Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Neurobiology

What must a global theory of cortex explain?

Leslie G Valiant

At present there is no generally accepted theory of how cognitive phenomena arise from computations in cortex. Further, there is no consensus on how the search for one should be refocussed so as to make it more fruitful. In this short piece we observe that research in computer science over the last several decades has shown that significant computational phenomena need to circumvent significant inherent quantitative impediments, such as of computational complexity. We argue that computational neuroscience has to be informed by the same quantitative concerns for it to succeed. It is conceivable that the brain is the one computation that does not need to circumvent any such obstacles, but if that were the case then quantitatively plausible theories of cortex would now surely abound and be driving experimental investigations.

Address
School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA

Corresponding author: Valiant, Leslie G (valiant@seas.harvard.edu)

Current Opinion in Neurobiology 2014, **25**:15–19

This review comes from a themed issue on **Theoretical and computational neuroscience**

Edited by **Adrienne Fairhall** and **Haim Sompolinsky**

0959-4388/$ – see front matter,

http://dx.doi.org/10.1016/j.conb.2013.10.006

Introduction

That computing is the right framework for understanding the brain became clear to many soon after the discovery of universal computing by Turing [1], who was himself motivated by the question of understanding the scope of human mental activity. McCulloch and Pitts [2] made a first attempt to formalize neural computation, pointing out that their networks were of equivalent expressive power to Turing machines. By the 1950s it was widely recognized that any science of cognition would have to be based on computation.

It would probably come as a shock to the earliest pioneers, were they to return today, that more progress has not been made towards a generally agreed computational theory of cortex. They may have expected, short of such a generally agreed theory, that today there would at least exist a variety of viable competing theories. Understanding cortex is surely among the most important questions ever posed by science. Astonishingly, the question of proposing general theories of cortex and subjecting them to experimental examination is currently not even a mainstream scientific activity.

It is not that the computational perspective was ever abandoned. It was well articulated by David Marr [3], who split the problem into three levels: '*Computational theory*: What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out? *Representation and algorithm*: How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation? *Hardware implementation*: How can the representation and algorithm be realized physically?' This widely quoted passage is, of course, very general and could pass as a mission statement for computer science itself.

Our review here is informed by the observation that since Marr's time computer science has made very substantial progress in certain quantitative directions. The following four phenomena are clearly critical for the brain: communication, computation, learning and evolution. Over the last few decades all four have been subject to quantitative analysis, and are now known to be subject to *hard quantitative constraints* (see [4] for a general exposition). First there is the obvious cost of communication: if we desire to be able to communicate *any n*-bit message we will need to be able to send *n* bits. Second there is computational complexity: if we have some information and can define what processing we wish done on it, that processing may have an unaffordable cost in terms of operations even if we have at hand all the information and can precisely define the desired processing. A third level is learning — even if the desired processing can be achieved by an efficient computation, acquiring a program for it from examples or other behavior presents further impediments. Fourth, if we wish to acquire this program by Darwinian evolution then we encounter even more obstacles.

We do not believe that there can be any doubt that the theory sought has to be computational in the general sense of Turing. The question that arises is: In what way does Marr's articulation of the computational approach fall short? Our answer is that, exactly as in any other domains of computation, a successful theory will have to show additionally, *how the quantitative challenges that need to be faced are solved in cortex*. If these challenges were nonexistent or insignificant then plausible theories would now abound and the only task remaining for us would be to establish which one nature is using.

An augmented computational framework

If, as we believe, cortex is addressing this quartet (computation, learning, evolution and communication) with subtlety, then two additional requirements need to be added to those of Marr for any successful theory. First, *it has to incorporate some understanding of the quantitative constraints that are faced by cortex*. Second, as in other domains of computing, this quantitative understanding has to be articulated in terms of *models of computation appropriate to the problems at hand and the chosen levels of analysis*.

This augmented set of requirements is quite complex in that many issues have to be faced simultaneously. We suggest the following as a streamlined working formulation for the present:

(i) Specify a candidate set of quantitatively challenging cognitive tasks that cortex may be using as the primitives from which it builds cognition. At a minimum, this set has to include the task of memorization, and some additional tasks that *use* the memories created. The task set needs to encompass both the learning and the execution of the capabilities in question.
(ii) Explain how, on a model of computation that faithfully reflects the quantitative resources that cortex has available, instances of these tasks can be realized by explicit algorithms.
(iii) Provide some plausible experimental approach to confirming or falsifying the theory as it applies to cortex.
(iv) Explain how there may be an evolutionary path to the brain having acquired these capabilities.

To illustrate that this complex of requirements can be pursued systematically together we shall briefly describe the framework developed for this by the author [5]. It targets a particular class of tasks called *random access tasks*, to be executed on the *neuroidal* model of computation, using a *positive representation* and a particular style of algorithms called *vicinal algorithms*. Other researchers who have sought to understand cortex have generally not placed quantitative computational goals center stage. We shall make references to some recent examples [6•,7•,8•,9•] in order to contrast some of the currently pursued alternatives.

Positive representations

In order to specify computational tasks in terms of input-output behavior one needs to start with a representation for each task. It is necessary to ensure that for any pair of tasks where the input of one is the output of the other there is a common representation at that interface. Here we shall take the convenient course of having a *common representation* for all the tasks that will be considered, so that their composability will follow.

In a *positive representation* [5] a real world *item* (a concept, event, individual, etc.) is represented by a set S of r neurons. A concept being processed corresponds to the members of S firing in a distinct way. More precisely, as elaborated further in [10], if more than a fraction b (e.g. 88%) of S fire then the concept is definitely being processed, if fewer than fraction a (say 30%) then the concept is not being processed, and the system is so configured that the intermediate situation almost never occurs. We note that for any computational theory with specific algorithms one needs some definition of representation as specific as this.

Positive representations come in two varieties, *disjoint*, which means that the S's of distinct concepts are disjoint, and *shared*, which means that the S's can share neurons. Disjointness makes computation easier but requires small r (such as $r = 50$) if large numbers of concepts are to be represented. The shared representation allows for more concepts to be represented (especially necessary if r is very large, such as several percent of the total number of neurons) but can be expected to make computation, without interference among the task instances, more challenging.

Random access versus local tasks

We believe that cortex is communication bounded in the sense that: (i) each neuron is connected to a minute fraction of all the other neurons, (ii) each individual synapse typically has weak influence, in that a presynaptic action potential will make only a small contribution to the threshold potential needed to be overcome in the post-synaptic cell, and (iii) there is no global addressing mechanism as computers have. We call tasks that potentially require communication between arbitrary memorized concepts *random-access tasks*. Such tasks, for example, an association between an arbitrary pair of concepts, are the most demanding in communication and therefore quantitatively the most challenging for the brain to realize. The arbitrary knowledge structures in the world will have to be mapped, by the execution of a sequence of random access tasks that only change synaptic weights, to the available connections among the neurons that are largely fixed at birth.

We distinguish between two categories of tasks.

Tasks from the first category assign neurons to a new item. We have just one task of this type, which we call *Hierarchical Memorization* and define it as follows: For any stored items A, B, allocate neurons to new item C and make appropriate changes in the circuit so that in future A and B active will cause C to be active also.

The second category of tasks make modifications to the circuits so as to relate in a new way items to which neurons have been already assigned. We consider the following three. *Association*: For any stored items A, B, change the

circuit so that in future when A is active then B will be caused to be also. *Supervised Memorization of Conjunctions*: For stored items A, B, C change the circuits so that in future A and B active will cause C to be active also. *Inductive Learning of Simple Threshold Functions*: for one stored item A learn a criterion in terms of the others. This third operation is the one that achieves generalization, in that appropriate performance even on inputs never before seen is expected.

The intention is that any new item to be stored will be stored in the first instance as a conjunction of items previously memorized (which may be visual, auditory, conceptual, etc.) Once an item has neurons allocated, it becomes an equal citizen with items previously stored in its ability to become a constituent in future actions. These actions can be the creation of further concepts using the hierarchical memorization operation, or establishing relationships among the items stored using one of the operations of the second kind, such as association. The latter operations can be regarded as the workhorses of the cognitive system, building up complex data structures reflecting the relations that exist in the world among the items represented. However, each such operation requires each item it touches to have been allocated in the first instance by a task of the first kind.

Random access tasks are the most appropriate for our study here since, almost by definition, they are the most challenging for any communication bound system. For tasks that require only local communication, such as aspects of low-level vision, viable computational solutions may be more numerous, and quantitative studies may be less helpful in identifying the one nature has chosen.

We emphasize that for the candidate set it is desirable to target from the start a *mixed set* of different task types as here, since such sets are more likely to form a sufficient set of primitives for cognition. Previous approaches have often focused on a single task type [11–13], such as the storage of bit patterns in classical associative memories.

The neuroidal model

Experience in computer science suggests that models of computation need to be chosen carefully to fit the problem at hand. The criterion of success is the ultimate usefulness of the model in illuminating the relevant phenomena. In neuroscience we will, no doubt, ultimately need a variety of models at different levels. The *neuroidal model* is designed to explicate phenomena around the random access tasks we have described, where the constraints are dictated by the gross communication constraints on cortex rather than the detailed computations inside neurons.

The neuroidal model has three main numerical parameters: n, the number of neurons, d the number of connections per neuron, and k, the minimum number of presynaptic neurons needed to cause an action potential in a postsynaptic neuron (in other words the maximum synaptic strength is $1/k$ times the neuron threshold). Each neuron can be in one of a finite number of states and each synapse has some strength. These states and strengths are updated according to purely local rules using computationally weak steps. Each update will be influenced by the firing pattern of the presynaptic neurons according to a function that is symmetric in those inputs. There is a weak timing mechanism that allows the neurons to count time accurately enough so stay synchronized with other neurons for a few steps.

Vicinal algorithms: incremental addition of functionality

In [5,14] it is shown that algorithms for the four random access tasks described above can be performed on the neuroidal model with realistic values of the numerical parameters. The algorithms used are all of the *vicinal* style. Their basic steps are all local in that they only change synaptic strengths between pairs of neurons that are directly connected. Yet they need to achieve the more global objectives of random access. In order that they be able to do this certain graph theoretic connectivity properties are required of the network. The property of expansion [15], that any set of a certain number of neurons have between them substantially more neighbors than their own number, is an archetypal such property. (This property, widely studied in computer science, was apparently first discussed in a neuroscience setting [16].) The vicinal algorithms for the four tasks considered here need some such connectivity properties. In each case random graphs with appropriate realistic parameters have it, but pure randomness is not necessarily essential.

The role of random graphs has been studied in a variety of neural models. Abeles [17], has hypothesized synfire chains for communication between different parts of cortex. In general, random access tasks as we have described them, where communication has to be established between arbitrary sets of neurons, are difficult to support in networks with plausible parameters by synfire chains of any significant depth.

We note that vicinal algorithms allow some modularity of operation, if the changes to the circuit are initiated at the neurons involved with the relevant memories. Alternative approaches, including classical associative memories [11–13,18,19], often have a more global memory mechanism where there is less identifiable modularity in the execution of a task instance.

Incremental lifelong learning

Over a lifetime humans can accomplish large numbers of learning tasks, whether of facts, concepts or skills. Further, they can preserve the functionality of tasks

Table 1

Some features deemed here critical for any general theory of cortical computation

Desirable	Undesirable
Incremental lifelong learning: new item can be learned without retraining on previous ones.	Retraining necessary even when only one item to be added.
Concepts may be defined hierarchically.	Flat memory space.
Arbitrary knowledge structures allowed.	Knowledge structures restricted by architecture.
Mixed set of basic tasks are supported simultaneously.	Only single task supported.
Resources of neuron numbers, synapse numbers, and synaptic strengths are all accounted for quantitatively	Resources assumed free.

learned decades earlier, even after tens of thousands of acts of learning in the intervening years. Expert knowledge, whether of vocabulary or chess openings, has been estimated to comprise hundreds of thousands of facts. Any theory of cortex needs to be able to explain how such numbers of individual acts of learning may be performed in succession without degrading the lasting effectiveness of the earlier ones. This is a quantitative question which we present as a challenge that any theory needs to meet to be considered viable. The study in [10] is the only one we know where large scale learning of such a mixed set of tasks has been shown feasible.

Hippocampus

Hippocampus in mammals is believed to be essential for the acquisition of new episodic or declarative concepts. One would expect that any theory of cortex would explain what indispensable computation hippocampus performs for cortex and the specifics of the interaction between the two. The theory posited in [20] is that for hierarchical memorization hippocampus identifies the set of neurons in cortex at which the new memory is to be located.

Evolution

A theory should not make requirements on cortex for which there is no plausible explanation for how any evolutionary path could have led to it. Theories that apparently require the interconnections to reflect the structure of the knowledge that is represented would appear to require additional explanations. In contrast, theories that explain how a randomly connected set of neurons running simple local algorithms can already compute something useful place a much lighter burden on evolution [10].

Experimental validation

For any theory to be useful there needs to be a path towards testing it experimentally. For each of the three random access tasks, Association, Supervised Memorization of Conjunctions, and Inductive Learning of Simple Threshold Functions, one can in principle test directly whether cortex is capable of it. Each of the task specifications requires for arbitrary sets of neurons such as A, B, C, that a certain training regimen (corresponding to the learning algorithm for the task instance) produces a certain subsequent behavior (e.g. that action potentials in one subset will cause action potentials in another.) Such specifications can be tested in cortex fairly directly. If such arbitrary sets A, B, C are stimulated according to a suitable training regimen, and subsequently when the sets are also recorded from the desired behavior observed, then one has demonstrated that cortex is at least capable of performing the specified task. Ref. [21] describes an existing relevant study.

Conclusion

We believe that neuroscience will go through a stage in which quantitative computational theories and their experimental validation will be pursued in a more integrated way than hitherto. The approach described here illustrates a possible way forward. It emphasizes that the concepts to be memorized and manipulated often have a hierarchical relationship with each other, that explicit means of allocating neurons to new concepts needs to be specified, that multiple kinds of tasks need to be explained within a consistent framework, and that executions of long sequences of such tasks need to be supported (Table 1). The particular tasks considered here, hierarchical memorization together with some further operations on memorized items, are suggested as a minimum basis. How more complex tasks can be built from these, or from an augmented basis set, remain subjects for future investigation.

Acknowledgements

I am grateful to Haim Sompolinsky for his valuable comments on a draft of this paper, and to Hagai Lalazar for pointing out reference [21]. This work was funded in part by National Science Foundation Grant CCF-09-64401.

References and recommended reading

Papers of particular interest describing alternative approaches, published within the period of review, have been highlighted as:

- of special interest

1. Turing AM: **On computable numbers, with an application to the Entscheidungsproblem**. *Proc London Math Soc* 1936, **2**:230-265.
2. McCulloch W, Pitts W: **A logical calculus of the ideas immanent in nervous activity**. *Bull Math Biophys* 1943, **7**:115-133.
3. Marr D: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman; 1982, .
4. Valiant LG: *Probably Approximately Correct*. New York: Basic Books; 2013, .
5. Valiant LG: *Circuits of the Mind*. New York: Oxford University Press; 1994, 2000, .

6. • Friston K: **The history of the future of the Bayesian brain**. *NeuroImage* 2012, **62**:1230.
A brief review of recent efforts to build a theory of cortex starting from Bayesian assumptions.

7. • Zylberberg A, Slezak DF, Roelfsema PR, Dehaene S, Sigman M: **The brain's router: a cortical network model of serial processing in the primate brain**. *PLoS Comput Biol* 2010, **6**:e1000765.
Explores the idea that there is a central router that controls information flow among the various modules in cortex.

8. • Eliasmith C, Choo X, Bekolay T, DeWolf T, Tang Y, Rasmussen D: **A large-scale model of the functioning brain**. *Science* 2012, **338**:1202-1205.
Describes a system that can learn multiple different functions from visual image sequences to physically modeled arm movements. Terrence C. Stewart.

9. • Probst D, Maass W, Markram H, Gewaltig MO: **Liquid computing in a simplified model of cortical layer IV: learning to balance a ball**. *Lecture Notes Comput Sci* 2012, **7552**:209-216.
Liquid computing is a neural network model that treats timed sequences as fundamental in neural computation.

10. Feldman V, Valiant LG: **Experience-induced neural circuits that achieve high capacity**. *Neural Comput* 2009, **21**: 2715-2754.

11. Marr D: **A theory for cerebral neocortex**. *Proc R Soc London B* 1970, **176**:161-234.

12. Graham B, Willshaw D: **Capacity and information efficiency of the associative net**. *Network: Comput Neural Syst* 1997, **8**:35-54.

13. Hopfield JJ: **Neural networks and physical systems with emergent collective computational abilities**. *PNAS* 1982, **79**:2554.

14. Valiant LG: **Memorization and association on a realistic neural model**. *Neural Comput* 2005, **17**:527-555.

15. Hoory S, Linial N, Wigderson A: **Expander graphs and their applications**. *Bull Am Math Soc* 2006, **43**:439-561.

16. Kolmogorov AN, Barzdin YaM: **On the realization of networks in three-dimensional space**. In *Selected Works of Kolmogorov*, vol 3. Edited by Shiryaev AN. Dordrecht: Kluwer Academic Publishers; 1993.

17. Abeles M: *Corticonics: Neural Circuits of the Cerebral Cortex*. New York: Cambridge University Press; 1991, .

18. Edelman G: *Neural Darwinism. The Theory of Neuronal Group Selection*. New York: Basic Books; 1987, .

19. Maass W, Markram H: **On the computational power of recurrent circuits of spiking neurons**. *J Comput Syst Sci* 2004, **69**:593-616.

20. Valiant LG: **The hippocampus as a stable memory allocator for cortex**. *Neural Comput* 2012, **24**:2873-2899.

21. Jackson A, Mavoori J, Fetz EE: **Long-term motor cortex plasticity induced by an electronic neural implant**. *Nature* 2006, **444**:56-60.