Songs to Syntax: Cognition, Combinatorial Computation, and the Origin of Language

Language comprises a central component of what the co-founder of modern evolutionary theory, Alfred Russell Wallace, called "man's intellectual and moral nature" – the human cognitive capacities for creative imagination, language and symbolism generally, a complex that is sometimes simply called "the human capacity." This complex seems to have crystallized fairly recently among a small group in East Africa of whom we are all descendants, distinguishing contemporary humans sharply from all other animals, with enormous consequences for the whole of the biological world, as well as for the study of computational cognition. How can we explain this evolutionary leap? On the one hand, common descent has been important in the evolution of the brain, such that avian and mammalian brains may be largely homologous, particularly in the case of brain regions involved in auditory perception, vocalization and auditory memory. On the other hand, there has been convergent evolution of the capacity for auditory-vocal learning, and possibly for structuring external vocalizations – apes lack the abilities that are shared between songbirds and humans. Language's recent evolutionary origin suggests that the computational machinery underlying syntax arose via the introduction of a single, simple, combinatorial operation. Further, the relation of a simple combinatorial syntax to the sensory-motor and thought systems reveals language to be asymmetric in design: while it precisely matches the representations required for inner mental thought, acting as the "glue" that binds together other internal cognitive and sensory modalities, at the same time it poses computational difficulties for externalization, that is, parsing and speech or signed production. Despite this mismatch, language syntax leads directly to the rich cognitive array that marks us as a symbolic species, including mathematics, music, and much more.

It seems appropriate to open a conference on cognitive informatics and computing with a talk about the origin and nature of that part of cognition that seems to be uniquely human, namely, language. There can be no doubt that language comprises a central component of what the co-founder of modern evolutionary theory, Alfred Russell Wallace, called "man's intellectual and moral nature" – the human cognitive capacities for creative imagination, language and symbolism generally, a complex that is sometimes simply called "the human capacity." In short, language makes us smart. In what follows I would like to sketch how this remarkable ability arose during the course of evolution, and exactly how language boosts our cognitive capacity beyond that of all other animal species. To do this, I will first have to outline a bit of what we know about the evolution of modern humans. This will give us some important clues as to what marks out language something uniquely human. This will lead naturally to a brief review of what it is that we humans have that other animals don't – a sweeping, floodlight intelligence, what the paleo-anthropologist Ian Tattersall calls "flexibility instead of specificity in our behavior." After all, ants or bees can easily beat us at navigation, and it seems from recent studies that songbirds can do better than us at auditory production and perception. As Tattersall (1998) notes, "Over millenia now, philosophers and theologians have made some- thing of an industry of debating the human condition. Even if inevitable, it is rather ironic that the very species that apparently so much enjoys agonizing over its own condition is, in fact, the only species that doesn't have one—or at any rate, whose condition, if any, is most difficult to define. Whatever condition is, it is surely a lot easier

to specify it in the case of an amoeba, or a lizard, or a shrew, or even a chimpanzee, than it is in our own." (1998:197).

Remarkably, it will turn out that human language seems to arise from just a single, small evolutionary innovation, built on two already-available cognitive substrates, present separately in other animals, but brought together for the first time in modern humans. So human language is not just 'more of the same', to use Tattersall's (2010) words, but involves something entirely new, "how we integrate" existing elements.

In fact then, contrary to what is sometimes thought, human language is *not* complex – on the contrary, it is far simpler than anyone may have thought, certainly simpler than what one reads about in standard linguistic textbooks. But it *is* novel. On reflection, this is not at all surprising, given the relatively short time scale involved in evolutionary terms – not millions of years but just 50 thousand years. Complicated evolutionary change typically occurs over the time span of many thousands or millions of generations. We would therefore expect any such that occurred in mere blinking of an evolutionary eye to be relatively small, since it seems to have occurred within the time of a few hundred generations, and a hundred generations already takes us back to the founding of the Roman Republic. There simply was not enough time to evolve something as radically new and complex as, say, the wings of birds. As always, evolution by natural selection had to make do largely with what is at hand. Once unleashed, language serves as a kind of *lingua franca*, a "cognitive glue" that lets all our other cognitive faculties talk to each other, in a way that is not available to other animals. And applied to other human, digital cognitive domains, it leads numbers, mathematics, and music. All this – the result of a

single evolutionary innovation, quite unlike any other animal. So how did this all happen?

Before beginning it is worthwhile to clear away two common misconceptions. First, I am not claiming that 'thought' is co-extensive with 'language' – that they are one and the same thing. Obviously they are not. Why? We all know that we can have language without thought, as is demonstrated to use every day by politicians. [SLIDE]. Conversely, there can be thought without language, as evidenced by this remarkable example of visual computation [SLIDE FEYNMAN diagram]. Nonetheless, it is clear that language plays a large role in our mental lives. Putting this matter to one side then, let me turn to a brief review of the paleontological record regarding us and our immediate ancestors, as we currently understand it.

This slide [SLIDE] shows a picture of the 'family tree' of our recent homin ancestors, distinct species, with time stretching back to 5-7 million years ago at the bottom, to the present at the top. There are two crucial things to take note of in this figure. First, like virtually any other tree of related species, all the family *Hominidae* – it is very bushy, just as Darwin taught us. In fact, there are approximately 49 other 'Hominidae' species that are not drawn in this particular diagram, simply because we don't have enough data currently to figure out where to put them. So, there have in fact been at least over 100 distinct species in our immediate family tree. Many of these die out, after making their brief appearance on the evolutionary stage, just as Darwin suggested. Second, at any one time in the past there have typically been several, often many hominids that co-existed – for instance, Homo sapiens (modern humans); Homo neandertalesis; and Homo erectus near the top here – what is unusual is that at there present time we have no other living

relatives – there is just a single Homo species left alive in the world after millions of years of co-existence: us.

Second, I would also like to emphasize that the emergence of language simply cannot have been due to something like this, that we have discovered using better brain imaging. [SLIDE – SIMPSON] While it is true that there has been a *general* increase in brain size throughout the primate lineage, as we shall see, language cannot be the result of brain size alone, since Neandertals were bigger brained than us.

So what has been the history of our hominidae ancestors? Following Tattersall (2011), perhaps most strikingly, the several-million year period before the appearance of clearly behaviorally modern Homo sapiens approximately 75 kYa is marked in general by a 'disconnect' (Tattersall, 2011) between the appearance of each new hominid species and new technologies as evidenced by differences in, e.g., stone tool making. That is, most often a new species appears on the scene (with a different body morphology and larger brain capacity, etc.), but *without* any concomitant innovative change in external behavior, a disconnect depicted in this [SLIDE]. For example, there is nearly a 1 million year 'gap' between the appearance of the first, Type 1 'scraping tools' about 2.5 million years ago and [SLIDE] and the type 2 ovoid Acheulean tools, appearing about 1.5 million years ago; crucially, these post-dated the apperance of the hominid that first made them, *Homo ergaster,* dating from about 1.9 million years ago. "As far as can be told, aside from the invention of the Acheulean in Africa (its spread beyond that continent occurred considerably later) the history of the genus Homo in the period between about two and and one million years ago the Old World, but without radical physical or as far as we can tell cognitive innovation. It is not until 600 thousand years ago that we find, again first in

Africa, a new kind of hominid with a significantly larger brain…This is *Homo heidelbergensis*."[SLIDE].

Similarly, even though "anatomically recognizable" *Homo sapiens* appears about 200 – 150 thousand years ago in Africa, they come "bearing a technology that was basically indistinguishable from those of its contemporaries and immediate predecessors", again a "disconnect" between anatomical and behavioral innovation. Remarkably, as depicted in this perhaps fanciful scene [SLIDE], at one time in the Levant, at least 3 distinct species of *Homo* lived side-by-side for a hundred thousand years – *Neandertals, heidelbergesis, and sapies* – all at apparently the same level of tool-making.

All this changed starting about 75 thousand years ago, with the appearance of behaviorally modern humans in Europe, also *Homo sapiens,* known informally as Cro-Magnons, whom almost certainly arose from a second wave of migration out of Africa, and who displaced Neandertals in site after site in Europe. Indeed, the contrast between the Neanderthals and Cro-Magnon provides perhaps the best picture of the differences between a cognitively adept species seemingly operating at the maximum level possible *without* something like language, the Neandertals, and a species – in effect us – who had already embarked upon a path of constant cultural and creative change that continues to the present day.

One side-by-side look at the skeletons of Neandertal vs. Homo sapiens suffices to point out the remarkably different skull, thorax, and pelvis shapes; and in this regard, Neandertal more closely represents the common ancestor, highlighting the extensive, higly derived changes that took place yielding modern Homo sapiens. Note, however,

that if one considers simply brain capacity, that in fact Neandertals had, if anything, a *larger* cranial capacity. They evidently hunted in groups, perhaps even more effectively than Cro-Magnons (consider the metabolic needs); used animal skins and built shelters, and much more. But, with the advent of Cro-Magnons and modern humans, something quite different appears on the scene. In contrast to the scant evidence of any symbolic behavior, with the coming of modern Homo sapiens, there is a virtual explosion of artifacts that all say that this species was just us, including the first sculptures of remarkable aesthetic skill [SLIDE]; sophisticated musical instruments [SLIDE]; the first 'written' records on plaques and bones; and the astonishing images on the Chauvet caves in France that you may have seen in Werner Herzog's new film [SLIDE].

"What may thus on the face of it seem more remarkable is that we do not see any convincing evidence of symbolic behavior – and certainly no indisputable symbolic artifacts – until long after anatomically recognizable Homo sapiens had arrived on the scene. In the Levant, anatomically distinctive early Homo sapiens behaved, as far as can be told, pretty much as the Neanderthals had done for tens of thousands of years both before and after the episode of occupation by early moderns; and earlier Homo sapiens or near- Homo sapiens in Africa are invariably associated with much more archaic stone tool industries than those characteristic of the Cro-Magnons in Europe. Modern human behaviors, then, began to be expressed only when modern anatomy had already been long established; and we thus have to make a conscious mental effort to distinguish between "behaviorally archaic" and "behaviorally modern" Homo sapiens. This is so even though there is no way to distinguish between these two forms of Homo sapiens in zoological taxonomy."

So what happened? What unleashed this astonishing creativity? Something that, as Alfred Wallace took pains to point out, was not something merely 'more of the same' on an incrementally graded linear climb up from the early bi-pedal apes. To answer that question, it is helpful to examine the abilities of other living animal species. Clearly, some other animals possess formidable cognitive skills. We saw that our pre-Homo sapiens ancestors made increasingly sophisticated tools, albeit at a glacial pace. It was once thought that this distinguished us from other species, but this has long been proved false. Birds, especially the corvids (ravens, crows, etc.) make sophisticated tools and can engage in what seems to be quite sophisticated causal reasoning.

[SLIDE western scrub jay]

The western scrub jay

[SLIDE carrion crow, Japan]

Finally, as Aristotle appreciated, song birds are clearly superb at vocal production, perception, mimcry, and learning. [SLIDE of birds passing on song.] In both birdsong and speech, auditory-vocal learning takes place during a sensitive period early in life, and there is a transitional phase of vocalization called 'babbling' in infants and 'subsong' in young songbirds. More recently, the parallels between speech and song have been extended to the neural and genetic levels. As you may know, the juvenile males acquire their songs by listening to con-specific adult male 'tutors', apparently molding an initial 'babbling' template into a progressively more accurate form. Here is one male zebra finch juvenile at 42, 60, and 127 days of age, where you can hear the finch closing in on the 'model' provide by its male tutor.

[SLIDE of progressively improving babbling birdsongs]

But as complex as this song is, it lacks an essential ingredient for human language: bird songs are songs without *words*. Birdsong, as varied and as sophisticated as it might be, is not varied to convey distinct meanings, but rather maps directly to some attentive/hormonal state: it is a monoblock signal to mark territory (Me, me, me!) or sexual availability (Ready, ready, ready!). While people quite easily morph "Obama likes Palin" into "Palin likes Obama" to mean something radically different, no bird juggles its song components in any comparable way. No words, no language.

What about the great apes, our closest living relatives? They are good at many cognitive tasks, including cooperative behavior, causal reasoning, and the like. Other primates probably have conceptual structures are found in other primates: probably actor-action-goal schemata, categorization, possibly the singular-plural distinction, and others. These were presumably recruited for language**,** though the conceptual resources of humans that enter into language use appear to be far richer.

Perhaps some of you have seen the recent documentary film "Project Nim" and almost certainly you are aware of the several past efforts to 'teach' chimpanzees or gorillas 'language' either by using sign language, in the case of Nim, or, in the case of pygmy chimpanzees, bonobos, by this means. [SLIDE].

What may not be so immediately apparent from the *Project Nim* film – and this applies to the other attempts as well – is that these efforts all *failed*, and failed miserably. No other living non-human ape or dolphin has attained anything close to human language. Rather, as Prof. Laura-Anna Petitto notes, "while apes can string one or two 'words' [or signs] together in ways that seem patterned, they cannot construct patterned sequences of three, four, and beyond…After producing [a] matrix of two words they then– choosing from only the top five or so most frequently used words that they can produce (all primary food or contact words, such as *eat* or *tickle*) – randomly constructing a grocery list. There is no rhyme or reason to the list, only a word salad lacking internal organization…." And "alas, the whole story is even worse than irregularities in the chimpanzees' syntax, morphology, or phonology: the very *meanings* of their words were "off." Chimps, unlike humans, use words (labels) in way that seems to rely heavily on the notion of global association. A chimp will use the same label *apple* to refer to the action of eating apples, the location where apples are kept, events and locations of objects other than apples that happen to be stored with an apple (the knife used to cut it), all simultaneously without apparent recognition of the apparent differences. [But] even the first words of the young human baby are used in a kind-concept way – kinds of events, kinds of actions, etc. …Chimps do not really have "names for things" at all. They have only a hodge-podge of loose associations… in effect, they do not ever acquire the *human* word *apple*." (2004, 85-86).

According to Jane Goodall, the closest observer of chimpanzees in the wild, for them "the production of a sound in the *absence* of the appropriate emotional state seems to be an almost impossible task" (Goodall, cited in Tattersall, 2002).

And as Povinelli (2004:33) notes,

> "Chimpanzees rely strictly upon observable features of others to forge their social concepts. If correct, [this] would mean that chimpanzees do not realize that there is more to others than their movements, facial expressions, and habits of behavior. They would not understand that other beings are repositories of private, internal experience."

And further, that chimpanzees, are:

> "intelligent, thinking creatures who deftly attend[ed] to and learn[ed] about the regularities that unfold[ed] in the world around them. But … they [did] not reason about unobservable things: they [had] no ideas about the 'mind,' no notion of 'causation.'"

So what do *we* have that the other animals don't? Here's the surprising answer.

[SLIDE – pencil]

That's right. While other animals make tools, there is apparently no other animal that makes a *combinatorial* tool. There is no other animal that stitches together separate 'bits' like an 'eraser' and a 'stick of lead' that then can be manipulated as if it were a new, single object, that can be labeled as such – a *pencil.*

So in the case of language, what's the novel *bricolage*? What was that neural change in some small group that was rather minor in genetic terms? To answer that, we have to

consider the special properties of language. I believe these come down to Humboldt's famous aphorism that language makes 'infinite use of finite means'. The most elementary property of our shared language capacity is that it enables us to construct and interpret a discrete infinity of hierarchically structured expressions: discrete because there are 5 word sentences and 6 word sentences, but no 5½ word sentences; infinite because there is no longest sentence; and hierarchical because what our language capacity assembles are *structures* not mere strings of sounds – what are called phrases.

Language is therefore based on some generative procedure that takes elementary word-like elements from a mental store, call it the lexicon, and applies repeatedly to yield structured expressions, without bound. Operating unfettered, such a system can even arrive at astonishing language combinations like this one [SLIDE da Vinci].

This is this ability we immediately recognize as the hallmark of human language (even if it's not good language) an ability to produce a discrete infinity of possible meaningful 'signs' integrated with the human conceptual system, the algebraic closure of a recursive operator over our 'dictionary.' No other animal has this combinatorial promiscuity, an open-ended quality quite unlike the frozen 10-20 'word' vocalization repertoire that marks the maximum for any other animal species. Such combinatory promiscuity seemingly permeates all of human mental life, from our lexicon, to mathematics and music, and to tools.

To account for the emergence of this ability we have to face two basic tasks. One task is to account for the "atoms of computation," the words – commonly in the range of 30–50,000. The second is to discover the computational properties of the language faculty.

This task in turn has several facets: we must seek to discover the generative procedure that constructs infinitely many expressions in the mind, and the methods by which these internal mental objects are related to two *interfaces* with language-external (but organism-internal) systems: the system of thought, on the one hand, and also to the sensorimotor system, thus *externalizing* internal computations and thought. This is one way of reformulating the traditional conception, at least back to Aristotle, that language is sound with a meaning. [SLIDE]

So what's the 'secret sauce' that lets us, but no other animal, grab any two individual 'words' and paste them together, assembling a new object that itself can be manipulated *as if* it were a single object? Whatever it is, it can take two words, for example, *the* and *apples*, and glue them together into a single new object, here written as *the-apples,* This combinatory operation can in turn paste together a verb with this newly formed object, selecting the verb as 'most prominent' and yielding a verb-like chunk that forever after acts like a verb-like object and so on, yielding *ate the apples, John ate the apples, I know John ate the apples,* etc., the familiar open-ended creativity we associate with human language, an infinite number of (sound, meaning) pairs.

If we assume, reasonably, that the human brain is finite, taking the computational theory of mind seriously, then all this must be produced by some *finite* number of rule or operators. But this logically entails that at least one of the operators or rules must apply to its own output, that is, the computational system must be *recursive*. What does this recursive, generative system look like?

The simplest assumption, hence the one we adopt unless counterevidence appears, is that the generative procedure emerged suddenly as the result of a minor mutation. In that case we would expect the generative procedure to be very simple. Various kinds of generative procedures have been explored in the past 50 years. One approach familiar to linguists and computer scientists is context-free phrase structure grammar, developed in the 1950s and since extensively employed. The approach made sense at the time. It fit very naturally into one of the several equivalent formulations of the mathematical theory of recursive procedures – Emil Post's rewriting systems – and it captured at least some basic properties of language, such as hierarchical structure and embedding. Nevertheless, it was quickly recognized that phrase structure grammar is not only inadequate for language but is also quite a complex procedure with many arbitrary stipulations, not the kind of system we would hope to find, and unlikely to have emerged suddenly.

Over the years, research has found ways to reduce the complexities of these systems, and finally to eliminate them entirely in favor of the simplest possible mode of recursive generation: an operation that takes two objects already constructed, call them *X* and *Y*, and forms from them a new object that consists of the two objects unchanged, hence simply the set with *X* and *Y* as members, along with a *label* for the new object. Call this operation *cons*, after the familiar Lisp operation *constructor*. Provided with conceptual atoms of the lexicon, the operation *cons* iterated without bound, yields an infinity of hierarchically constructed expressions. If these can be interpreted by conceptual systems, the operation provides an internal "language of thought." Notice that there is no room in this picture for any precursors to language – say a language-like system with only short sentences. There is no rationale for positing such a system: to go from seven-word

sentences to the discrete infinity of human language requires emergence of the same recursive procedure as to go from zero to infinity, and there is of course no direct evidence for such "protolanguages."

Not any two arbitrary objects can be combined: we cannot have *the the*, or *ate ate* for instance; this implies that one of the two objects *X, Y* glued together by *cons* has have what we might call and 'edge' feature, like the notch in a jigsaw puzzle piece, that matches up with the other object, as per this [SLIDE]

Further, this tells us what the basic structure of human language is, akin to the spiral structure of DNA. But instead of DNA, the basic structure of language is this kind of skeleton shape: an asymmetrical, hierarchical template. [SLIDE]

These internal 'mental objects' are all *hierarchical structures*. The right way to picture them is like this: as a pair of coat hangers stuck together, that are free to turn, in a mobile-like fashion, around the vertical axis. Thus, in this structure for *ate-the-apples*, the coat hanger unit corresponding to *the apples* is free to rotate around the higher coat hanger *ate*. So left to right order does not matter: indeed, in some languages, like German or Japanese, *the apples–ate would* be the right order.

How do we know that these objects are hierarchical 'chunks' (rather than, say, just flat strings)? Because the constraints and operations of human language respect hierarchical structure, not linear order. [SLIDE]

Example: In the sentence, *Obama likes him*, 'him' cannot refer to Obama. But if a chunk of hierarchical structure intervenes between 'him' and 'Obama', as in this example, *Obama thinks Palin likes him*, now 'him' *can* refer to Obama (but of course need not). It

does not matter whether Obama is ‘to the left’ or ‘to the right’ of ‘him’; what matters is the relationship between the two of *hierarchical* structure. But there is more.

As an important example, while in the previous cases of *cons*, the two items we combined, *X* and *Y,* were disjoint sets, suppose we have the case where *Y* is a subset of *X* (or vice-versa), as shown in this slide, where the set object *Y* is the structure corresponding to *the apples*, while the set object *X* corresponds to the structure associated with *John ate the apples*. Then *cons*(*X, Y*) yields the new set structure shown on the right, corresponding to *the apples John ate the apples.* In effect, we have ‘copied’ the object of the verb *ate* to a position that is sometimes called the ‘focus’ of the sentence, to draw attention to it in the discourse. There is an additional principle at work that suppresses the pronunciation of the second copy of *the apples* when it is passed to the speech (or sign language) output machinery to get “flattened” onto a set of instructions to the speech apparatus in a left-to-right-fashion. So in fact what gets said is, *the apples, John ate*, noting that internally the object of *ate* is in the proper place for interpretation. This is an important point: note that *the apples* must appear in <u>two</u> distinct places: one, the position for proper interpretation of *the apples* as the object of the verb (namely, directly after the verb); the second, the position for the proper interpretation of *the apples* as a ‘focused’ item for intonation (at the front of the sentence). The representation built by the generative apparatus is thus optimal in this regard: it yields exactly the right structure and no more.

More generally, the operation *cons* yields the familiar *displacement* property of language: the fact that we pronounce phrases in one position, but interpret them somewhere else as well. Thus in the sentence “guess what John is eating,” we understand “what” to be the

object of "eat," as in "John ate the apple," even though it is pronounced somewhere else. This property has always seemed paradoxical, a kind of "imperfection" of language. It is by no means necessary in order to capture semantic facts, but it is ubiquitous. It surpasses the capacity of context-free phrase structure grammars, requiring that they be still further complicated with additional devices. But it falls within *cons* automatically, as we have seen. For, suppose that *cons* has constructed the mental expression corresponding to "Did John eat what." A larger expression can be constructed by *cons* by adding something from within the expression, so as to form "what did John eat what". In "what did John eat what," the phrase "what" appears in two positions, and in fact those two positions are required for semantic interpretation: the original position provides the information that "what" is understood to be the direct object of "eat," and the new position, at the edge, is interpreted as a quantifier ranging over a variable, so that the expression means something like "for what thing *x*, John did eat the thing *x*."

These observations generalize over a wide range of constructions. The resulting representations are in the exact form needed for semantic interpretation: these *interior* mental representations yield a kind of 'logical form'. If you again remember the lambda calculus, or better, programming languages built on *cons* like Scheme (or Lisp), then the representations are precisely those would we would expect to fins if language takes *interior* representations, the interface to semantics, to be primary, so that these representations are 'easy' and transparent to process, involving no extra work. As a more complicated example, consider the question:

Which of his pictures did they persuade the museum that every painter likes best?

The answer to this question might be, 'his first one', crucially a *different* picture for each painter (Picasso, Manet, Rembrandt, ….)

Now, this kind of answer is possible only if the human system of inference and interpretation constructs a representation that builds *two* instances of "his pictures", one that is logically present as the object of *likes* (and therefore hierarchically underneath), but is not pronounced, and one copy of "his pictures" that is the one you hear.

However, this dual representation, while making *semantics* easy does *no*t yield representations that are equally transparent or easy to process for *external* computations like parsing or production. We do not pronounce "guess what John is eating what," but rather "guess what John is eating," with the original position suppressed. That is a universal property of displacement, with minor (and interesting) qualifications that we can ignore here. The property follows from elementary principles of computational efficiency. In fact, it has often been noted that serial motor activity is computationally costly, a matter attested by the sheer quantity of motor cortex devoted to both motor control of the hands and for oro-facial articulatory gestures.

To externalize the internally generated expression "what did John eat what," it would be necessary to pronounce "what" twice, and that turns out to place a very considerable burden on computation, when we consider expressions of normal complexity and the actual nature of displacement *cons*. With all but one of the occurrences of "what" suppressed, the computational burden is greatly eased. The one occurrence that must be pronounced is the most prominent one, the last one created *cons*: otherwise there will be no indication that the operation has applied to yield the correct interpretation. It appears,

then, that the language faculty recruits a general principle of computational efficiency for the process of externalization. The suppression of all but one of the occurrences of the displaced element is computationally efficient, but imposes a significant burden on interpretation, hence on communication. The person hearing the sentence has to discover the position of the gap where the displaced element is to be interpreted. That is a highly non-trivial problem in general, familiar from parsing programs. Sometimes the resulting sentence can be ambiguous [SLIDE], and indeed, there are cases where externalization is impossible, even though the meaning is perfectly clear. [SLIDE]

There is, then, a conflict between computational efficiency and interpretive-communicative efficiency. Universally, languages resolve the conflict in favor of computational efficiency. That is, the system makes life easy for the *internal* system, rather than making life easy for the *external* system of parsing. These facts at once suggest that language evolved as an instrument of *internal thought*, with externalization a secondary process.

There are independent reasons for the conclusion that externalization is a secondary process. One is that externalization appears to be modality-independent, as has been learned from studies of sign language in recent years. The structural properties of sign and spoken language are remarkably similar. Additionally, acquisition follows the same course in both, and neural localization seems to be similar as well. That tends to reinforce the conclusion that language is optimized for the system of thought, with mode of externalization secondary.

The individual first endowed with *cons* would have had many advantages: capacities for complex thought, planning, interpretation, and so on. We might call this new *Homo,* not *Homo sapiens,* but *Homo combinans.* The capacity would be partially transmitted to offspring, and because of the selective advantages it confers, it might come to dominate a small breeding group. What it implies is that the emergence of language in this sense could indeed have been a unique event, accounting for its species-specific character. Such 'founder effects' in population bottleneck situations are not uncommon.

When the beneficial mutation has spread through the group, there would be an advantage to externalization, so the capacity would be linked as a secondary process to the sensorimotor system for externalization and interaction, including communication as a special case. It is not easy to imagine an account of human evolution that does not assume at least this much, in one or another form. Any additional assumption requires both evidence and rationale, not easy to come by.

Returning to the general theme of cognitive computing, it is important to see what this new combinatorial ability unleashed – that is, how language lit a bonfire under the rest of cognition. How? Recall that there is plenty of evidence for specialized cognitive "modules" in other animals – like bee navigation, or bat echolocation. But characteristically, these 'laser-like' modules are not able to "talk" to one another – bats cannot press their echolocation abilities into the service of solving some *other* cognitive task. But this is quite different from the "Swiss army knife" character of human cognition – the ability to cobble together novel mental representations of complex events, well beyond the power of any single 'module' [SLIDE] So I would like to put forth the claim, made by others such as Spelke, Tattersall, and Chomsky, that this is a direct result

of the infiltration of *cons* into every aspect of our conceptual life: language acts as a kind of 'cross-module' cognitive glue that links all *other* representations together In other words, *cons* lets us hook together words into 'chunks' that can then act as single units, but we should recall what stands behind the words, namely, *concepts*. By enabling the construction of extremely complicated, novel conceptual objects and events, language enables the internal construction of representations *of* representations, cross-wiring other mental modules. What other species could come up with an event description like this? What is more remarkable, is that this description was produced by a deaf-blind person – so, a person with clearly impoverished *input* to the language faculty. Yet, they are able to attain quite sophisticated knowledge about 'seen' objects, actions, and the world – a rich inner mental life.

Evidence for this cross-modal coupling comes from recent brain evidence regarding the interaction between the brain regions often cited to be active during syntactic processing, e.g., Broca's area (Brodmann's area 44/45), a "phylogenetically younger" part of the cortex (Friederici et al., 2011), and areas involved recognizing events semantically or in relevant motor processing. Two brain tracts as shown in this slide seem to form two streams, a top, *dorsal* tract or bundle of fibers, that is involved in coupling sound to its articulation, and a separate, bottom, ventral tract that connects syntax to the retrieval of stored representations of objects and actions. Note how this corresponds exactly to the two interfaces we mentioned earlier, as well as to similar dorsal/ventral processing streams found in the visual system. Further, it exhibits explicitly how visual representations are cross-wired by language. For example, Zadeh *et al.* (2006) showed that when adults read about action sentences such as "eating a peach", the brain areas that

lit up were not only the classical language ones but also the same areas activated when just viewing the action visually Both Zadeh and Pervmuller and Fadiga (2010) have demonstrated that particular networks are activated for distinct body parts and actions, e.g., arm vs. legs. One interpretation of activation patterns such as these is that it is language that 'binds' vision and action together.

And here is a second example of what I mean by 'cross-modal' enabling. It is taken from the work of Prof. Elizabeth Spelke at Harvard. [SLIDE]. Room geometry + color of wall to find hidden object. Blindfold child, spin them around, they look for object, etc. (Toy or stock certificate). BUT children before they learn language cannot seem to integrate the information from these two sources. AND furthermore, both children and adults after they've learned language CAN use these two sources together – they can construct the representation that combines geometry with color, and it seems, by means of language. The evidence? If we 'overload' the language system when carrying out the search task, by having the subjects perform a simultaneous language task, like reciting a poem, then their performance degrades back to their 'pre-linguistic' state (and this can be shown to be something other than a pure memory effect).

With *cons*, humans are liberated from the here-and-now to develop ever richer descriptions – Human cognoscitive powers provide us with a world of experience, different from the world of experience of other animals. Being reflective creatures, thanks to the emergence of the human capacity, humans try to make some sense of experience. These efforts are called myth, or religion, or magic, or philosophy, or in modern English usage, science. For science, the concept of reference in the technical sense is a normative ideal: we hope that the invented concepts *photon* or *verb phrase* pick

out some real thing in the world. Human cognoscitive powers provide us with a world of experience, different from the world of experience of other animals. Being reflective creatures, thanks to the emergence of the human capacity, humans try to make some sense of experience. These efforts are called myth, or religion, or magic, or philosophy, or in modern English usage, science. For science, the concept of reference in the technical sense is a normative ideal: we hope that the invented concepts *photon* or *verb phrase* pick out some real thing in the world.

But there is more. Applied repeatedly to the domain of a single element, *cons* acts like the successor function of Peano arithmetic: *cons*(*cons*(*cons*(x))) = 3. This immediately yields the number system of integers, with all its familiar properties. And sure enough, as soon as children begin to acquire the rudiments of syntax, and *cons*, they apparently an open-ended quantificational ability with large numbers, unlike any other animals. While crows and chimpanzees can seemingly 'count' up to 5-7, beyond that, they deal with large numbers as though they were quantities of 'stuff' weighed on a scale – a more-or-less affair. But children by age 5, or as soon as they have acquired language syntax, seem to easily grasp that if there's a number 100, then there can be 101, as anyone who's ever played the game with a child, "I can find a bigger number than you can" might attest. This kind of ability lies beyond the read of any other species we know.

Yet a third domain where the 'grouping' operations of *cons* has to do with this domain: [SLIDE] both rhythm and music. Consider the BEAT STRUCTURE of a line of metrical poetry, "Tell me not in mournful numbers". (Wordsworth). The pattern of strong and weak beats, and in fact that the strongest beat is first on TELL, then NOT, and then MOURN, can be explained very simply by the *cons* theory. In this case, *cons* works its

way left to right through the string, grouping together pairs of elements, here syllables, as before, just like *ate* and *the apples*, collecting them into a new group that is then supposed to be labeled as such. We then SELECT one of the two units we have collected as the new label of the grouped hierarchical structure, obtaining a second level of representation. After one pass through the initial syllables, we make a second pass through this next level, as before, as shown, successively collecting groups of two, until we can do no more. Then the strongest 'beat' is just the stack with the greatest depth, and so on. You will note that there is a CRUCIAL and KEY difference here between full-fledged language and beat structure. Note that the 'lexical items' (words) are just denoted as asterisks, because there are no words in beat structure with features like 'verb' or 'noun.' There are just the 'marks' of the beats. Thus, when two marks are collected together to form a new, higher level unit, there are no features to *label* the new unit, as we labeled *ate the apples* as a 'verb phrase'. That is, beat structure is what one gets if one applies the same *cons* operation as in the rest of language, but to a system *without* word features. One and the same innovation leads to both language syntax and metrical structure. While there is no time to demonstrate it here, one can show that just a few different ways of 'passing through' the initial string of syllables – from right to left as opposed to right to left, perhaps alternating between levels, gives rise to all the possible metrical patterns shown in all human languages. We might even think of this 'beat structure' as the first glimmerings of language syntax – the platform for *cons* that existed before words were wired into language. If so, then a primitive ability like *cons* might actually be apparent in other species that exhibits metrical patterning to their vocal output – in particular, songbirds. And in fact, there is some suggestive evidence from genomic

data (the famous FOXP2 gene) that this is the case, but I will have to leave aside a detailed commentary on this point for now. What it suggests is that the ability to carry out *cons* might have been available either many hundreds of millions of years ago, but that true language did not appear because there were no words to wire it to, or else that *cons* arose in songbirds by means of convergent evolution. It's simply impossible to tell at the moment.

What we do know is that there is yet another domain of human cultural activity where *cons* surfaces, related to metrical structure, and it is this one [SLIDE]: music. First, the beat structure of music works exactly like what we showed for metrical poetry: the same grouping-and-projection (without words). Second, extending this to a 'lexicon' that consists of melodic notes, leads to an analysis of 'musical syntax' that looks a lot like language (with obvious differences again because melodic elements are not words). But as you can see from this example of a Mozart piano sonata as analyzed Pesetksy, the two are quite close, though of course the 'atoms' now differ, where the "I"s and "IVs" now represent dominant and tonic chord progressions. So perhaps this was Mozart's deep secret: for whatever reason, the generative faculty for language that makes speaking for us so effortless, was somehow cross-wired in Mozart's brain at a very early age so that literally, for him, making music was just as natural and easy as speaking is for us.

Let us just summarize briefly what seems to be the current best guess about unity and diversity of language and thought. In some completely unknown way, our ancestors developed human concepts, as opposed to what chimps, birds, and bees possess. At some time in the very recent past, maybe about 75,000 years ago, an individual in a small group of hominids in East Africa underwent what was likely a small mutation that

provided the operation *cons*– an operation that takes human concepts as computational atoms, and yields structured expressions that provide a rich language of thought. The innovation had obvious advantages, and took over the small group. At some later stage, the internal language of thought was connected to the sensorimotor system, a complex task that can be solved in many different ways and at different times, and quite possibly a task that involves no evolution at all. In the course of these events, the human capacity took shape, yielding a good part of our "moral and intellectual nature," in Wallace's phrase (1871), extending far beyond language to mathematics and musics, indeed, all of what is distinctive about the human condition. So, while a bird might be able to make a tool like this [SLIDE] and use it to stir a drink, no other animal we know can SAY this.