
welcome
to Lecture 11
Computation and the Brain

What happened last Wednesday

Remarkable lecture by Larry Abbott

- He grounded for us the Churchland et al paper on recent results with monkeys learning a circular hand motion
- The motor cortex dynamical system is cyclic but not in the geometric sense
- It is the necessary “disentangling cycle,” always clockwise
- The actual geometric cycles must exist in the spinal

Reinforcement Learning

Classical conditioning (1960s): The delta rule

Stimulus u, prediction x, weights w

(w, u possibly vectors)

Reward R

δ

$x = u \cdot w$

Rescola-Wagner plasticity: $\mathbf{w \rightarrow w + \varepsilon \cdot (R - x) \cdot w}$

Note that this is gradient descent with error $2(R - x)2$

But: does not model foresight

Animals choose actions
by looking beyond
the present reward…

It works! [Tesauro 1992] backgammon

before there was Alpha Go, there was TD-Gammon

Reinforcement Learning in the brain

Spot:

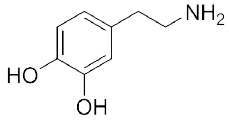
The δ calculator

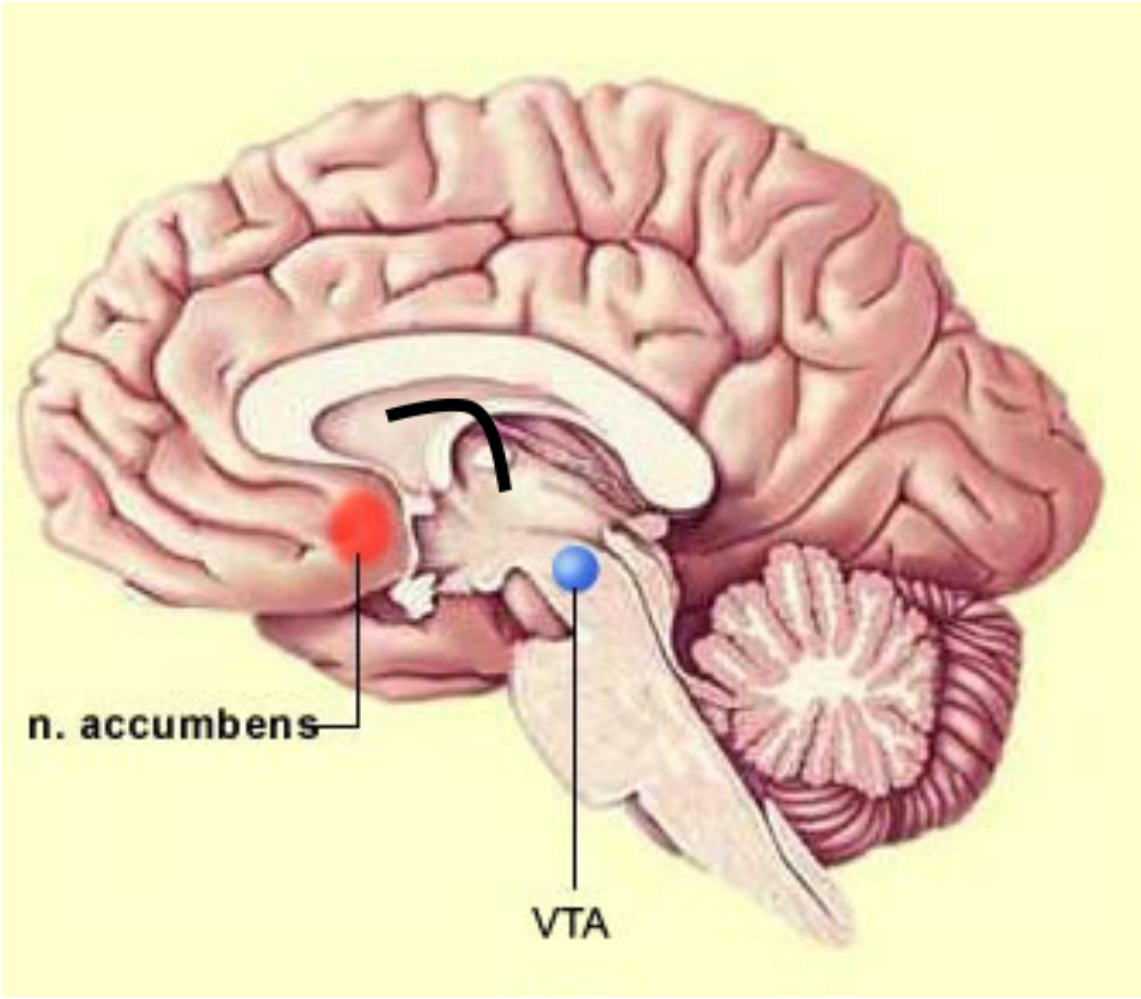
The weights w of the stimuli

The reward delivery system

The reward circuit

Btw: the reward is

dopamine 



Important model in RL: multi-armed bandits

m actions a = 1…m

Each has an unknown reward distribution Da

You are stuck at the casino for a very long time T

How would you play?

Exploration vs. exploitation

- One machine has the best expectation, M*
- Best thing to do is play this forever – but you don't know which…
- So, at time t you choose A(t)
- Every machine i has a (unknown) gap Ga = M* - Ma
- You want to minimize regret: Regret(A) = Σt GA(t)

Exploration vs. exploitation: is linear regret inevitable?

- Lower bound: regret is at least log T · (Σa Ga/KL(a, a*))

UCB: Estimate the expectation of every action as

current mean + √(log t/sample size)

- **Theorem**: UCB1 has regret at most 10 log T· (Σa Ga)
- **Theorem:** Thompson's algorithm (parametric representation of the Di's) achieves lower bound

More advanced model: time-dependent bandits

Suppose now the machines are known Markov chains with a reward at each transition

We know the current states

Discounted rewards

Gittins index of a Markov chain

- Gittins index of M: the smallest reward for which, if you had a choice of any number of rounds with M and then R forever, you would not touch M
- **Gittins Theorem:** The optimum strategy is always play the machine with the highest Gittins index

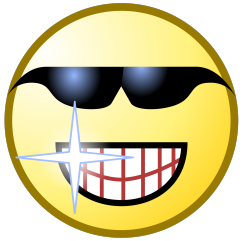
Markov decision processes (MDPs)

- Gittins allows you to switch Markov chains
- In MDPs the Markov chains all share the same states
- You have a choice of action at each state
- Your choice changes the transition probabilities and the rewards
- Strategy A: You choose, once and for all, what to do at each state, and then follow the resulting Markov chain
- What is the best strategy?

Value of a state: Bellman's equation

V[state] = maxA [R(state,A) + γ·EA [V[next state])

Can be solved three different ways as LP



Problem is, Chess has 1050 states, Go has 10170 and an automated driver may have more…

(Also, rewards are infrequent…)

Deep reinforcement learning

- Policy A is a parametrized function of the state
- Action at state s is $A(s,\theta)$
- $J(\theta)$ = the expected reward of $A(s,\theta)$
- Maximize by stochastic gradient ascent: calculate $\nabla\theta$ by sampling paths of the Markov chain (roll-outs)

Today: language

- Talk by Dan Mitropolsky!
- Talk by Mike Collins!
- In the remaining time:
- Some history and math of language
- Language and the Brain
- (incl. auditory cortex)

Next weeks

- Evolution and Development
- Discussion of Consciousness.

- Preliminary project presentations (please volunteer)

How did language come about?

- Some 3 MYBP the homo group separated from the chimps
- Habilis, floresiensis, neanderthalensis, heidelbergensis, denisovan, erectus, ergaster, sapiens
- Only sapiens seems to have had language
- Evidence: lack of trappings of symbolic behavior such as figurative art
- 80 KYBP: first figurative art in Africa
- Did language come about the same time?

How language came about: the cons

- 1866: French Academy bans discussion of the origin of language
- “It is an irresistible question because it is about us. But this does not make it a scientific question. If there is no way to find out by science, then there is no way.” Noam Chomsky
- "If you don't have a related species with a similar trait you have the problem of novelty.'' Richard Lewontin
- Is human fascination with language an exercise in specism?
- Pinker: imagine elephants admiring their trunk…

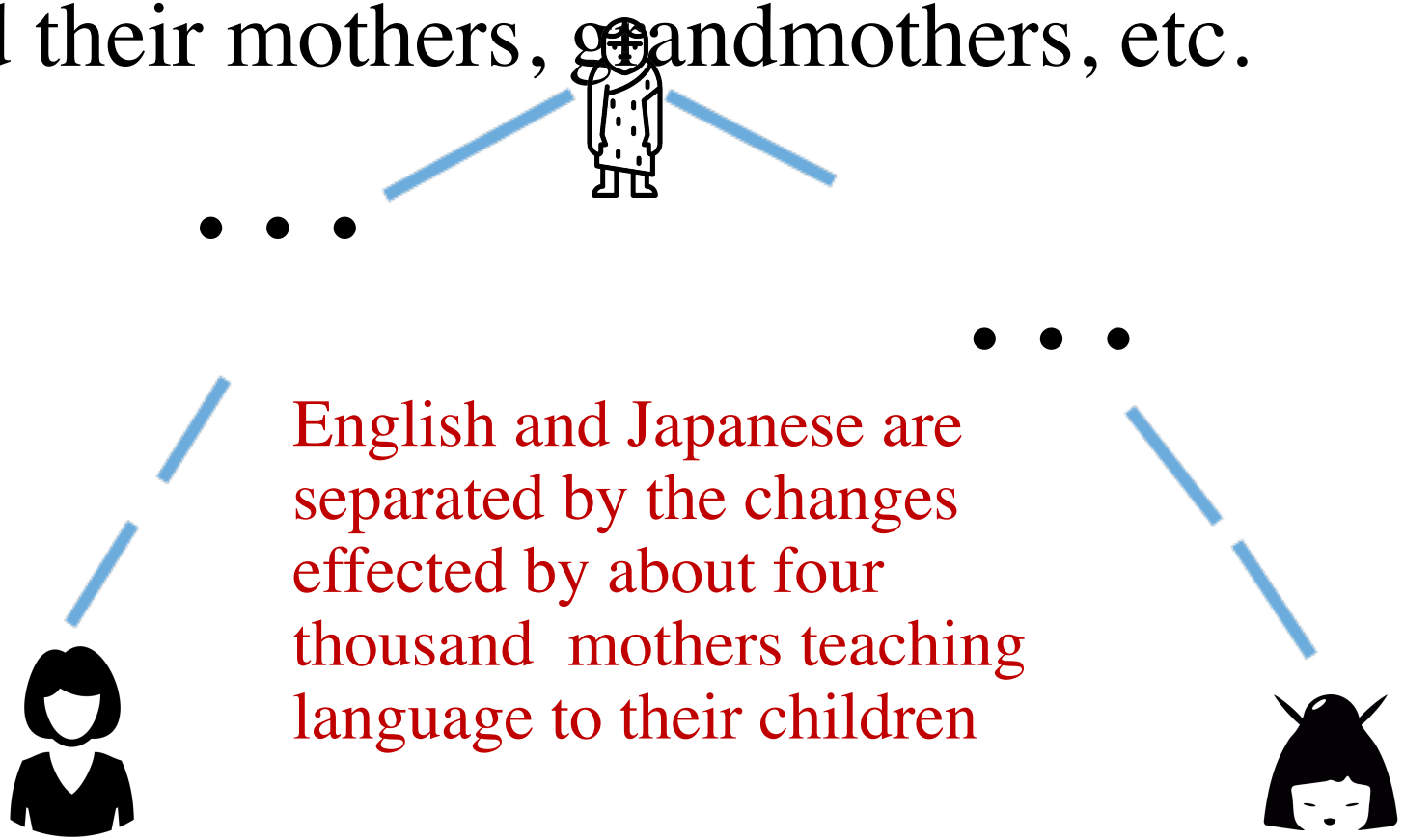
How language came about: Speculation

- Was there a cognitive/neural Big Bang?
- Or gradual progress?
- Perhaps we had language (in some weak sense) far before 80 KYBP
- Maybe around 500 KYBP
- Homo sapiens
- Perhaps we gestured for a very long time
- C Corballis “The gestural origin of language” Wiley 2010

What is needed for language?

- Is a prohibitively sophisticated muscular system in the mouth/face/tongue necessary?
- (Or is it all in the Mind?)
- When did we become capable of competent speech?
- Biologists seem to think that many apes are ready…
- MC Corballis thinks that we may have switched to speech long after we could have…
- Why was language of thought an evolutionary advantage?

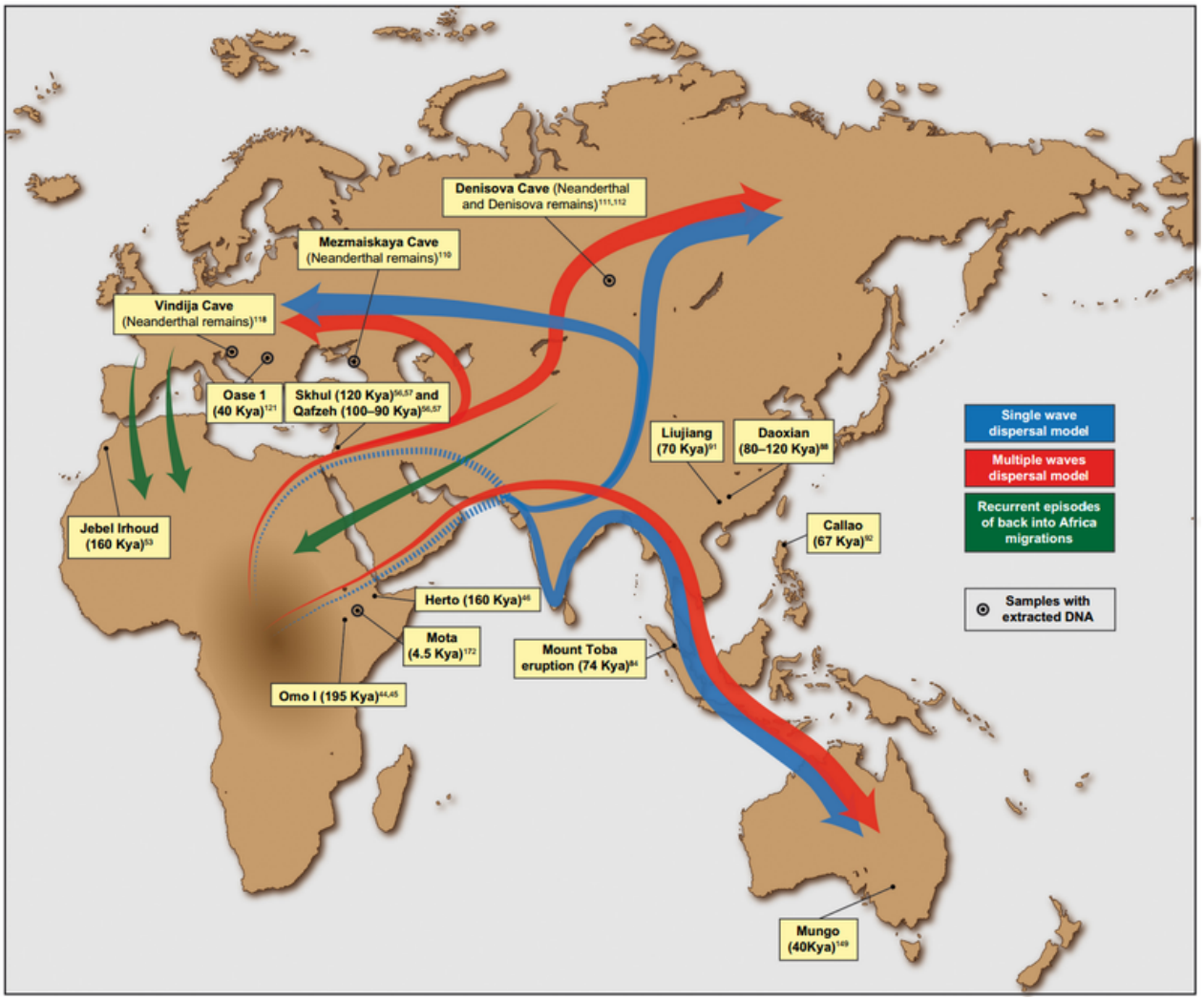
A tale of two island girls
and their mothers, grandmothers, etc.



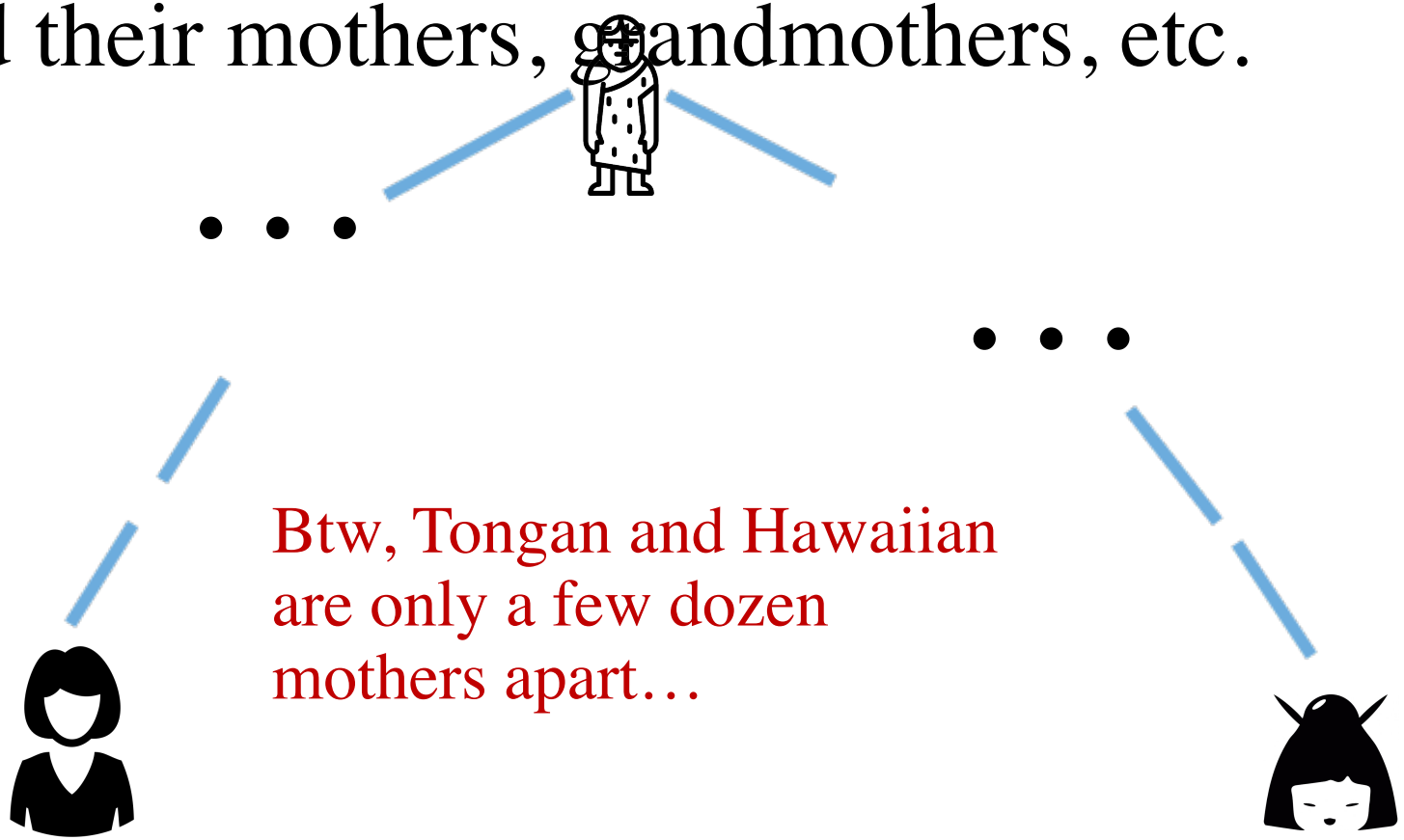
English and Japanese are separated by the changes effected by about four thousand mothers teaching language to their children

Out of Africa

- 80 - 70 KYBP
- Europe: 40 KYBP
- Americas: 20 KYBP
- Arctic: 10 KYBP
- Polynesia: 2 KYBP
- **Single migration?**
- **Or multiple?**



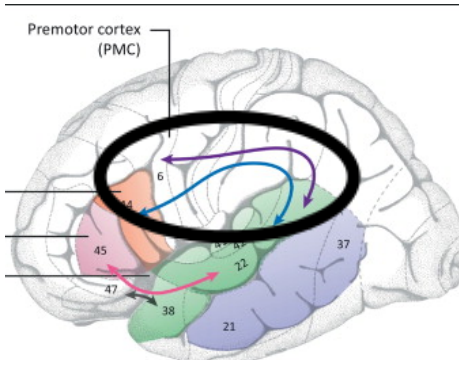
A tale of two island girls
and their mothers, grandmothers, etc.

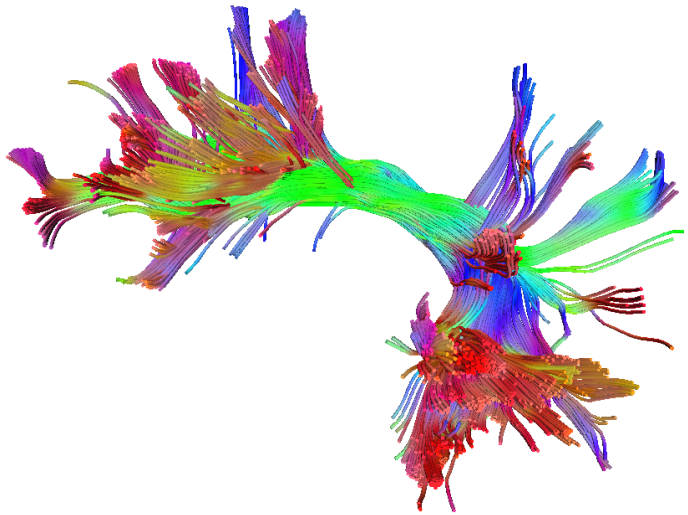


Btw, Tongan and Hawaiian
are only a few dozen
mothers apart…

Was there a neural Bing Bang?

- Our left and right hemispheres are anatomically asymmetric
- Seems to be specific to humans
- Main locus of asymmetry: the ***arcuatum fasciculum***

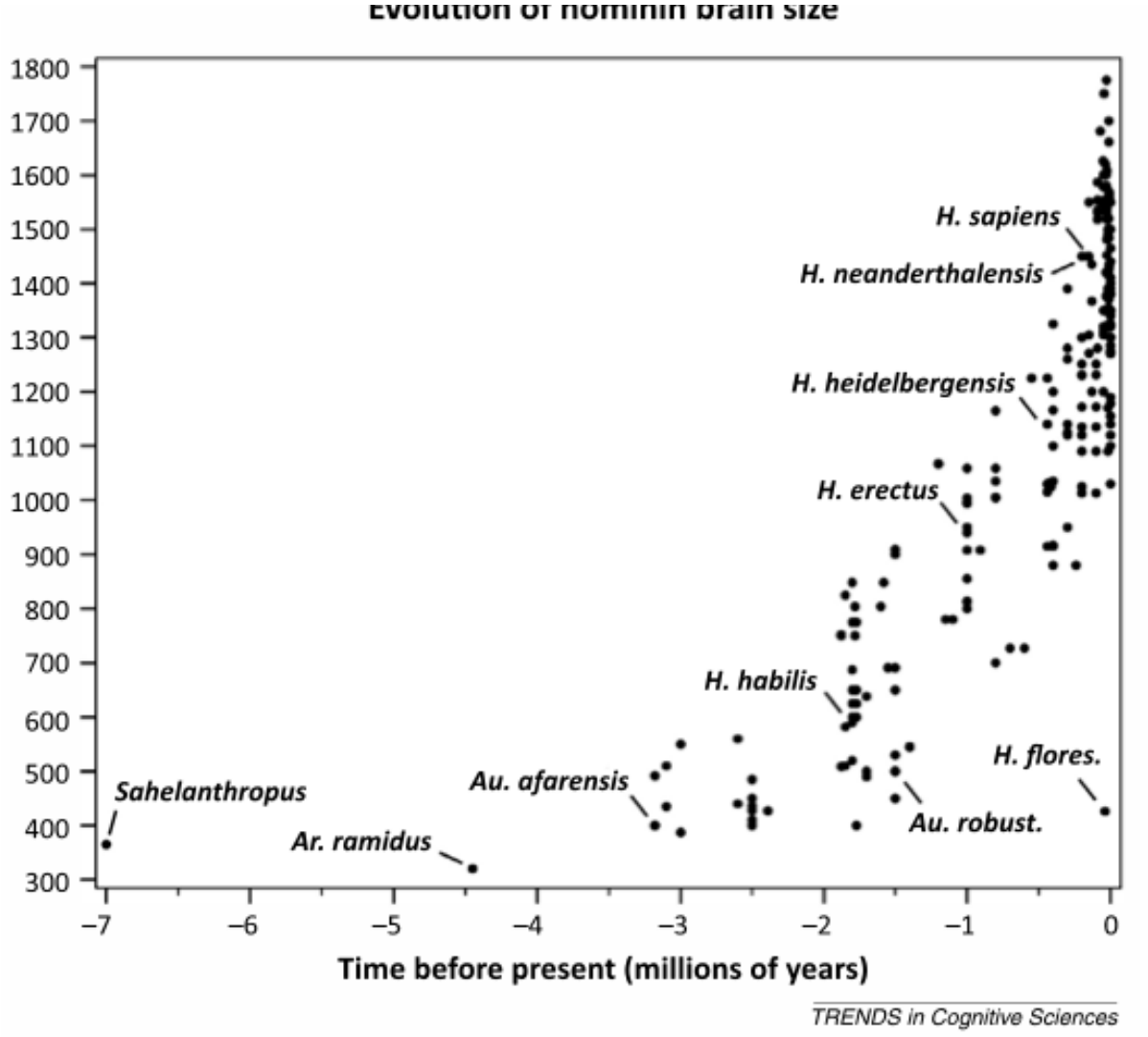




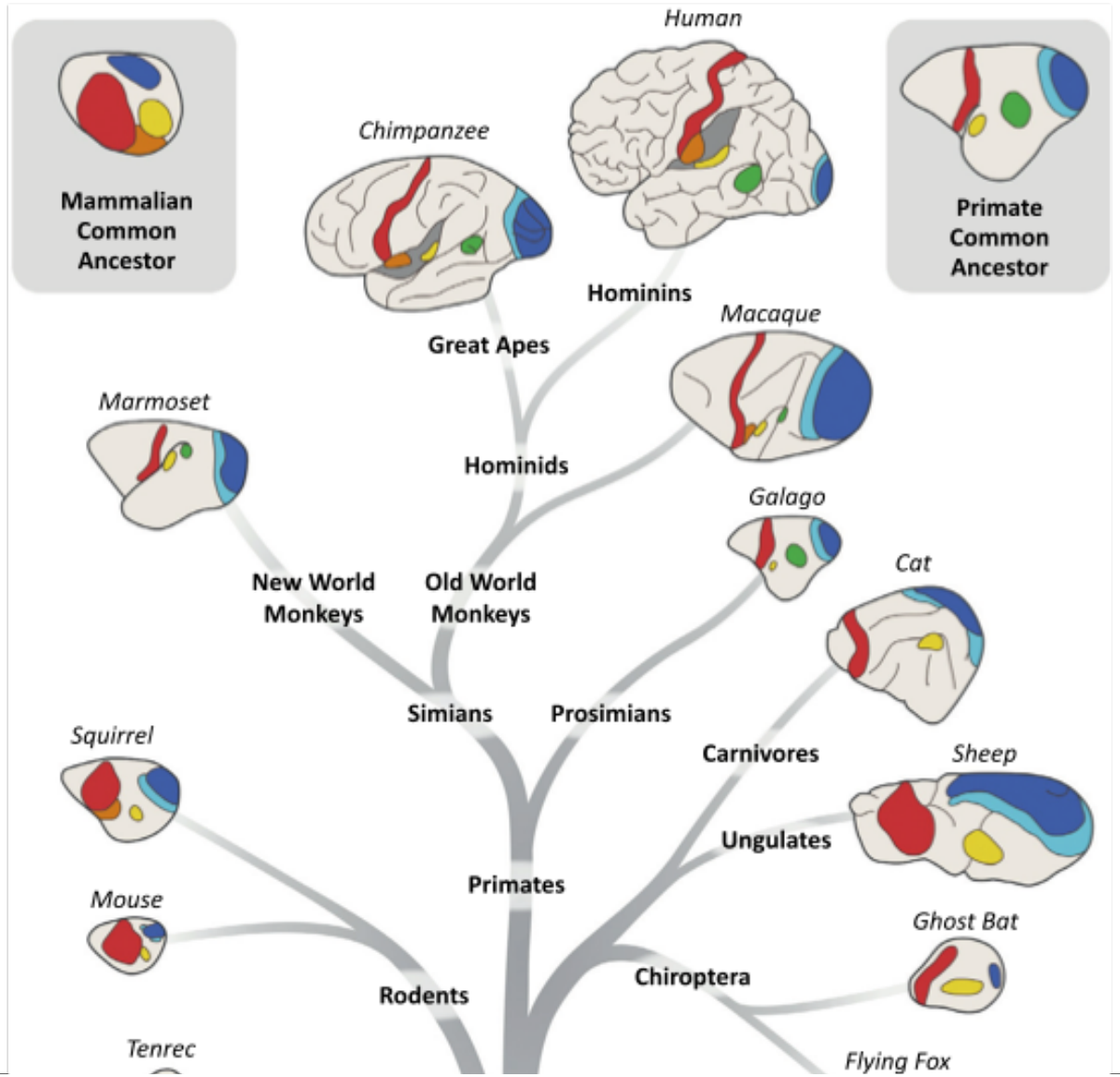
Was there a neural Bing Bang?

- Our left and write hemispheres are physiological asymmetric
- Main locus of asymmetry: the ***arcuatum fasciculum***
- Left is dominant for 90% of right-handers, the opposite for *female* left-handers
- Degree of lateralization: 60% extreme, 20% mild, 20% almost none
- What happened since the chimp?

Brain size evolution [Buckner and Krienen, 2013]

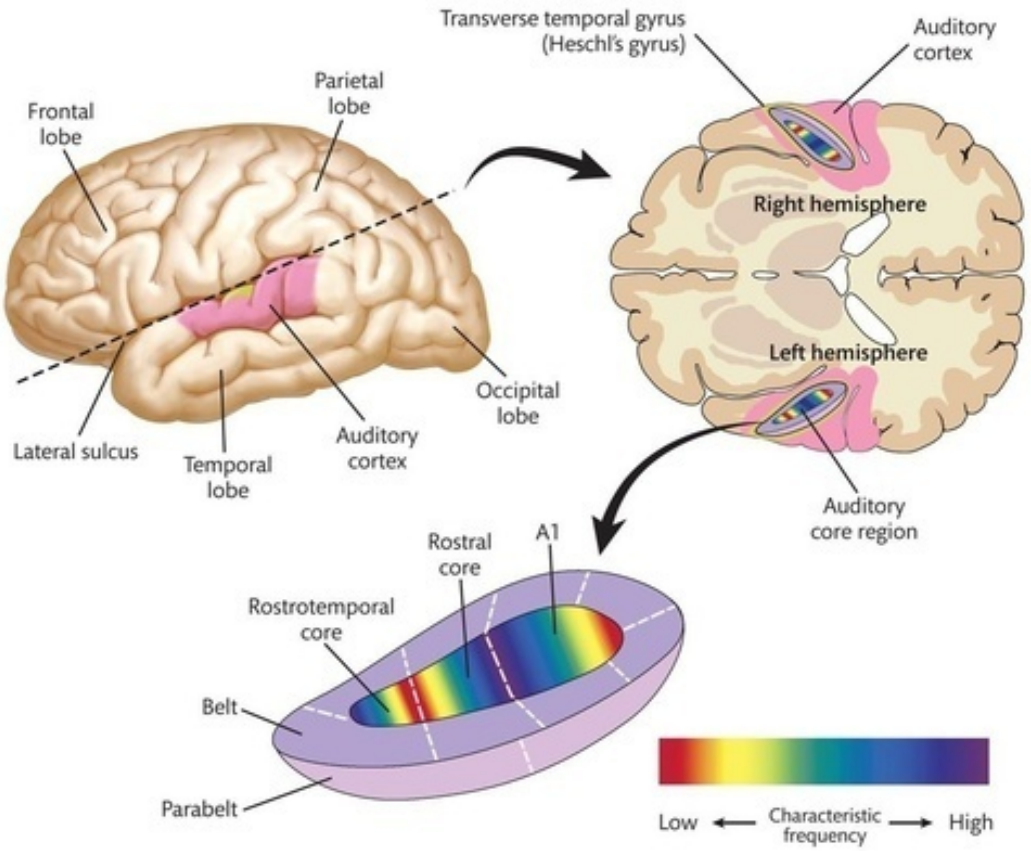


The rise of the association cortex



Btw: auditory cortex

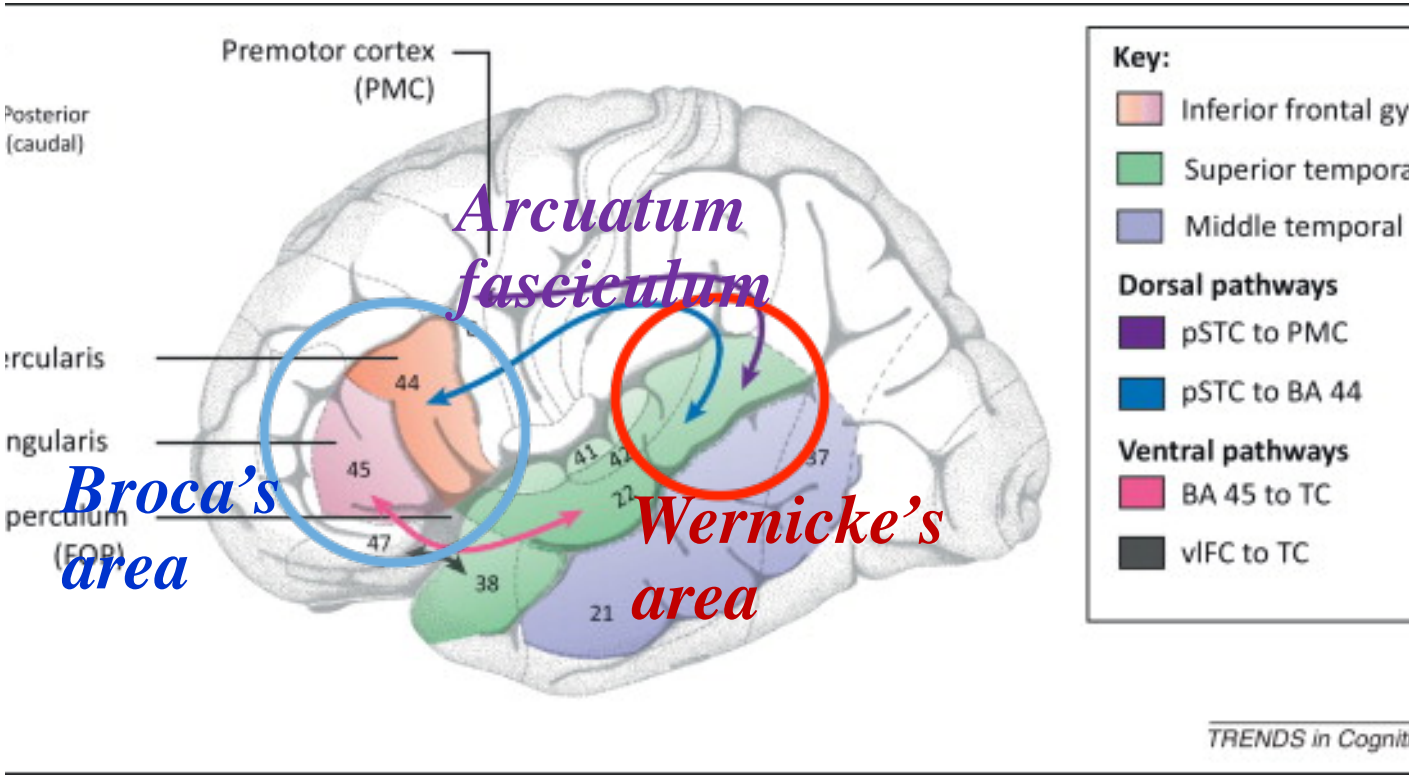
- Near the ear
- A1, A2, belt, parabelt
- Still cochlea signal goes through thalamus (MGN)
- Specializes in frequencies
- Right: tonal, music
- Left: temporal aspects, ***speech***



Language and the Brain: history

- 1830s: language must be in the left hemisphere
- 1870s: Broca (production), Wernicke (comprehension)

The Language hemisphere



Language and the Brain: history

- 1830s: language must be in the left hemisphere
- 1870s: Broca (production), Wernicke (comprehension)
- Revised in 1970s: Broca (syntax, grammar), Wernicke (word meaning and selection)

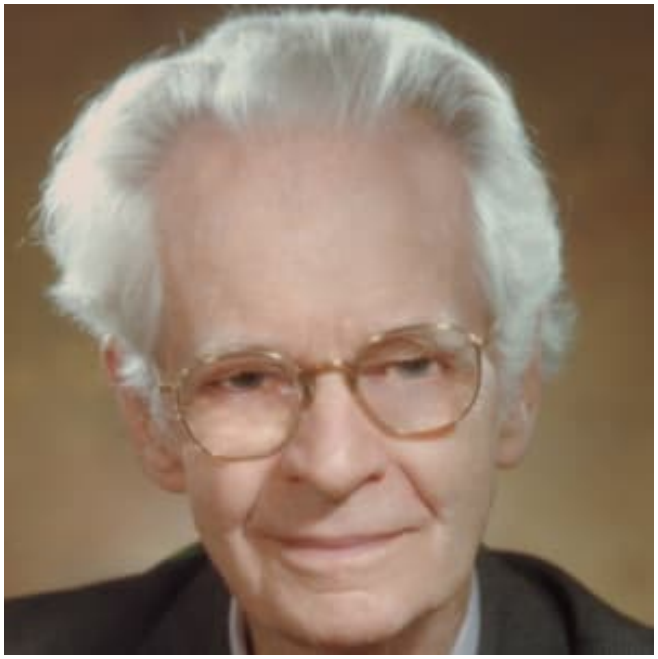
Broca aphasia

- poor or absent grammar
- difficulty forming complete sentences
- “Cup, me” instead of “I want the cup”
- more difficulty using verbs than nouns correctly
- difficulty repeating what has been said by others
- trouble with writing sentences, reading
- problems with full comprehension
- difficulty following directions
- frustration

Wernicke aphasia

- string words together to make sentences that don't make sense
- make up words that have no meaning
- unaware of the mistakes in their speech
- deliver words in a normal melodic line, even though the content may not make any sense
- articulate their words normally
- have difficulty repeating phrases
- add words when trying to repeat someone
- interrupt others and speak rapidly

Meanwhile in Cambridge, Mass , ca 1960 : The Skinner – Chomsky debate



The Skinner - Chomsky debate

- Or the behaviorist – structuralist debate
- Skinner believed, and showed by experiments, that behavior is the result of received reinforcements
- Influenced by J. Locke's tabula rasa
- His 1958 book "Verbal behavior" went a bridge too far
- Chomsky's devastating critique is far better known and accepted than the book

Chomsky's theory

- Language is innate
- Enormous gap between stimulus and competence
- Grammar is innate and universal
- Children only have to tune it with "details"
- What makes us different: Recursion and infinity
- The minimalist program: all you need is merge

Grammar

S → A V P (agent, verb, patient)

A → Alice | Bob | Chris |David

V → loves | hates | collects | enjoys

P → children | jewels | animals | toys

The sentence generation algorithm:

Start with S

Keep replacing a symbol in the current string with rhs of rule that has the symbol on lhs, until no such possibility

Recursion and infinity

S → A V P

S → S and S

A → Alice | Bob | Chris | David

V → loves | hates | collects | enjoys

P → children | jewels | animals | toys

Recursion and infinity

S → A V P

S → A said that S

S → S and S

A → Alice | Bob | Chris | David

V → loves | hates | collects | enjoys

P → children | jewels | animals | toys

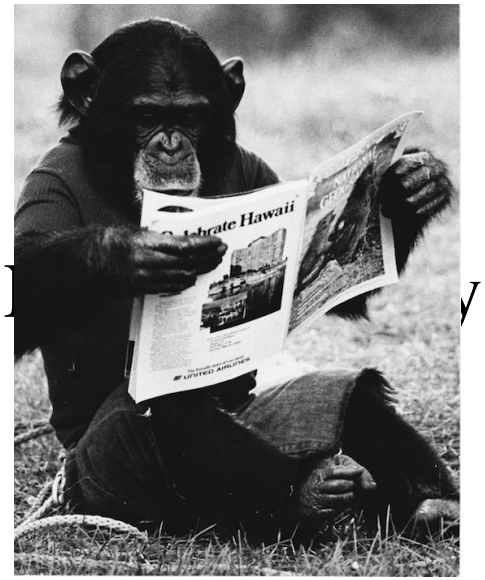
A propos recursion and infinity: the Pirahã controversy

- No number words, eg for “one”
- “few” and “more”
- No color words
- Daniel Everett 2010:

“No recursion!”

Besides universality:

how about exclusivity?



The Chomsky hierarchy

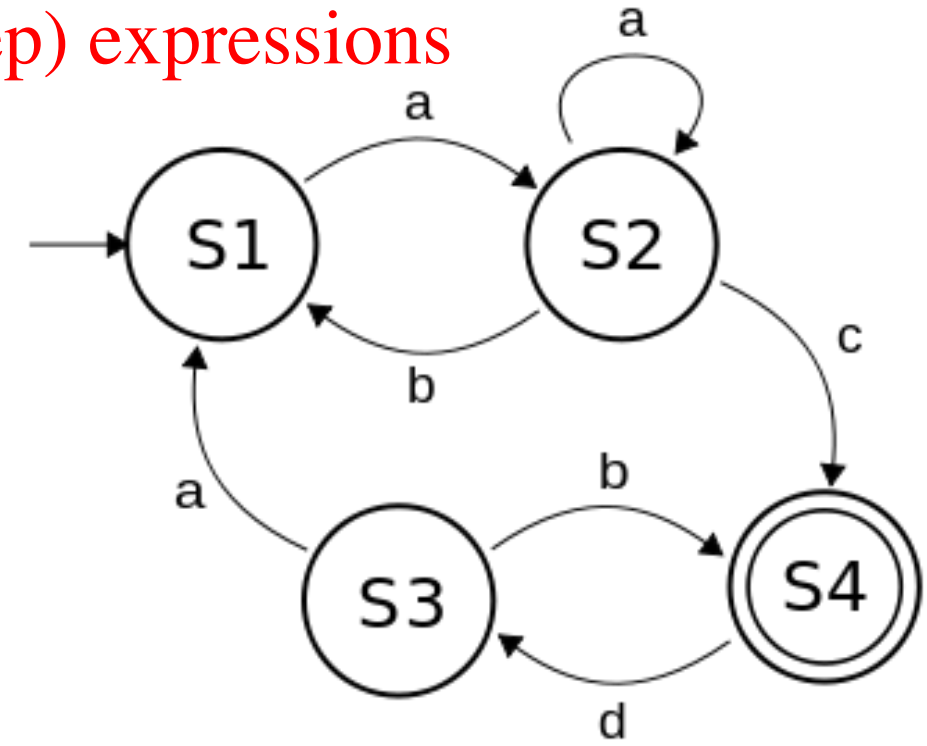
Four increasingly general kinds of (recursive, infinitary) grammars

(Assume only caps occur in lhs)

1. Regular or right-linear: A → a b B
2. Context–free: S → a B b S
3. Context-sensitive: SB → SabB (rhs no shorter than lhs)
4. General: ScD → Ba

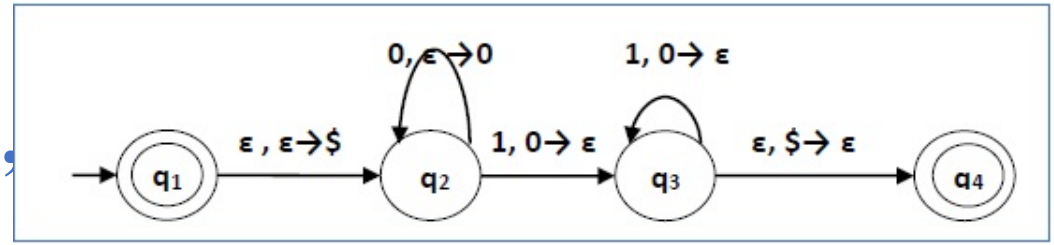
Grammar type hierarchy induces a machine/recognizer hierarchy

- Right linear grammars A → abB correspond to finite state machines or regular (grep) expressions
- One-way reading, finite states
- E.g., (a+b)aa(bb+ab)*
- **But not {anbn : n ≥ 0}**



The Chomsky hierarchy II

- Context-free (or phrase-structure) grammars A→ bBA correspond to nondeterministic pushdown automata
- Finite states, one-way reading, but can also look at/modify the top of a stack
- E.g. {anbn : n ≥ 0}, palindromes, …
- **But not {ww: w a word }, or {anbn cn : n ≥ 0}**



PDA for L= {0ⁿ 1ⁿ | n≥0}

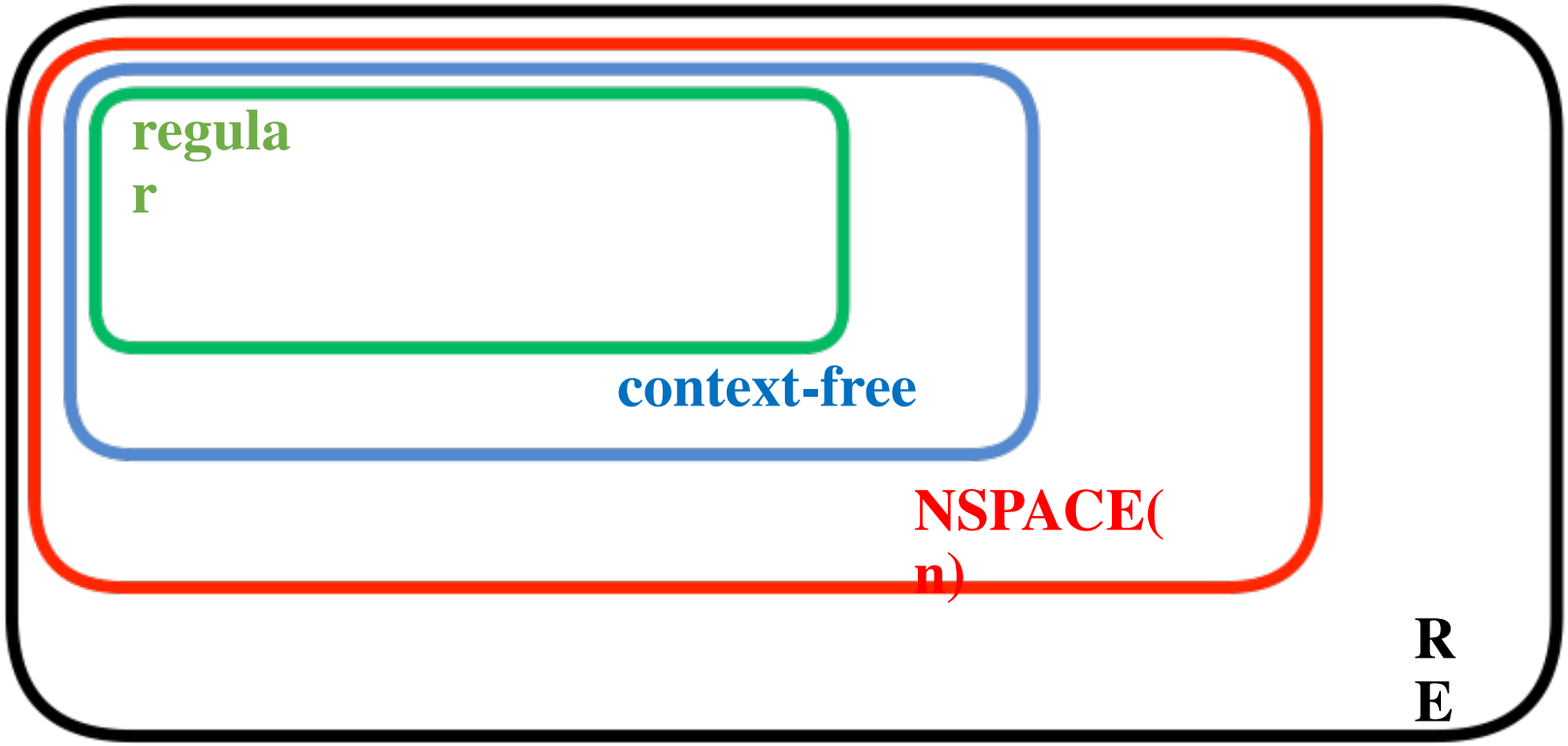
The Chomsky hierarchy III

- Context-sensitive (or non-shrinking) grammars Ab → bBab correspond to nondeterministic Turing machines that run in linear space (LBA's)
- E.g. {ww: w a word }, {anbn cn : n ≥ 0},…
- **But not {M: M is (the description of) a Turing machine destined to halt}**

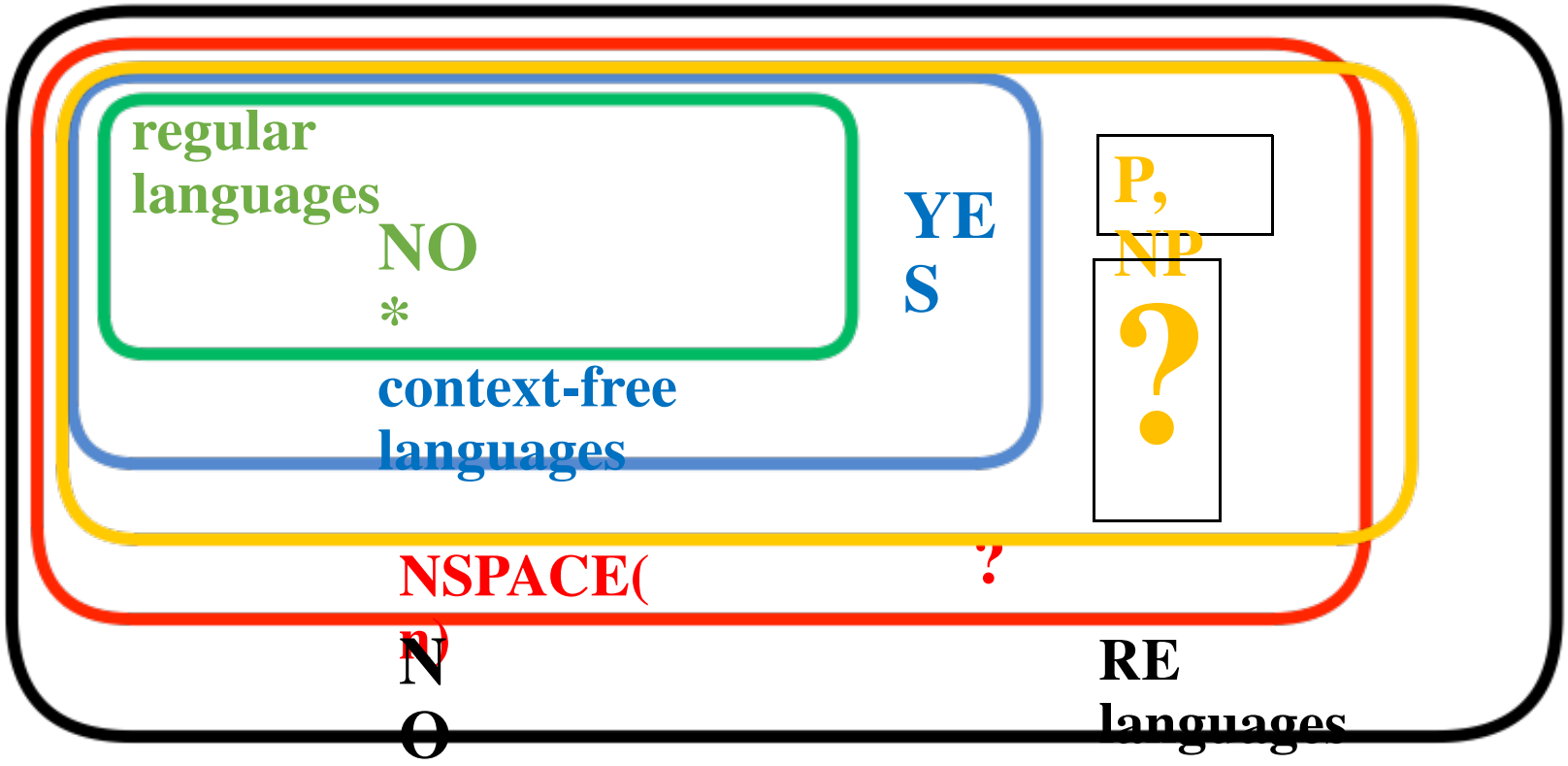
The Chomsky hierarchy IV

- Unrestricted grammars correspond to Turing machines that accept by halting
- All recursively enumerable (RE) languages: decision problems "solved" by Turing machines that reject by never halting…

The Chomsky hierarchy



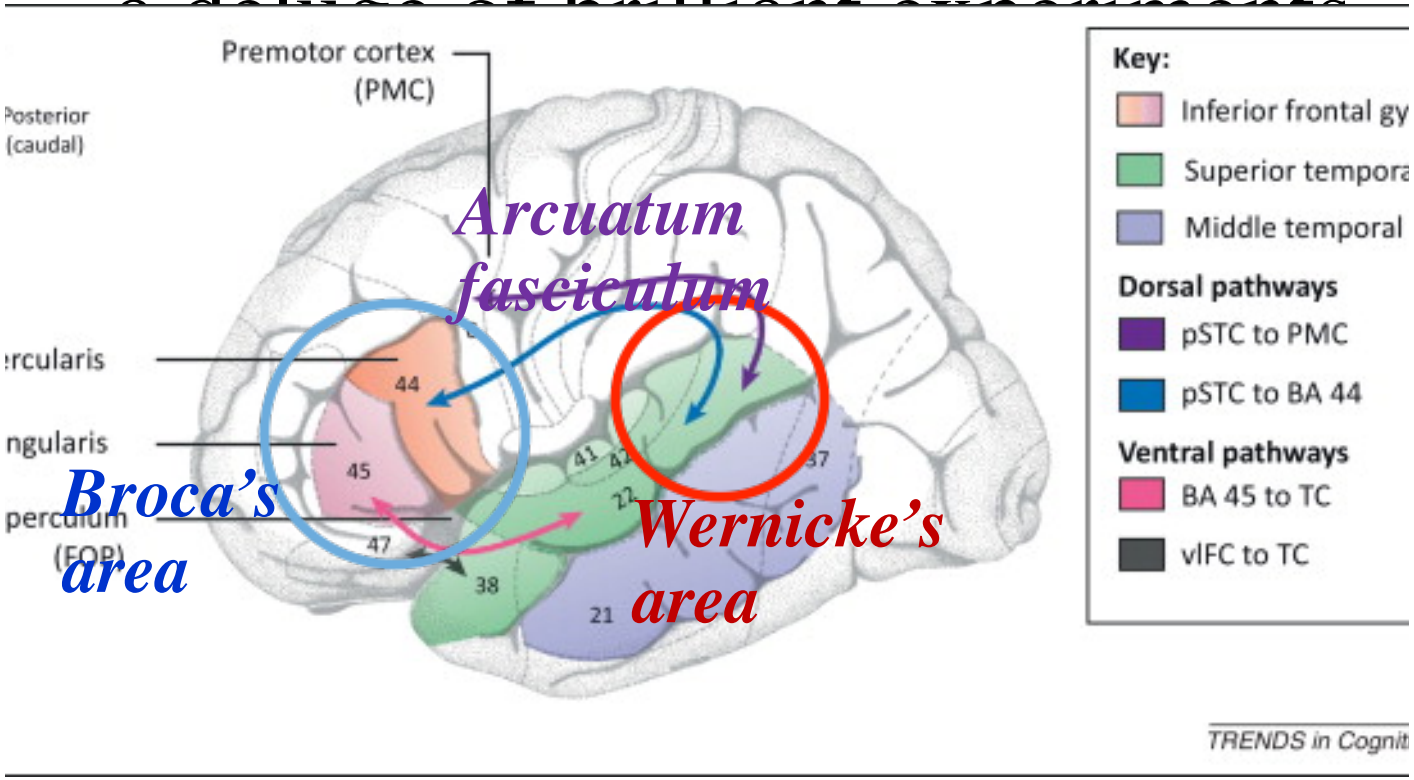
The Chomsky hierarchy:
Nondeterminism makes a difference?



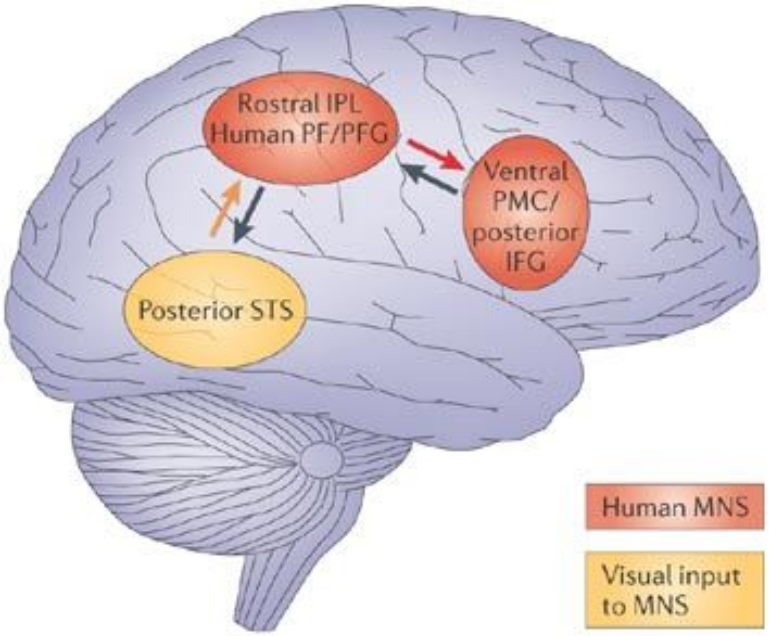
The Chomsky hierarchy

- Tremendous influence on CS
- Much of the research agenda in the 1960s TCS
- Helped us understand how to write compilers
- Trained us for the real problems to come

Meanwhile, in the language hemisphere:



Incidentally:
the mirror neural system



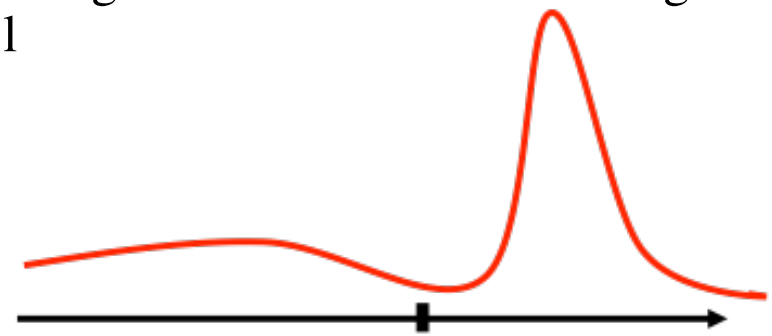
- Helps us understand the actions and intentions of others (and imitate)
- In humans and primates
- (and a little in songbirds…)
- Seems closely related to language (look where it sits…)
- Cf: Corballis, Lakoff

…this text pulsates at four
Hertz, the rhythm of four
beats per second, and I
believe that you may find this
rhythm a bit familiar,
because it coincides with the
rhythm of speech, and I don’t
mean my speech but speech
in general, by all speakers, in
all languages…

θ

The [Poeppel 2016] experiment

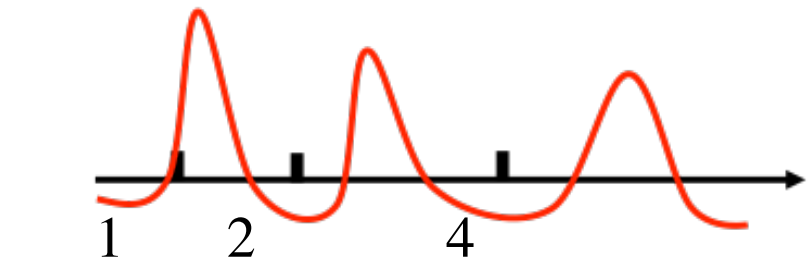
fret ship hil give true melt fans blue guess hits then cats
l



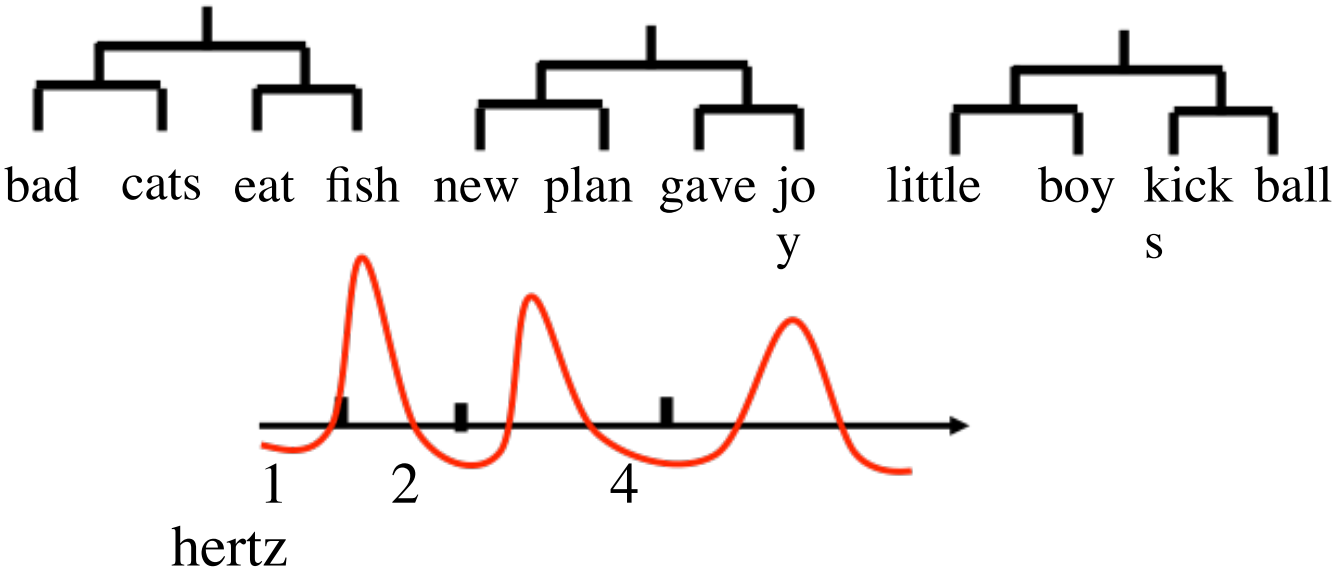
4 hertz

The [Poeppel 2016] experiment, stage II

bad cats eat fish new plan gave jo y little boy kick s ball



My interpretation



[Frankland & Greene PNAS 2015]

"The ball hit the truck"
vs
"The truck hit the ball"

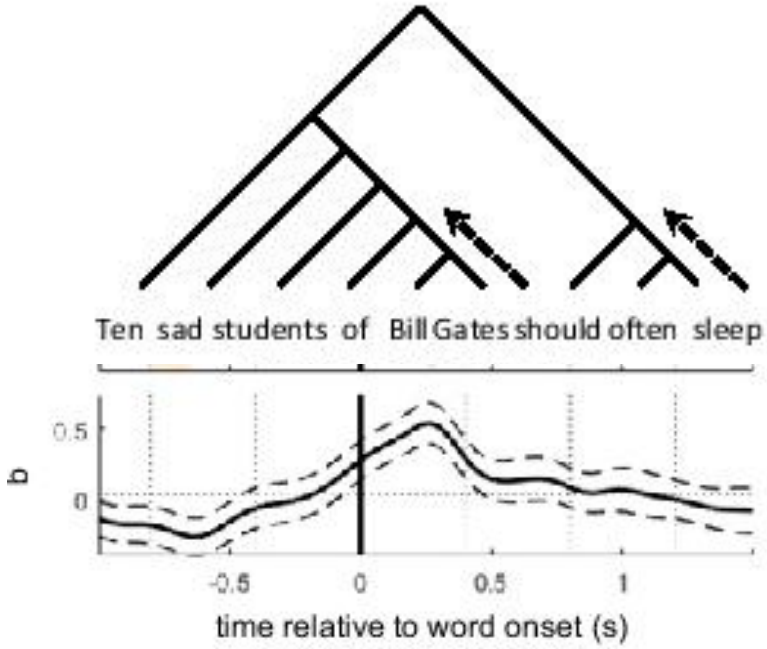
Different areas of the STG responded to "truck" in the two sentences *[Recall relations…]*

The first area also responded to
"The truck was hit by the ball"

But…

- By what mechanism can each tree-building step be carried out by a dozen or so spikes?
- 12 ~ gamma / theta
- Recall our discussion of Projection and Merge of Assemblies

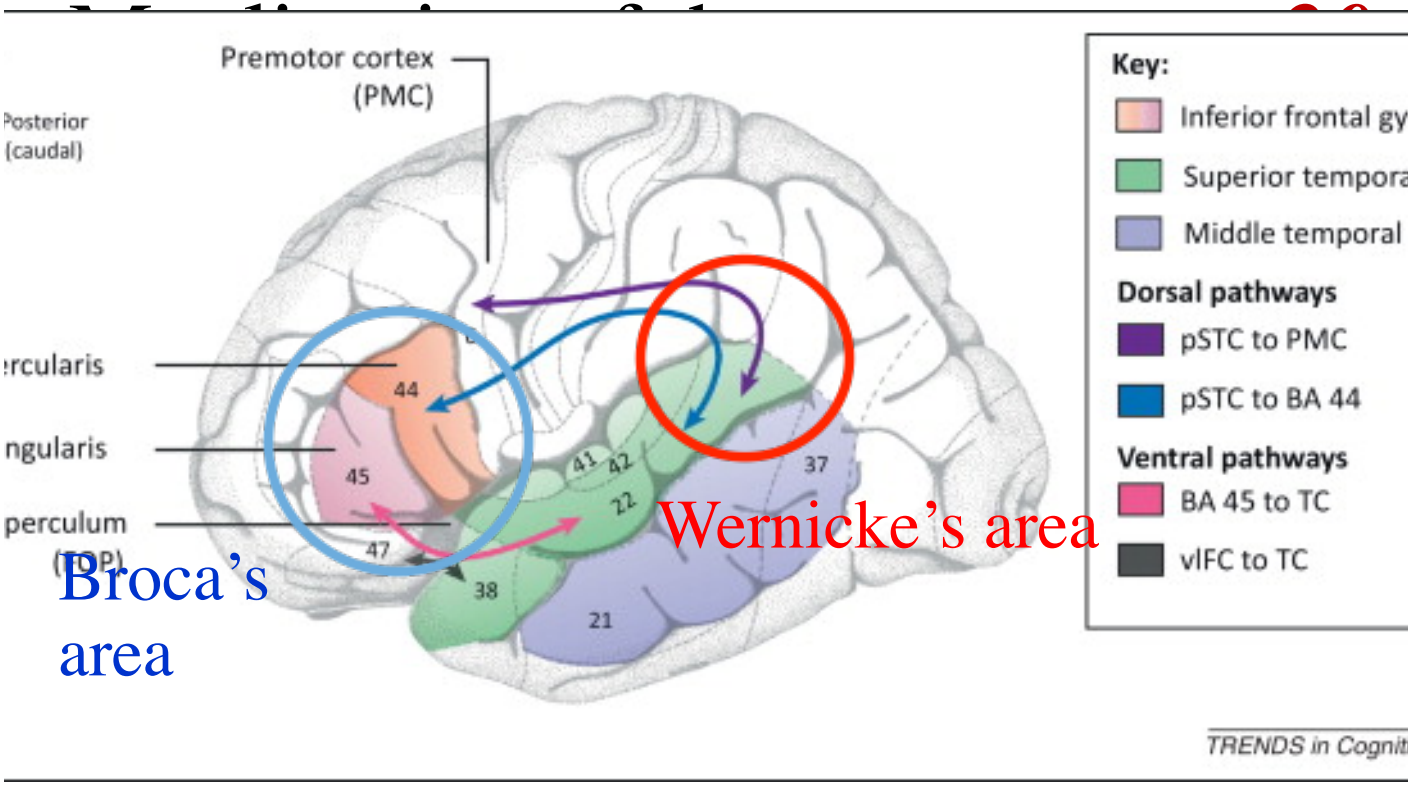
[Nelson…Dehaene, PNAS 2017]



Zaccarella & Friedericci "Merge in the human Brain" *Front. Psych. 2015*

- The completion of phrases, and especially of sentences, lights up parts of Broca's area

[ZF 2010]: Neural pathways for syntax?



But what is the neural gadget that does all this?

- Let us consider ***assemblies of neurons***
- They seem to encode concepts, presumably also words
- They can be be projected to other areas
- They associate/intersect to reflect affinity
- They can create trees by Merge

a brain architecture for syntax