The Song and the Word
AAAS 2008

Paraphrasing Dick Lewontin's opening line in my favorite earlier written version of his remarks, "If it were my job to say what is wrong about Dick's I could stop at the end of this sentence." So my job here today is really going to be two-fold, an opera in two acts: Act I, to provide some hard detail, some DNA-on-the-bones as it were, to press home Dick's point that in the case of cognition, specifically human language, one would be hard pressed to come up with an example less amenable to evolutionary study. Specifically, I am going to drill down into the case of "FOXP2", a transcription factor whose two amino acid differences between us and other primates (and whose commonality with, now, Neandertals) has been touted as part of the 'push' to language. And I'll try to show that the argument here falls apart statistically and biologically on a point that Dick has so often stressed: namely, that when there are so few differences at such a far remove from a phenotype, and so much intervening time, it sometimes can be just plain impossible to tell why something happened, whether it was 'selected for' or not. There are so many parameters, and it's so chaotic, it's like weather forecasting. Assumptions have to be made – about population size and changes, generation times, selection coefficients. This is well known – but. What happens if the assumptions are off? So that's what I've looked at: I cast a broader net, re-did the weather forecasts. Used a finer mesh. And the scientists in this case might have been better served to have simply reported, 'cloudy with a 50% chance or rain." Or, this being Boston, perhaps they should have said nothing at all. One simply cannot tell whether Foxp2 has the stamp of selection on it. The noise overwhelms the signal. Game over. End of any interesting story. What's more, as soon as one digs into the biochemistry of a regulatory component, a transcription factor, like Foxp2, it's just as easy to spin an explanatory evolutionary tale that

And Act II:

To begin. Nowhere is the evolutionary explanatory challenge more pointed than in the case of human language. What even do we mean by 'human language?' I'll proceed to what Noam Chomsky's view is about this in just a moment, but, whatever we think about language, I think that it's fair to say that it's what biologists call an 'autapomorphy' or a trait unique to a single lineage. As Dick says, have no close living relatives among species, so comparative analysis, the mainstay of the evolutionary program, becomes extremely difficult.
I begin with language evolution redescribed as hypotheses about characters and character states (made explicit with cladograms). Using this representation and terminology of cladistics invites a discussion of when selectionism a good explanation (i.e. a discussion of multiple independent autapomorphies). I stress that what can you say about a unique autapopmorphy is whatever you want to or nothing. And nothing is more compact.

To explain differences we should choose exemplars that are as close as possible, to explain similarities we should choose exemplars that are as far apart as possible. Why? Well, what would the best evidence for the adaptationist account of language be? What character state distributions since the common ancestor of gorillas, chimps and humans would shed light on language evolution? Finally I raise the possibility that deep homologies (as found increasing in developmental genetics) annihilate most current selectionist accounts.

Much excitement has followed from the full genome sequencing of our nearest living relative, the chimpanzee, with other primate genomes to come. However, the special problems of evolutionary inference given close neighbors suggests this may tell us very little about human cognitive faculties such as language. The well-known example of a putative 'language gene', FOXP2, is a prime example: the differences between us and chimps could be due to chance alone. Where then can we look for insight? The most recent research by Halle in language sound systems combined with Chomsky's most recent model of syntax provides a possible and so far unexplored connection to birdsong.

An Opera in 3 Acts.
Act I, 15 The Incredible Lightness of Being an Evolutionary Argument.
Would like to clear up some possible misunderstandings about what Dick likes to call "Vulgar Darwinism" – something you may have picked up like a virus from reading certain popular nonfiction books in airports – about evolution, genes, natural selection, and language.
How to go about properly arguing for the 'selection for' some trait, why we don't want to look at chimpanzees or other primates, and why Foxp2 story looks at first blush like its headed for disaster. However, the heroine will come try to rescue the fox in the last act.

What's wrong with the FOXP2 story

The basics: Ka/Ks. Why it fails. All the parameters!
The simulation: re-read Enard. Do they use PAML? Yes, for likelihood ratio. But Nature chimp-human genome article says no diff from mouse. Have large negative Tajima D. but…
Show all the paramters: transition-transversion, Show they have a grid, it's like weather forecasting, chance of 'rain' across 100,000 years. Simulation based on T= 1000 generation grid. Change to finer Grid, T=100, get different result. Like doing better weather simulation. Draw picture. Also coincides w/ putative human population explosion. No stochastic equilibrium. Too sensitive to parameters: change selection to more reasonable value and change grid simulation, get difft results.

The complexity (from slides) of FOXP2. What is the lesson here? Fact: FOXP2 is expressed in different tissues besides the brain. Is FOXP2 somehow involved in diet?? (Q: is it expressed in the digestive tract? -- don't know.) Maybe when proto-humans became hunters and started to have more meat in their diets there was selective pressure for the Asn325 --> Ser mutation? [Note: forkhead box-type transcription factors are thought to sometimes have one role during development and a totally different and separate role in adult tissues.] Thus, did the Ser325 mutation in FOXP2 happen first and "prime" the system for the presumed language enhancing effects of the Thr303 --> Asn mutation (e.g., because of stabilizing side chain - side chain interactions in the tertiary structure), or did the Asn303 mutation just arise co-incidentally?? (not likely, but possible) If so, will the Asn303 change eventually be found in other species, too, separate from the Ser325 mutation?

Are forkhead box-type transcription factors part of the "on-signal" apparatus coordinating temporally-spaced developmental programs? (What does the bird knock out say about this?) For example, a slow ramp-up in the intracellular (intranuclear) concentration of a FOX transcription factor might sequentially turn on different genes by acting first on those for which it had the highest affinity to the promoter regions, and subsequently on those for which it had progressively lower affinities. We need to know the 'circuitry'. Emphasize this picture from

the webpage. For FOXP2 mutants, some of the delayed development phenotypes seen for language (in humans) or vocalization (in mice) -- and the dominant, haploinsufficiency characteristic of the phenotypes --are consistent with this, but many more experiments will have to be done to shore up this hypothesis.

FOXP2 as "third factor" principle: language (NFL), syntax is strictly internal.
Linearization. Simple example: Mary likes pizza and John likes beer. Projected onto time sequence stream. Recruitment of sensorimotor systems: this is FOXP2.

Birdsong: hierarchy here depends on structure. What is hierarchy in birdsong? Unclear.

Act II: A Venetian Fairy tale

Act II: Optimal Parsing Just is Merge: If we associate parsing operations transparently with merge operations, phase by phase, we replicate the model that Weinberg and I proposed nearly (gulp) 30years ago, and it fits Luigi's bill perfectly: namely, it will allow us to compute necessary relations locally, and with the optimal time complexity, namely, deterministic and in real time (what does this mean – essentially, no arb. long pauses) This is the best possible result. Details to be worked out, I think in essence it all works. (inverse of linearization in Kayne's sense).
Antisymmetry in Kayne paper: sequence in time IS a PF interface property – third factor, so whole FOXP2 story is about third factor principle, NOT 'interior' syntax, or language.
Have to re-read bird knockout to see how they fail.

Act III: the return of the fox and the birds. We've seen Anna Maria tell us that Merge operates at the level of morphology. I want to take that one step further, and articulate the Syntax-SM interface into several layers, show that Merge operates at the level of Halle-Idsardi metrical structure, but with a twist: that will possibly give us a new insight into the last, and biggest question, namely:

Where did the language "Big Bang" come from? – a Venetian fairy tale. (a continuation of Noam's tale). that more and more organisms have the 'bricks and mortar' for language – talents like making tools, 'reasoning' about causality, singing, and so forth. - so what do we got that they don't? Answer: well yes, Merge, BUT Merge combined with these other parts in a novel way.

Act I.

In order to understand language from an evolutionary perspective we need to begin by understanding in what way evolutionary research differs from the model of most scientific work. The study of evolutionary processes must confront special difficulties in both the conceptual and the methodological aspects of research. On the conceptual side, unlike for molecular, cellular, and developmental biology, there is no basic mechanism that evolutionists are attempting to elucidate. There is no single cause of the evolutionary change in the properties of members of a species. Natural selection may be involved but so are random events, patterns of migration and interbreeding, mutational events, and horizontal transfer of genes across species boundaries. The change in each character of each species is a consequence of a particular mixture of these causal pathways.

Second, evolutionary biology is particularly plagued by the methods that are available to answer the questions posed. This has not been simply a matter of waiting until an appropriate methodology is developed, although there are indeed such cases, but for many problems the difficulty lies in the relationship between processes that have occurred in the past and the evidence about those processes that can only be recovered from presently existing organisms, with a little help from fossils. In the first place, the fossil record is so sparse that no quasi-continuous process of change and diversification can be followed as a detailed dynamic process. Second, the fossil record provides evidence primarily of morphological change, and nothing can be reconstructed about underlying genetic change and only rarely about physiological and biochemical changes. Nor can living species be used, except under exceptional circumstances, to follow evolutionary changes now occurring. Processes of change and divergence of species are usually extremely slow compared not only to the lifetime of an investigator but of science as an institution. This low speed is, in turn, a consequence of the weakness of most evolutionary forces most of the time. It is now generally agreed that selection differences in nature for most of the gene variants segregating at most loci are likely to be of the order of 10^{-3} or less. (That is: roughly a 1 out of 1000 difference, where one individual produces 1000 offspring, another 999.)

The consequence of the weakness of selective and random forces is that the processes of evolution in living species cannot, except very rarely, be followed as a dynamic process in time. Instead, the evolutionary biologist must depend on static data, observations of patterns of variation within and between species, to infer the dynamic processes that could not be directly observed. The other methodological apparatus for reconstructing past events from present organisms is meant to estimate the actual magnitude of forces of natural selection and random events that have led to genetic differences between closely related species or to genetic differences between populations of the same species. It involves a mixture of the mathematical theory of population genetics and statistical theory. It begins with a complex mathematical apparatus that is designed to carry the state of a population forward in time from some initial condition. It predicts rates of genetic change from an initial state and possible equilibria that will result from selection, mutation, migration, and recombination. This must be a stochastic, rather than a deterministic, theory to account for random changes that result from genetic drift in finite populations, so that the form of the prediction is not a unique state at future time, but a probability distribution of states. This is the stochastic theory of population genetics originally produced by Wright and Fisher and further elaborated by Kimura & Ohta . [picture here]

Second, a probabilistic theory is needed that can reverse the deductions of the first theory and infer backwards in time from a particular observed state at present what the most likely dynamical forces were that have led to the actual present situation. But a difficulty arises here. A dynamical theory that predicts the present state generally requires that we know not only the nature and magnitude of the forces that have operated, but also the initial condition and how long the process has been in operation. That means that if we wish to use a backward inference from the present state – the data we can collect, in this case, the pattern of amino acid differences between us and chimps and mice – to estimate the forces that have operated, we would need to know the initial condition and how long the process has been going on as well as assuming that the forces have not changed during the process. But this is precisely what we cannot know. Either we assume that we know the forces, in which case we can make probability statements about the initial conditions, or else we assume that we know the initial conditions, in which case we can make estimates of the forces that have led to the present. We cannot do both. There is one solution to this dilemma, and that is what researchers are forced to assume, even though this assumption usually goes completely unstated in the usual tests for selection: If the evolutionary process has gone on for a sufficiently long time with no changes in the forces, then there is an equilibrium probability distribution of the present states, the so-called steady-state distribution, that is reached irrespective of the original state of the population. What this means is that all the effects of the initial state have disappeared – like dropping a pebble into a pond at time 0, and saying that the ripples have died away by the present day. So, if we can observe many genetic variations all of which can be assumed to be the result of the same forces, then the distribution of those variations can be used to estimate those
forces.

Two methods: (1) count # survivors. Old subjacency and sex problem. Etherized drosophila.
(2)

Variation is central to evolution. For evolutionary biology, it is of the essence. That was Darwin's achievement; not 'evolution' (known back to Aristotle, Lamarck); not descent w/ modification. Rather, as Lewontin stresses: notion that

Point of contradiction with minimalism: there is no such thing as an 'ideal type' to which an individual or set of individuals strives to attain. So single, perfect 'crystal'.

Conclusions: what did we try to show?
1. Caution: one cannot do 'armchair evolutionary theory' no more than one can do armchair linguistics. Many folk intuitions flat-out wrong.
In particular,

2. Showed there is no substantial conflict between Merge and optimal parsing: while true that there are senses in which parsing cause lots of search (familiar filler-gap), if we identify merge phase by phase w/ a particular transparent design of a strictly left to right, deterministic parser, we can obtain the best possible match between merge and the external parser interface, namely nearly identity, plus real-time (better than linear time) parsing - the optimal result. There is no way computational way to do better than this.

3. Articulated the S-M interface by unraveling PF spell-out into several layers, and then push merge down to the level of PF. In particular, the Halle-Idsardi system for building metrical structure just is merge, with the additional a single, minimal new predicate, adjacency. Key idea: pure syntax, merge operating w/ no features, plus adjacency
Bonus: new theory of parameter setting for metrical structure

4. Extra bonus: if true, allows us to possibly rescue the Foxp2 story in a dramatic, though perhaps fanciful way.