The atoms of neural computation

Gary Marcus
Professor of Psychology and Neuroscience
NYU

CEO and Founder
Geometric Intelligence, Inc.

Ontology is the philosophical study of the nature of *being*, *becoming*, *existence*, or *reality*, as well as the basic categories of being and their relations. Traditionally

Neural ontology is, or should be, the interdisciplinary study of the basic neural components of thought.

before digging into neurontology…

a reflection on what if anything AI and ML are teaching us about the brain

We can probably all agree that
AI has yet to live up to its promise

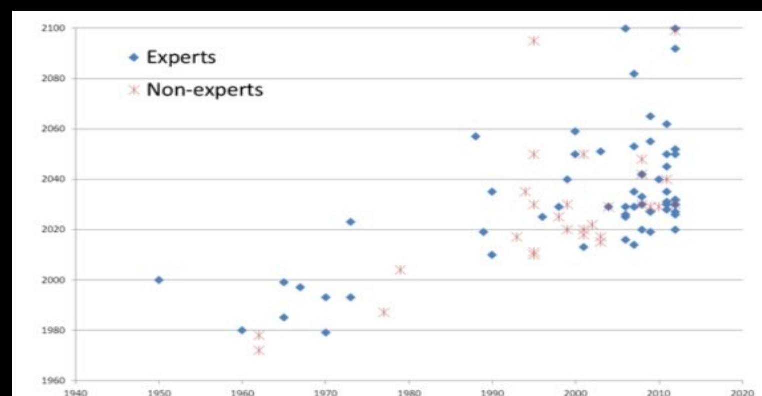
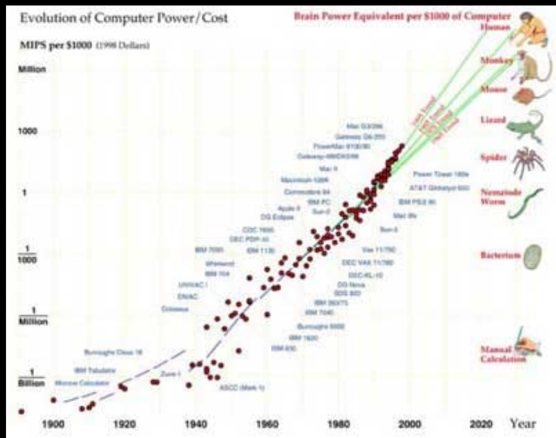


Figure 1: Median estimate for human-level AI, graphed against date of prediction

When will AI come? A metaanalysis of predictions
Stuart Armstrong and Kaj Sotala (2012)

Exponential Growth in Computer Power/$



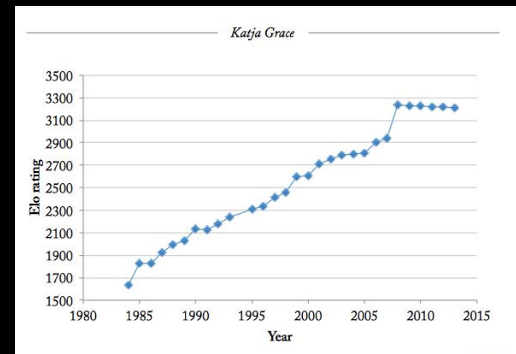
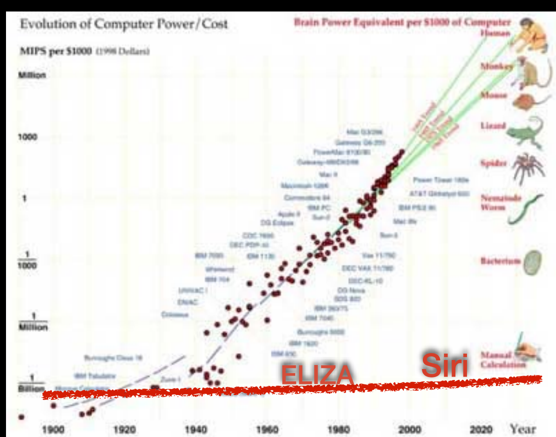


Figure 15: Elo ratings of the best program on SSDF at the end of each year. Data from Wikipedia (2013).

In narrow domains like chess, computers are getting exponentially better

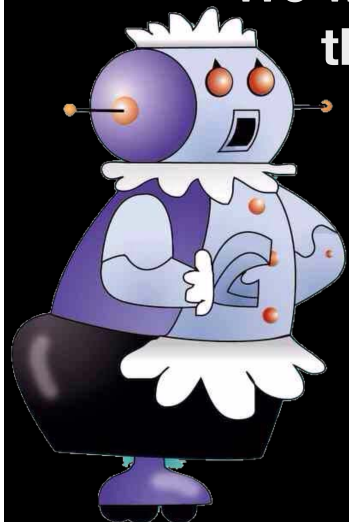


but in strong AI, there has been very little progress

“We wanted flying cars,
instead we got 140 characters”

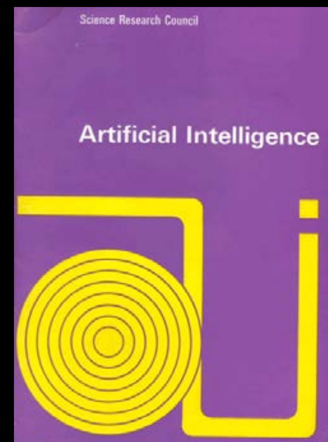
—Peter Thiel

We wanted Rosie the Robot



instead we got Roomba

AI nearly died in 1973



The Lighthill Report

- said AI only worked in narrow domains
- unlikely to scale up
- would have limited applications
- basically led to an end of funding in British AI research
- “The First AI Winter”

Current systems are *still* narrow

- Chess computers (that can’t do anything else)
- Driverless cars (that can’t do anything else)
- Language translators (that can’t answer questions about what they are translating)
- etc.

AI is still basically a collection of idiot savants

A diagnosis

- AI has fallen in love with statistics
- AI has fallen in love with Big Data
- AI has forgotten its roots

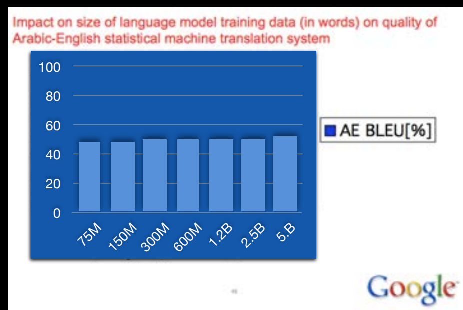
Consider Google Translate

stats + big data = state of the art

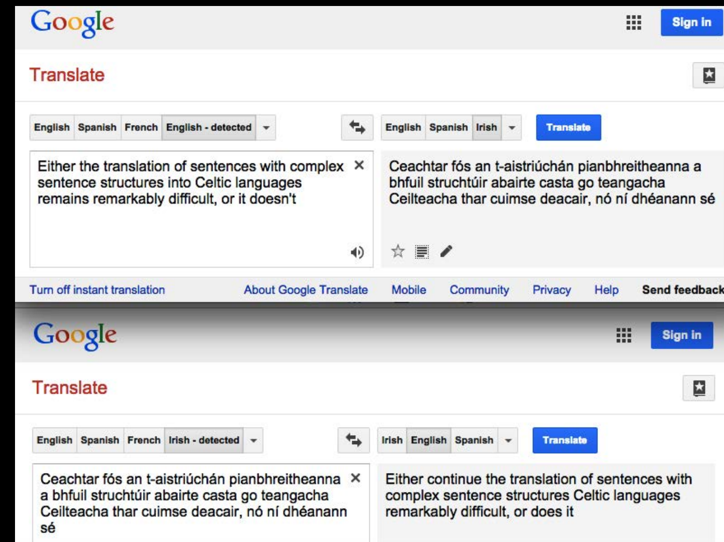
Minimum loss decision rule; MAP decision rule

(General) Minimum loss decision rule:

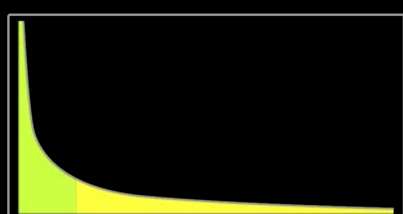
$$\hat{e}(f) = \text{argmin}_e \sum_{e'} L(e, e', f) \cdot Pr(e|f)$$



ten years later, there are still a few bugs in the system



The Long Tail Problem



- Lots of corpus data for a few common examples, little data for less common examples
- Common examples are easy for many systems
- Rare examples are hard

Going beyond the data

“Are we at onety-one?”

AI’s roots

- were partly in studying *natural* intelligence
    - how do brains work?
    - how do minds work?
    - such discussion is mostly absent for current ML work


NEUROSCIENCE

The atoms of neural computation

Does the brain depend on a set of elementary, reusable computations?

By Gary Marcus, Adam Marblestone, Thomas Dean

“...a monetary valuation of nature should be accepted only where it improves environmental [and] socioeconomic conditions...”

one candidate neurocognitive ontology

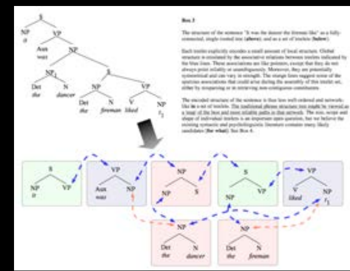
Marcus, 2001, MIT Press

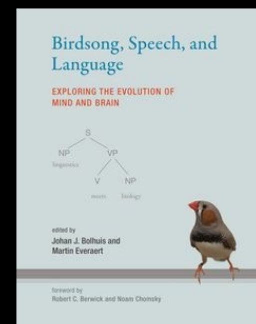
1. ∃ a neurally-realized way of representing **symbols**
2. ∃ a neurally-realized way of representing **variables**
3. ∃ a neurally-realized way of representing **operations over variables**
4. ∃ a neurally-realized way of representing **distinguishing types from tokens**
5. ∃ a neurally-realized way of representing **ordered pairs (AB ≠ BA)**
6. ∃ a neurally-realized way of representing **structured units** (treelet C composed of elements A and B)
7. ~~∃ a neurally-realized way of representing **arbitrary trees**~~

at least one of my 2001 claims was probably wrong

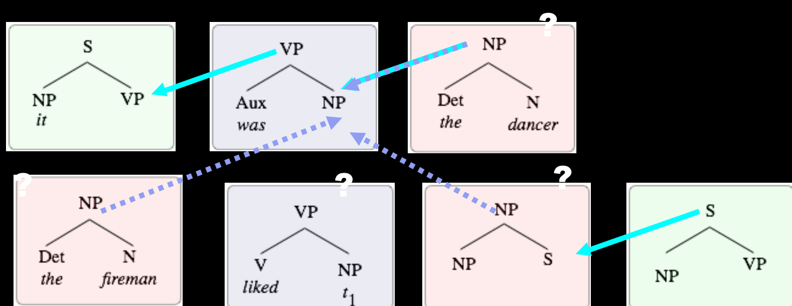
Marcus, 2001, MIT Press

1. ∃ a neurally-realized way of representing **symbols**
2. ∃ a neurally-realized way of representing **variables**
3. ∃ a neurally-realized way of representing **operations over variables**
4. ∃ a neurally-realized way of representing **distinguishing types from tokens**
5. ∃ a neurally-realized way of representing **ordered pairs (AB ≠ BA)**
6. ∃ a neurally-realized way of representing **structured units** (treelet C composed of elements A and B)
7. ~~∃ a neurally-realized way of representing **arbitrary trees**~~





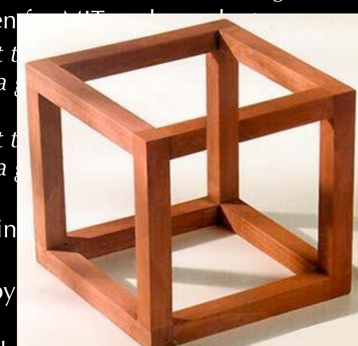
Parsing with treelets



Connections between tree fragments (treelets) are expensive, because the brain is poor at tracking > 4 short-term bindings

Some evidence that people don't have a perfect stack

- We are good at remembering *gist*, poor at remembering verbatim structure (Jarvella, 1971).
- Parsing is vulnerable to interference
  - *It was the dancer that liked the fireman before the argument began*
- Center-embedding is hard, even
  - *The ancient manuscript that t* … *w card catalog had confused a* … *=???*
    The ancient manuscript that t … *w card catalog had confused a* … *rary was missing a page.*
- Elements of discarded parses lin … *by slept…*
- We often get easily confused by … *the cabinet is/are)*
- *Linguistic illusions* (More people … *)*

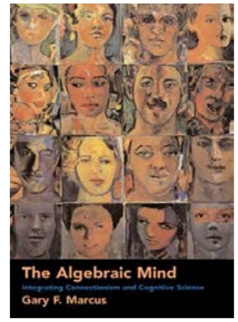


But even though we probably can't veridically represent full trees, I stand by the other claims

especially the one about operations over variables

1. ∃ a neurally-realized way of representing **symbols**
2. ∃ a neurally-realized way of representing **variables**
3. ∃ a neurally-realized way of representing **operations over variables**
4. ∃ a neurally-realized way of representing **distinguishing types from tokens**
5. ∃ a neurally-realized way of representing **ordered pairs (AB ≠ BA)**
6. ∃ a neurally-realized way of representing **structured units** (treelet C composed of elements A and B)
7. ~~∃ a neurally-realized way of representing **arbitrary trees**~~

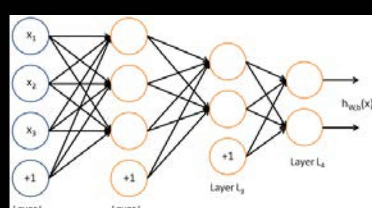
variables and operations



- variables [x, y]
- instances [7, eat, the happy coincidence]
- bindings [x=2, verb stem = *eat*; *np* = *the happy coincidence*]
- operators [+, concatenate]
- hence functions (f(x)=x; s->NP VP

at the opposite extreme: a minimalist neurocognitive ontology

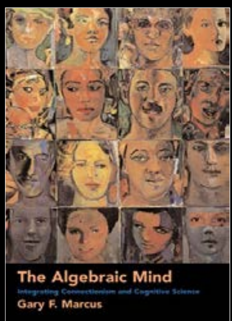
- ∃ nodes
- ∃ connections
- ∃ an activation function
- nodes are grouped into layers and otherwise randomly interconnected
- that's all there is.



these are ultimately *empirical* questions

- Can one can capture the richness of human language and thought from a reduced set of neurocognitive primitives (e.g. the set involved in backprop networks, or LSTM's, or recurrent nets + Hinton_Stacks)?
- Do people behave empirically as if they are symbol-manipulators?

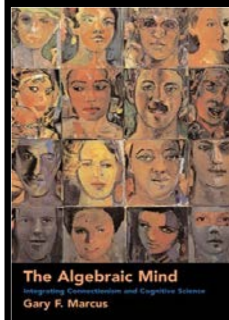
variables and operations



- variables [x, y]
- instances [7, eat, the happy coincidence]
- bindings [x=2, verb stem = *eat; np = the happy coincidence*]
- operators [+, concatenate]
- hence functions (f(x)=x; s->NP VP

(Marcus, 2001, MIT Press)

operations over variables afford open-ended generalization



A rose is a rose
A tulip is a tulip
A lilac is a lilac
_____
A lily is a ____

0110 - > 0110
1100 -> 1100
1010 -> 1010
_____
1111 -> ____

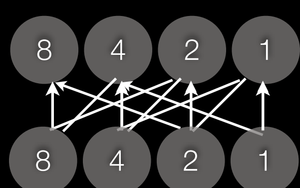
la ta ta
ga na na
_____
wo fe wo
vs *wo fe fe*

Marcus et al, 1999, Science)
w 7-month-olds
later replicated by Gervain et al, 2012
w newborns

(Marcus, 2001, MIT Press)

identity pose problems for backprop nets

0110 - > 0110
1100 -> 1100
1010 -> 1010
_____
1111 -> ____



0110 -> 0110
1100 -> 1100
1010 -> 1010
_____
1111 -> 1110

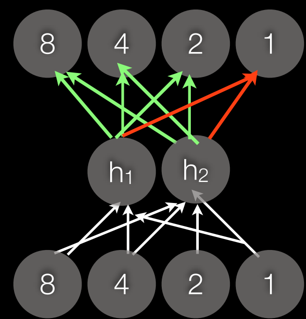
Failure to generalize a universally quantified one-to-one mapping,

(Marcus (1998; 2001 Chapter 3)

Same problem extended to infinitely many functions

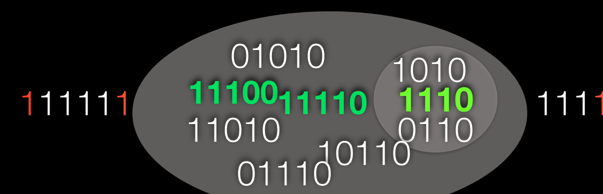
- String reversal: 1110 -> 0111
- Bit inversion: 0000 -> 1111
- Sequence of words over time with repetition (A rose is a __)
- f(x) = 2x, f(x) = 4x, etc
- *Universally quantified one-to-one mappings in general*

Hidden units didn't help



Marcus (1998; 2001 Chapter 3)

Training Space



PDP nets: good at generalizing **within** the space of training examples
poor at generalizing **outside** the space of training examples

other networks: YMMV

Marcus (1998; 2001 Chapter 3)

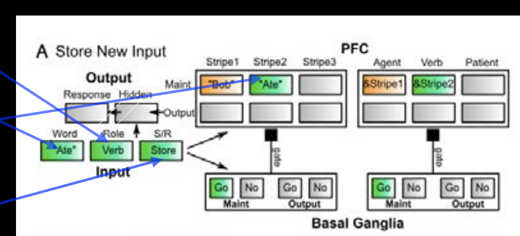
This does *not* mean that you couldn't design a neural network to operate over variables

- It just means that if you want to build a neural network that extrapolates in the right way, you will need something *maps onto* apparatus of symbol-manipulation

- A way of representing variables (**x, y**, **stem**, **noun-phrase**, etc.)
- A way of representing instances (7, *sing*, *the man in the parking lot*)
- A way of representing the instantiation of a given variable (**x** = 7, **stem** = *sing*, etc.)
- A means for performing operations (`add, store, concatenate`, etc)

some neural networks *can* be understood in this way

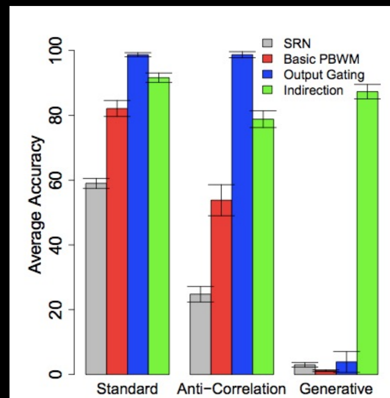
- Variables (**x, y**, **stem**, **noun-phrase**)
- Instances (*7, sing, the man in the parking lot*)
- A way of representing the instantiation of a given variable (**x** = 7, **stem** = sing, etc.)



Kriete, Noelle, Cohen, and O'Reilly's Indirection Network (PNAS 2013), analyzed in terms of the claims of Marcus, 2001

and they do better on tasks of generalization

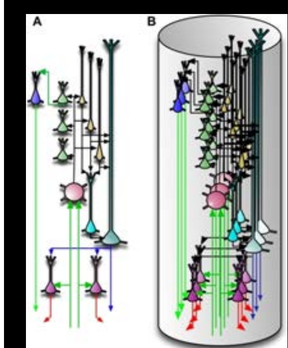
- Kriete et al's results (see right panel, "Generative" task) confirm the key prediction of Marcus (2001, Chapter 3): systems that represent variables, instances, binding, and operations over variables significantly outperform systems that lack such mechanisms.



a classic view: the canonical cortical computation

- "The neocortex .. can be understood as a cooperative network that acts as a nonlinear spatiotemporal filter with adaptive properties (memory) and that transforms afferent signal flow. **It is assumed that these filter properties are identical for all neocortical areas.[the] functional role of a circumscribed cortical area depends exclusively on its position within a certain functional circuit and is defined by it."**
  - Otto Creutzfeldt, 1979

What Many People Are Looking For



"All parts of the neocortex [might] operate based on **a common principle**, with the cortical column being the unit of computation"
- Vernon Mountcastle (1978)

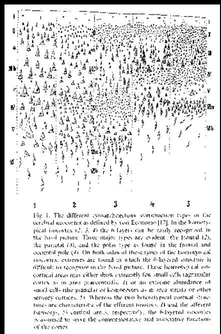
"Functionally heterogeneous cortical areas can be generated by **only a few computational principles**" with "the **variability of the input signals [yielding] functional specialization"**, Wyss et al (2006)

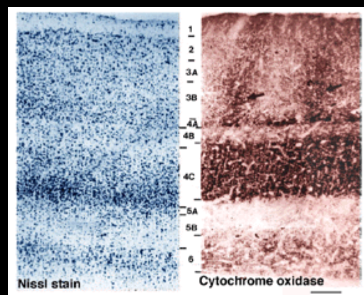
"**The concept of a canonical circuit,** like the concept of hierarchies of processing**, offers a powerful unifying principle** that links structural and functional levels of analysis across species and different areas of cortex."
— Douglas and Martin (2010)

Five Reasons One *Might* Take the Canonical Microcircuit View* Seriously

* some versions of the view focus on common circuitry, others on shared learning rules. for present purposes I will collapse the two.

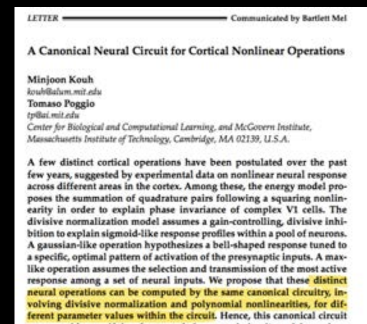
1. the cortex is surprisingly uniform between areas, and across species





adapted from von Economo (1927)

2. Ostensible functional differences can sometimes be obtained through parametric changes in a single underlying architecture

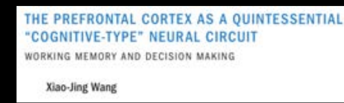

LETTER — Communicated by Bartlett Mel

A Canonical Neural Circuit for Cortical Nonlinear Operations

Minjoon Kouh
kouh@alum.mit.edu
Tomaso Poggio
tp@ai.mit.edu
Center for Biological and Computational Learning, and McGovern Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

A few distinct cortical operations have been postulated over the past few years, suggested by experimental data on nonlinear neural response across different areas in the cortex. Among these, the energy model proposes the summation of quadrature pairs following a squaring nonlinearity in order to explain phase invariance of complex V1 cells. The divisive normalization model assumes a gain-controlling, divisive inhibition to explain sigmoid-like response profiles within a pool of neurons. A gaussian-like operation hypothesizes a bell-shaped response tuned to a specific, optimal pattern of activation of the presynaptic inputs. A max-like operation assumes the selection and transmission of the most active response among a set of neural inputs. We propose that these distinct neural operations can be computed by the same canonical circuitry, involving divisive normalization and polynomial nonlinearities, for different parameter values within the circuit. Hence, this canonical circuit

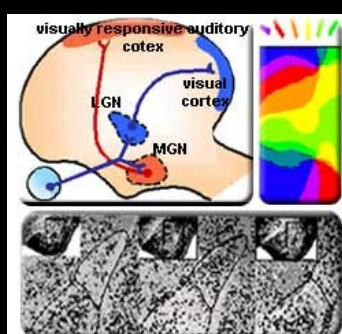
Kouh and Poggio (2008)


THE PREFRONTAL CORTEX AS A QUINTESSENTIAL "COGNITIVE-TYPE" NEURAL CIRCUIT
WORKING MEMORY AND DECISION MAKING
Xiao-Jing Wang

In our model, both working memory and decision making rely on slow reverberatory dynamics that gives rise to persistent activity and time integration, and inhibitory circuitry that leads to selectivity and winner-take-all competition. … At a fundamental level, these studies point to a unified view of why and how "cognitive" cortical area can serve both internal representation (active working memory) and processing (decision, action selection, etc.)

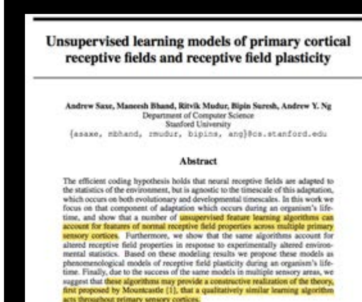
X.-J. Wang (2013)

3. The Apparent Interchangeability of Cortex



Sur et al's studies of rerouting visual input to auditory cortex

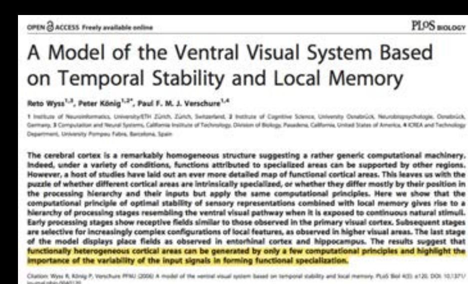
4. The Apparent Success of Hierarchical Feature Detectors and Unsupervised Learning


Unsupervised learning models of primary cortical receptive fields and receptive field plasticity

Andrew Saxe, Maneesh Bhand, Ritvik Mudur, Bipin Suresh, Andrew Y. Ng
Department of Computer Science
Stanford University
{asaxe, mbhand, rmudur, bipins, ang}@cs.stanford.edu

Abstract

The efficient coding hypothesis holds that neural receptive fields are adapted to the statistics of the environment, but is agnostic to the timescale of this adaptation, which occurs on both evolutionary and developmental timescales. In this work we focus on that component of adaptation which occurs during an organism's lifetime, and show that a number of unsupervised feature learning algorithms can account for features of normal receptive field properties across multiple primary sensory cortices. Furthermore, we show that the same algorithms account for altered receptive field properties in response to experimentally altered environmental statistics. Based on these modeling results we propose these models as phenomenological models of receptive field plasticity during an organism's lifetime. Finally, due to the success of the same models in multiple sensory areas, we suggest that these algorithms may provide a constructive realization of the theory, first proposed by Mountcastle [1], that a qualitatively similar learning algorithm acts throughout primary sensory cortices.


OPEN ACCESS Freely available online — PLoS BIOLOGY

A Model of the Ventral Visual System Based on Temporal Stability and Local Memory

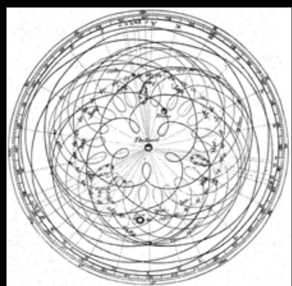
Reto Wyss[1,3], Peter König[1,2*], Paul F. M. J. Verschure[1,4]

The cerebral cortex is a remarkably homogeneous structure suggesting a rather generic computational machinery. Indeed, under a variety of conditions, functions attributed to specialized areas can be supported by other regions. However, a host of studies have laid out an ever more detailed map of functional cortical areas. This leaves us with the puzzle of whether different cortical areas are intrinsically specialized, or whether they differ mostly by their position in the processing hierarchy and their inputs but apply the same computational principles. Here we show that the computational principle of optimal stability of sensory representations combined with local memory gives rise to a hierarchy of processing stages resembling the ventral visual pathway when it is exposed to continuous natural stimuli. Early processing stages show receptive fields similar to those observed in the primary visual cortex. Subsequent stages are selective for increasingly complex configurations of local features, as observed in higher visual areas. The last stage of the model displays place fields as observed in entorhinal cortex and hippocampus. The results suggest that functionally heterogeneous cortical areas can be generated by only a few computational principles and highlight the importance of the variability of the input signals in forming functional specialization.

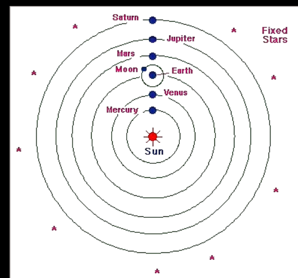
Saxe et al (2011)

Wyss et al (2006)

5. Parsimony



Epicycles



Heliocentric universe

why stipulate a multiplicity of circuits, if one would suffice?

thus the quest to characterize a (single) “canonical microcircuit”

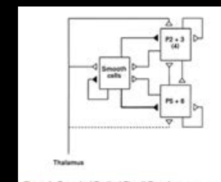


Figure 4. Canonical Cortical Circuit Based on Electrophysiological and Modeling Studies in the Cat Visual Cortex
From Douglas and Martin (1991).

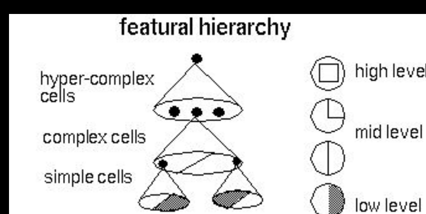
A Canonical Microcircuit for Neocortex

Rodney J. Douglas
Kevan A.C. Martin
David Whitteridge
MRC Anatomical Neuropharmacology Unit, Department of Pharmacology, South Parks Road, Oxford OX1 3QT, England

Douglas et al. (1989) *Neural Computation*

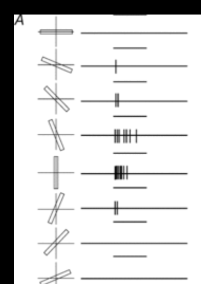
The Canonical Canonical Microcircuit

“the greatest single influence on the ways neuroscientists think about the brain during much of the second half of the twentieth century.”

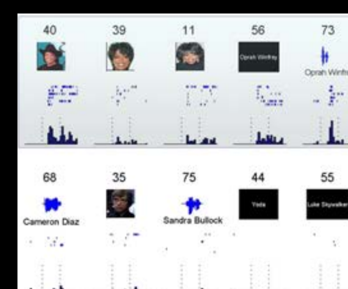


The Hubel-Wiesel tradition

very simple neural ontology: simple and complex cells, arranged in a hierarchy

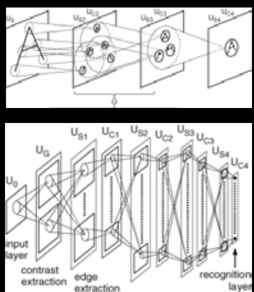


Hubel and Wiesel (1959)

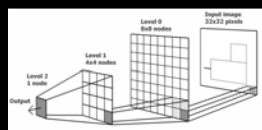


Quian-Quiroga et al. (2009)

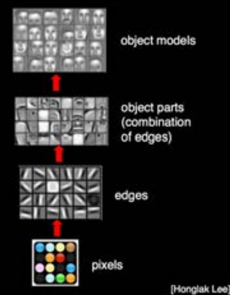
And the impetus behind a wide range of models



Neocognitron
Fukashima (1980)



Hierarchical temporal memory
Hawkins (2004)



Deep learning
e.g., Lee (2012)

Part II. Some reasons to doubt the canonical circuit view

---

1. After 40 year, shere is no satisfactory account of what the canonical circuit might be, if there is one

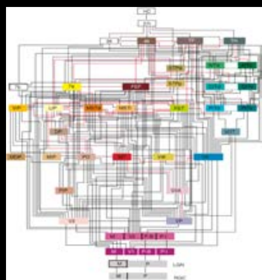
"One simplifying hypothesis that has existed since Cajal is that the neocortex consists of repeated copies of the same fundamental circuit. However, finding that fundamental circuit has proved elusive" - Douglas & Martin (2007)

"I still haven't found what I am looking for" - Bono (1987)

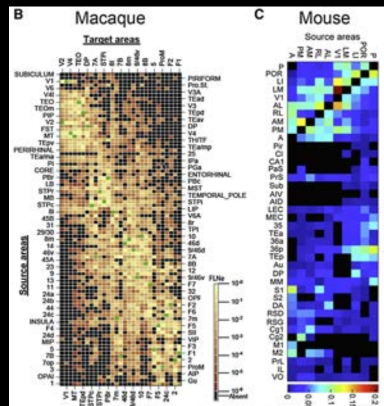
---

2. The canonical circuit view offers no account of why cortical diversity is so pervasive

Complexity in cortical areas and their connectivity

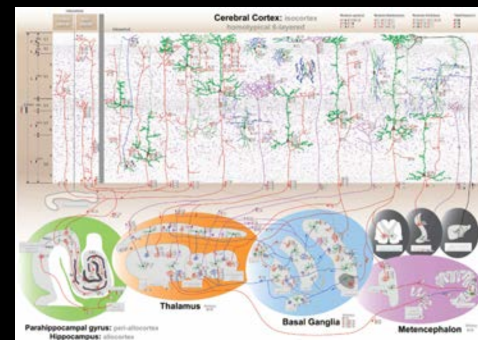


Felleman and van Essen (1991)



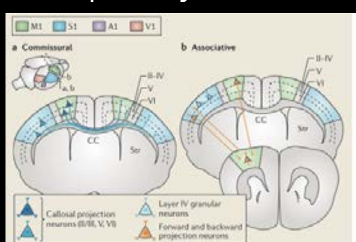
Markov et al. (2012)/adapted in van Essen (2013

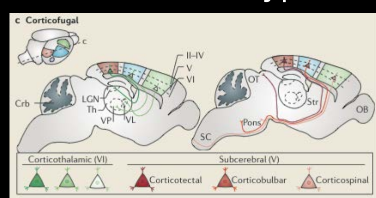
Complexity in how the six layers connect to each other and brain areas

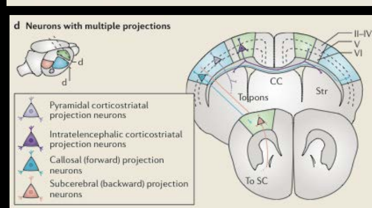


Solari & Stoner (2011)
Cognitive Consilience

Complexity at the level of neuronal subtypes







The division of “The mammalian neocortex … into only six histologically distinct layers belies an extraordinary diversity of neuronal subtypes“ Greig et al. (2013) *Nature Neurosci*

BRAIN STRUCTURE

Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq

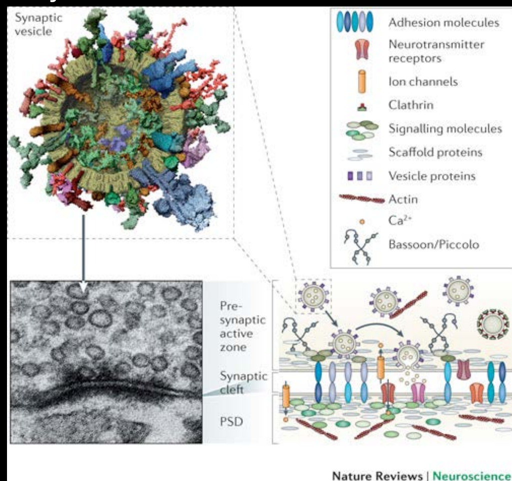
Amit Zeisel,[1*] Ana B. Muñoz-Manchado,[1*] Simone Codeluppi,[1] Peter Lönnerberg,[1] Gioele La Manno,[1] Anna Juréus,[1] Sueli Marques,[1] Hermany Munguba,[1] Liqun He,[2] Christer Betsholtz,[2,3] Charlotte Rolny,[4] Gonçalo Castelo-Branco,[1] Jens Hjerling-Leffler,[1†] Sten Linnarsson[1†]

The mammalian cerebral cortex supports cognitive functions such as sensorimotor integration, memory, and social behaviors. Normal brain function relies on a diverse set of differentiated cell types, including neurons, glia, and vasculature. Here, we have used large-scale single-cell RNA sequencing (RNA-seq) to classify cells in the mouse somatosensory cortex and hippocampal CA1 region. We found 47 molecularly distinct subclasses, comprising all known major cell types in the cortex. We identified numerous marker genes, which allowed alignment with known cell types, morphology, and location. We found a layer I interneuron expressing *Pax6* and a distinct postmitotic oligodendrocyte subclass marked by *Itpr2*. Across the diversity of cortical cell types, transcription factors formed a complex, layered regulatory code, suggesting a mechanism for the maintenance of adult cell type identity.

The brain is built from a large number of specialized cell types, enabling highly refined electrophysiological behavior, as well as fulfilling brain nutrient needs and defense against pathogens. Functional specialization allows fine-tuning of circuit dynamics and decoupling of support functions such as energy supply, waste removal, and immune defense. Cells in the nervous system have historically been classified using location, morphology, target specificity, and

march 5, 2015

Complexity at the level of the individual synapse



"At least 410 different proteins have been identified in synaptic vesicles"
O'Rourke et al. (2012) *Nature Reviews Neuroscience*

"Unfortunately, nature seems unaware of our intellectual need for convenience and unity, and very often takes delight in complication and diversity."

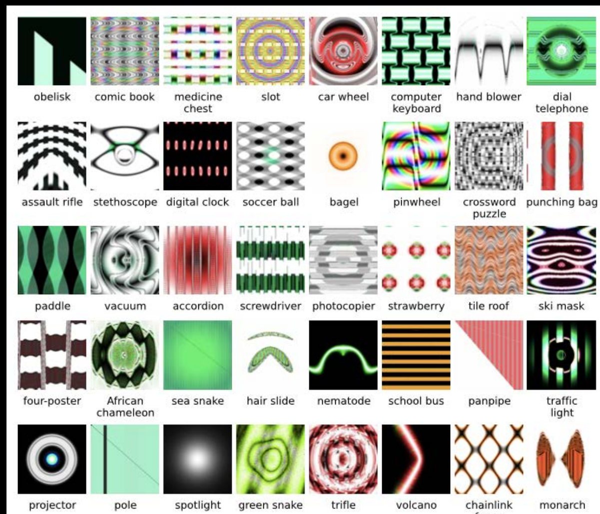
Santiago Ramón y Cajal (1906)

[I have] sometimes heard it said that the nervous system consists of huge numbers of random connections. Although its orderliness is indeed not always obvious, I nevertheless suspect that those who speak of random networks in the nervous system are not constrained by any previous exposure to neuroanatomy.

—David Hubel, *Eye, Brain, and Vision*

---

3. Hierarchies of feature detectors can only get you so far



Nyugen et al 2014

Hierarchies of features

- Not on par with human performance - part of the solution, not the the full solution
- Not wrong, but nor are they sufficient, neither for AI nor for neuroscience

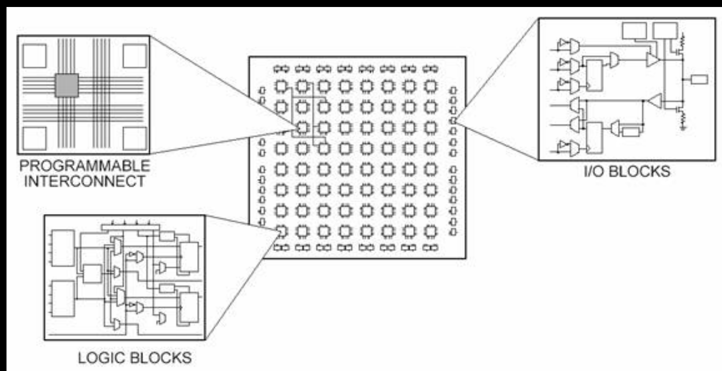
Part III: A conjecture

inspired by digital circuit design

The Conjecture

The cortex consists not of a single repeated element that performs a single computation, but a heteregenous set *of* basic circuit types (possibly evolved from a common origin)

The inspiration



The FPGA - ostensibly uniform at macro level, but precisely configured at the micro level

	Anatomy	Computations	Wiring
Canonical cortical microcircuits tuned by experience	Essentially uniform	Identical, differing only to the extent that they are tuned by different inputs	Tuned by experience
fMRI literature on functional specialization	Often implicitly presumed to be heterogeneous	Often implicitly presumed to be heterogeneous	Not specified
Cortex as an array of reconfigurable computational elements	Largely shared, but with important molecularly-guided fine-tuning for individual blocks	Tinkered variations on theme	Prewired by molecular cues, shaped by experience

Computation	Algorithmic/ representational realization	Neural implementation(s)	Brain location(s)
Rapid perceptual classification	Receptive fields, pooling and local contrast normalization [58,61]	Hierarchies of simple and complex cells[62]	Visual system
Complex spatiotemporal pattern recognition	Bayesian belief propagation [19,63]	Feedforward and feedback pathways in cortical hierarchy[19]	Sensory hierarchies
Learning efficient coding of inputs	Sparse coding[64]	Thresholding and local competition[65]	Sensory and other systems
Working memory	Continuous or discrete attractor states in networks[66,67]	Persistent activity in recurrent networks[68]	Prefrontal cortex
Decision making	Temporal-difference reinforcement learning algorithms[69,70]; actor-critic models[71]	Cortically implemented Bayesian inference networks combined w td reinforcement learning …	Prefrontal cortex
	Winner-take-all networks[73]	Recurrent networks coupled via lateral inhibition[73]	Prefrontal cortex
Gating of information flow	Context-dependent tuning of activity in recurrent network dynamics[74]	Recurrent neural networks implementing line attractors and selection vectors[74]	Prefrontal cortex
	Shifter circuits[64,75]	Divergent excitatory relays and input-selective shunting inhibition in dendrites[76]	Visual system
Gain control	Divisive normalization[35]	Shunting inhibition in networks or balanced background synaptic excitation and inhibition [77]	Common across many cortical areas
Sequencing of events over time [78]	Feed-forward cascades; Serial working memories[79]	Synfire chains[80-82]; Thalamo-cortico-striatal loops[83,84]	Common across many cortical areas
Representation and transformation of variables	Population coding [85]	Time-varying firing rates of cosine-tuned neurons representing dot products with encoding vectors	Motor cortex
Variable binding	Holographic reduced representations [56,86]	Circular convolution of vectors represented by neural population codes	Cortical areas involved in sequential or symbolic processing
	Dynamic binding[87,88]	Neural synchronization[89]	

Some apparently conflicting evidence

Sur and collaborators' "rewiring" experiments

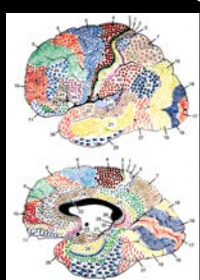
- Visual inputs to primary visual cortex (V1) were rerouted to the primary auditory cortex (A1), which in turn was shown to be capable of processing visual stimuli.

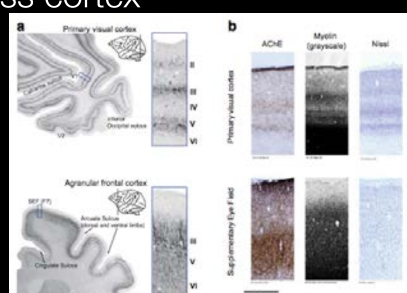
But

- Such results have only been demonstrated between primary sensory cortices
- The "rewired" auditory cortex still retains some of its intrinsic properties, and the resulting "visual" system was not perfect
- In the subsequent two decades, there appears not to have been any successful attempts to reroute visual inputs to other areas that seem more different (e.g., prefrontal cortex)

Some evidence that is consistent with our view

anatomical differences across cortex



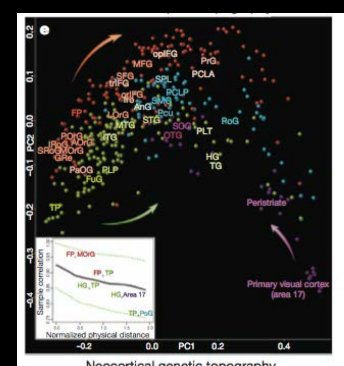


Godlove et al 2014

gene expression differences across cortex


An anatomically comprehensive atlas of the adult human brain transcriptome

"The neocortex displays a relatively homogeneous transcriptional pattern, but with distinct features associated selectively with primary sensorimotor cortices and with enriched frontal lobe expression"

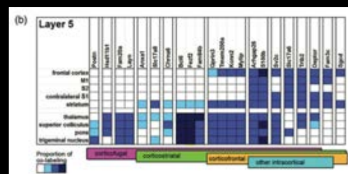


Hawrylycz et al. (2012) *Nature* and Konopka et al (2012)

There exist ways of configuring the microcircuitry of individual blocks in appropriately customized ways

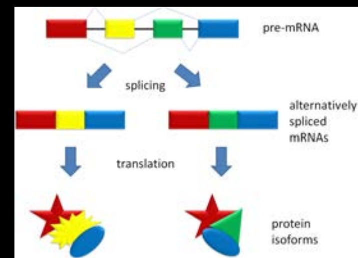
- Combinatorial genetic codes offer one way of specifying fine-grained molecular detail (e.g. Drosophila “stripes”)
- New evidence suggests that Individual neuron types, e.g., mouse S1 layer 5 projection neurons, can have molecularly-defined subclasses that project to different destinations (Sorenson et al, 2013)

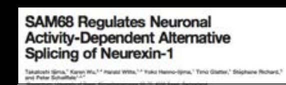

Correlated Gene Expression and Target Specificity Demonstrate Excitatory Projection Neuron Diversity



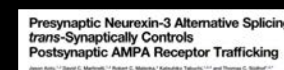
Combinatorial logic of S5 L5 projection neurons
Sorenson et al 2013, *Cerebral Cortex*

Ways of configuring the microcircuitry of individual blocks in appropriately customized ways (2)




SAM68 Regulates Neuronal Activity-Dependent Alternative Splicing of Neurexin-1

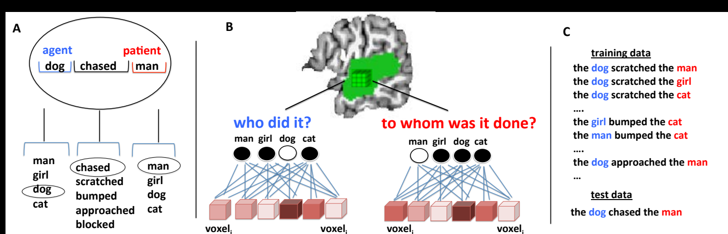
Ijima et al. (2011) *Cell*

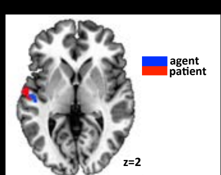

Presynaptic Neurexin-3 Alternative Splicing *trans*-Synaptically Controls Postsynaptic AMPA Receptor Trafficking

Aoto et al. (2013) Cell

Alternative splicing offers another mechanism, by which small molecular differences could lead to critical synaptic differences

Frankland and Greene (PNAS)





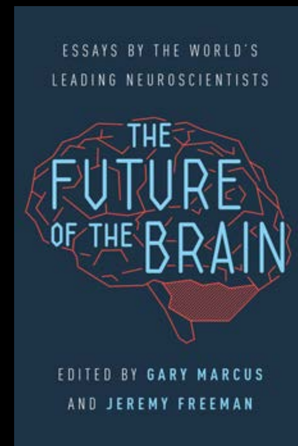
- “dog bites man” vs “man bites dog”
- MVPA in fMRI points to reliably separate “registers” for agent vs patient, consistent across subjects
- Conclusion “[this result] supports an intriguing possibility (Marcus, 2001): that the explicit representation of abstract semantic variables in distinct neural circuits plays a critical role in enabling human brains to compose complex ideas out of simpler ones.”

Some parallel questions, for biologically-inspired AI

- Is better AI really about the quantity and quality of data? Or the nature of the representations we extract from the data?
- Why is there so much diversity in the brain? What does it tell us about the underlying algorithms?
- How much do we need to enrich our computational ontology? MLPs vs LSTMS; stacks; ; FOPC? Are their probabilistic programs among our neurons, and if so how are they realized?
- Is memory best understand as sets of vectors? Are their useful higher level constructs, eg akin to data structures in which binary bits are organized (jpgs, mp3s, linked lists, etc)?
- How can we *bridge* between the language of nodes to the language of propositions, trees, and abstractions?

Summary

- Comparatively little attention has been paid in computational neuroscience to models that incorporate *a rich set of basic computational circuit types* (as opposed to only one or a few elementary operations)
- Yet such architectures are a natural choice - given our knowledge of the brain's development and function
- The conjecture provides a conceptual framework for *bridging* between neural structures and computational function
- New tools mean that conjectures like ours may soon be testable
- But this is early days for our approach — and we would love help!!!



featuring
Christof Koch
George Church
May-Britt and Edvard Moser
and more

gary.marcus@nyu.edu