

**Genetics and evolution of  
vocal learning and language**

**Language  
Evolution**

# **PART II**

A Brief Critique of Standard  
Approaches

# Remember...

- If you want to model the evolution of the Adaptive Immune System...
- The least you need is a theory of Adaptive Immunity.
- It would be crazy to proceed otherwise,
- Or to make up your own theory on the fly...
- And there is no a priori reason to think that the language faculty is any simpler than AI.

# Syntactic Boundary Conditions

- a. Syntactic dependencies arrange themselves in terms of formal objects that can be quite *high within the Chomsky Hierarchy*.
- b. Context-sensitive dependencies are generally *triggered, structure-dependent*, and limited by *locality* considerations.
- c. Semantic dependencies are determined by syntactic dependencies and obey definite mapping principles (*Full Interpretation, Compositionality, Conservativity*).
- d. Morphological variation, of the sort patent across languages, in many instances involves *uninterpretable* elements.
- e. Core language acquisition involves the fixation of a few fixed, normally morphological, *syntactic options* ('parameters').

# **In terms of the theory of automata...**

- The language faculty requires an automaton in between the Push-Down Automata or PDA (for phrases) and the Linear-Bounded Automata or LBA (for generalized entanglements).
- We are calling that the PDA+.
- A theory of the evolution of syntax ought to say something about how the PDA+ could have emerged (with its memory regime and other specifications...).

# That's a bare minimum...

- Surely the language faculty has other properties, some still not understood.
- But the ones just outlined are pretty much agreed upon.
- Only property (e) is still under debate –not so much because it being in doubt, but because it isn't clear yet what is the best way to flesh the idea out.
- If your evolutionary model doesn't have at least some of these notions in sight it would be pretty useless.

# Pinker and Bloom (1990)

- Classic adaptationist piece.
- It virtually ignores syntax, even as understood at the time of its writing.
- It admits that language variation is a mystery and has nothing to say about uninterpretable morphology and how the system deals with it.
- It focuses on showing how one could tell a story about ‘language’ in a loose sense, which Chomsky had refused to do for decades.

# Arbib, Deacon, Lieberman...

- All major players, but absolutely no discussion of the facts of syntax, presupposing a PDA+ architecture.
- In fact, they are even hostile (for unclear reasons) to anything having to do with transformational grammar.
- Evolving the Adaptive Immune System by *denigrating* theories of Adaptive Immunity?



# More serious attempts (I)

- Carstairs-McCarthy (1999) takes phrasal syntax to be an outgrowth of syllabification (exaptation).
- For Calvin and Bickerton (2000) it is an exaptation from pre-existing thematic relations.
- Unfortunately, it is unclear how either approach bears on whatever goes beyond such local dependencies.
- Stories would remain unchanged even if the language faculty did not present discontinuous dependencies, or they obeyed conditions that are the inverse of what actually holds (for instance, anti-c-command).

# More Serious Attempts (II)

- Kirby (2000) explores the learners' capacity to segment utterances and generalize over chance coincidences in their meanings of identical segments.
- A similar approach is taken by Hurford (2000), still emphasizing broad generalizations and regularizations,
- or Nowak et al. (2002), which shows –also through modeling, which we return to– that beyond a threshold of active word usage it is advantageous for a system to deploy phrases.

# However...

- We knew that important fact already, in formal terms discussed in Berwick (1982).
- And, again, such a system does not go beyond phrasal associations and the meaning compositionality thereof, at best reaching PDA capabilities.
- But again: we need PDA+ resources...

## And given the Chomsky Hierarchy...

- A transition to a PDA+ *presupposes* the PDA describing mere constituents,
- and so an evolutionary change taking an organism into PDA+ territory *automatically carries it over intermediate realms.*

## **In other words...**

- We don't know whether syntax got to be PDA+ directly or through PDA stages, as these studies assume.
- Of course, it may well be that a PDA+ automaton is an even more effective way to compress the sort of dependencies that arise with linguistic symbols.
- But if so, why should the grammar go through a “mere” PDA stage?

# The evolution of syntax

- won't be modeled by ignoring syntax,
- or by asserting that it should follow from effective packing or similar considerations.
- The latter claim would move professionals if we were shown how any of the basic properties we routinely work with do indeed emerge, in detail...
- and moreover why they are unique to this system.
- (Alternatively: show the properties in other organisms)

# Hauser et al 2002

- Conceptually divides those aspects of language that humans share with other species (the Broad Faculty of Language, FLB)
- from those that are unique, if any exist (the Narrow Faculty of Language, FLN).
- The piece furthermore raises the empirical challenge of deciding which features of syntax fall within each.

# Hauser et al 2002

- For example, *recursion* is seen as a property of information-exchange systems that has not, so far, been encountered in non-human species, and so is hypothetically declared part of FLN.
- In contrast, some phonetic and semantic features are shown to have been spotted, in some form, in other creatures, and are thus taken to be part of FLB.
- The paper concretely suggests that syntax may be part of FLN.



# Jackendoff and Pinker 2005

- (i) There is more that is unique to language than recursion.
- (ii) The model framing the evolutionary picture in Hauser et al –nothing but the dominant paradigm– is not of their liking.
- We can't say much about (ii), but (i) can be addressed.

# Let's construct a real challenge

- Hauser, Chomsky and Fitch (2004) (CHF) argue that *recursion* is the key to the evolution of the narrow faculty of language.
- But what are they talking about? Don't ants (certainly chimps) need PDAs for their thought processes?
- How can you prevent a PDA from giving you recursion?!
- If the system allows  $X \rightarrow Y Z$  it will automatically allow  $X \rightarrow \dots X \dots$





## One form of the puzzle:



- Nature shows many instances, even in behavior, where the best modeling available arguably implies a PDA (think of foraging strategies, for instance) –though this is hard to prove, let's grant it for the sake of argument.
- Many apparently creative, non-encapsulated, animal behaviors have been reported, especially in the ape literature.
- So doesn't that immediately show recursion outside of language?

# However...

- FL is not just a system of thought. It's a bunch of things, and surely one of those:
- Is a system that is characteristically *used publicly*.
- The issue is whether *that* imposes any architectural conditions.
- It can be shown that it does.

# The Compression Problem



- Given our understanding of FL, the system is at least three-dimensional.
- If nothing else: ‘higher’ grammar involves:
  - A predication dimension (thematic stuff, adjuncts of various sorts...).
  - c-command dimension (‘chain’-related dependencies, binding, control,...).
  - An antecedence dimension (weak-cross over, context-confinement,...).

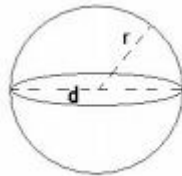
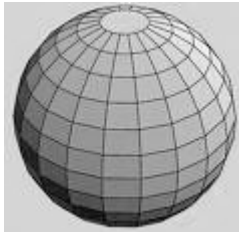
NO NEED TO UNDERSTAND ANY OF THESE NOW

# The Compression Problem

- That's fine: thought is multi-dimensional...
- But how do you translate thought to something sharable?
- If telepathically, end of problem...
- If not, you'll have to involve the motor system.
- But that seems to be generally one dimensional (at best one-and-a-half dimensional, for those of you who can dance...).

# The Compression Problem

- Wanna pack a 3D object into a 1d line?!
- It's like asking me to figure out a sphere from a dot!
- Not even flat-liners could do that, and they got to see a whole line segment!



- Still, with ingenuity some nifty tricks can be played to solve (parts of) the puzzle...
- We'll return to that when we talk about birds!

## But all of that said...

- The issue is no longer evolving recursion.
- The real concern should be evolving a *mechanism to make recursion usable* as a public system.
- That is a very tough and narrow problem.
- In all likelihood *it requires more than a standard PDA* to be executed.
- **So again you need PDA+ resources!**



# Keep in mind...

- For  $n > m$ , to go from  $nD$  to  $mD$  is a solvable problem, for which you can devise an algorithm.
- But to go from  $mD$  to  $nD$  is not a solvable problem in general if  $n > m$ , and an algorithm can thus exist only for a simplified version of this problem.



# **This matters in performance...**

- The production problem in terms of the Compression Problem...
- Is much simpler than the corresponding parsing problem.
- You after all know what your thought was, hence can compress it somehow.
- But I have to attempt a reconstruction from the scraps you give me.

# In particular in the case of interpreted recursion...

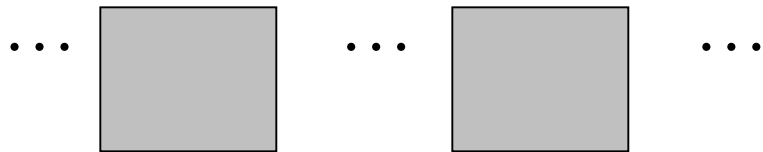
- When you compress  $[_X \dots [_X \dots$  into something involving just the dot, dot, dot...
- You know how you're doing this.
- But I just get to hear your dots.
- And even if I use a PDA to figure out that there ought to be an X in my parse...
- I still need to make decisions about *how to attach* that X...

# Think of the logic...

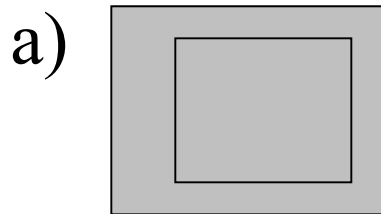
- If I parse X, but treat it as a conjunct, ‘paratactically’ associated to whatever else I had parsed up to that point...
- I haven’t figured out your recursive structure.
- For that I must go ‘hypotactic’ with my X, hence making it part of the X I had already parsed.
- Then I can actually fully understand the (true) recursivity of your original thought.
- Nothing forces me to go paratactic or hypotactic.
- A decision procedure involves enriching the PDA.

# Visually

- You hear...



- Now how do you know whether to parse what you heard as (a) or (b)?



But understanding (a) is crucial to parsing recursion!

## So there you go:

- To make FL (with full complexity?) usable in customary fashion, you need (a bit) more than a standard PDA gives you –a PDA+.
- And once you have one (or we should say: IF you manage to get one): What new tricks can you perform with it?

## That said...

- It is possible to devise some thought experiments with both the proto-language we had prior to the PDA+ emergence (if there was one).
- ... and also the system of thought at that time.



## As for proto-language (if it existed)

- Prior to the linearization boost, it must have been essentially Markovian.
- It may well have been quite complex in the thoughts it entertained, but lacking a way to linearize them is akin to not having anything usable beyond two to four symbols, simply because you get  $n!$  combinations –bedlam as  $n$  grows.
- Perhaps Nim Chimksy had something like this...





# Undecided in this scenario...

- As a system *of thought*, was whatever we had at the time of proto-language something with the capacity of current LF?
- Or, instead, did we just have a system merely using PDA capacities (recursive though those may be)...
- Which got expanded once we required a PDA+ to parse its recursive structures?

# Regardless of how we answer...

- The matter of evolving a system with extended PDA+ capabilities is fully testable:
- a) Do other species have such systems?
- b) Does the archeological record allow us to reconstruct behaviors requiring a PDA+?
- c) Do neural circuits bearing on PDA+ memory exist –are they where expected?
- d) Can genes implicated in language be remotely linked to PDA+ memory?

GOD THROWS THE CREATIONISTS A CURVE...



LET THERE BE  
EVOLUTION.

RUTH

**Genetics and evolution of  
vocal learning and language**

**Language  
Evolution**

# **PART III**

Some Simple Questions to Bear in Mind

## **a) Do other species have PDA+ systems?**

- If we mean by that ‘communication systems’ (with that computational power) the answer is –from what is known– almost certainly ‘No’.
- Familiar systems in, say, rhesus macaques and the like have a very small repertoire of signals with no obvious syntax.
- Bee dances and other such behaviors are almost certainly finite-state (with loops).
- Bird songs are potentially more interesting, though we return to these in more detail.
- Although we should be attentive to new and surprising instances, that’s the general tendency.

# **Then again, why concentrate on communication systems only?**

- From a computational perspective, it doesn't really matter whether the computational system is made public (communication) or not.
- And within 'private' systems of thought we may come closer to something of interest.
- Every new thing we learn about animal behavior shows us how smart they really are.

# If we have time (and we can connect)

- See some videos for yourselves:
  - a) Higher apes.
  - b) Other mammals.
  - c) Some birds.
  - d) Even nematodes...

Granted, there is no comprehensive theory of any of this, but it does look very impressive, from a sheer computational perspective.



# Field of ‘computational behavior’?

- How can you ascertain the complexity of a given behavior from a computational angle?
  - Does the computation group units?
  - Is it recursive?
  - Does it make reference to internal states in the computation itself?
  - Does it require non-trivial memory?
  - Is it plastic?

# My hunch

- As far as I can tell, from what I've seen...
- Computational thought is a *very* ancient trait.
- Very possibly pre-Cambrian, given the diversity of where you find it.
- Perhaps even more, as all eukaryotes would seem to exhibit *some* degree of computation, virtually as a definition of what it means to react to the environment in organized fashion.
- Don't know about bacteria, etc...

# That said...

- Even if a mind goes high in the CH in terms of the complexity of its thought processes...
- That has no consequences, in itself, for whether all that richness is ‘communicable’.
- The Gregor Samsa Syndrome: you may be a genius trapped inside your brain (in fact, for a variety of reasons).
- Computational thought is probably not even a *pre-requisite* to communication.
- You could communicate relatively trivial stuff, via pheromones, etc...

# So what's interesting in the faculty of language, in the end...

- The fact that you have *both* a) computational richness, and b) a way to make it public.
- In fact, it is almost certain that you have the computational richness at least in all higher apes (if not before).
- And on the other hand that you have the ability to make public –more on this when we study this in detail– in a variety of ‘vocal learners’ at least.

# The first key, in my view:

- A priori complex thought and vocal learning (say –or any such system to make things public) are orthogonal to one another.
- In all likelihood the vocal-learning systems in birds, at least, are not even used to carry complex mental computations (even if birds have them, independently).
- And the complex thoughts in apes are in general quite solipsistic.

# Logic of the HCF paper, coupled with Gallistel's findings ...

- suggests FLN had something to do with making language public.
- FLB was probably very rich, but 'trapped' as mere mentalese.

.....



# Which is not to say...

- That, even if humans had a rich computational system prior to language...
- When for some reason they did gain the ability to make language public...
- The event didn't entail getting an even richer (higher in the CH) computational system as a result...
- In fact, in sheer computational terms (exact numerosity, knots, diasporas, plasticity, etc.) our behavior as humans is obviously unique.

# **So the issue of whether language is (not) continuous with other systems**

- Is very tough... (And a matter of perspective.)
- In one respect there is obviously a drastic jump, as a ‘communication system’.
- But if you mean as a ‘system of thought’, it is less obvious (setting aside exact numerosity).
- Then again, as a ‘system of culture’ you again have discontinuities, again obviously.
- Do those affect ‘thought’, though? Who knows!



## **b) Does the archeological record allow us to reconstruct behaviors requiring a PDA+?**

- This is also a very difficult question, since behavior as such doesn't fossilize.
- But the issue is to look at behavior from a computational perspective, if we can...
- Which up to very recently hasn't even been attempted.

**Mount (1989): Modeling knots  
requires a context-sensitive system**



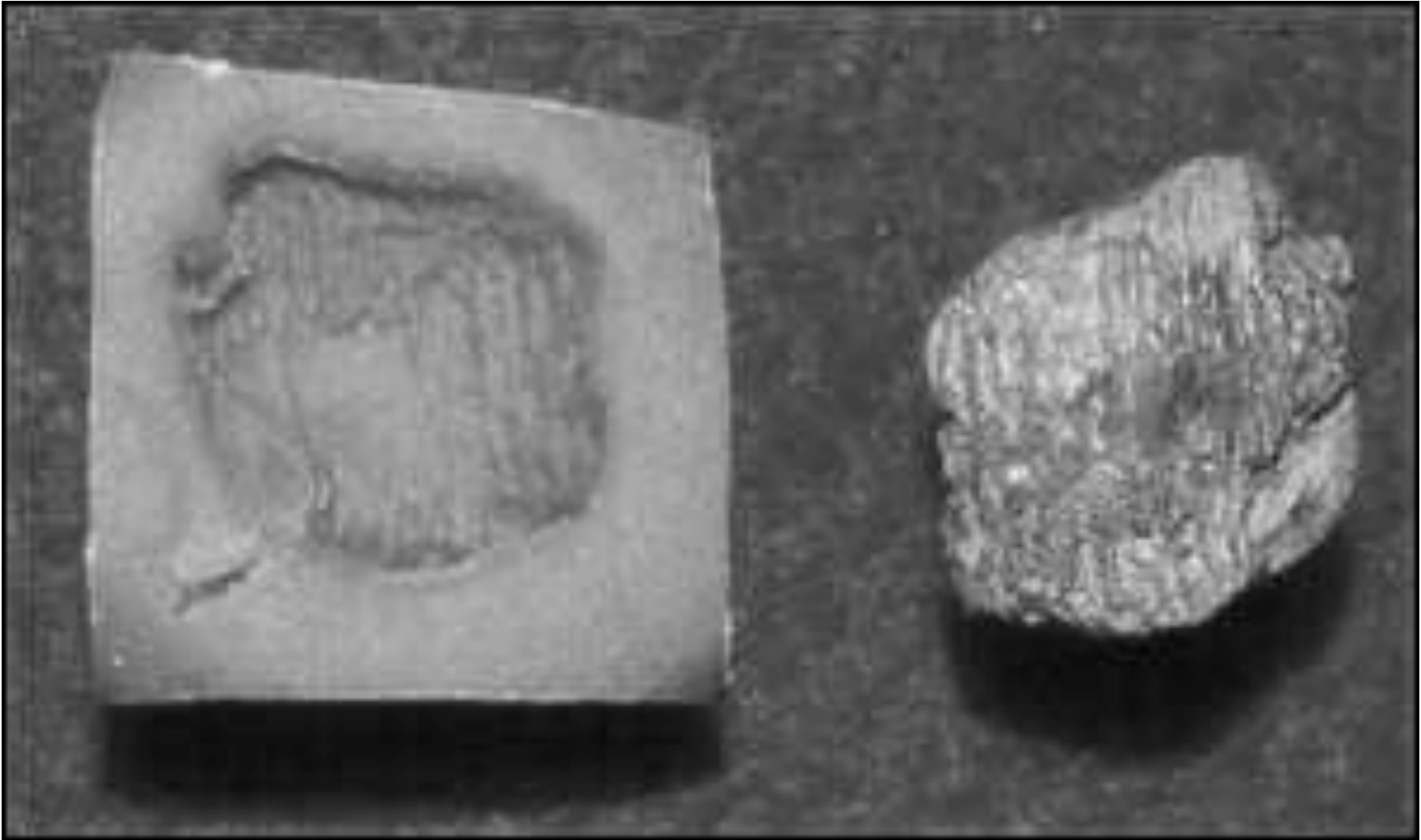
# Important caveat:

- A mess is not the same as a knot.
- A cat playing with wool can stumble into a knot: but can he undo it?
- One could argue (perhaps) that some nests have pseudo-knot characteristics (still very interesting, given other cognitive properties of birds) –but can they undo them? If not, not a real knot.

# **Possible cognitive bases for knots (similarly for other tasks):**

- (A) There is a separate grammar for knots
- (B) Knotting is parasitic on context-sensitive grammars of the sort needed for language
- (C) A deeper organ underlies two systems
  
- Possibility (A) is unlikely, if only because no non-human species uses knots, their development arguably correlates with language development, and they can be selectively impaired in certain language disphasias.
- Both (B) and (C) are relevant to our purposes.

**Woven cloth (27,000 BP), Europe,  
from Soffer 2000**



# European figurines (25,000 BP), hats or hairstyles?



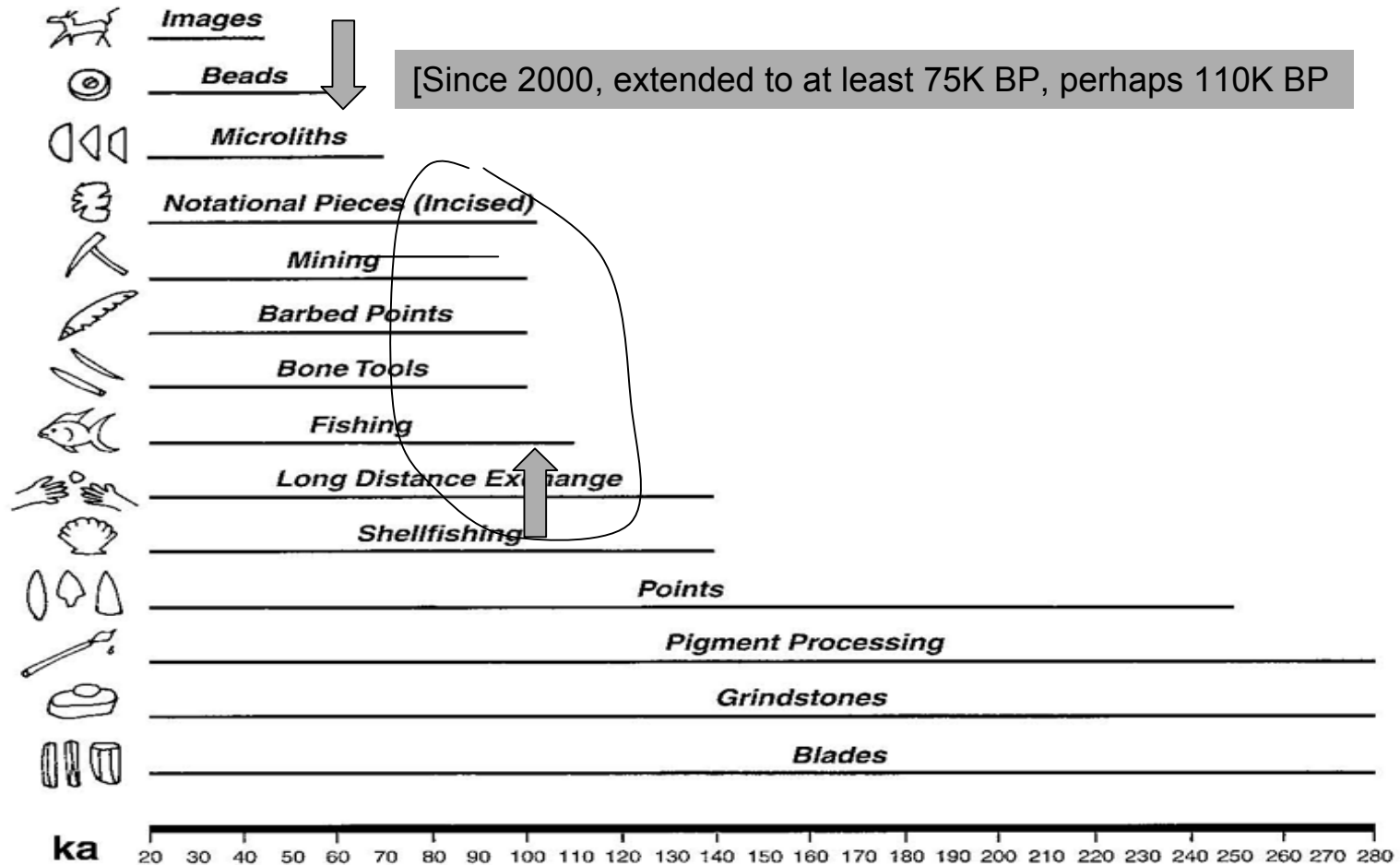
Note also belly-button!

# **But how early can we find knots?**

- (I) An argument from points
  - (II) An argument from beads
  - (III) An argument from shoes?
- 
- Note how carefully executed (therefore valuable) the following pieces are

# From McBrearty & Brooks 2000

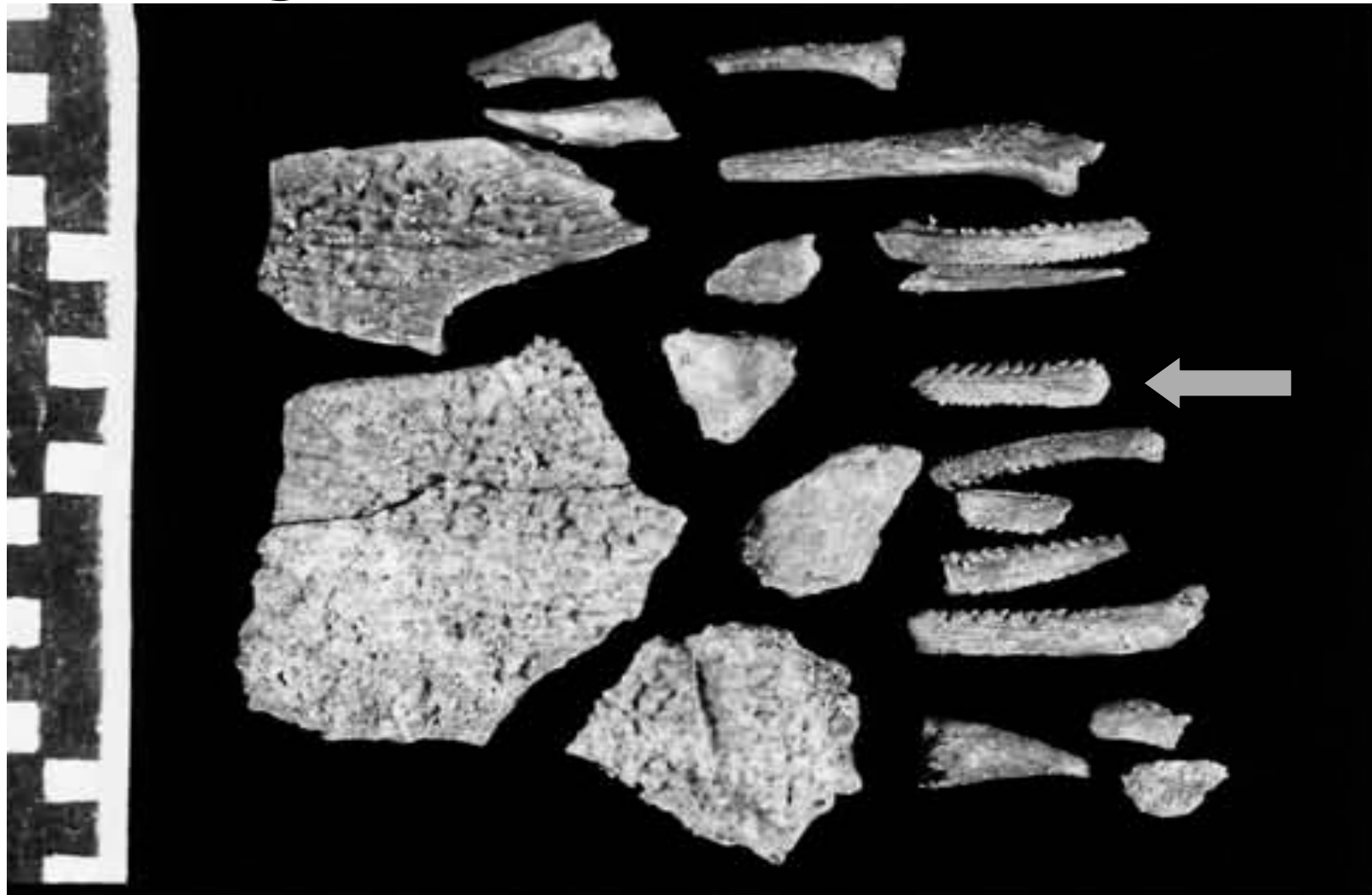
## Behavioral Innovations of the Middle Stone Age in Africa



Observe period from 100K to 60K BP



**(I) Fish bones from Katanda  
(Congo), 90,000 BP, scale in cm.**



# **(I) Harpoon technology**

- Points made of bone in order not to sink.
- Barbed points hard to come out of fish.
- Apart from careful planning, time required to manufacture points.
- If points not tied to shaft (in turn possibly tied to string) process pointless.

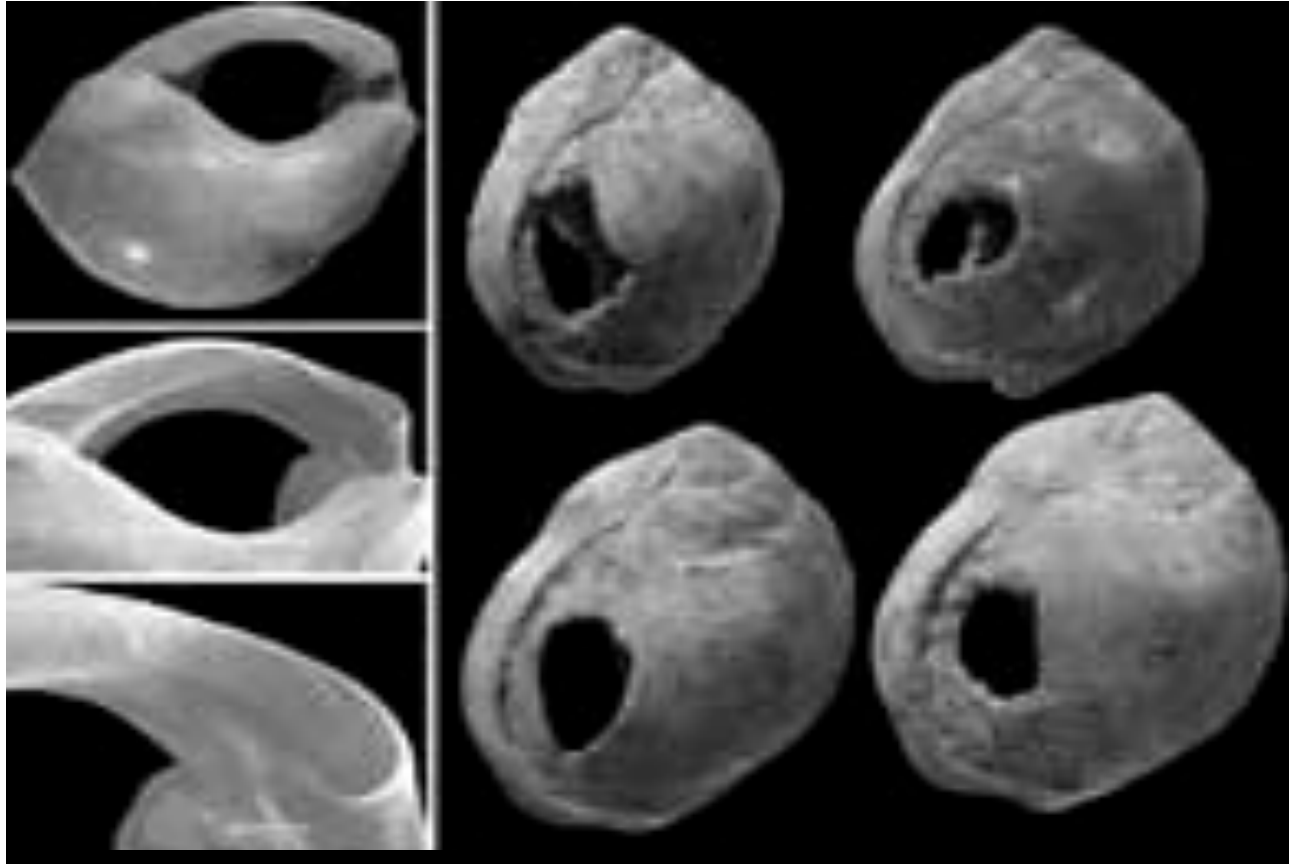
# **(I) Microliths are useless unless**

- They are firmly mounted
- Implying tighter connection to shaft than mere binding & sap achieves
- Possibly useless unless relevant point is shot with a thrower or bow
- Implying very reliable knotting at bow's ends and at thrower.

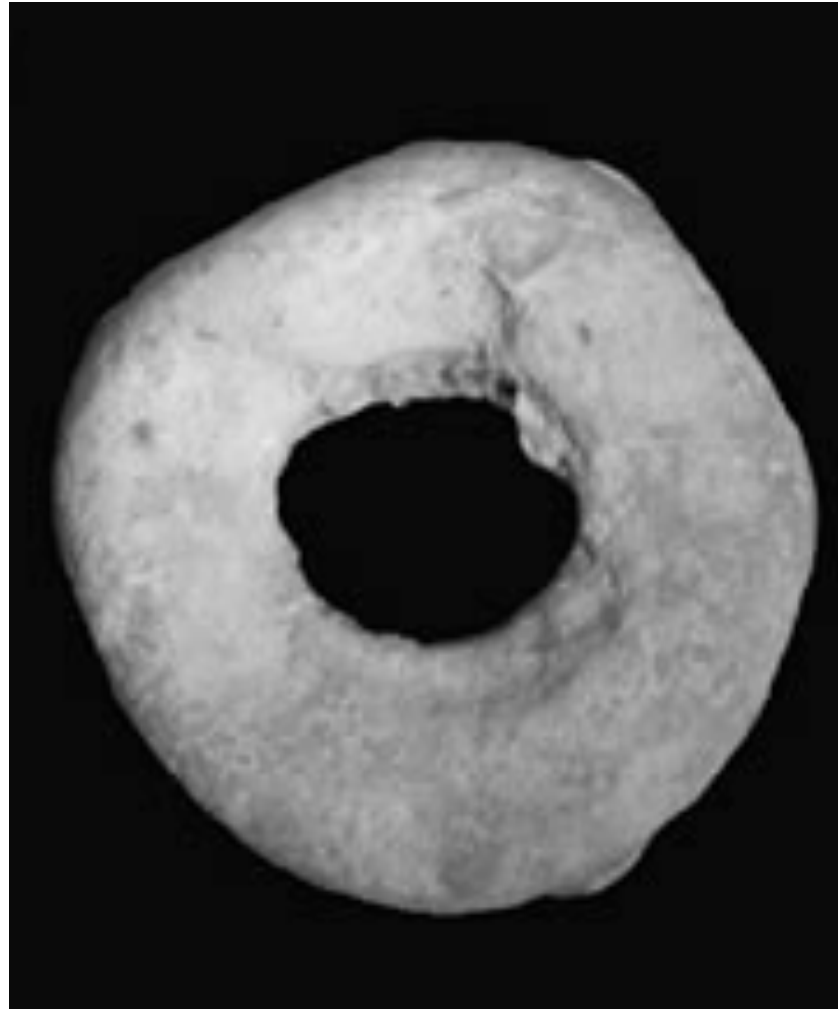
## **‘untying’ test...**

- Almost certain in the case of arches.
- Almost certain in the case of the assembly of at least microlithic projectiles.
- And by same reasoning, earlier lithic projectiles may well have involved mere entanglements (not knots).

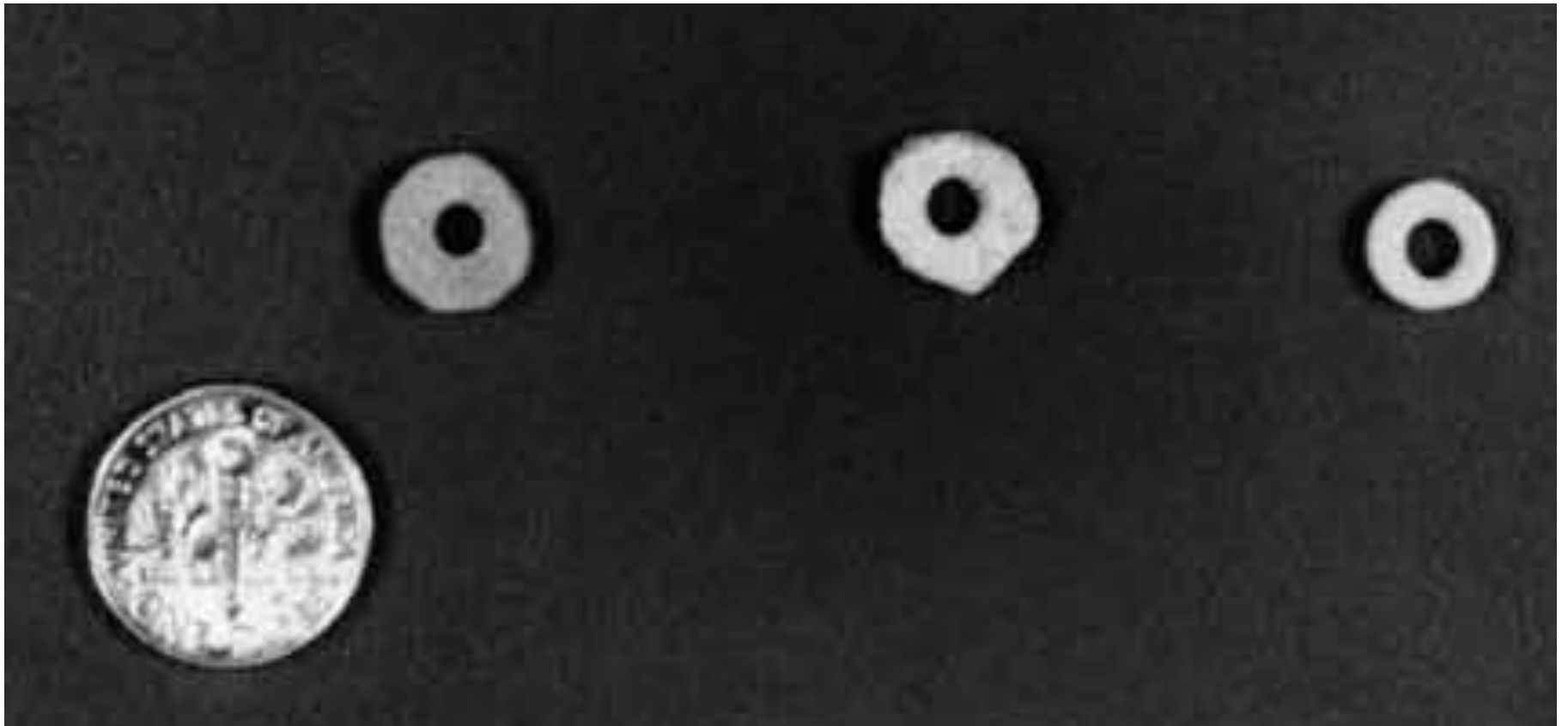
## **(II) Blombos Beads, South Africa (75,000 BP)**



**(II) Loiyangalani bead, Tanzania  
(70,000 BP)**



## (II) Ostrich eggshell beads

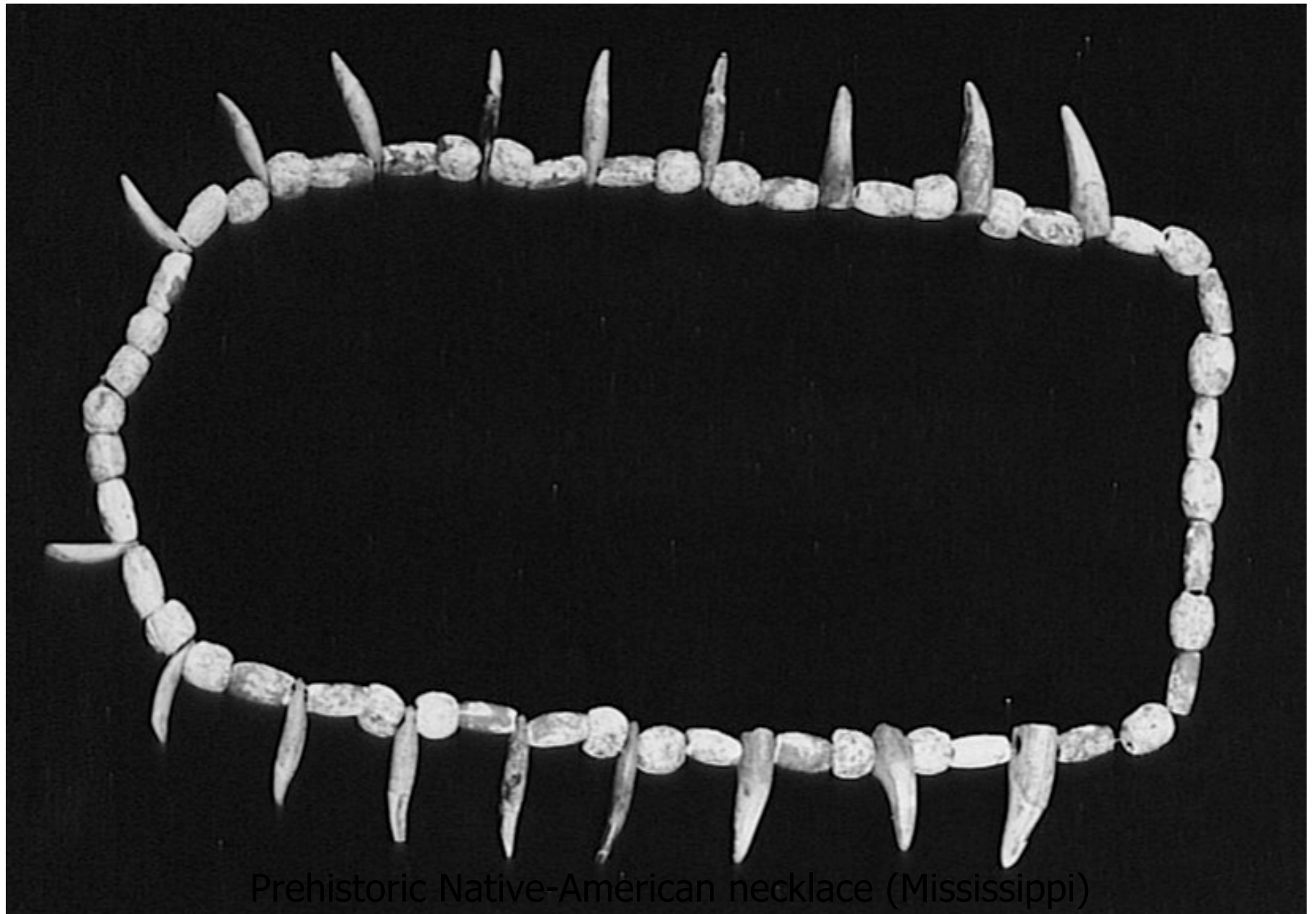


Observe  
necklace  
(25KB.P.)



Massive  
jewelry  
implies  
secure  
knotting





Prehistoric Native-American necklace (Mississippi)

## **(II) Beads imply:**

- Mastery of microlithic technology, including perforating strong needles.
- Knotting for corresponding necklaces or ornaments to hang from clothes.
- A hierarchical, large, society where elaborate ornamentation plays a role.
- NOTE: Latest dating possibly at 110,000 BP (Middle East).

# Untying test

- In part it depends on how elaborate the ornament arrays were.
- Judging from historic situations, it is unlikely that these objects weren't mounted and dismounted (not just for sleeping, for bathing, for battle, for hunting, etc.).
- By the same reasoning, unreliability of merely entangled Neanderthal ornaments (the very few ones found, all of them in times of co-existence with us).

## **(III) Shoes:**

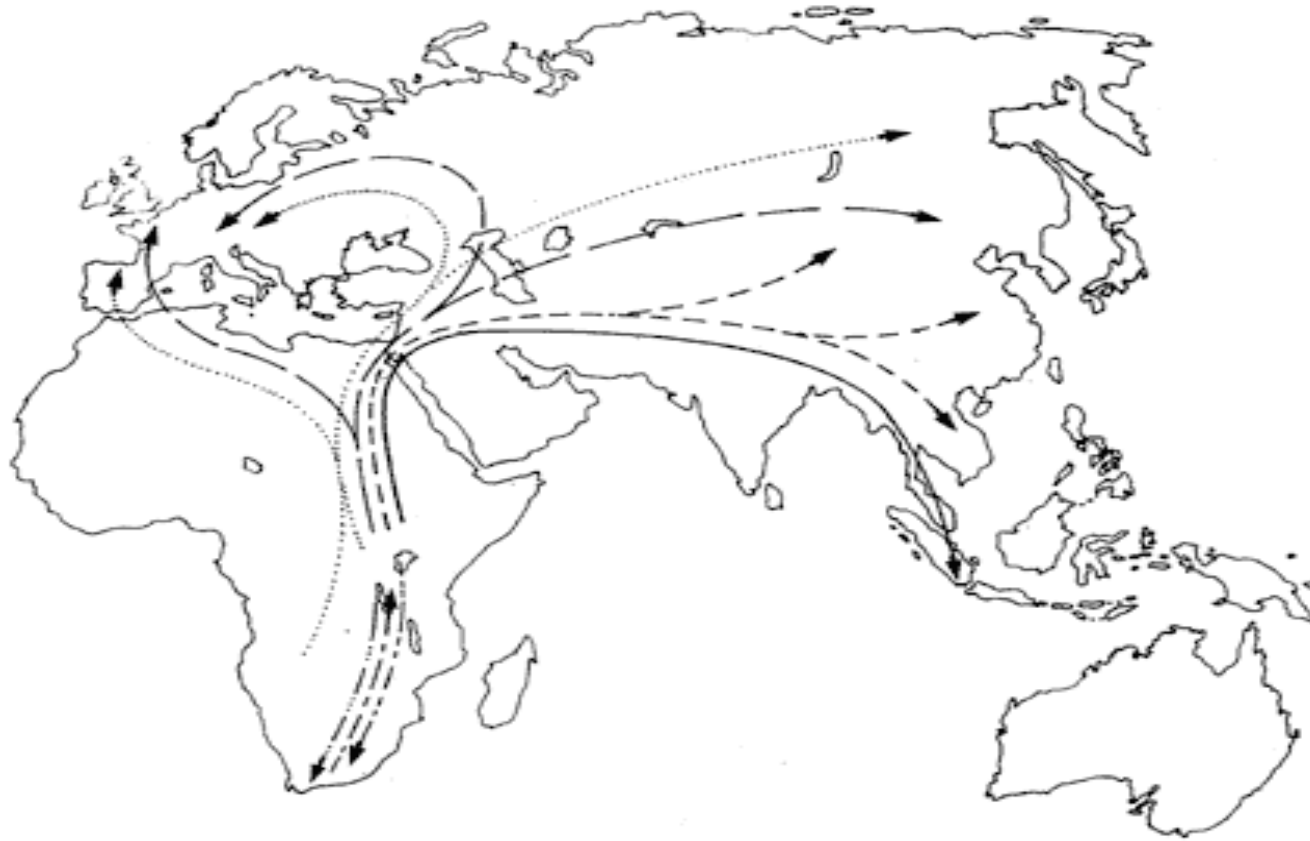
- Recent findings regarding toe shape allow us to determine possible footwear (Trinkaus).
- This passes test of full knot (or you won't ever take off your shoes!)
- NOTE: Latest dating possibly at 42,000 BP (near Beijing, China)

## **Additional Fact IV: Blombos engraving, South Africa dated at 77,000 BP**



Can this pattern of crossing be reasonably generated with Less than a context-free grammar?

# **Additional Fact V: Rapid last human diaspora, between 100,000 / 60,000 BP**



Note also, Chomsky's argument: after diaspora, all human languages share basic UG. Ergo in all likelihood UG was already in place by that time.

## ***Pattern* of eventualities that provide the lower boundary for modern language:**

- Certainly by the time navigation and art are in place (60,000 BP).
- And very likely even by the time when the last diaspora starts and knots can be inferred in multiple domains (110,000 BP?).
- Modern humans are behaving in ways unimaginable without human syntax (in particular context-sensitivity).

# Interesting convergence: *Wynn and Coolidge 2004 on Neanderthals*

- Expert performance via long-term working memory (centerpiece of problem solving).
- Working memory capacity (ability to hold a variety of information in active attention) may not have been as large as that of modern humans.



## **c) Do neural circuits bearing on PDA+ memory exist –are they where expected?**

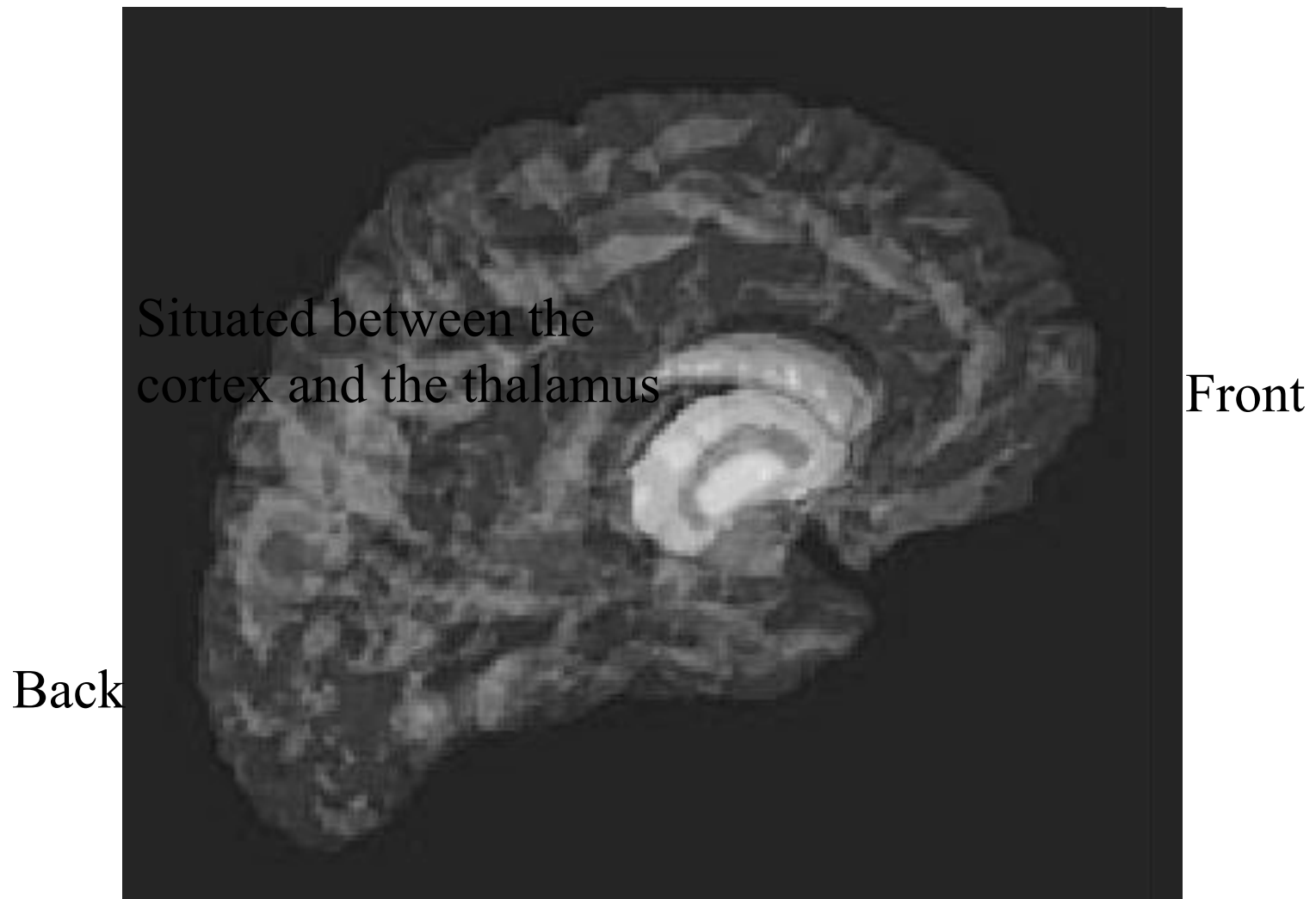
- The short answer is: Nobody knows.
- But recall the Wynn & Coolidge speculation about memory in Neanderthals vs. AMHs...
- Moreover, there's the very intriguing case of FOXP2 in humans...
- It's worth a quick reflection on that (prior to our looking at this more seriously).

# ULLMAN'S CONJECTURE

**Genetic, brain-imaging, and behavioral patterns:**

- SLI-style impairments are associated with dysfunctions of basal ganglia (caudate nucleus) and frontal cortex (Broca's area).
- Frontal/basal-ganglia circuits play a core role in 'procedural' memory.

## The basal ganglia (in color)



<http://www.psychiatry.uiowa.edu/ipl/pdf/globus.pdf>

# ULLMAN'S CONJECTURE

## Procedural Memory

- Rule-governed (vs. idiosyncratic) linguistic mappings can be captured by distinguishing procedural memory vs. declarative memory [Declarative/Procedural (DP) model].
- Idiosyncratic mappings stored in memorized 'mental lexicon' dependent on declarative memory.
- Rule-governed computations involve procedural-memory-dependent grammar.

# ULLMAN'S CONJECTURE

## **Procedural Deficit Hypothesis (PDH)**

- SLI patients [enough to concentrate on KE family, JU]: afflicted with procedural system brain abnormalities resulting in grammatical impairments, lexical retrieval deficits.
- Individuals with Procedural Language Disorder (PLD) should also have impairments of 'non-linguistic functions' that depend on affected brain structures of procedural system.

# If Ullman's conjecture is right...

- Can we connect knots findings in archeology to anything seriously related to language?
- Knots too require serious procedural memory for their execution...
- Again, this is in principle testable.
- Much will be clarified if/when methods to detect gene activity *in vivo* are developed...

# But meanwhile...

- We need ‘model organisms’.
- We need AL models.
- We need to look at these various strands in an imaginative way.
- Asking, to begin with, whether there is anything out there with the right format, so that we understand what the emergence of ‘a system’ might be in other instances.
- Which moves us to...

**d) Can genes implicated in language  
be remotely linked to PDA+  
memory?**

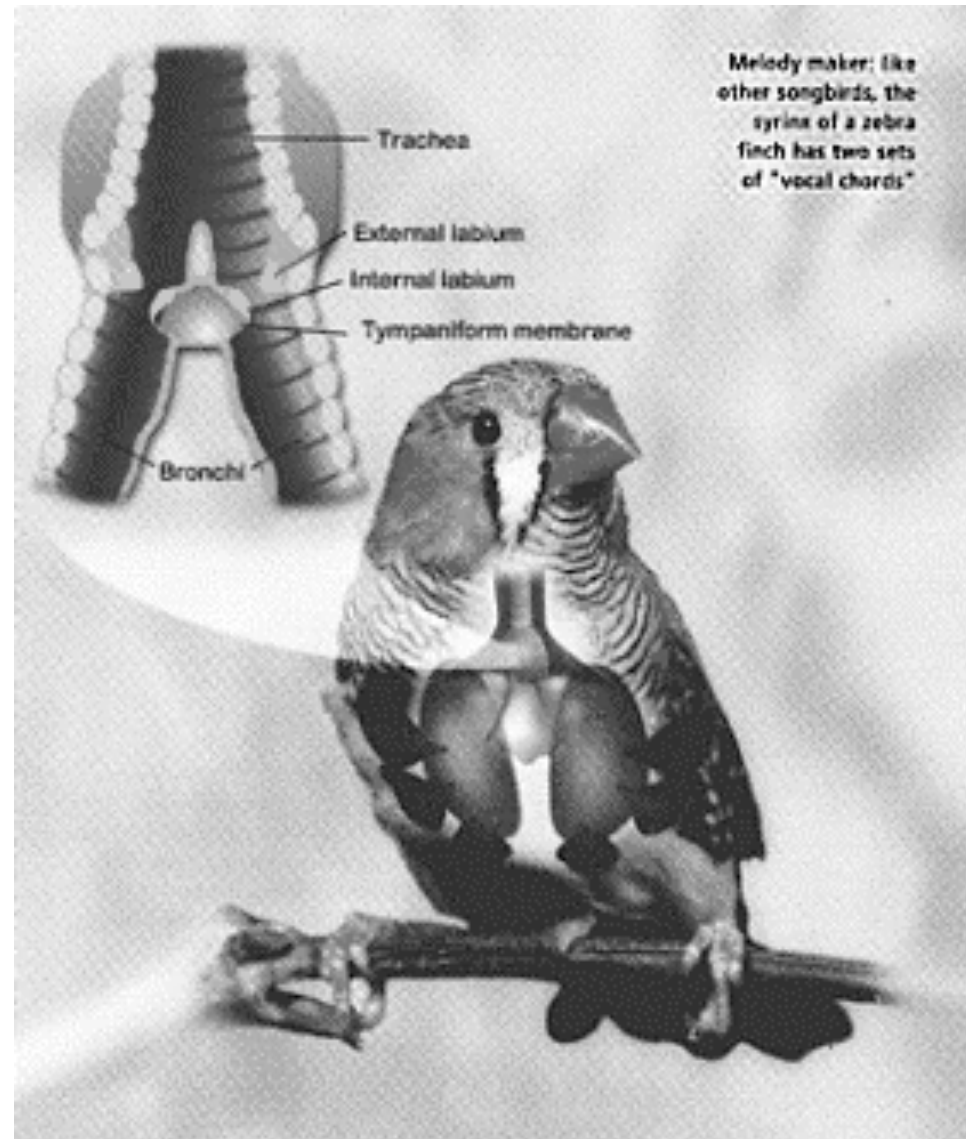
- Who knows –perhaps that computational characterization is at right angles with real memory.
- Then again, is there a better alternative to modularizing the problem and then asking what sorts of computational boundary conditions the problem has?



# What we're talking about...

- Zebra Finch syrinx:

- House Finch song:



**Wouldn't it be great if there were species out there to analyze for some of the basic components of language?**

- Possible (obvious) objections:

- “But isn't language unique?”

Well, sure, but WHAT is the unique part? It could be a combination of non-unique elements, to start with (Gould).

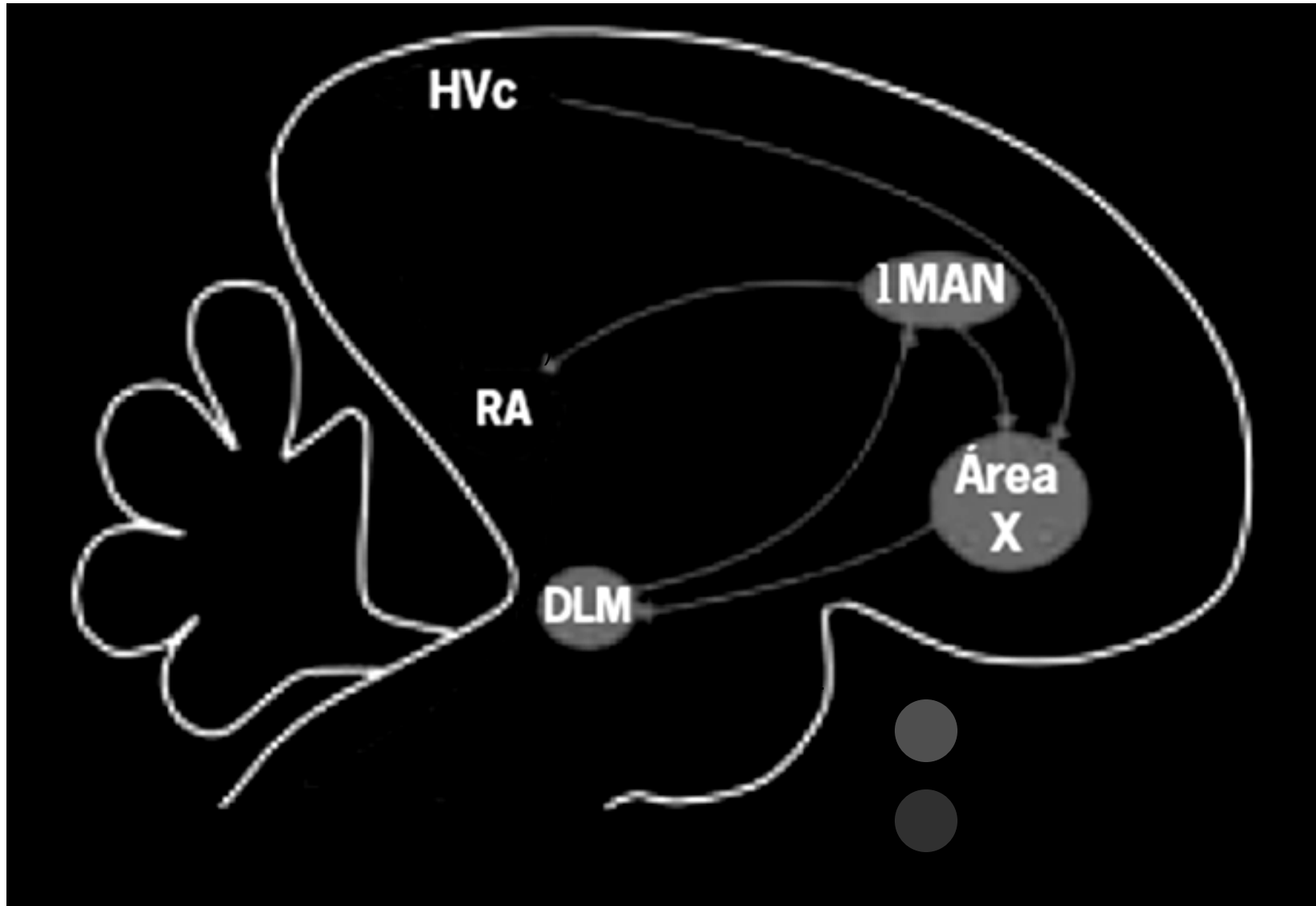
- “But didn't language evolve very recently?”

Ah, but perhaps some of the elements it involves have an ancestral base, which for some reason got to be re-used.

# FoxP2 in songbirds

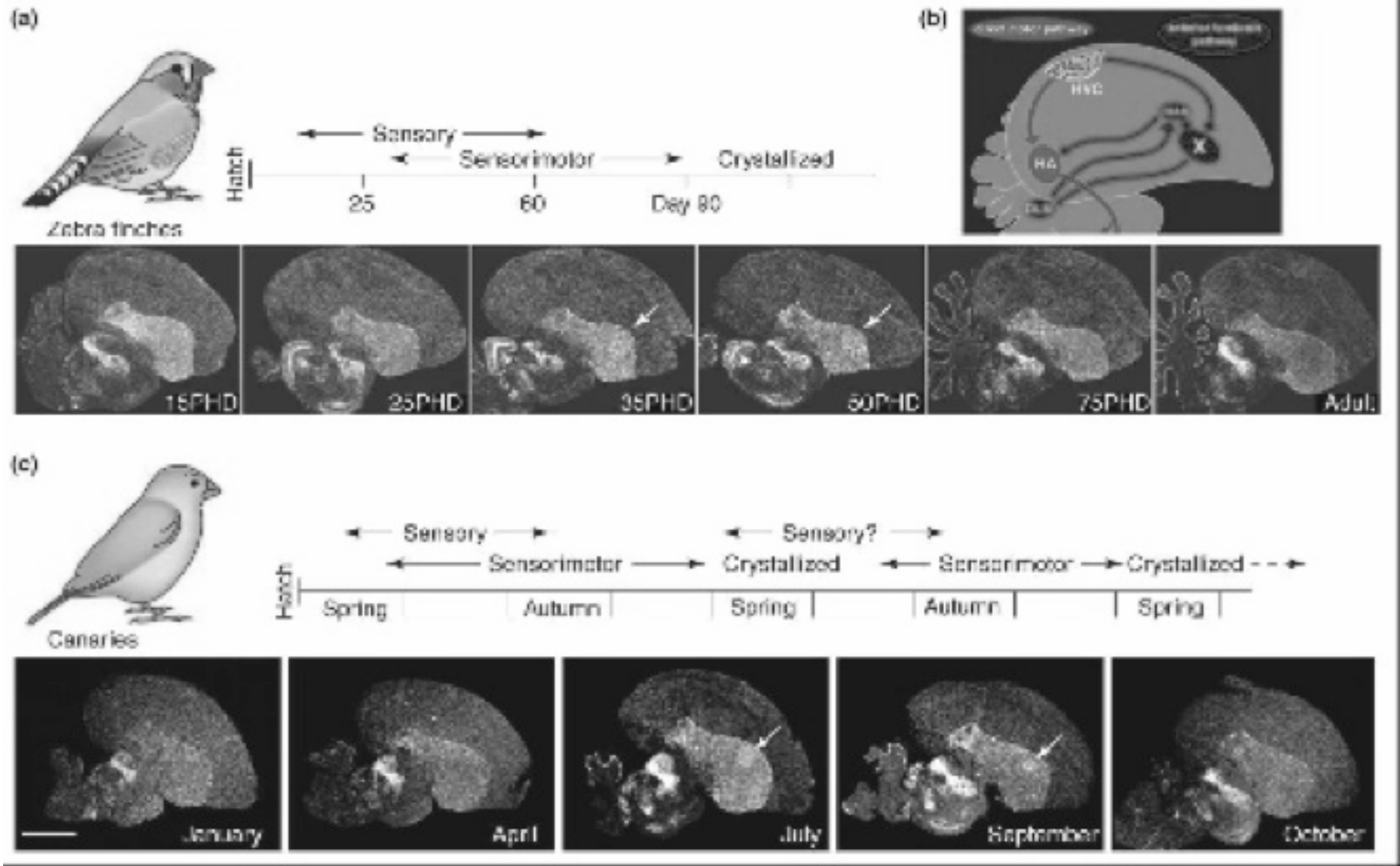
- Although separated from mammals some 300 million years ago, the FoxP2 protein in the zebra finch differs from the FoxP2 protein in mice at only five amino acid positions,
- From human FOXP2 at 8 out of 200,000 positions;
- Moreover, pattern of FoxP2 expression in zebra finch brain is remarkably similar to mammalian brain pattern, including human fetus.
- Consider FoxP2 expression in avian song circuit.

# The basic circuitry in song-birds



# FoxP2 in songbirds

- Not only is FoxP2 expressed in basal ganglia area X and thalamic region DLM,
- Expression in area X **increases** during the critical age (post-hatch days 35-50) at which the bird **acquires song** (Scharff).
- Adult canaries also have higher *FoxP2* expression in Area X during those months of the year when they remodel song.



*Blue = Motor Pathway*

*Red = Anterior Forebrain Pathway*

# Teramitsu & White's results:

- The mRNA of FoxP2 is **sub-regulated** in Area X as the male sings to himself, practicing variants of the song;
- In contrast, when the singing is directed to a female, the mRNA slightly **up-regulates**.
  
- Note: in both instances the same motor control –ergo FoxP2 not motor control.

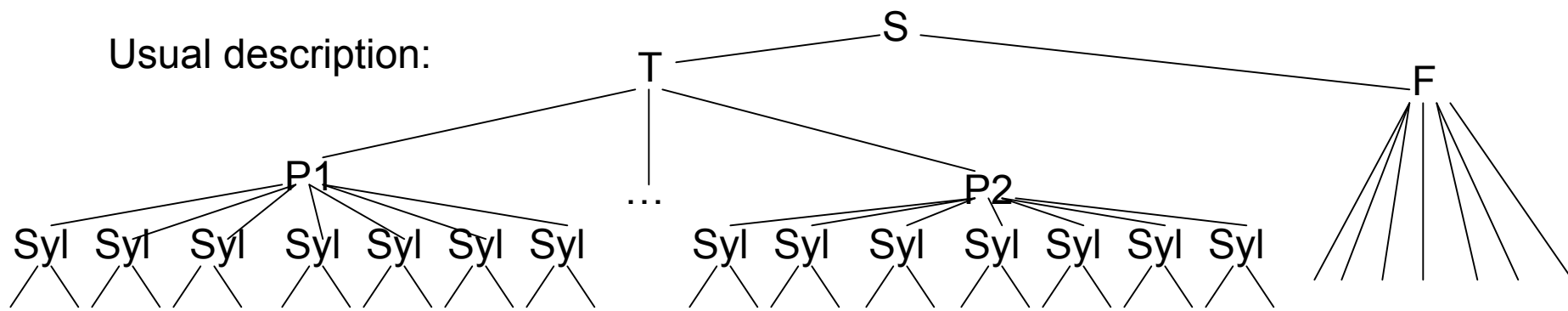
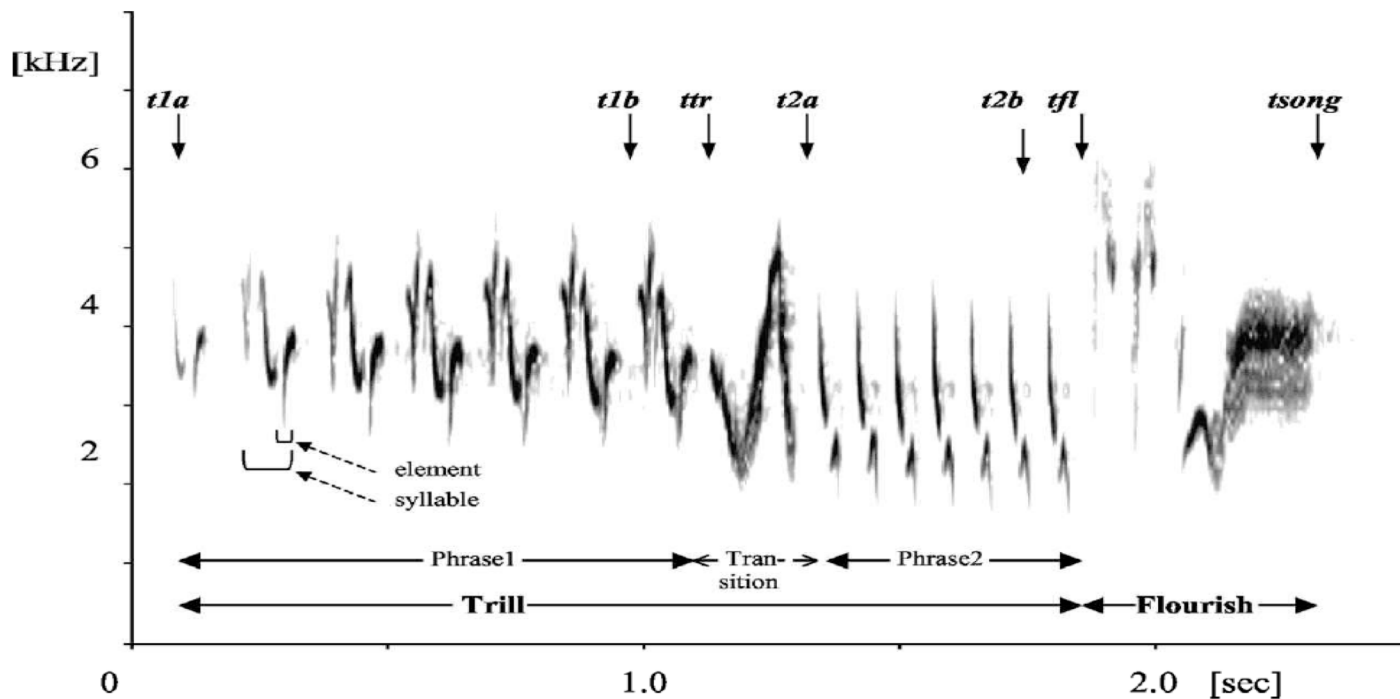
# What is reasonable to conclude?

- Foxp2 is not only active in development
- But also during active (female-directed) song in the adult male.
- It does not regulate breathing as such
- Or the emission of single notes,
- But the **rhythmico-melodic** structure.
- Now: What does that mean?



# A quick look at the song pattern...

(chaffinch song data from Riebel and Slater 2003)



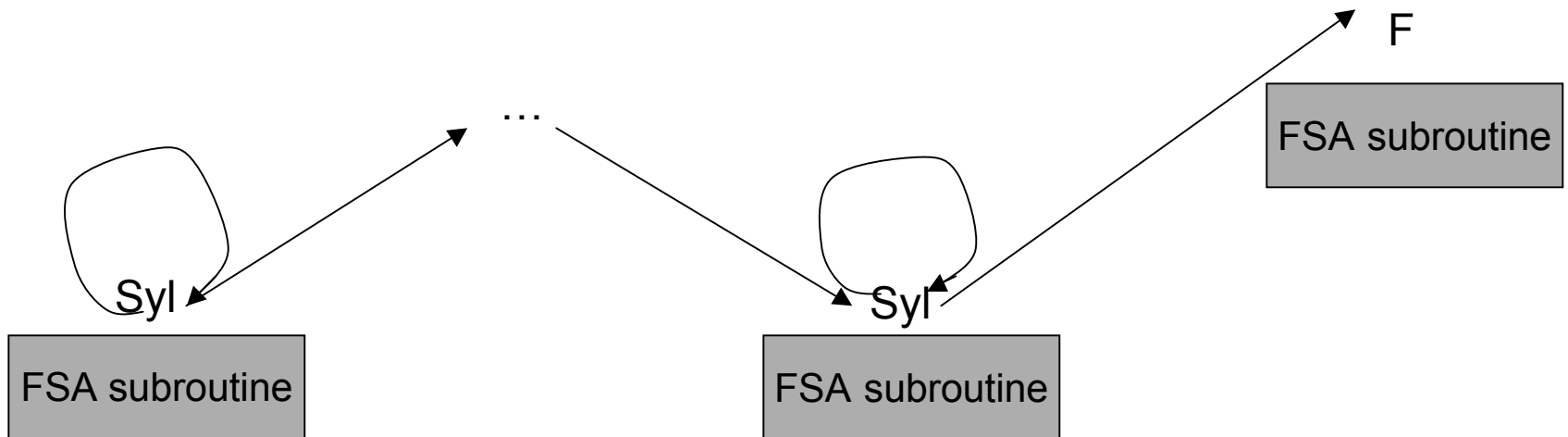
# Typical song in chaffinch:

- Two to four phrases in *Trill*, plus *Flourish*.
- Number of repeated syllables within phrases in *Trill* varies for same song type.
- *Flourish*: complex but roughly fixed sequence of non-repeated elements.
- Song type: if all phrases within *Trill* are the same in shape and order (ignoring number of syllables) and followed by same *Flourish*.

# (Mere) Repetitions suggest a different graphic representation...

Could there be several FSA components?

Note: the loops are just shorthands here (we could devise a more complex FSA subroutine):



# Perhaps, after all...

- We need a PDA only when we demonstrate (center) recursion.
- And how can you demonstrate recursion in a system you don't understand?
- (The trick we use to demonstrate our own recursion involves realizing you can contain an element of type  $X$  *within another*  $X$ .)

# Three important issues:

- Male must acquire syllables during critical period, and basic structure (both Trill/Flourish and overall duration).
- But he develops his own song(s), which female responds to.
- Both male *and female* show a preference for tutor (normally father) song in community (even if they may be focusing on different aspects).
- SO: what is female parsing, and moreover how is male parsing his own positively reinforced song-type, to keep (re)producing it.

# Hard to know what female parses.

## Could it be...

- Number of syllables in (portions of) the Trill?
  - Something in corresponding Flourish?
  - A relation between how the Trill comes out and how the corresponding Flourish does?
  - Something else entirely?
- 
- In any case, she must be parsing, in some sense, a song-type.
  - Moreover, somehow this information comes back to the male.

## **This would be directly possible if...**

- Birds have some sort of PDA, of course.
- Alternatively, some FSA enriched with enough operational/procedural memory to gain required globality, presupposed in reasoning (they need access to more or less whole song).

# Some considerations about FSAs

- Finite-state methods allow the encoding of some memory in the states themselves.
- You can build FSA where each state has no memory of the previous state, memory of one state, memory of two states, and so on –up to any fixed  $n$  states.
- Those are called unigram, bigram, trigram, etc., up to  $n$ -gram (for  $n$  the number of states in the FSA. They each have a larger memory ‘window’.



# Some considerations about FSAs

- The more memory of context, the more accurately a model will capture the distributions underlying the data.
- For machine trained on Shakespeare works, randomly using model:
  - Unigram:
    - Every enter now severally so, let
    - Hill he late speaks; or! A more to leg less first
  - Bigram:
    - What means, sir. I confess she? then all trim, captain.
    - Enter Menenius, if it so many good direction thou art a strong upon command of fear not largess given away, Falstaff! Exuent.
  - Trigram:
    - Sweet prince, Falstaff shall die. Harry of Monmouth's grave.
    - This shall forbid it be branded, if renown made it empty.
  - ...
- N-gram:
  - [the entire Hamlet].

# Possible approach to the bird-song facts:

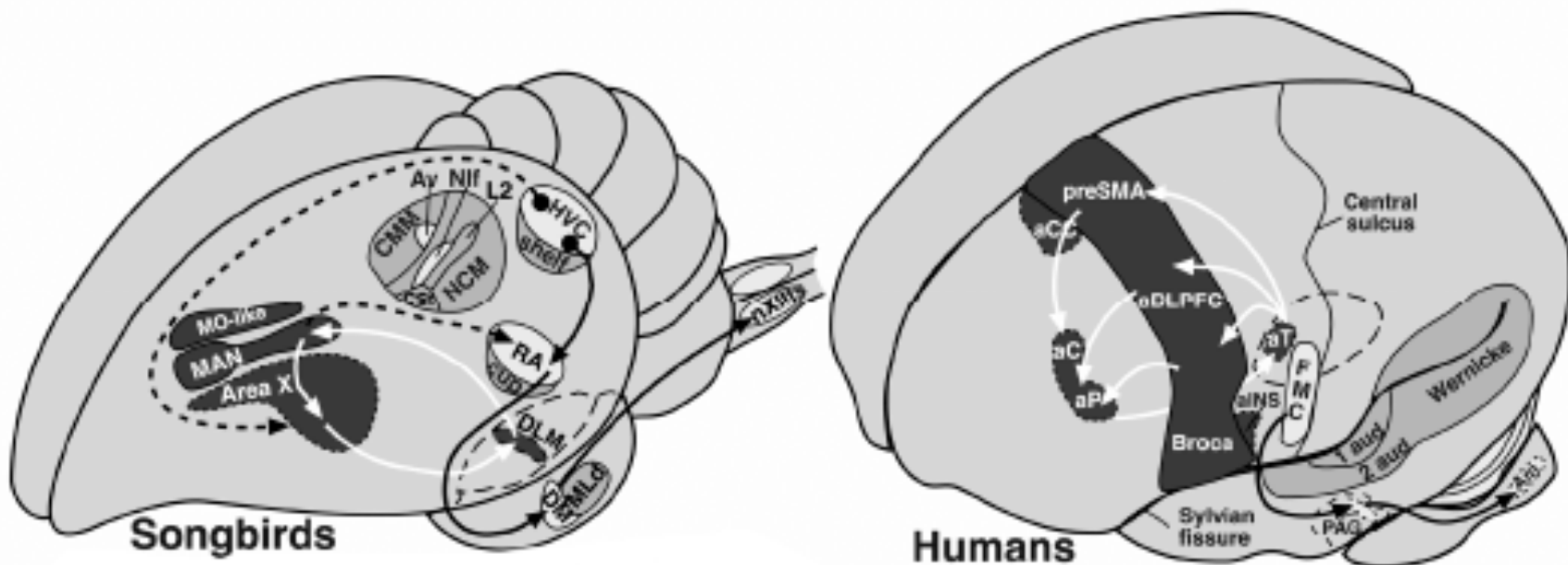
- Upregulating FoxP2 in production circuit, birds make use of *more context*.
- Therefore they produce something that is easier to categorize in those contexts (like a 4-gram or higher pseudo-Shakespeare instance).
- In contrast, downregulating FoxP2 in the production circuit, birds use *less context*.
- Then the output sequence is more variable.
- From the perspective of reception (in the female and feed-back for the male), using more context would lead to more accurate classification of an input sequence.

# If anything like this is at issue

- We should be able to see a FoxP2 up-regulation *in the female* (non-singing) finch, as she listens to a song.
- And in principle there may be more than one phase in the male singing, perhaps with various degrees of regulation:
  - (i) A ‘try-this/try-that’ phase.
  - (ii) A ‘gocha!’ phase after success.

# Comparable brain areas between songbirds and humans (left hemisphere)

- White regions and black arrows: posterior vocal pathways.
- Dark gray regions and whiter arrows: anterior vocal ones.
- Dashed lines: connections between the two.
- Light gray: auditory regions. (From Jarvis 2004)

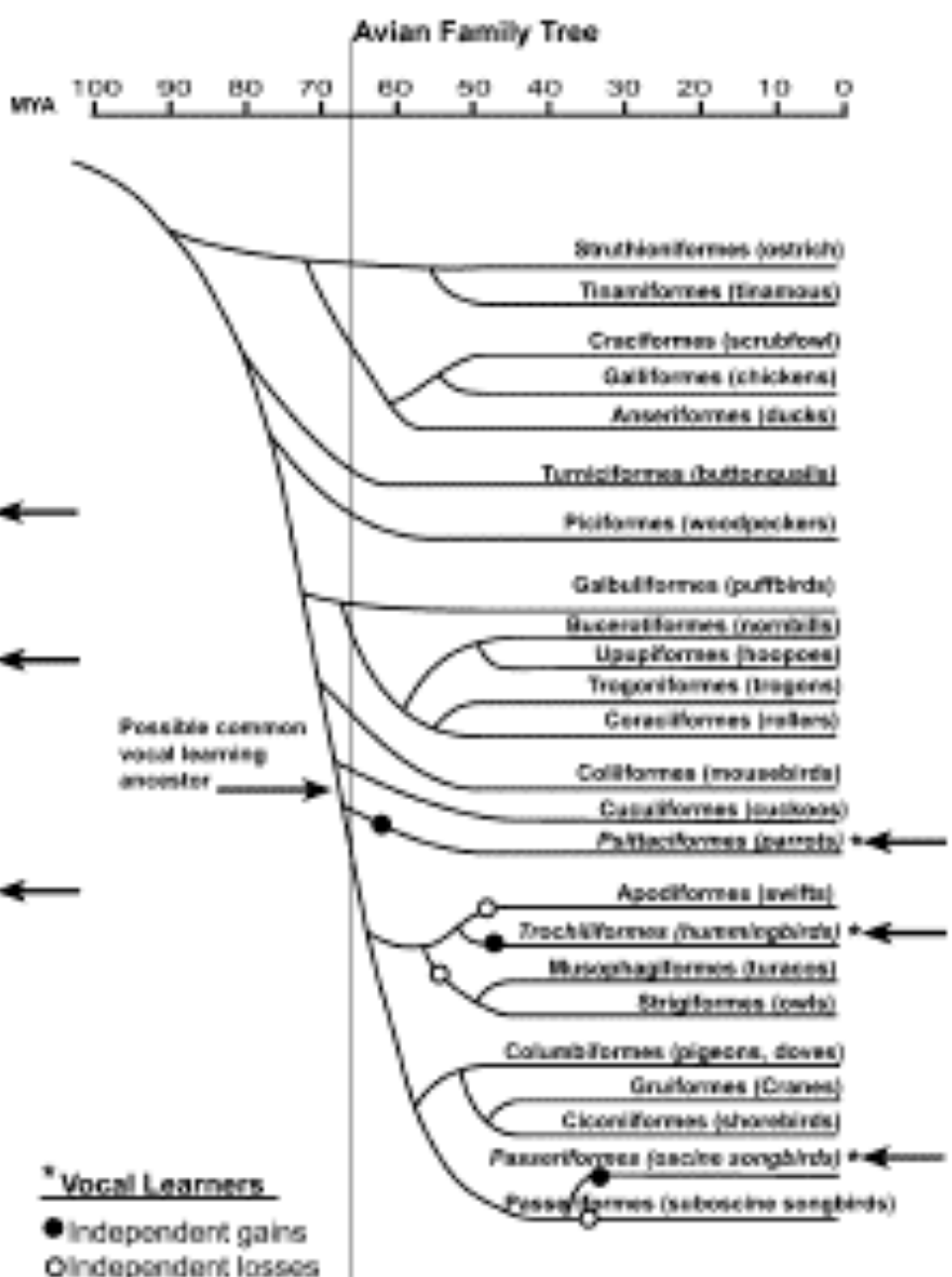
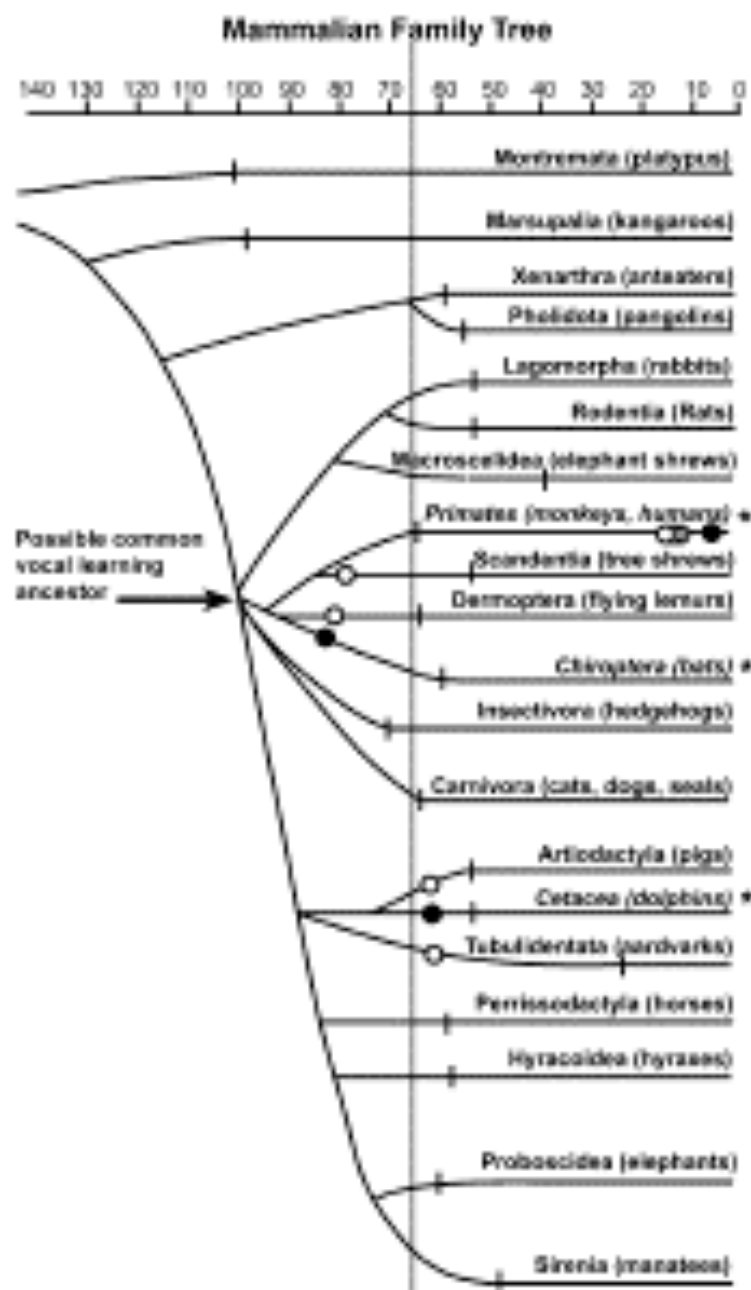


# **Could the analogies between these brain circuits be accidental?**

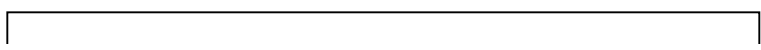
- If not, what is the role that FoxP2/FOXP2 is playing in the brain circuitry?
- Both at development and (at least in the case of the birdsong) at performance.
- And bear in mind that acquisition (in both birds and humans) presupposes a parser of as much complexity as the one used for performance.

# Could Ullman's procedural memory be of the sort the presupposed parser needs?

- After all, we were independently reaching the possibility that, at least in song birds, FoxP2 regulates a given parsing window.
- If so, regulation of the parsing window will affect both acquisition and performance.
- This is testable in principle (we expect Area X, connecting both circuits, to be key).
- And bear in mind that the relevant role for FoxP2 was turned on and off various times.



\* **Vocal Learners**  
 ● Independent gains  
 ○ Independent losses  
 - Everybody has it to varying degrees



**Don't Teach Our Kids  
They Come From Apes**







*Our true evolutionary lineage!*