since more powerful neural models, such as the Izhikevich [23] model, can imitate the behaviour of simpler models (such as integrate and fire) then these more powerful models will have the same behaviour.

LIF has been used for very large scale models of the brain, such as that of Eliasmith [9]. Some limitations, which appeared in the simulations herein are discussed below.

Equation (1) shows the model for one neuron, where $R$ is resistance, $I$ the input current, $u$ the membrane voltage and $\tau$ the time constant:

$$\frac{du}{dt} = -\frac{u}{\tau} + \frac{IR}{\tau} \qquad (1)$$

Synaptic activation is represented by an alpha function with another time constant $\tau_s$:

$$\varepsilon(t) = \frac{1}{\tau_s} e^{1-t/\tau_s} \qquad (2)$$

The two simulation models use the same type of neuron, although the time constants are not the same.

#### A. Concept and Inhibition

Local inhibitory circuits have been studied in the visual system and hippocampus amongst others, which through inhibitory GABAnergic neurons and serotonergic neurons allow rapid information processing, in essence priming cells that leads to changes in membrane potential and more rapid response rates. Co-transmission and utilization of both G-protein coupled as well as ion channel gating enhances response rates. These structural specialisations presumably overcome synaptic delays in information processing due to neurotransmitter release, passage and binding between connected neurons [24], [17], [55].

In Model 1 we use a local inhibitory circuit, shown in Figure 1.

Since an eye fixation takes around 200 msec [29], we assume this represents the minimum time, for which a concept would remain active. The inhibitory circuit requires around 20 msec. It does not matter if input spikes come in as a single volley or as some Poisson process; if the maximum spike rate is around 100 spikes per second, the concept cell can see about 2 spikes in 20 msec, and should it see a spike from every cell, then it takes 40 msec to turn the input cells off. This would represent an spike saving of around a factor of 5.

#### B. Model 2: Prior Knowledge and Intention

There is abundant evidence of the use of Bayesian information processing throughout sensory and cognitive processing [26], [33]. For the purposes of this paper, the implication is that only a small subset of feature detectors need to fire to recognise something, given the assumption that something is going to appear.

Sometimes a single cue, such as hair color, might be enough to distinguish two people. So, if we know that the person coming up the driveway is one of two people who look very similar, then hair colour might be enough to distinguish them. In this case, it is not necessary to wait for all feature cells to fire. Just a few cells may suffice, in which case inhibition can start sooner. This is the essence of Model 2, illustrated in Figure 2. The prior neuron represents the assumption of what will appear: as soon as it has its minimal set of features it activates the output neuron, so, in turn, reducing the input activity early.
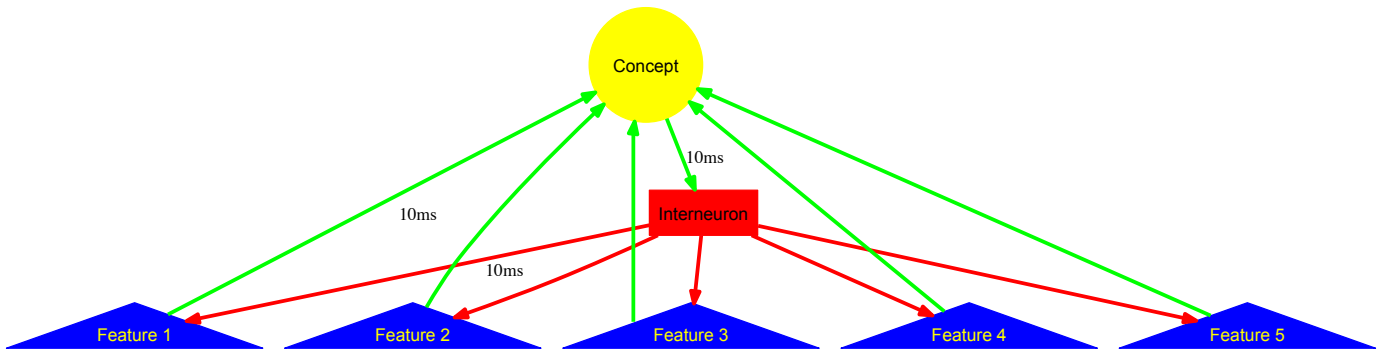
Fig. 1. The basic model. Sensory signals are the features in blue, of which there may be many more than 5. Connections from features to concept and concept to inhibitory interneuron are excitatory. Connections from the interneuron to the features are inhibitory. Parameters are in Table I

<table>
<tr><td colspan="4" align="center">TABLE I<br>PARAMETERS FOR MODEL 1</td></tr>
</table>

| neuron | $\tau$ | $\tau_s$ | R |
|---|---|---|---|
| Feature | 1 | 2 | 20 |
| Concept | 50 | 50 | 20 |
| Inhibitory | 1 | 4 | 20 |

<table>
<tr><td colspan="4" align="center">TABLE II<br>PARAMETERS FOR MODEL 2</td></tr>
</table>

| neuron | $\tau$ | $\tau_s$ | R |
|---|---|---|---|
| Feature | 1 | 2 | 20 |
| Concept | 50 | 50 | 10 |
| Inhibitory | 1 | 2 | 20 |
| Prior | 50 | 50 | 20 |

Now, assume that we have attentional control or a mindset that one is going to see K5 or K7 represented by the cells labelled *prior* in Figure 2. The facilitating cell is activated from higher up, but is agnostic as to whether K5 or K7 appears. It fires slowly with a long recovery time and brings a small subset of features closer to threshold. This costs a small number of spikes and synaptic events, since on average only one facilitating cell will fire. Now, only this small number of features needs to be activated for the concept to trigger. But, since these features lead over the remainder, only they will be allowed to fire.

## IV. METHODOLOGY

We begin by carrying out a theoretical analysis of the possible energy savings in Section V. This makes use of a previously described framework [2] for estimating energy costs and compares the two models described above.

We then carry out computer simulations of both models. Simulations were carried out in Matlab using the Biological Neural Network toolbox [43]. The toolbox uses Matlab's integration routines for solving differential equations, for a variety of single neuron models. Since this is a proof-of-concept simulation, the choice of neuron model is not critical. Thus, the results presented here use the simple LIF model, described above.

Table I gives the parameters for the Model 1 simulation. Model 2 adds an additional neuron, the prior (in a Bayesian sense). Table II shows the parameters used for the simulation.

## V. THEORETICAL ANALYSIS

The most primitive biological model is Attwell and Laughlin's model A [2]. This is a concept-neuron cell model where $x$

neurons sit latent and 1 neuron fires in response to the stimuli, for which it codes. This has an energy expense, $E_A$ given by:

$$E_A = (x + 1)r + a \tag{3}$$

where $r$ is the resting energy usage, $a$ the additional excitatory energy usage. For $N$ concepts there are $N$ neurons required.

Attwell and Laughlin's more complex model (B) is distributed representation, in which a subset of neurons represent each possible stimulus condition, or concept. Thus, we now have to look for the possible combinations, $N_D$, we can form If there are $x$ inactive neurons while sets of $y$ neurons encode a minimum number of conditions. i.e., $x, y$ are integers such that the number of conditions, $N_D$ is given by:

$$N_D = (x + y)!/(x!y!) \tag{4}$$

Its energy cost function, $E_B$, is given by:

$$(x + y)r + ya \tag{5}$$

Our Model 1 (denoted M1) has a concept neuron, an interneuron and a set $n_i$ of feature neurons, the size of this set is $x = |n_i|$. For any concept, some number of feature neurons will fire to represent the concept, say eyes, nose and mouth to indicate a face. In the following analysis we make the simplifying assumption that the number of features is the same across all concepts. The count of the feature neurons $x$ determines the maximum number of conditions that need to be encoded. A single concept neuron requires $m_i \subset n_i$ neurons to fire (i.e., $|m_i| = y, y < x$). Attwell and Laughlin [2] assert that there is a need to encode each concept as sparsely as possible, but now there is an overhead of $+2$ neurons (the concept and
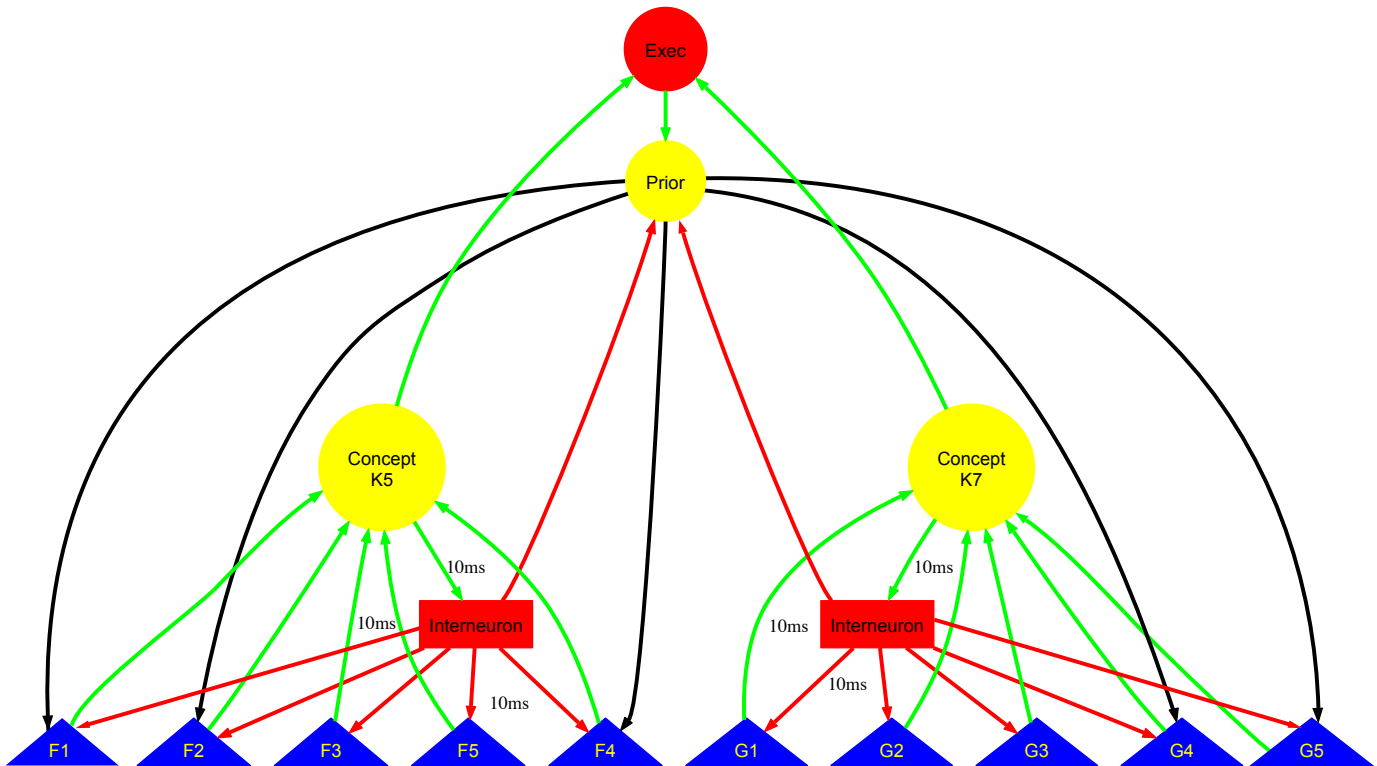
Fig. 2. The model with attention. Features, concepts and inhibitory interneuron are the same as in Model 1. Here we have two concepts and a single attention/prior neuron. This has excitatory connections to a small number of feature detectors.

the interneuron), but these two enforce the minimum sparse coding required.

$$x!/((x-y)!y!) \tag{6}$$

where $y$ is the size of the minimal set of feature neurons necessary to encode a given concept. Its energy expense (Equation (5)) is now:

$$(x+y+2)r + (y+2)a \tag{7}$$

We assume that the energy costs of inhibition are the same as excitation, although they might be slightly less [2]. If $y \gg 2$, i.e., there are many features needed to encode a concept sparsely, then the energy cost approximates Laughlin's Model B.

The Prior Knowledge Model, Model 2 (M2) model has a further overhead of a *prior neuron*. This adds additional neurons to the energistic overhead of M2 over and above Attwell and Laughlin's Model B. But the prior neuron also reduces the number of feature neurons that need to fire in order to activate the concept neuron. If the prior neuron is pre-exciting $z$ feature neurons, and one feature is sufficient to uniquely distinguish between the other competing possibilities (such as hair colour or glasses etc.), then we can have optimal encoding using a prior neuron in the sense that two neurons, one encoding black hair and one encoding blonde hair, can encode the difference in the two concepts, so we

(nearly) recover Attwell and Laughlin's optimal solution for the number of different, possible concepts, $N_c$

$$N_c = x!/((x-2)!2!) \tag{8}$$

(not conditions) for a cost, $E_p$, of

$$E_p = (x+1+4)r + (4+1)a \tag{9}$$

Note that the interneuron will still fire in this model, this increases the fidelity of the signal propagating to the concept neuron by suppressing possible confounding signals (just as it usually does). Although in this paper a single prior neuron is used, the pre-narrowing of a set of conditions, in a Bayesian framework is a well substantiated model of perception [26], [33].

Returning to the example above, where the selected feature is to distinguish between either black or blonde hair, this feature is encoded in two separate feature neurons that connect to two different concept neurons. When the prior pre-excites these two feature neurons they need to provide a voltage potential to the axon hillock sufficient to fire the concept neuron by themselves. If all feature neurons are connected to a concept neuron with the same synaptic strength, then doubling the firing rate of a single feature neuron doubles the energy that neuron expends but only provides the same amount of voltage potential (less some leakage) to the axon hillock as two other feature neurons. If, however, it is a strongly connected

neuron, then doubling its firing rate may equate to adding the equivalent voltage potential of four lesser connected feature neurons.

We can make this a stronger statement. The prior neuron pre-excites a feature neuron by reducing the potential threshold the feature neuron needs to achieve in order to fire. For example, a useful ballpark figure of 100mV is needed to fire an excitatory neuron, a prior neuron can provide the first 80mV in the form an excitatory voltage that essentially lowers the threshold in the feature neuron to only 20mV. This means the feature neuron fires $m = 5$ times more readily than before (assuming a linear integrate and fire profile with limited neural voltage leakage), i.e.,

$$m = \frac{100mV}{(100mV - 80m)}V \qquad (10)$$

If the prior neuron continues to provide this excitatory signal then potentially the feature neuron could fire 5 times more frequently (original frequency times 5), again, assuming linearity and limited leakage.

Energy is saved only if the synaptic weights connecting the feature neuron to the concept neuron are inhomogeneous. To see this assume that the pre-excited feature neuron has a synaptic weight $w_1 = 1$ and there are $n$ other neurons also connected to the concept neuron with weights

$$w_i = \frac{1}{(n+k)} \qquad (11)$$

for some $k$ a free parameter, as discussed below. Assume that all $n + 1$ neurons have approximately the same noisy, leaky integrate and fire profile (weakly leaking and linearly integrating over time). Then each neural pulse received by the concept neuron from the strongly connected and pre-excited feature neuron is stronger than the weakly connected neurons by a factor, $q$, given by:

$$q = \nu_{nat} \frac{100mV}{(100mV - v_{pre})} \frac{w1}{(n+k)} \qquad (12)$$

with $\nu_{nat}$ a natural frequency. and $v_{pre}$ the pre-excited potential. This simply says that a neural pulse received by the concept neuron from the pre-excited feature neuron is both stronger and occurs more frequently than the signal from the weakly connected neurons. As $k \to 0$ the expected firing of the feature neurons that are not pre-excited has an expected aggregate contribution to the firing of the concept neuron that is equivalent to the signal provided by the pre-excited neuron. At the point $k = 0$ it is just as likely that all feature neurons have fired as it is the pre-excited feature neuron fires, indicating the onset of an energy usage transition point where many neurons start firing at $k < 0$ because the aggregate signal from the rest of the neural population begins to fire before the pre-excited neuron. The neurons that are not pre-excited are unlikely to fire when $k > 0$ and there is an energy saving of $na$ non-firing neurons where $a$ is the additional excitatory (firing) energy over and above the resting energy as introduced

earlier. The cost of this is two neurons (the prior neuron and the pre-excited neuron) that fire with an increased rate of:

$$\nu_{nat} \frac{100mV}{100mV - v_{pre}} = \Delta f \qquad (13)$$

for a total cost of

$$E_{total} = 2\Delta f \qquad (14)$$

Note that the term $\frac{100mV}{100mV - v_{pre}}$ is the ratio of the un-excited neural firing frequency to the pre-excited neural firing frequency.

Bringing all of this together we can discuss the energy savings available in an example. Assume that 1,000 feature neurons encode two different concepts, each concept being uniquely composed of 500 features each, each concept has inhibitory interneurons. We consider two examples, one, in which there is a prior, and one, in which there is not:

- **No prior neuron:**
  500 feature neurons contribute a noisy, linear integrate and fire signal to each of the concept neurons. The concept neuron that fires first has its winner takes all signal passed directly to the executive. At the same time the interneurons are activated and inhibit 450 feature neurons from firing. In this case M1 saves $a(450 - 2)$ ATP units per concept over the simpler model of single neurons encoding whole concepts in an excitatory fashion. A total of $2 \times 52a$ ATP are used by excited neurons, the factor of two accounts for the energy usage for both concepts.

- **Prior neuron:** The prior neuron pre-excites two neurons, one for concept neuron A and one for concept neuron B, pre-excitement occurs before any other external information excites the feature neurons. When information does impinge upon the feature neurons, pre-excitement increases the frequency of firing of both feature neurons by a factor of 5, i.e., the firing threshold was 100mV, it is now 20mV for both neurons. This costs $2\Delta fa = 10a$ for both concept neurons, i.e., $20a$ total because pre-excitement instigated firing happens regardless of whether or not the concept neuron fires. The neural savings achieved through this cost is $2a$ ATP , the net savings in energy consumption is $(88+1)a$ ATP (one of the concept neurons does not fire), nearly an order of magnitude better than without a prior neuron.

The above analysis is minimalist, but still requires a few assumptions:

- It requires the connectivity between the non-pre-excited feature neurons to be lower than the inverse of the number neurons. Without this assumption the prior neuron is unable to boost the signal from the pre-excited neuron sufficiently to activate before the rest of the feature neuron population can activate.
- The value of $a$ is arbitrary, but Attwell and Laughli [2] give estimates for rat grey matter. All calculations are relative, relative frequencies and relative energy consumption, as this is a purely comparative analysis. The
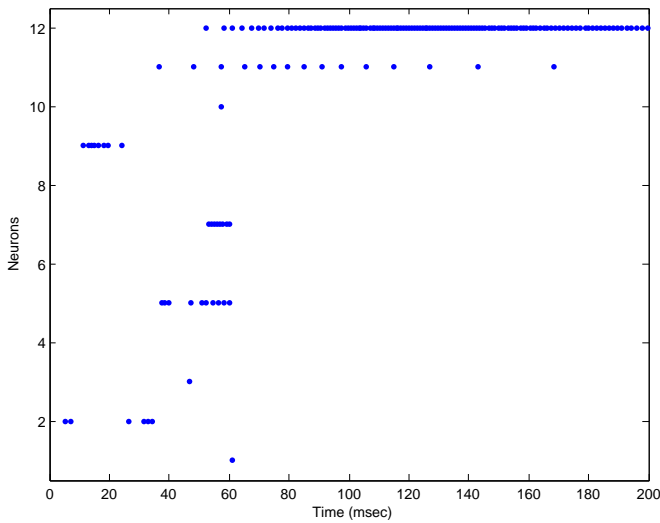
Fig. 3. Activity of the network of model 1. Cells are laid out along the y-axis. The top cell is the inhibitory interneuron, the next cell down is the concept and the remainder are the features. Each dot represents a spike event. The inhibition in neuron 12 sets in after the concept neuron has started to fire.
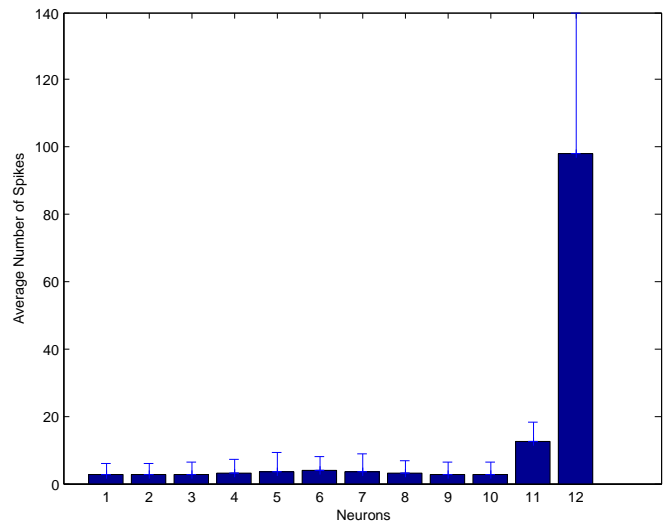


Fig. 4. The average number of spikes for each neuron. Neurons 1–10 are the input features, neuron 11 the concept and neuron 12 the inhibitory neuron.

actual ratio $\frac{r}{a}$ depends on the temporal resolution, which determines the firing rate, $R$. For rat this varies by an order of magnitude from $a = 6.4R$ at a low firing rate of 4Hz to $a = 64R$ at a high firing rate of 40Hz, with an approximately linear increase of energy usage with firing rate.

- There has been no explicit discussion of the synaptic costs, with a focus on action potentials. The costs vary depending upon species, but can be as much as a third of the energy budget [2].

In both M1 and M2, the concept neurons are single cells, but this is not contrary to the efficient distributed model. The inhibitory interneurons may be activated by a collection of cells (a distributed concept). But for simplicity we have confined the model to just one or two concept cells.

## VI. SIMULATION RESULTS

Figure 3 shows the spiking patterns for Model 1. The features are suppressed for the duration of activation of the concept, representing at least a substantial decrease in energy usage. Whereas the activity of the concept and inhibitory neurons are maintained throughout the 200msec simulation, activity of the feature neurons rapidly dies away. Without the inhibition, their firing would also be maintained. Figure 4 shows the average number of spikes in each neuron over 100 runs.

Turning to Model 2, Figure 5 shows the activity with input (from the executive) to the prior neuron only. There is no activity from the feature neurons showing that they do not get over threshold from the prior alone.

Figure 6 shows the effect of adding a small input to the feature detectors. The neurons sensitised by the prior (2,8,10, here, but selected at random in each run) now fire and activate the concept neuron (number 11).
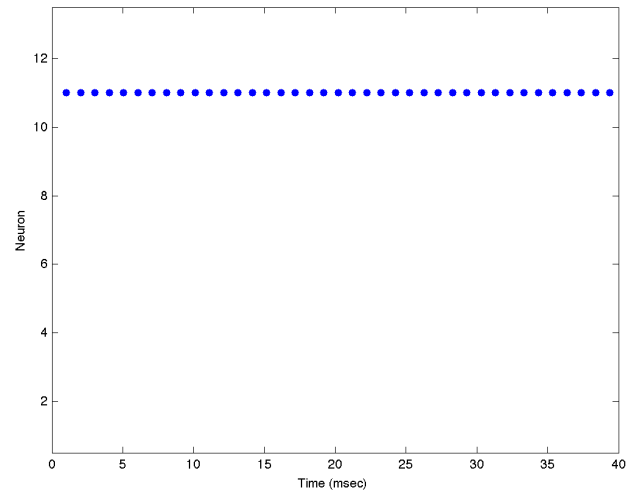


Fig. 5. Input to prior neuron only. The feature neurons do not get over threshold

Figure 7 shows increased input to the feature neurons, which now all get over threshold and fire.

Finally, Figure 8 introduces the inhibitory neuron. The feature neurons and the prior are suppressed. The concept neuron stops firing as the input dries up.

In this article, only one concept neuron appears, but a single prior could in principle pre-activate any number of feature neurons, subserving more than one concept, but biasing the outcome to some subset of possible concepts, which might occur in a given context.

## VII. DISCUSSION

The simulation model presented here demonstrates the feasibility of saving energy through inhibition of lower levels. Since this is a new conjecture to the best of the authors' knowledge,
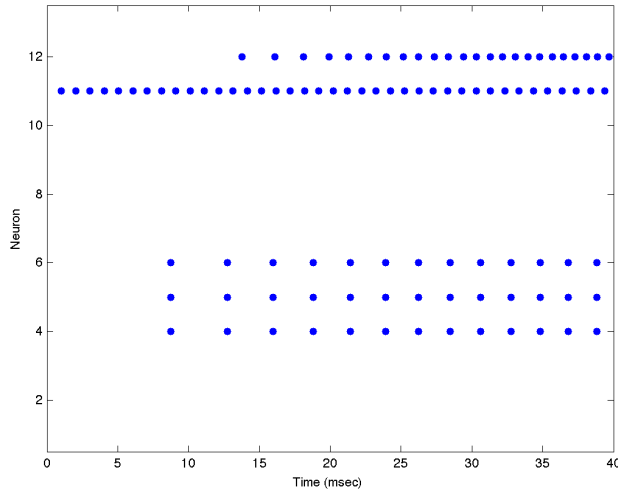
Fig. 6. Simulation with low input to the feature neurons, enough to drive those pre-sensitised by the prior over threshold
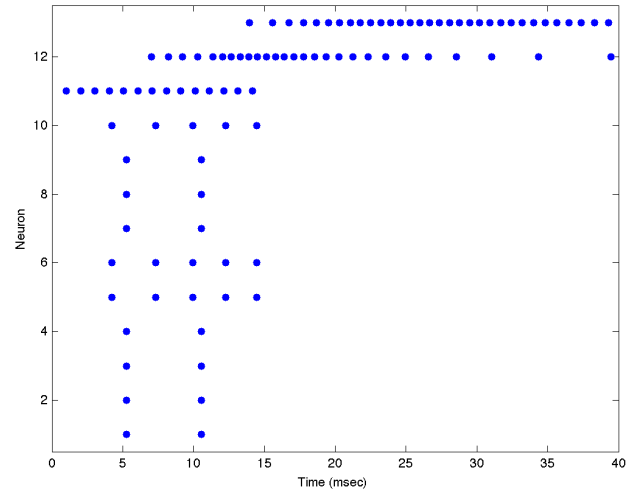


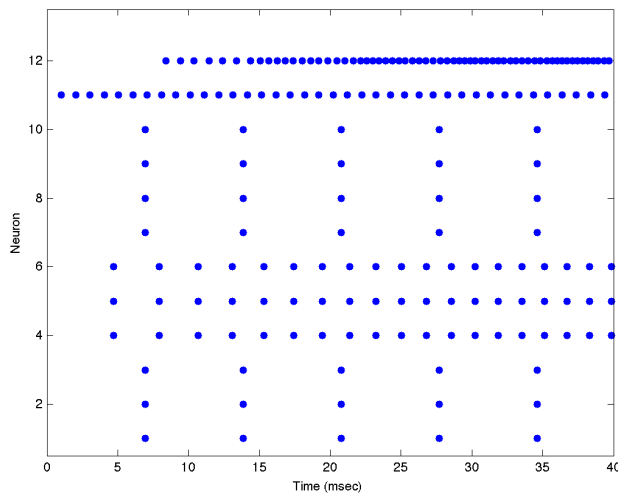Fig. 8. Full Model 2 with prior neuron and inhibition



Fig. 7. High input to feature neurons resulting in all of them getting over threshold. Higher firing rate is seen for the prior-sensitised neurons.

there are no prior results for comparison. The model uses the simplest spiking neuron, on the basis that any more complex model would be able to reproduce the same effects if they exist for the LIF neuron.

However, one limitation of this neuron did become apparent. Ideally the concept neuron should be able to maintain its firing rate until attention switches to something else. However, to maintain firing rate in LIF would require increasing $\tau_s$. This has the consequence of either causing the concept firing to build very slowly, or to allow the firing rate to grow unnecessarily high. The simple LIF does not saturate easily. Further work would look at the more sophisticated Izhikevich models [23].

The conjecture that it is possible to reduce the spikes

generated by a feature might seem surprising. There is, however, substantial work demonstrating that a single spike per neuron may be enough for pattern recognition. Thorpe et al. [54] discovered that people can make *very* rapid decisions on whether pictures contain animals – so rapid that they are likely to be able to use only a single spike along the path from retina to associative cortex. Subsequent computational models demonstrate the feasibility of the single spike model.

The information overload argument suffers from a lack of understanding of what the brain can actually do on a large scale. We know something about the capacity of simple neural networks, such as the number of patterns storable in a Hopfield net or the Vapnik-Chervonenkis Dimension of feedfoward networks. But on the scale of the cortex we have only the most rudimentary of measures.

There has been some interesting progress made in understanding the macroscopic (i.e., psychological) aspects of such limits in real cognitive processes. As early as 1956 Miller [34] proposed on information theoretical grounds that our ability to store and retrieve information is limited. Recent work has extended this idea for complex games [21] showing that *chunks* have a finite information capacity that can be inferred from game data. Chunks can be thought of approximately as the feature neurons of the cognitive models discussed in this work. These chunks have been extensively studied by Simon, Gobet and colleagues [48], [47], [42], [6] in order to explain the rapid, efficient and quite remarkable talents of world leading experts in diverse specialist domains such as chess, nursing and physics.

Beyond these chunks a further mechanism, called *templates* has been suggested as a way to aggregate and contextualise chunks [13], [16], [19], [20]. Using these notions of chunks and templates there is some empirical evidence to suggest that experts may pass through transition points in their development [18] as first suggested on theoretical grounds

by Ericsson [10]. These cognitive results provide high level principles with, which to inform the neuro-cognitive models such as that presented in this work.

It is of no small significance that the prior neuron as described in this work might be identifiable with a *concept neuron* (or a small set of neurons encoding the same or a very similar concept). Recent neurological studies on monkeys and humans [28] have shown that categories of objects, candidates for the instantiation of templates, have a remarkable similarity between species. A theoretical model [46] of feedforward processes for generating and implementing such categories or concepts has been put forward for more general processes such as searching a scene or recognising objects. Such neurological and psychological studies continue to provide the substantive empirical support for the theoretical frameworks such as the one proposed in this work.

Darwin [8] famously remarked: *to suppose that the eye [...] could have been formed by natural selection, seems, I freely confess, absurd in the highest degree.* A century later, Nillson and Pelger [37] showed that evolving an eye was actually relatively easy. By the same token, without a very good model of the computational limits of the brain, the information-overload argument is hard to substantiate.

On the other hand, people are good at blocking out stimuli. The noise of a busy road, the drone of the engines in an aircraft cabin, the buzz of other speakers in a cocktail party – all demonstrate our remarkable capacity to shut out interference when we so desire. But this blocking is reversible and we *can* turn our attention to the distractions themselves. Koechlin [27] shows that the pre-frontal cortex can select one context and block others in choosing an action.

A model that allow specific context retrieval that combines alternate views of top-down guidance by prefrontal cortex or selection of goal-relevant information has been proposed by Badre et al. [3] It proposes that different areas of the prefrontal cortex are specialised for different functions that combine to enhance cognitive processing speed.

The blocking of sensory detail seems to be hardwired and is *not* switchable. To turn off this inhibition would require additional circuits to turn off conceptual information. In general, such circuits do not seem to have evolved, and external techniques such as TMS are required for their inhibition. This would make sense: strategies to save energy would be likely to have evolved much earlier than the expansion of the cortex and its sophisticated filters and control mechanisms.

## VIII. Conclusions

One of the most remarkable findings of the last two decades has been the discovery of the way higher level concepts inhibit low level detail in most individuals, although not in high-functioning autistic savants. The building of conceptual structures on top of the the raw sensory detail is essential for advanced cognition, and is illustrated by the many difficulties experienced in Asbergers and autism, where this conceptual building seems to be impaired [53].

This paper shows that there is an energetic cost to maintaining access to this this raw detail. Thus, the reason we cannot have both a conceptual system and a savant-like raw detail system, might arise from the need to conserve energy.

## References

[1] T. Bossomaier, L. Barnett, V. Thiruvarudchelvan, and H. Jelinek, "Energy saving accounts for the suppression of sensory detail," in *Proc. Fourth International Conference on Advanced Cognitive Technologies and Applications, COGNITIVE12, Nice*, 2012, p. 14–18.

[2] D. Attwell and S. B. Laughlin, "An energy budget for signaling in the grey matter of the brain." *J. Cereb. Blood Flow Metab.*, vol. 21, pp. 1133–1145, 2001.

[3] D. Badre, R. Poldrack, E. Paré-Blagoev, R. Insler, and D. Wagner, "Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex," *Neuron*, vol. 47, pp. 907–918, 2005.

[4] P. S. Boggio, F. Fregni, C. Valasek, S. Ellwood, R. Chi, J. Gallate, A. Pascual-Leone, and A. W. Snyder, "Temporal lobe cortical electrical stimulation during the encoding and retrieval phase reduces false memories," *PLoS One*, vol. 4, no. 3, p. e4959, 2009.

[5] T. Bossomaier and A. W. Snyder, "Absolute pitch accessible to everyone by turning off part of the brain?" *Organised Sound*, vol. 9, pp. 181–189, 2004.

[6] P. Chassy and F. Gobet, "Visual search in ecological and non-ecological displays: Evidence for a non-monotonic effect of complexity on performance," *PLoS one*, vol. 8, no. 1, p. e53420, 2013.

[7] R. P. Chi and A. W. Snyder, "Facilitate insight by non-invasive brain stimulation," *PLoS One*, vol. 6, no. 2, p. e16655, 02 2011.

[8] C. Darwin, *On the Origin of the Species*. John Murray, 1859.

[9] C. Eliasmith, T. Stewart, X. Choo, T. Bekolay, T. DeWolf, C. Tang, and D. Rasmussen, "A large-scale model of the functioning brain," *Science*, vol. 338, pp. 1202–1205, 2012.

[10] K. A. Ericsson and N. Charness, "Expert performance: Its structure and acquisition." *American Psychologist*, vol. 49, no. 8, p. 725, 1994.

[11] K. Gaschler, "One person, one neuron?" *Scientific American*, vol. 17, pp. 77–82, 2006.

[12] C. Gilbert, M. Ito, M. Kapadia, and G. Westheimer, "Interactions between attention, context and learning in primary visual cortex," *Vision Research*, vol. 40, pp. 1217–1226, 2000.

[13] F. Gobet, P. C. Lane, S. Croker, P. C. Cheng, G. Jones, I. Oliver, and J. M. Pine, "Chunking mechanisms in human learning," *Trends in Cognitive Sciences*, vol. 5, no. 6, pp. 236–243, 2001.

[14] F. Gobet, A. Snyder, T. Bossomaier, and M. Harr, "Designing a better brain: Insights from experts and savants," *Frontiers in Psychology*, vol. 5, no. 470, 2014. [Online]. Available: http://www.frontiersin.org/cognition/10.3389/fpsyg.2014.00470/full

[15] P. S. Goldman-Rakic, L. D. Selemon, and M. L. Schwartz, "Dual pathways connecting the dorsolateral prefrontal cortex with the hippocampal formation and parahippocampal cortex in the rhesus monkey," *Neurosci.*, vol. 12, no. 3, pp. 719–743, 1984.

[16] A. Guida, F. Gobet, H. Tardieu, and S. Nicolas, "How chunks, long-term working memory and templates offer a cognitive explanation for neuroimaging data on expertise acquisition: A two-stage framework," *Brain and Cognition*, vol. 79, no. 3, pp. 221–244, 2012.

[17] Z.-S. Han, E. H. Buhl, Z. Lrinczi, and P. Somogyi, "A high degree of spatial selectivity in the axonal and dendritic domains of physiologically identified local-circuit neurons in the dentate gyms of the rat hippocampus." *Eur. J. Neurosci*, vol. 5, pp. 395–410, 1993.

[18] M. Harré, T. Bossomaier, and A. Snyder, "The development of human expertise in a complex environment," *Minds and Machines*, vol. 21, no. 3, pp. 449–464, 2011.

[19] ——, "The perceptual cues that reshape expert reasoning," *Nature Scientific Reports*, vol. 2, 2012.

[20] M. Harré and A. Snyder, "Intuitive expertise and perceptual templates," *Minds and Machines*, pp. 1–16, 2012.

[21] M. Harré, T. Bossomaier, A. Gillett, and A. Snyder, "The aggregate complexity of decisions in the game of go," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 80, no. 4, pp. 555–563, 2011.

[22] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *Journal of Physiology*, vol. 117, pp. 500–544, 1952.

[23] E. Izhikevich, "Which model to use for cortical spiking neurons," *IEEE Trans. Neural Networks*, vol. 15, pp. 1063–1070, 2004.

[24] R. L. Jakab and P. Goldman-Rakic, "Segregation of serotonin 5-HT2A and 5-HT3 receptors in inhibitory circuits of the primate cerebral cortex," *J. Comparative Neurology*, vol. 417, no. 3, pp. 337–348, 2000.

[25] H. F. Jelinek and G. N. Elston, "Dendritic branching of pyramdial cells in the visual cortex of the nocturnal owl monkey: A fractal analysis," *Fractals*, vol. 11, no. 4, pp. 391–396, 2003.

[26] D. Knill and W. Richards, *Perception as Bayesian Inference*. Cambridge University Press, 1996.

[27] E. Koechlin, C. Ody, and F. Kouneiher, "The architecture of cognitive control in the human prefrontal cortex," *Science*, vol. 302, no. 5648, pp. 1181–1185, 2003.

[28] N. Kriegeskorte, M. Mur, D. A. Ruff, R. Kiani, J. Bodurka, H. Esteky, K. Tanaka, and P. A. Bandettini, "Matching categorical object representations in inferior temporal cortex of man and monkey," *Neuron*, vol. 60, no. 6, p. 1126, 2008.

[29] M. Land and B. Tatler, *Looking and Acting: Vision and Eye Movements during Natural Behaviour*. Oxford University Press, 2009.

[30] S. B. Laughlin, R. d. Ruyter van Steveninck, and J. Anderson, "The metabolic cost of neural computation," *Nature Neuroscience*, vol. 1, no. 1, pp. 36–41, 1998.

[31] S. B. Laughlin and T. J. Sejnowski, "Communication in neural networks," *Science*, vol. 301, no. 5641, pp. 1870–1874, 2003.

[32] P. Lennie, "The cost of cortical computation," *Current Biol.*, vol. 13, no. 493–497, 2003.

[33] W. J. Ma, J. M. Beck, P. E. Latham, and A. Pouget, "Bayesian inference with probabilistic population codes," *Nat. Neurosci*, vol. 9, no. 11, pp. 1432–1438, 2006.

[34] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information." *Psychological review*, vol. 63, no. 2, p. 81, 1956.

[35] R. Näätänen, M. Tervaniemi, E. Sussman, P. Paavilainen, and I. Winkler, "Primitive intelligence in the auditory cortex," *Trends in Neurosci*, vol. 24, no. 5, pp. 283–288, 2001.

[36] A. Navarette, C. P. v. Schaik, and K. Isler, "Energetics and the evolution of human brain size," *Nature*, vol. 480, pp. 91–94, 2011.

[37] D.-E. Nillson and C. Pelger, "A pessimistic estimate of the time required for an eye to evolve," *Proc. Royal Soc. Lond. B*, vol. 256, pp. 53–58, 1994.

[38] B. Olshausen and D. Field, "Sparse coding with an overcomplete basis set: A strategy by V1?" *Vision Research*, vol. 37, no. 3, pp. 3311–3325, 1997.

[39] P.-Y. Plaçais and T. Preat, "To favor survival under food shortage, the brain disables costly memory," *Science*, vol. 339, pp. 440–442, 2013.

[40] R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, and L. Fried, "Invariant visual representation by single neurons in the human brain," *Nature*, vol. 435, pp. 1102–1107, 2005.

[41] M. E. Raichle and D. A. Gusnard, "Appraising the brain's energy budget," *PNAS*, vol. 99, no. 16, pp. 10 237–10 239, 2002.

[42] H. B. Richman, F. Gobet, J. J. Staszewski, and H. A. Simon, "Perceptual and memory processes in the acquisition of expert performance: The KP AM Model," in *The Road to Excellence*, K. Ericsson, Ed. Mahwah, NJ: Erlbaum, 1995.

[43] A. Saffari, "Biological Neural Network Toolbox for Matlab." [Online]. Available: http://www.ymer.org/amir/software/biological-neural-networks-toolbox

[44] R. Schäfer, E. Vasilaki, and W. Senn, "Perceptual learning via modification of cortical top-down signals," *PLoS Comput Biol*, vol. 3, no. 8, p. e165, 2007.

[45] B. Sengupta, M. Stemmler, S. B. Laughlin, and J. Niven, "Action potential energy efficiency varies among neuron types in vertebrates and invertebrates," *PLoS Comput Biol*, vol. 6, no. 7, p. e1000840, 07 2010.

[46] T. Serre, A. Oliva, and T. Poggio, "A feedforward architecture accounts for rapid categorization," *Proceedings of the National Academy of Sciences*, vol. 104, no. 15, pp. 6424–6429, 2007.

[47] H. A. Simon, "What is an "explanation" of behavior?" *Psychological Science*, vol. 3, no. 3, pp. 150–161, 1992.

[48] H. A. Simon and W. G. Chase, "Skill in chess: Experiments with chess-playing tasks and computer simulation of skilled performance throw light on some human perceptual and memory processes," *American scientist*, pp. 394–403, 1973.

[49] E. Simoncelli and B. Olshausen, "Natural images statistics and neural representation," *Annual Rev. Neurosci*, vol. 24, pp. 1193–1216, 2001.

[50] A. W. Snyder, H. Bahramali, T. Hawker, and D. Mitchell, "Savant-like numerosity skills revealed in normal people by magnetic pulses," *Perception*, vol. 35, no. 6, pp. 837–845, 2006.

[51] A. W. Snyder, T. Bossomaier, and D. Mitchell, "Concept formation: object attributes dynamically inhibited from conscious awareness," *Journal of Integrative Neuroscience*, vol. 3, pp. 31–46, 2004.

[52] A. W. Snyder and D. J. Mitchell, "Is integer arithmetic fundamental to mental processing?: The mind's secret arithmetic," *Proc. Royal Soc. London B*, vol. 266, pp. 587–592, 1999.

[53] A. W. Snyder, E. Mulcahy, J. L. Taylor, D. J. Mitchell, P. Sachdev, and S. Gandevi, "Savant-like skills exposed in normal people by supressing the left fronto-temporal lonbe," *Journal of Integrative Neuroscience*, vol. 2, no. 2, 2003.

[54] S. Thorpe, A. Delorme, and R. v. Rullen, "Spike-based strategies for rapid processing," *Neural Networks*, vol. 14, pp. 715–725, 2001.

[55] R. Yuste, J. N. MacLean, J. Smith, , and A. Lansner, "The cortex as a central pattern generator," *Nat. Rev. Neurosci*, vol. 6, no. 6, pp. 477–483, 2005.

[56] Y. Zhang, E. Meyers, N. Bichot, T. Serre, T. Poggio, and R. Desimone, "Object decoding with attention in inferior temporal cortex," *PNAS*, vol. 108, pp. 8850–8855, 2011.