



[Amy Perfors](#) (2002)

Simulated Evolution of Language: a Review of the Field

Journal of Artificial Societies and Social Simulation vol. 5, no. 2

<<http://jasss.soc.surrey.ac.uk/5/2/4.html>>

To cite articles published in the *Journal of Artificial Societies and Social Simulation*, please reference the above information and include paragraph numbers if necessary

Received: 27-Sep-2000

Accepted: 11-Mar-2002

Published: 31-Mar-2002

Abstract

This is an overview of recent computational work done in the simulated evolution of language. It is prefaced by an overview of the broader issues in linguistics that computational models may help to clarify. Is language innate - genetically specified in the human organism in some way, a product of natural selection? Or can the properties of language be accounted for by general cognitive capabilities that did not develop as a consequence of language-specific selective pressures? After a consideration of the intellectual background surrounding these issues, we will examine how recent computational work sheds light on them.

Keywords:

Computational Simulations; Innateness; Language Evolution

Introduction

1.1

What evolutionary forces propelled the development of language? Are the language abilities of humans the result of an innate, language-specific portion of the brain, or do they result from a more general application of our cognitive abilities? These questions are some of the oldest and the most difficult for linguists to answer. For a long time they were restricted to philosophers. It is only within the last century (especially the last few decades) that the sciences of evolutionary biology, computation, psychology, and cognitive science have begun to provide a direction and a focus in our search for answers. Verbal theorizing and mathematical modeling, both guided by rigorous empirical study, are now the backbone of linguistic thought in this realm.

1.2

However, both verbal theorizing and mathematical modeling have clear limitations. In asking about the evolution and innateness of human language, we are attempting to understand the behavior of dynamical systems involving multiple variables that interact in potentially very complex ways. The result of this is that even the most apparently obvious intuitions can easily be led astray, and even the most complicated mathematics can be too simple. This is where more recent developments in computational simulation come in. These developments may provide an additional direction and focus that linguists have so far only begun to take advantage of. Computational simulations -- while clearly most useful in tandem with rigorous mathematical and verbal reasoning -- provide a means to bypass the difficulties inherent in following either approach alone.

1.3

In this series of papers I will discuss some of the major theoretical issues in evolutionary linguistics. What are the big questions, and why are they significant? What is the shape of current linguistic thought? And how has work on computational simulations of language use and evolution helped to guide that thought so far? In Part 1 (Theories of Language Evolution) of this paper I provide the intellectual background to recent work on language simulation, discussing some of the principle lines of research on the evolution of language that have been pursued by linguists during the last thirty years. It is important for non-specialists to realize that although a relatively low proportion of work by linguists has focused on early evolution of language, by comparison with (for instance) work on theoretical syntax or phonology, attitudes towards the early evolution of language have largely shaped the field. In particular, a mainstream view in linguistics has been that much of our linguistic knowledge is innate, and thus that the most significant aspects of language evolved through classical biological means and not merely culturally. In Part 1 (Theories of Language Evolution) this view and some of the evidence (both for and against it) is discussed. Part 2 (Simulations) will follow with a critical overview of some recent computational work on the simulation of the evolution of language, with coverage of a range of recent simulations and a discussion of their significance in the context of linguistic theory. This work includes studies of the evolution of syntax and the evolution of semantics, and both symbolic and neural network inspired techniques.

I. Theories of Language Evolution

The Questions

2.1

What is the heart of human uniqueness? For as long as recorded history -- and possibly much longer -- we have pondered what characteristics are responsible for the apparently large gap separating us from other animals. In one sense, the question is a silly one, since "human uniqueness", if such a concept has any meaning, is undoubtedly due to a multiplicity of factors that overlap considerably with one another. In another sense, however, the question is important: the search for possible answers has resulted in dramatic improvement in our understanding of topics ranging from human origins to the nature of identity to language representation in the brain.

2.2

It is increasingly evident that one of the most important factors separating humans from animals is indeed our use of language. The burgeoning field of linguistic research on chimpanzees and bonobos has revealed that, while our closest relatives can be taught basic vocabulary, it is extremely doubtful that this linguistic ability extends to syntax. (Fouts [1972](#); Savage-Rumbaugh [1987](#)) Chimps like Washoe can be taught (not easily, but reliably) to have vocabularies of up to hundreds of words, but only humans can combine words in such a way that the meaning of their expressions is a function of both the meaning of the words as well as the way they are put together. Even the fact that some primates can be tutored to have fairly significant vocabularies is notable when one considers that such achievements come only after considerable training and effort. By contrast, even small children acquire much larger vocabularies -- and use the words far more productively -- with no overt training at all. There are very few indications of primates in the wild using words referentially at all (Savage-Rumbaugh [1980](#)), and if they do, it is doubtful whether vocabularies extend beyond 10 to 20 words at maximum (Cheney [1990](#)).

2.3

Humans are noteworthy for having not only exceptional linguistic skills relative to other animals, but also for having significantly more powerful intellectual abilities. This observation leads to one of the major questions confronting linguists, cognitive scientists, and philosophers alike: to what extent can our language abilities be explained by our general intellectual skills? Can (and should) they really be separated from each other?

2.4

This question can be rephrased as whether language ability is somehow "pre-wired" or "innate," as opposed to being an inevitable by-product of the application of our general cognitive skills to the problem of communication. A great deal of the linguistic and psychological research and debate of the latter half of this century has been focused on analyzing this question. The debate has naturally produced two opposing schools of thought. Some (like Noam Chomsky and Steven

Pinker) claim that a great deal, if not all, of human linguistic abilities are innate: children require only the most basic of environmental input in order to become fully functioning fluent speakers. Others suggest that our language competence is either a byproduct of general intellectual abilities (e.g. Tomasello [1992](#); Shipley & Kuhn [1983](#)) or an instance of language adapting to human minds rather than vice versa. (Deacon [1997](#))

2.5

The controversy over the innateness of language touches on another of the most unexplored and controversial areas of linguistics: the domain of language evolution. How did a system with the enormous complexity of natural language first take root? When attempting to answer this question, we are confronted with a seemingly insurmountable paradox: in order for language to be adaptive, communicative skill would need to belong to many members of a population. Yet, in order for multiple members to have language ability, that skill would need to be adaptive enough to spread through the population.

2.6

Over the past century, many theories have been proposed seeking to explain language evolution. In order to be plausible, a theory needs to account for two main things: the evolution of referential communication and the evolution of syntax. The former refers to the phenomenon within all human languages of using an arbitrary sound to symbolize a meaning. The power of a completely unrelated symbol to stand for a thought or thing - even when that referent is not present -- is one of the most powerful characteristics of language. The latter, syntactic ability, is apparently unique to humans. Syntax is the root of our ability to form and communicate complex thoughts and productively use sentences that have never before been stated. Language with syntax appears to be qualitatively different than language without it. We are thus faced with what Derek Bickerton has called the Paradox of Continuity: language *must* have evolved from some precursor, and yet no qualitatively similar precursors exist. What can explain this?



The Innateness of Language

3.1

Possibly the most fundamental issue debated among linguists is the extent to which our ability to communicate is "innate." This controversy is as old as modern linguistics -- both the controversy and the approach were started by Noam Chomsky in the middle of the twentieth century when he published his ideas regarding the biological basis of language. Since then, much of the history of linguistics has been a response to him, with more and more refined viewpoints being brought to bear on the issue over time. The viewpoints have naturally tended to fall into two extremes, which I refer to here as the nativist and non-nativist approaches. All reasonable scholars today believe that some combination of the extremes is correct, but the issues under debate are clearest when examined in the context of the polarization of the two camps. Therefore, I shall consider the strongest arguments for each position, beginning with the nativist viewpoint and concluding with the non-nativist approach.

3.2

Before moving on, however, it is important to consider carefully a distinction that until now we haven't clarified fully. This is the difference between innateness -- the extent to which our language capacities are built in biologically -- and domain specificity, the extent to which our language capabilities are independent of other cognitive abilities. Logically, it would be coherent to hold a position whereby linguistic ability was innate but not domain specific. For example, it could happen that highly developed signal processing abilities we use for senses such as hearing and vision formed the core of an innate language ability. [It is more difficult to coherently suggest that an ability can be domain-specific but not innate; such a thing is not logically impossible, but it is probably rarer than the alternative]. However, it is generally the case that when linguists espouse a nativist view, they are usually supposing that humans are born with a highly specific ability for processing language, one that functions at least semi-independently of other cognitive abilities. I discuss the instances when the distinction between innateness and domain specificity grows hazy in the evidence that follows, but it is not important for most formulations of the question.

The Nativist View

3.3

The pure nativist believes that language ability is deeply rooted in the biology of the brain. The strongest nativist viewpoints go so far as to claim that our ability to use grammar and syntax is an instinct, or dependent on specific modules ("organs") of the brain, or both. The only element essential to the nativist view, however, is the idea that language ability is in some non-trivial sense directly dependent upon the biology of the human organism in a way that is separable from its general cognitive adaptations. In other words, we learn language as a result of having a specific biological adaptation to do so, rather than because it is an emergent response to the problem of communication confronted by ourselves and our ancestors, a response that does not presume the existence of certain traits or characteristics specified by our biology.

3.4

There are a variety of reasons for believing the nativist view: the strongest come from genetic/biological data and research in child acquisition. Chomsky's original argument was largely based on evidence from acquisition and what he called the "poverty of the stimulus" argument. The basic idea is that any language can be used to create an infinite number of productions -- far more productions and forms than a child could correctly learn without relying on pre-wired knowledge. For example, English speakers learn early on that they may form contractions of a pronoun and the verb *to be* in certain situations (like saying "he's going to the store"). However, they cannot form them in others; when asked "who is coming" one cannot reply "he's," even though semantically such a response is correct. Unlike many other learning tasks, during language acquisition children do not hear incorrect formulations modeled for them as being incorrect. Indeed, even when children might make a mistake, they are rarely corrected or even noticed. Morgan & Travis (1989; Pinker 1995; Stromswold 1995) This absence of negative evidence is an incredible handicap when attempting to generalize a grammar, to the point that many linguists

dispute whether it is possible at all without using innate constraints. (e.g. Chomsky [1981](#); Lenneberg [1967](#))

3.5

In fact, nativists claim that there are many mistakes that children *never* make. For instance, consider the sentence *A unicorn is in the garden*. To make it a question in English, we move the auxiliary *is* to the front of the sentence, getting *Is a unicorn in the garden?* Thus a plausible rule for forming questions might be "always move the first auxiliary to the front of the sentence". Yet such a rule would not account for the sentence *A unicorn that is in the garden is eating flowers*, whose interrogative form is *Is a unicorn that is in the garden eating flowers?*, NOT *Is a unicorn that in the garden is eating flowers?* (Chomsky, discussed in Pinker, [1994](#)) The point here is not that the rule we suggested is incorrect -- it is that children never seem to think it *might* be correct, even for a short time. This is taken by nativists like Chomsky as strong evidence that children are innately "wired" to favor some rules or constructions and avoid others automatically.

3.6

Another reason linguists believe that language is innate and specific in the brain is the apparent existence of a critical period for language. The claim of the existence of a critical period suggests that children -- almost regardless of general intelligence or circumstances of environment -- are able to learn language fluently if they are exposed to it before the age of 6 or so. Yet if exposed after this date, they have ever-increasing difficulty learning it. We see this phenomenon in the fact that it takes a striking amount of conscious effort for adults to learn a second language, and indeed they often are never able to get rid of the accent from their first. The same cannot be said for children.

3.7

Additionally, those very rare individuals who are not exposed to language before adolescence (so-called "wild children") never end up learning a language that even *approaches* full grammaticality. (Brown [1958](#); Fromkin et. al. [1974](#)) One should not draw too hasty conclusions about wild children; there are very few and these children usually suffered extraordinarily neglectful early conditions in other respects, which might mitigate the results. Nevertheless, it is noteworthy that some wild children who were found and exposed to language while still relatively young ultimately ended up showing no language deficits at all. (Pinker [1994](#))

3.8

Deaf children are especially interesting in this context because they represent a "natural experiment" of sorts. Many of these children are cognitively normal and raised in an environment offering everything except language input (if they are not taught to sign as children). Those exposed to some sort of input young enough will develop normal signing abilities, while those who are not will have immense difficulty learning to use language at all. Perhaps most interesting is the case of Nicaraguan deaf children who were thrown together when they went to school for the first time. (Coppola et. al. [1998](#); Senghas et. al. [1997](#)) They spontaneously formed a pidgin

tongue -- a fairly ungrammatical ``language" combined from each of their personal signs. Most interestingly, younger children who later came to the school and were exposed to the pidgin tongue *then* spontaneously added grammatical rules, complete with inflection, case marking, and other forms of syntax. The full language that emerged is the dominant sign language in Nicaragua today, and is strong evidence of the ability of very young children to not only *detect* but indeed to *create* grammar. This process -- children turning a relatively ungrammatical protolanguage spoken by older speakers (a pidgin) into a fully grammatical language (a creole) -- has been noted and studied in multiple other places in the world. (Bickerton [1981](#), [1984](#))

3.9

Evidence that language is domain-specific comes from genetics and biology. One can find instances of individuals with normal intelligence but extremely poor grammatical skills, and vice versa, suggesting that the capacity for language may be separable from other cognitive functions. Individuals diagnosed with Specific Language Impairment (SLI) have normal intelligence but nevertheless seem to have difficulty with many of the normal language abilities that the rest of us take for granted. (Tallal et. al. [1989](#); Gopnik & Crago [1991](#)) They usually develop language late, have difficulty articulating some words, and make persistent, simple grammatical errors throughout adulthood. Pinker reports that SLI individuals frequently misuse pronouns, suffixes, and simple tenses, and eloquently describes their language use by suggesting that they give the impression ``of a tourist struggling in a foreign city." ([1994](#))

3.10

The opposite case of SLI exists as well: individuals who are demonstrably lacking in even fairly basic intellectual abilities who nevertheless use language in a sophisticated, high-level manner. Fluent grammatical language has been found to occur in patients with a whole host of other deficits, including schizophrenia, autism, and Alzheimer's. One of the most provocative instances is that of William's syndrome. (Bellugi et. al. [1991](#)) Individuals with this disease generally have mean IQs of 50 but speak completely fluently, often at a higher level than children of the same age with normal intelligence. Each of these instances of having normal intelligence but extremely poor grammatical skills (or vice versa) can be shown to have some dependence on genetics, which suggests again that much of language ability is innate.

The Non-Nativist View

3.11

All of this evidence in support of the nativist view certainly seems extremely compelling, but recent work has begun to indicate that perhaps the issue is not quite as cut and dried as was originally thought. Much of the evidence supporting the non-nativist view is therefore actually evidence *against* the nativist view.

3.12

First, and most importantly, there is increasing indication that Chomsky's original ``poverty of the

stimulus" theory does not adequately describe the situation confronted by children learning language. For instance, he pointed to the absence of negative evidence as support for the idea that children had to have some innate grammar telling them what was not allowed. Yet, while overt correction *does* seem to be scarce, there is a consistent indication of parents implicitly "correcting" by correctly using a phrase immediately following an instance when the child misused it. (Demetras et. al. [1986](#); Marcus [1993](#), among others) More importantly, children often pick up on this and incorporate it into their grammar right away, indicating that they are extremely sensitive to such correction.

3.13

More strikingly, children are incredibly well attuned to the statistical properties of their parent's speech. (Saffran et. al. [1997](#); De Villiers [1985](#)) The words and phrases used most commonly by parents will - with relatively high probability -- be the first words, phrases, and even grammatical structures learned by children. This by itself doesn't necessarily mean that there is no innate component of grammar - after all, even nativists agree that a child needs input, so it wouldn't be *too* surprising if they were especially attuned to the most frequent of that input. Yet additional evidence demonstrates that children employ a generally conservative acquisition strategy; they will only generalize a rule or structure after having been exposed to it multiple times and in many ways. Pinker ([1994](#)) These two facts combined together suggest that a domain- general strategy that makes few assumptions about the innate capacities of the brain may account for much of language acquisition just as well as theories that make far stronger claims. In other words, children who are attuned to the statistical frequency of the input they hear who also are hesitant to overgeneralize in the absence of solid evidence will tend to acquire a language just as certainly, if not as quickly, as those who come "pre-wired" in any stronger way.

3.14

Other evidence strongly indicates that children pay more attention to some words than others, learning these "model words" piece-by-piece rather than generalizing rules from few bits of data. (Tomasello [1992](#); Ninio [1999](#)) For instance, children usually learn only one or a few verbs during the beginning stages of acquisition. These verbs are often the most typical and general, both semantically and syntactically (like *do* or *make* in English). Non-nativists (such as Tomasello) suggest that only after children have generalized those verbs to a variety of contexts and forms do they begin to acquire verbs *en masse*. Quite possibly, this is an indication of a general-purpose learning mechanism coming into play, and the use of an effective way to learn the rules of inflection, tense, and case marking in English without needing to resort to a reliance on pre-wired rules.

3.15

There is also reason to believe that language learning is an easier task than it first appears: children get help on the input end as well. People speaking to young children will automatically adjust their language level to approximately what the child is able to handle. For instance, Motherese is a type of infant- directed (ID) speech marked by generally simpler grammatical forms, higher amplitude,

greater range of prosody, and incorporation of basic vocabulary. (Fernald & Simon [1984](#)) The specific properties of Motherese are believed to enhance an infant's ability to learn language by focusing attention on the grammatically important and most semantically salient parts of a sentence. Babies prefer to listen to Motherese, and adults across the world will naturally fall into ID speech when interacting with babies. (Fernald et. al., [1989](#)) They are clearly quite attuned to the infant's linguistic level; the use of ID speech subsides slowly as children grow older and their language grows more complex. This sort of evidence may indicate that children are such good language learners in part because parents are such good instinctive language teachers.

3.16

The evidence considered here certainly seems to suggest that perhaps the nativist viewpoint isn't as strong as originally thought, but what about the points regarding critical periods, creolization, and the genetic bases of language? These points might be answered in one of two ways: either they are based on suspect evidence, or draw conclusions that are too strong for the evidence we currently have. Consider the phenomenon of critical periods. Much of the research on wild children is based on five or fewer individuals. A typical example is the case of Genie, who was discovered at the age of 13. (Fromkin et. al. [1974](#); Curtiss [1977](#)) She had been horribly abused and neglected for much of her young life and could not vocalize when first found. After extensive tutoring, she could speak in a pidgin-like tongue but never showed full grammatical abilities. However -- as with most wild children -- any conclusions one might reach are automatically suspect because her early childhood was marked by such extreme abuse and neglect that her language deficits could easily have sprung from a host of other problems.

3.17

Even instances of individuals who are apparently normal in every respect but who were not exposed to language are not clear support for the critical period notion. Pinker ([1994](#)) considers the example of Chelsea, a deaf woman who was not diagnosed as deaf until 31, at which point she was fitted with hearing aids and taught to speak. Though she ultimately was able to score at a 10-year old level on IQ tests, she always spoke quite ungrammatically. Pinker uses this to support the nativist view, but it's not clear that it does. A 10-year old intelligence level is approximately equal to an IQ of 50, so it is quite plausible that Chelsea's language results are conflated with generally low intelligence. Even if not, both nativists and non-nativists would agree that the ability to think in complex language helps develop and refine the ability to *think*. Perhaps the purported "critical period" in language development really represents a critical period in intellectual development: if an individual does not develop and use the tools promoting complex thought before a certain age, it becomes ever more difficult to acquire them in the first place. If true, then, the existence of critical periods does not support the domain-specific perspective of language development because they do not show how language is separable from general intelligence.

3.18

Another reason for disbelieving that there is a critical period for language development lies in second language acquisition. While some adults do never lose an accent, many do -- and in any case it is far from obvious that because there might be a critical period for phonological

development (which would explain accents), there would necessarily be a critical period for grammatical development as well. Indeed, the fact that adults can and do learn multiple languages - eventually becoming completely fluent -- is by itself sufficient to discredit the critical period hypothesis. The biological definition of critical periods (such as the period governing the development of rods and cones in the eyes of kittens) requires that they not be reversible *at all*. (Goldstein, [1989](#)) Once the period has passed, there is no way to acquire the skill in question. This is clearly not the case for language.

3.19

The existence of genetic impairments like Specific Language Impairment seem to be incontrovertible proof that language ability must be domain-specific (and possibly innate as well), but there is controversy over even this point. Recent research into SLI indicates that it arises from an inability to correctly perceive the underlying phonological structure of language, and in fact the earliest research suggested this. (Tallal et. al. [1989](#); Wright et. al. [1997](#)) This definitely suggests that *part* of language ability is innate -- namely, phonological perception -- but this fact is well accepted by both nativists and non-nativists alike. (Eimas et. al. [1971](#); Werker [1984](#)) It is a big leap from the idea that phonological perception is innate to the notion that syntax is.

3.20

What about the opposite case, that of "linguistic savants" like individuals with Williams Syndrome? As Tomasello ([1995](#)) points out, there is evidence suggesting that Williams syndrome children have much less advanced language skills than was first believed. (e.g. Bellugi, Wang, and Jernigan 1994, discussed in Tomasello [1995](#)) For instance, the syntax of Williams syndrome teenagers is actually equivalent to typical 7-year olds, and some suggest that the language of Williams syndrome individuals is quite predictable from their mental age. (Gosch, Städing, & Pankau [1994](#)) Williams' syndrome individuals appear proficient in language development only in comparison with IQ- matched Down's syndrome children, who language abilities are actually *lower* than one would expect based on their mental ages.

3.21

Even if there *were* linguistic savants, Tomasello goes on to point out, that wouldn't be evidence that language is innate. There are many recorded instances of other types of savants - "date-calculators" or even piano-playing savants. Yet few would suggest that date calculation or piano playing is independent of other cognitive and mathematical skills, or that there is an innate module in the brain assigned to date calculating and piano playing. Rather, it is far more reasonable to conclude that some individuals might use their cognitive abilities in some directions but not others.

3.22

The final argument for the non- nativist perspective is basically just an application of Occam's Razor: the best theory is usually the one that incorporates the fewest unnecessary assumptions. That is, the nativist suggests that language ability is due to some specific pre-wiring in the brain. No plausible explanation of the nature of the wiring has been suggested that is psychologically

realistic while still accounting for the empirical evidence we have regarding language acquisition. As we have seen, it is possible to account for much of language acquisition without needing to rely on the existence of a hypothetical language module. Why multiply assumptions without cause?

The Evolution of Language

4.1

There is considerable overlap between questions regarding the innateness of language and questions regarding the evolution of language. After all, if the evolution of language can be explained through the evolution of some biological capacity or genetic change, that would be strong evidence for its innateness. On the other hand, if research revealed that language evolved in a way that did not rely crucially on any of our genetic or biological characteristics, that would suggest that it was *not* innate.

4.2

Any scientist hoping to explain language evolution finds herself needing to explain two main "jumps" in evolution: the first usage of *words as symbols*, and the first usage of what we might call *grammar*. For clarity, I will refer to these issues as the "Evolution of Communication" and the "Evolution of Syntax," respectively.

4.3

For each concern, scientists must determine what counts as good evidence and by what standard theories should be judged. The difficulty in doing this is twofold. For one thing, the evolution of language as we know it occurred only *once* in history; thus, it is impossible to either compare language evolution in humans to language evolution in others, or to determine what characteristics of our language are accidents of history and what are necessary parts of any communicative system. The other difficulty is related to the scarcity of evidence available regarding the one evolutionary path that *did* happen. "Language" doesn't fossilize, and since many interesting developments in the evolution of language occurred so long ago, direct evidence of those developments is outside of our grasp. As it is, scientists must draw huge inferences from the existence of few artifacts and occasional bones -- a process that is fraught with potential error.

4.4

In spite of these difficulties, a significant amount of theorizing and research has been done. To some extent the dominant theories of thought in this field parallel the dominant theories of thought discussed in the last section: a portion of scientists strongly adhere to a more nativist perspective, while others argue against it.

4.5

In the following sections I will examine and discuss three of the dominant theories of language evolution, especially with regard to views on the Evolution of Communication and the Evolution of Syntax. For each theory (Bickerton, Pinker and Bloom, and Deacon) I will discuss both

supporting and disconfirming evidence. Finally, I will end the section with a commentary tying together research to date about both the innateness and the evolution of language, leading to suggestions of where to go from here.

Bickerton: Incorporating the fossil record

4.6

In 1990, Derek Bickerton authored one of the first and most ambitious attempts to explain the evolution of human languages. The basic idea his theory is based on is the notion of the Primary Representation System (PRS). According to him, the way in which humans represent the world -- the PRS -- forms the basis for the structure of human language, which evolved in stages. Bickerton hypothesizes that our ancestors as far back as *Homo erectus* (1.5 to 1 million years ago) could speak some sort of protolanguage (which is roughly similar to a typical two-year old's capabilities or a pidgin tongue). However, language as we know it - the Evolution of Syntax -- did not develop until as recently as 40,000 years ago, due to a mutation affecting the brain.

4.7

What specifically is meant by a PRS? A representational system can be roughly defined as the system linking those things in the world with those things that we believe we perceive. We cannot have access to "things in the world" *except* as they are filtered through our representation system: as Bickerton states, "there is not, and cannot in the nature of things ever be, a representation without a medium to support it in." (Bickerton [1990](#)) The question is, what are the properties of that representation system?

4.8

Bickerton proposes that the PRS of humans is fundamentally binary and hierarchical. In other words, the concepts in our minds that seem to correspond to notions in the world are defined *vertically* (superordinate or subordinate to other concepts) as well as *horizontally* (by the bounds of other concepts). For example, a spaniel can be defined horizontally by associating it with other types of dogs (beagle, dachshund, collie, etc). It can also be defined vertically by identifying it with its superordinate concept (a kind of dog) or the subordinate concepts (it has a tail). According to Bickerton, we classify all of our concepts in this hierarchical manner.

4.9

What does this have to do with language? Quite simply, the lexicon reflects this hierarchical structuring. Every word in every language can not only be defined in terms of other words in the same language, but exists as part of a sort of "universal filing system" that allows for rapid retrieval of any concept. Bickerton suggests that this filing system, as it were, was achieved *before* the emergence of language (or at least before the emergence of language much beyond what we see in animals today). Thus, meaning was originally based on our functional interaction with other creatures; only as our general cognitive abilities grew strong enough did we gain the skills to arbitrarily associate symbols with those basic meanings. Eventually, of course, language was used to generate its own concepts (like *unicorn*), but initially, language merely labeled these

protoconcepts that were already in our heads as part of our PRS.

4.10

Where did these concepts come from, then, and why are they fundamentally binary branching? As might be expected, the categories that constitute the PRS of any species are the categories that are necessary for the survival of that species. Thus, humans do not naturally distinguish between (say) low-pitched sonar pings and high-pitched sonar pings, while bats might; it is just not evolutionarily relevant for humans to make that distinction. Notably, all distinctions are of the form ``X and not-X". Bickerton suggests that this is because, at root, all of our knowledge stems from cellular structures (like neurons) that only distinguish between two states. Hence the binary branching nature of our PRS.

4.11

At this point one is inclined to object that humans *are* perfectly able to represent even things that are not direct evolutionary adaptations, like low-pitched pings and high-pitched pings. (After all, I just did it in the last paragraph). That is, we can represent the difference between them even though we have never ``heard" the difference and almost undoubtedly never will. The skill of generalizing beyond what has been directly selected for is what Bickerton proposes is the main advantage of language. As our secondary representation system (SRS), language makes it possible to conceptualize many things we otherwise couldn't have represented except after many years of direct biological evolution. A highly developed SRS would therefore be highly advantageous in the evolutionary sense.

4.12

Thus, the evolution of ``protolanguage" -- a language marked by fairly large vocabulary but very little grammar -- formed concurrently with the gradual expansion of our general intelligence. Speakers of pidgin tongues, children at the two-word stage, and wild children are all considered to speak in protolanguage. Why not believe that full language (incorporating nearly modern grammatical and syntactic abilities) evolved at this time, not just protolanguage? There are two primary reasons. First of all, there is strong indication that the vocal apparatus necessary for rapid, articulate speech did not evolve until the advent of modern *Homo sapiens* approximately 100,000 years ago. (Johanson & Edgar [1996](#); Lieberman [1975](#), [1992](#)) Language that incorporated full syntax would have been prohibitively slow and difficult to parse without a modern or nearly-modern vocal tract, indicating that it probably did not evolve until then. This creates a paradox: such a vocal tract is evolutionarily disadvantageous unless it is used for the production of rapid, articulate speech. Yet the advantage of rapid, articulate syntactic speech does not exist without a properly shaped vocal tract. Which came first, then, the vocal tract or the syntax? Bickerton's proposal solves this paradox by suggesting that the vocal tract evolved gradually toward faster and clearer articulation of protolanguage, and only then did fully grammatical language develop.

4.13

The other reason for believing that full language did not exist until relatively recently is that there is little evidence in the fossil record prior to the beginning of the Upper Paleolithic (100,000 to

40,000 years ago) for the sorts of behavior presumably facilitated by full language. (Johanson & Edgar [1996](#); Lewin [1993](#)) Although our ancestors before then had begun to make stone tools and conquer fire, there was little evidence of innovation, imagination, or abstract representation until that point. The Upper Paleolithic saw an explosion of styles and techniques of stone tool making, invention of other weapons such as the crossbow, bone tools, art, carving, evidence of burial, and regional styles suggesting cultural transmission. This sudden change is indicative of the emergence of full language in the Upper Paleolithic, preceded by something language-like but far less powerful (like protolanguage), as Bickerton suggests.

4.14

Not surprisingly, then, the final part of Bickerton's theory concerns the emergence of full syntax during the Upper Paleolithic. According to him, this emergence was sudden, caused by a mutation affecting the brain. Since there is no evidence from the fossil record that brain *size* altered at this point, Bickerton argues that the mutation must have altered structure only.

4.15

Why doesn't he suggest that syntax emerged more gradually? Primarily, because such a view is not in keeping with the empirical evidence he considers. We have already seen that the flowering of culture in the Upper Paleolithic was sudden as well as pronounced. Thus, it is more easily explainable by the rapid emergence of full language rather than a gradual development of syntax. Additionally, Bickerton has drawn strong parallels between protolanguage and pidgins or the language of very young children. He notes that in both of those cases, the transformation to full grammars is sudden and pronounced. Creoles arise out of pidgins within the space of a generation or two, and children move from the two-word stage to long and surprisingly complex sentences within the space of a few months. If ontogeny recapitulates phylogeny, this is strong evidence for a view that syntax emerged rapidly.

4.16

The initial part of Bickerton's theory has much to recommend it. First of all, it coincides with much of the fossil record. The brain size of our ancestors doubled between 2 million and around 700,000 years ago, most quickly when late *Homo erectus* evolved into pre-modern *Homo sapiens*. (Johanson & Edgar [1996](#)) This change in size was matched by indirect indications of language usage in the fossil record, such as the development of stone tools and the ability to control fire. Admittedly, an increase in cranial capacity may not *necessarily* coincide with greater memory and therefore the ability to manipulate and represent more lexical items. Yet that, in combination with the glimpses of behavioral advancements that would have been much facilitated by the use of protolanguage, is compelling.

4.17

However, there is one glaring drawback to Bickerton's theory. The problem with an explanation relying on a sudden genetic mutation (or even a slightly more probable fortuitous recombination) is that on many levels it is no explanation at all. It takes an unsolved problem in linguistics (the emergence of syntax) and answers it by moving it to an unsolved problem in biology (the nature of

the mutation). Still unknown is *what* precisely such a mutation entailed, how one fortuitous mutation could be responsible for such a complex phenomenon as syntax, and how such a mutation was initially adaptive given that other individuals, not having it, could not understand any grammaticalization that might occur.

4.18

Additionally, the relatively quick emergence of syntax may be explainable by routes other than biological mutation, such as rapid cultural transmission and adaptation of language itself. It is also not immediately obvious that ontogeny recapitulates phylogeny in the emergence of language, nor that it should. So even if the flowering of language is sudden in the cases of children and creolization -- itself a debated point - - that doesn't mean that the original emergence of language was also sudden.

4.19

Bickerton provides a highly original and thought-provoking theory of the evolution of language that is nicely in accord with much of what we know from the fossil record. Nevertheless, the implausibility of the emergence of such a fortuitous mutation is a fundamental flaw that makes many theorists unable to accept this account. In the next section, I will consider an alternative justification of the "nativist" perspective on the evolution of language, one that attempts to avoid the problems Bickerton falls prey to.

Pinker and Bloom: On Natural Selection

4.20

Pinker and Bloom ([1990](#)) argue that human language capacities must be attributed to biological natural selection because they fulfill two clear criteria: complex design and the absence of alternative processes capable of explaining such complexity. This argument, therefore, is less based on consideration of the characteristics of human history than is Bickerton's and more based on a theoretical understanding of evolution itself.

4.21

The first criterion, complexity, is a characteristic of all human languages. Just the fact that an entire academic field is devoted to the description and analysis of language is enough to suggest that! Less facetiously, Pinker and Bloom demonstrate its complexity by pointing out that grammars must simultaneously fulfill a variety of complicated needs. They must map propositional content onto a serial channel, minimize ambiguity, allow rapid and accurate decoding and encoding, and distinguish a range of potentially infinite meanings and combinations. Language is a system of many parts, each mapping a characteristic semantic, grammatical, or pragmatic function onto a certain symbol sequence shared by an entire population of people. The idea that language is incredibly complex is usually considered so obvious that it is taken as a given.

4.22

The second criterion, demonstrating that there are no processes other than biological natural selection that can explain the complexity of natural language, entails more than may appear on first glance. Pinker and Bloom must first demonstrate that processes not relating to natural selection as well as processes related to *non*-biological natural selection are both inadequate to explain this complexity. And finally they must demonstrate that biological natural selection *can* explain it in a plausible way.

4.23

Most of Pinker and Bloom's argument is devoted to demonstrating that processes not related to natural selection in general are inadequate to explain the emergence of a system as complex as language. They primarily discuss the possibility of spandrels, traits that have emerged during evolution but for other reasons than selection (genetic drift, accidents of history, exaptation, etc). Genetic drift and historical accidents are inadequate as explanations: a system as complex as language is, biologically speaking, quite unlikely to emerge spontaneously. This is essentially the main argument against Bickerton's "mutation" hypothesis, and is equally strong here. It is ridiculously absurd to suggest that something so complex could emerge by genetic drift or simple accidents during the short past of human history.

4.24

Exaptation is a bit more difficult to explain away. It refers to the process of coopting parts that were originally adapted to one function for another purpose. In this case, language could be a spandrel resulting from the exaptation of more general cognitive mechanisms that had evolved for other reasons. There is some reason for believing in an exaptationist explanation: the underlying neural architecture of the brain is highly conservative across mammalian brains, with no clear novel structures. This strongly argues that either language developed far earlier than we have supposed from the fossil record, language competence is biologically rooted but (implausibly) not visible in brain structure, or that language competence does not have much biological basis. Additionally, areas of the brain such as Broca's area or Wernicke's area, which are typically viewed to be specially adapted for speech, are probably modifications of what was originally the motor cortex for facial musculature. (Lieberman [1975](#))

4.25

However, the argument against the exaptation view is also strong. If general cognitive mechanisms were coopted to be used for language ability, the process of exaptation would have to have been through either *modified* or *unmodified* spandrels. If language is a modified spandrel, it is built on a biological base that was originally intended for another purpose, but then was modified to the purpose of communication. If this is correct, however, it is not an argument against the idea that language stems from biologically-based natural selection. After all, selection plays a crucial role in modifying the spandrel. In fact, the structure and location of Broca's area is often taken as evidence *supporting* viewpoints like that of Pinker and Bloom. It is quite easy to interpret the language-specific abilities of those areas of the brain as modifications of their original motor functions.

4.26

Unmodified spandrels are more interesting to consider, since if language were one -- say, an application of our general cognitive skills -- then it would clearly not have arisen through selection specifically *for language*. Yet as Pinker and Bloom point out, unmodified spandrels are usually severely limited in how well they can adapt to the function they have been coopted for. For instance, a wing used as a visor is far inferior for blocking the sun than something specially suited to that purpose that would simultaneously allow the bird in question to fly around. As the use to which the spandrel is put gets more and more complex, it is more and more improbable that the spandrel would be a useful adaptation, completely unmodified.

4.27

If the mind is indeed a multipurposive learning device then Pinker and Bloom suggest that it certainly must have been *overadapted* for its purpose before language emerged. They point out that our hominid ancestors faced other tasks like hunting, gathering, finding mates, avoiding predators, etc, that were far easier than language comprehension (with its reliance on memory, recursivity, and compositionality, among other things). It is unreasonable to assume that general intellectual capacity would evolve far beyond what was necessary *before* being coopted for language.

4.28

Additionally, scientists have thus far been unable to develop any psychologically realistic computational inference mechanism that is general purpose and can learn language as a special case. While this is not conclusive by itself -- it may, after all, merely reflect the fact that this field is still very young -- they argue that it is somewhat suspicious that most computational work suggests that complex computational abilities require rich initial design constraints in order to be effective. And it is incredibly implausible to assume that constraints that are effective for general intelligence are the exact same constraints necessary for the emergence of complex language.

4.29

The evidence considered here seems to argue convincingly that language must be the product of biological natural selection, but there are definite drawbacks. Most importantly, Pinker and Bloom do not suggest a plausible link between their ideas and what we currently know about human evolution. As noted before, they do not even consider possible explanations for the original evolution of referential communication, even though that is a largely unexplained and unknown story. As for the evolution of syntax, Pinker and Bloom argue that it must have occurred in small stages as natural selection gradually modified ever-more viable communication systems. The problem with this is that they do not suggest a believable mechanism by which this might occur. The suggestion made is, in fact, highly implausible: a series of mutations affecting the brain, each corresponding to grammatical rules or symbols. As they say, "no single mutation or recombination could have led to an entire universal grammar, but it could have led a parent with an n-rule grammar to have offspring with an n+1-rule grammar." ([1990](#))

4.30

This is unlikely for a few reasons. First of all, no neural substrates corresponding to grammatical rules have ever been found, and most linguists regard grammatical rules as idealized formulations of brain processes rather than as direct descriptions of a realistic phenomenon. Given this, how could language have evolved by the addition of these rules, one at a time, into the brain?

4.31

Secondly, Pinker and Bloom never put forth a believable explanation of how an additional mutation in the form of one more grammar rule would give an individual a selective advantage. After all, an individual's communicative ability with regard to other individuals (who don't have the mutation) would not be increased. Pinker and Bloom try to get around this by suggesting that other individuals could understand mutated ones in spite of not having the grammatical rule in question. It would just be more difficult. However, such a suggestion is at odds with the notion of how innate grammatical rules work: much of their argument that language is innate is based on the idea that individuals could not learn it without these rules. They can't have it both ways: either grammatical rules in the brain are necessary for comprehension of human language, or evolution can be explained by the gradual accumulation of grammatical rules, driven by selection pressure. But not both.

4.32

Another problem with the Pinker/Bloom analysis is that it relies on what Richard Dawkins terms the Argument from Personal Incredulity. They basically suggest that it is impossible for general intellectual functioning to be powerful enough to account for language because so far no cognitive scientists have yet succeeded in making a machine that powerful. Yet such an analysis says far more about the state of artificial intelligence than it does about the theoretical plausibility of the idea in question. There is no theoretical reason for believing that such a limit exists at all. Indeed, Pinker and Bloom may have unwittingly established a reason to believe that general intelligence might have evolved to be so powerful and flexible that it could quite easily be coopted into use for language. They point out that humans evolved in a social environment composed largely of other humans, all quite intelligent and devious. If our ancestors were competing with each other for limited resources, then there would have been a premium on abilities such as the skill to remember cooperation, detect and punish cheating, analyze and attend to the dynamics of social networks, and cheat in undetectable ways. The resulting pressures set the stage for a "cognitive arms race" in which skills such as increased memory, ever more subtle cognitive skills, and (in the case of social structure) hierarchical representation are strongly and rapidly selected for. Given this, it is indeed quite plausible to believe that a rich general intelligence may have evolved before language and only later been coopted to serve the purpose of language.

4.33

We have seen Pinker and Bloom suggest some strong arguments for believing that language is indeed the result of biologically-based natural selection. However, these arguments are vulnerable in some areas to counter-arguments based primarily on the notion that general intelligence may be the key to our language understanding. We have not yet considered the notion that language itself --

not human brains - - adapted over time for communicative purposes. Terence Deacon's presentation of this idea is the subject of the next section.

Deacon: The Natural Selection of Language Itself

4.34

One of the most plausible arguments to the viewpoint that language is the product of biologically-based natural selection is the idea that rather than the brain adapting over time, language itself adapted. (Deacon [1997](#)) The basic idea is that language is a human artifact -- akin to Dawkin's ideational units or "memes" - that competes with fellow memes for host minds. Linguistic variants compete among each other for representation in people's minds. Those variants that are most easily learned by humans will be most successful, and will spread. Over time, linguistic universals will emerge -- but they will have emerged in response to the already-existing universal biases inherent in the structure of human intelligence. Thus, there is nothing language-specific in this learning bias; languages are learnable because they have evolved to be learnable, not because we evolved to learn them. In fact, Deacon proposes that languages have evolved to be easily learnable by a specific learning procedure that is initially constrained by working memory deficiencies and gradually overcomes them. ([1997](#))

4.35

This theory is powerful in a variety of respects. First of all, it is not vulnerable to many of the basic problems with other views. For one thing, it is difficult to account for the relatively rapid (evolutionarily speaking) rise of language ability reflected in the fossil record with an account of biologically-based evolution. But cultural evolution can occur much more rapidly than genetic evolution. Cultural evolution also fits in with the evidence showing that brain structure itself apparently did not change just before and during the time that full language probably developed. This is difficult to account for if one wants to argue that there was a biological basis for language evolution, but it is not an issue if one argues that language itself is what evolved.

4.36

Another powerful and attractive aspect of Deacon's theory is its simplicity. It acknowledges that there can (and will) be linguistic universals -- as well as explains how these might come about - without postulating ad hoc mechanisms like sudden mutations in the process. It also fits in quite beautifully with another powerful theoretical idea, namely the idea that language capabilities are in fact an unmodified spandrel of general intelligence. The language that adapted itself to a complex general intelligence would of necessity be quite complex itself -- much like natural language appears to be.

4.37

Nevertheless, there are drawbacks to Deacon's idea. For one thing, he does not show a clear understanding of the importance of abstract language universals that may not have obvious surface effects. In other words, some properties of language - for example, whether it exhibits cross-serial

or arbitrarily intersecting dependencies like those found in the formal language $a^n b^n c^n$ - are not easily accounted for by the surface descriptions of language Deacon is most fond of. As a consequence, the theoretical possibility of a genetically assimilated language-specific rule system cannot be ruled out, even if 'surfacy' abstract universals can be accounted for by non-domain-specific factors. (see Briscoe [1998](#) for a discussion of this objection in more detail)

4.38

There are more general problems, too. Deacon's viewpoint is strongly anti-biological; he believes that language ability can be explained entirely by the adaptation of language itself, not by any modification of the brain. Yet computational simulations in conjunction with mathematical theorizing strongly suggest that -- even in cases where language change is significantly faster than genetic change -- the emergence of a coevolutionary language-brain relationship is highly plausible. (e.g. Kirby [1999b](#); Kirby & Hurford [1997](#); Briscoe [1998](#)) This is not a crippling criticism of his entire theory, but it the possibility of coevolution is something we would do well to keep in mind.



Comments and Discussion

5.1

The issues in both language acquisition and language evolution overlap with each other a great deal, and still have not been resolved completely. Indeed, if the work to date demonstrates anything, it is probably that -- in both areas -- the best answer is probably some combination of the two extremes under debate.

5.2

Many of the arguments against innateness are strong and have not been fully answered by the non-nativist approach. The existence of disorders such as SLI and Williams Syndrome suggests that there is some genetic component to language ability, and the fact that children never make certain mistakes indicates that the mind is initially structured in such a way as to bias them during language acquisition. Furthermore, the arguments suggesting that language had to arise out of some sort of natural selection are powerful and persuasive.

5.3

Nevertheless, there are significant gaps in this nativist account. First, and most importantly, the arguments supporting natural selection apply equally to biological natural selection and to selection of language itself. The completely biologically-based accounts considered here are tenuous and implausible since they rest on the assumption of fortuitous mutations, either one large one or many small ones. Additionally, there is evidence from the acquisition side indicating that the "poverty of the stimulus" facing children is not nearly as severe as originally thought. There are a host of compensatory mechanisms that may indicate how children can learn language by appropriate usage of their general intelligence and predispositions, rather than by relying on pre-wired rules. It is at this point that the fuzziness of the line between innateness and non-innateness

(and domain-specificity vs. non-domain-specificity) becomes so pronounced: could such predispositions be considered pre-wiring of sorts? Where should we draw the line? Should a line be drawn in the first place?

5.4

In any case, it is fairly clear that the explanation of the nature of human language representation in the brain must fall somewhere between the two extremes discussed here. Accounts of brain/language coevolution are quite promising in this regard, but they often can lack the precision and clarity common to more extreme viewpoints. That is, it is often difficult to specify exactly what is evolving and what characteristics of the environment and the organism are necessary to explain the outcomes. It is here that the power of computational tools becomes evident, since simulations could provide a rigor and clarity that is difficult to achieve in the course of abstract theorizing.

II. Simulations



Introduction

6.1

All of the inquiries above seek to understand the behavior of dynamical systems involving multiple variables and interacting in potentially very complex ways -- ways which often contradict even the most simple and apparently obvious intuitions. As such, verbal theorizing - though very valuable -- is not sufficient to arrive at a complete understanding of the issues.

6.2

Many researchers have therefore begun to look at computational simulations of theorized models and environments. Computational approaches are valuable in part because they provide a nice middle ground between abstract theorizing on one hand and rigorous mathematical approaches on the other. Additionally, computer implementation enforces rigor and clarity while still incorporating the simplification and conceptualization that good theorizing requires. Finally, and probably most importantly, computer simulations allow researchers to evaluate which factors are important, and under what circumstances, to any given outcome or characteristic. This evaluative ability is usually quite lacking in purely verbal or mathematical approaches of analysis.

6.3

While the benefits of computer simulations are recognized, there is still a relative paucity of research in this area. To begin with, knowledge involving neural nets or genetic programming (which is usually key to a successful simulation) is much more recent than is knowledge about the mathematical and abstract models used by other theorists. Additionally, it is quite difficult to program computer simulations that are realistic enough to be interesting while still generating interpretable results.

6.4

The work that has been done is roughly separable into three general categories: simulations exploring the nativist vs. non-nativist perspectives, simulations investigating details of the evolution of syntax and simulations investigating details of the evolution of communication in general. This category distinction is made in order to clarify the issues being looked at. In the remainder of this article, I will overview some of the most promising and current work in each category, discussing strengths as well as weaknesses of each approach. By the end I will bring it all together to discuss general trends and decide where to go from here.

6.5

First, however, it is useful to discuss the computational approaches used by the simulations we will be reviewing: genetic algorithms, genetic programming, and A-Life. Since these approaches are the basis of almost all of the work discussed, it is vital to have some understanding of the nature of the theory behind them.



GA, GP, and A-Life

7.1

Genetic algorithms (GA), genetic programming (GP), and Artificial Life (A-Life) are all approaches to machine learning and artificial intelligence inspired by the process of biological evolution. GAs were originally developed by John Holland in 1975. The basic idea of GAs is as follows: "agents" in a computer simulation are randomly generated bit-strings (e.g. '0101110' might be an agent). They compete with each other over a series of generations to do a certain task. Those individuals are most successful at accomplishing the task are preferentially reproduced into the next generation. Often this reproduction involves mating two highly fit individuals together, so that their bit strings become changed (just as genetic crossover occurs in biology). In this way, it is possible to create a population of agents that is highly competent at whatever task it was evolved to do.

7.2

This is most clear with a trivial example. Suppose you want to create a population that can move towards a prize located four steps away. The agents might correspond to a four-digit bit string, with each digit standing for 'move straight ahead' (1) or 'stay put' (0), and each bit stands for one move. Thus, the bitstring '1111' would be the most effective -- since it actually reaches the prize -- while the bitstrings '1011' and '1110' would be as effective as each other, since they both end up one step away from the prize. This effectiveness is scored by a fitness function. In the initial population of agents, the perfectly performing bitstring '1111' might not be created (which in fact usually happens when problems are not as trivial as this one). But some agents would be better performers than others, and these would be more likely to reproduce into the next generation. Furthermore, they may create more fit offspring via crossover, which combines two agents at once. For instance, the agents '1110' and '1011' might be combined by crossover at their midpoint

to create the agents `1111' and `1010'. In this way, optimal agents can be created from an initially low-performing population.

7.3

Artificial Life is quite similar to GA except for a different philosophical emphasis. Most GA scenarios are created to be fairly minimal except for the thing being studied. For instance, in the above example there was no attempt to immerse the agents in an artificial ``environment" of their own in which they must find their own ``food", in which reproductive success is partially dependent on finding available mates, etc. In short, A-Life attempts to model entire environments while GA (and GP) concentrate on smaller problems and compensate by, for instance, using imposed fitness functions rather than implicit ones. There are advantages and disadvantages to each. Most obviously, A-Life is promising in that it is generally a far more complete -- and, if done well, realistic -- model of evolution in the ``real world." On the other hand, it is more difficult to do well, it incorporates assumptions that may or may not be warranted, and it is also more difficult to interpret results because they are so dependent on all the conditions of the environment.

7.4

GP is just like GA except that instead of bit-strings, what is evolving are actual computer programs. Basic function sets are specified by the programmer, and initial programs are created out of random combinations of those sets. By chance, some individuals will do marginally better at a given task than others, and these will be more likely to reproduce into the next generation. Operations of mutation and crossover of function trees with other fit individuals create room for genetic variation, and eventually a highly fit population of programs tends to evolve.

7.5

Genetic programming has multiple advantages over other approaches to machine learning. Most importantly for our purposes, it is strongly analogous to natural selection and Darwinian evolution; since these phenomena are what we are most interested in studying, it makes a great deal of sense to use a GP approach. Even beyond that, there are other advantages. Genetic Programming implicitly conducts parallel searches through the program space, and therefore can discover programs capable of solving given tasks in a remarkably short time. (Goldberg [1989](#)) Additionally, GP incorporates very few assumptions about the nature of the problem being solved, and can generate quite complex solutions using very simple bases. This makes it very powerful as a tool for theorists, who often wish to explain quite complicated phenomena in terms of a few simple characteristics.

7.6

Specific details of the GP approaches used here will be discussed individually. For references about GP, please see Goldberg ([1989](#)) or Koza ([1992](#), Koza [2000](#)).

The Nativist vs. Non-Nativist Dilemma

8.1

As we saw, one of the largest issues in linguistics is the question of to what extent language need be explained through a biological evolutionary adaptation resulting in an innate language capacity. Computational simulations are especially useful in this domain, since they allow researchers to systematically manipulate variables corresponding to assumptions about which abilities are innate. By running simulations, linguists can form solid knowledge and ground assumptions about what characteristics need to be innate in order to have human-level language capacity. In this section, I will review some of the research involving computational simulations that sheds some light on this issue.

The Role of Learning

8.2

One of the simulations that most directly studied the issue of the innateness of language is work by Simon Kirby and James Hurford. ([1997](#)) In it, they argue that the purely nativist solution cannot work: a language acquisition device (LAD) cannot have evolved through biologically-based natural selection to properly constrain languages. Rather, they suggest, the role of the evolution of language itself is much more powerful -- and, surprisingly, this language evolution can bootstrap the evolution of a functional LAD after all.

Method

8.3

The two points of view being contrasted in this simulation are the non-nativist and nativist approach. The specifics of the nativist view are implemented here with a parameter-setting model. Individuals come equipped with knowledge about the system of language from birth, represented as parameters. The role of input is to provide triggers that set these parameters appropriately. This is contrasted with the Deacon-esque view, which holds that languages themselves adapt to aid their own survival over time.

8.4

In order to explicitly contrast these two views, individuals are created with an LAD. The LAD is coded as a genome with a string of genes, each of which has three possible alleles: 0, 1, or ?. The 0 and 1 alleles are specific settings ensuring that the individual will only be able to acquire grammars with the same symbol in the same position. Both grammars and LADs are coded as 8-bit strings, creating a space of 256 possible languages. A ? allele is understood to be an "unset" parameter -- thus, an individual with all ? alleles would be one without any sort of LAD, and an individual with all 0s and 1s would have a fully specified LAD without needing any input triggers at all.

8.5

What serves as a trigger? Each utterance is a string of 0s, 1s, and ?s. Each 0 or 1 can potentially

trigger the acquisition of a grammar with the same digit in the corresponding position of the LAD, and each ? carries no information about the target grammar. (1997) When "speaking," individuals will always produce utterances that are consistent with their grammars but informative only on one digit (so seven digits are ?s, and one is a 0 or 1). When listening, individuals learn according to the following algorithm:

Trigger Learning Algorithm: If the trigger is consistent with the learner's LAD:

1. If the trigger can be analyzed with the current grammar, score the parsability of the trigger with the current grammar.
2. Choose one parameter at random and flip its value.
3. If the trigger can be analyzed with the new grammar, score the parsability of the trigger with the new grammar.
4. With a certain predefined frequency carry out (a), otherwise (b):
 - (a) If the trigger can be analysed with the new grammar and its score is higher than the current grammar, or the trigger cannot be analysed with the current grammar, adopt the new grammar.
 - (b) If the trigger cannot be analysed with the current grammar, and the trigger can be analysed with the new grammar, adopt the new grammar.
5. Otherwise keep the current grammar

8.6

Basically, then, the learning algorithm only acquires a new parameter if the input cannot be analyzed with the old setting, or if the new settings improve parsability (with some probability). Thus, the algorithm does not explicitly favor innovation (since the default behavior is to remain with old settings). At the same time, it provides a means to innovate if that proves necessary.

8.7

Over the course of the simulation, individuals are given a critical period during which language learning takes place (in the form of grammatical change). This is followed by a period of continued language use, but no grammatical change -- it is during this latter period that communicative fitness is measured. During this period, each individual is involved in a certain number of communicative acts -- half as hearer and half as speaker. Fitness is scored based on both success as speaker and as hearer -- based on how many of the utterances spoken were analyzable, and on how many utterances that were heard were successfully analyzed.

8.9

With each generation, all individuals in a population are replaced with new ones selected according to a rank fitness measure of reproduction. However, the triggers for each successive generation are taken from the set of utterances produced by the adults of the previous generation; in this way, adult-to-child cultural transmission between generations is achieved. Adult-to-adult natural selection is simulated by the interaction of individuals within each generation "talking" to each other.

Results

8.10

In the first part of the simulation, Kirby and Hurford attempted to figure out if the nativist theory was sufficient by itself to explain the origin of communication. They arbitrarily set the parsability scoring function to prefer 1s in the first four bits of the grammar. Therefore, if language was successfully acquired, we would expect to find that by the end of the simulation a grammar of the form [1,1,1,1,...] would have been preferred. Results indicated that evolution completely failed to respond to the functional pressure on a purely cultural-evolutionary (rather than 'biological') level. The makeup of the average LAD after evolution varied widely from run to run, and never converged on grammars that assisted in parsing the grammar [1,1,1,1...].

8.11

Next, linguistic selection was enabled by scoring the parsability of triggers 10 percent of the time and the parsability of utterances 10 percent as well. After only 200 generations, two optimally parsable languages predominated: this is Deacon's cultural adaptation of languages in action. Interestingly, natural selection seemed to follow the linguistic selection here. After 57 generations - a point at which there are already only a few highly parsable languages in existence -- this had been formalized as only one principle (gene set to 0 or 1) in the LAD. Yet as time went on and linguistic selection had already occurred, LADs eventually seemed to evolve to partially constrain learners to learn functional languages. The interesting thing about this is that "biological" emerged only after substantial cultural evolution, rather than vice-versa as predicted by most nativists.

Comments and Discussion

8.12

Kirby and Hurford suggest that we can understand this surprising result when considering what fitness pressures face the individual. That is, selection pressure is for the individual to correctly learn the language of her speech community -- thus the most imperative goal is to achieve a grammar that reflects that of its peers. With this goal in mind, it is senseless to develop an LAD until the language has already been selected for -- otherwise, any parameters that have been set will be as likely as not to contradict those of others in the population, making them incapable of learning the language in question. Once a language has already been pretty much settled on, it makes sense to codify some of those changes into the innate grammar; before then, however, it is actually an unfit move.

8.13

This is a very thought-provoking simulation providing strong evidence not only for a coevolutionary explanation for the origins of language, but also for how in practice such a thing might have occurred. It is a nice balance to the abstract theorizing on these issues discussed in the last chapter, and demonstrates how linguistic evolution and natural selection might work in tandem to create the linguistic skills that nearly every human possesses.

8.14

The only issue with this research is that it is only directly relevant to nativist accounts that presume something like a parameter- flipping model of language acquisition. While many nativists are admittedly vague on the mechanics of language acquisition, there are other accounts that do not require as many assumptions as the parameter-flipping model. That is, the parameter-setting model implies that as soon as one parameter is set, it cannot be changed (or can only be changed a very few times). Other nativists might not be that extreme, suggesting only that people are born with innate biases inclining them to be more prone to consider one interpretation over another without immediately excluding some. Since the conclusions rely to some extent on the notion that the drawback of the nativist view is that it creates individuals that cannot parse certain languages, it is unclear how far this can be generalized to human evolution.

8.15

That said, it is at least true that it can apply to the most extreme views. And to some extent, all nativist views have to maintain that whatever is innate in the brain compels some languages to be so unlikely that they are essentially unlearnable because they take so long. To that extent, then, these results can generalize to cover most nativist positions.

Learning of Context-Free Languages

8.16

John Batali ([1994](#)) did a similar simulation to Kirby and Hurford, except that his involved the initial settings of neural networks. He discovered that in order for the neural networks to correctly recognize context-free languages, they had to begin with properly initialized connection weights (which could be set by prior evolution). Yet this should not be taken as evidence supporting a rigidly nativist approach: all that seems to be required is a general-purpose learning mechanism. To understand how Batali arrives at this conclusion, we must first take a look at some of the details of his experiment.

Methods

8.17

Batali used a combination of genetic algorithms and neural networks in a structure very similar to the work discussed above. Basically, the initial weights of a network are selected by a genetic algorithm and correspond to real numbers between -1 and +1. Each network is a recurrent neural network with three layers of units (one input, one output, and one hidden), with 10 units in its hidden layer.

8.18

In each generation, networks were trained on a set of inputs. After this, fitness (based on ability on the context-free language $anbn$) was assessed, and the top third of networks passed unchanged into the next generation (although their inputs were set to the initial values they had before training in

order to prevent Lamarckian inheritance). The other two thirds of individuals were offspring of the first third with a random vector added to the initial weights.

8.19

Training occurred with the context-free language ab^n . The networks were presented with strings from this language, preceded and terminated by a space character. Each network was trained for approximately 33,000 strings, and since they were never presented with incorrect data, the issue of lack of negative evidence in child acquisition was simulated here as well.

Results

8.20

Randomly initialized networks ultimately learned something, but performance was never high. Given the language that training occurred on, successful networks should have "known", once encountering a b rather than an a , that there would only be as many a 's as there were b 's. In general the networks did switch to predicting b 's, but it was fairly common for them to continue to predict b 's even when the string should have ended. In other words, many networks were not "keeping track" of the number of b 's that were encountered. Average per-character prediction error in this situation was 0.244.

8.21

When 24 of these networks were used as the initial generation of an evolutionary simulation, error was down to an average of 0.179, 2.2 standard deviations better than the randomly initialized networks achieved. This is interpreted as strong indication that initial weights were very helpful in acquisition of the target grammar. Similar results were achieved when networks were trained on a separate language in which individual letters stood for different actions: some symbols caused an imaginary "push", others a "pop", etc. It was the network's job to predict when the end of a string came along. Results indicate that the most successful networks were the ones that gradually developed an innate bias toward learning the languages in that class.

Comments and Discussion

8.22

This work demonstrated that in order for networks to achieve success in recognizing context-free languages, they had to begin learning with proper sets of initial connection weights. The values of those weights were learned through simulation of evolution, providing support for the idea that innate biases in human language comprehension may have been useful as well as biologically selected for.

8.23

Yet, as Batali cautions us, we cannot conclude from this experiment that this is an instance of language-specific innateness. Since the individuals involved here are neural networks, it is unclear

whether their initial settings are representative of language-specific learning mechanisms, or just general purpose ones. That is, any ``rules" the network might possess are represented only implicitly in the weights of the network -- so it is very hard to conclude that these weights represent language-specific rules at all.

8.24

There are additional problems as well. First of all, it is neither surprising nor especially interesting that improved initial weights made it more likely for the evolved neural networks to recognize the language they were trained on. After all, that is what neural networks do -- learn from examples. In order to make truly interesting claims about the nature of human language acquisition, one would need to at least demonstrate that these results hold even when the networks are trained on a class of languages rather than strings from only one. After all, human languages are believed to belong to a restricted class, thanks to the existence of linguistic universals. An LAD serves to ``pre-wire" the brain to only consider languages from that class. Thus, in order to be appropriately parallel, the neural networks would need to demonstrate that initial settings helped classify strings from a particular language in a class of languages, rather than strings from just one language. If it did that, we would still not know whether to ascribe this to general-purpose learning or language-specific inference, however.

Discussion of Language Innateness

8.25

Though it is highly debated in the theoretical literature, the issue of how ``innate" language is is directly studied in the computational literature less often than are issues like the Evolution of Syntax. Yet there has been some work done, and much of the work that is primarily geared to other areas touches upon these issues along the way. For instance, Briscoe's work on the evolution of parameter settings (which we will review later in this section) strongly implies that even under conditions of high linguistic variation, LADs evolve. Yet, it suggests that in all circumstances, evolution does not primarily consist of developing absolute parameters that cannot be changed over the course of an organism's life. Rather, default parameters (which begin with an initial value but could ultimately change if necessary) dominate, and there is an approximately equal number of set and unset parameters. This is strong evidence of a partially innate, partially non-nativist, view.

8.26

Hurford and Kirby's work is fascinating because it offers strong evidence for a more balanced, coevolutionary view of the origins of language. This is especially intriguing because it matches the tentative conclusion we arrived at by the end of Section. There is evidence and arguments supporting both extremes of the nativist debate, indicating that any answers probably lie somewhere in the middle. Hurford and Kirby, by suggesting that genetic evolution may work by capitalizing and bootstrapping off of linguistic evolution, clarify that insight into something that is finally testable and codifiable. However, questions about the innateness of language can find further elucidation in considering the larger issues of the Evolution of Syntax and the Evolution of Communication, for the details of those have large implications for the nativist and non-nativist.

Thus, it is to the first -- Evolution of Syntax -- that we now turn.

Evolution of Syntax

Parameter Setting Models

9.1

Parameter setting models of the evolution of syntax are based on the parameter setting framework of Chomsky (1981), in which agents are "born" with a finite set of finite-valued parameters governing various aspects of language, and learning involves fixing those parameters according to the specific language received as input. For instance, languages can be characterized according to word order (SOV, VOS, etc). According to the parameter setting model, a child is born with the unset parameter "word order", which is then set to a specific word order (say, SOV) after a given amount or type of input is heard.

9.2

In general, the computational simulations involving parameter setting that we will discuss here (Briscoe 1998, 1999a, 1999b) are motivated by one of four goals. First of all, the work demonstrates that it is possible to create learners who effectively acquire a grammar given certain "triggers" in their input. Secondly, the work examines to what extent environmental variables such as the nature of the input, potential bottlenecks, and nature of the population affect the grammar that is learned. Thirdly, the work discusses the extent to which this grammar induction is encoded "biologically" (in the code of the agent) as opposed to in the languages that are most learnable. And, finally, the work examines how selection pressures such as learnability, expressivity, and interpretability interact with each other to constrain and mold language evolution.

Method

9.3

The basic method followed by Ted Briscoe is essentially the same for all his research. Agents are equipped with an LAD (language acquisition device) made up with 20 parameter settings (p-settings) defining just under 300 grammars and 70 distinct full languages. Agents with default p-settings can be understood as those with some innate language abilities, whereas those without any default presets do not have any innately specified language capacities. While the subset of grammars made possible by the LAD in this study is clearly a subset of any universal grammar that might exist, it is proposed as a plausible kernel of UG, since it can handle (among other things) long-distance scrambling and even generate mildly context-sensitive languages (Briscoe 1998; Hoffman 1996). In later research (Briscoe 1999a, 1999b) LADs are implemented as general-purpose Bayesian learning mechanisms that update probabilities associated with p-settings. Thus, the actual parameters themselves have the ability to change and respond to triggers, making it possible to increase grammatical complexity/expressiveness corresponding to a growth in complexity of the LAD.

9.4

In addition to the LAD, agents contain a parser -- a deterministic, bounded-context, stack-based shift-reduce algorithm. Essentially, the parser contains three steps (Shift, Reduce, Halt); these steps modify the stack containing categories corresponding to the input sentence. An algorithm keeping track of working memory load (WML) is attached to the parser, and this algorithm is used to rank the parsability of sentence types (and hence indirectly languages).

9.5

Finally, agents contain a Parameter Setting algorithm which alters parameter settings in certain ways when the available input cannot be parsed. (see Gibson & Wexler [1994](#) for the research this algorithm is based on) Each parameter can only be reset once during the learning algorithm. Basically, when parse failure occurs, a parameter is chosen to be (re)set based on its location within the partial ordering of the inheritance hierarchy of parameters. Since each parameter can be reset only once, the most general are reset first, the most specific reset last.

In the simulation, agents within a population participate in interactions which are successful if their p-settings are compatible; that is, an interaction is successful if both agents use their parser to map from a given surface form to the same logical form. Agents reproduce through crossover of their initial, beginning-of-generation p-settings in order to avoid creating a scenario of Lamarckian rather than Darwinian inheritance.

Results

9.6

The basic conclusion from this research is that it is possible to evolve learners that can effectively acquire a grammar given certain input. Simulations revealed that learners with initially unset p-settings could converge on a language after hearing approximately 30 triggers. They converged on the correct grammar more quickly if they began with default p-settings that were largely compatible with the language triggers they heard, and more slowly if they were largely incompatible. (Briscoe [1998](#), [1999a](#))

9.7

Briscoe's work incorporates a great deal of experimentation including issues such as how the population makeup (heterogeneity, migrations, etc) affect acquisition ([1998](#), [1999a](#), [1999b](#)), how creolization may be explained using a parameter-setting approach ([1999b](#)), how an LAD and language might coevolve rather than be treated as separable processes ([1998](#), [1999a](#)), and how constraints on learnability, expressibility, and interpretability drive language evolution ([1998](#)). These are all important and interesting problems, but many fall out of the bounds of what is directly relevant to what we are studying here.

9.8

Therefore, I will limit myself to discussing only the latter two of Briscoe's results in more detail. These problems -- the extent of coevolution of language and the LAD, as well as how to what extent constraints on learnability, expressibility, and interpretability drive language evolution -- are the most directly relevant to the concerns discussed in Section 1.

9.9

First, let us consider what Briscoe's work demonstrates about the coevolution of language and the LAD. First of all, as we have seen, all learners -- even those with no prior settings -- were able to learn languages (given enough triggers and little enough misinformation). However, default p-settings did have an effect: for very common languages, default learners were more effective than unset learners, and for very rare languages, they were less effective. This promotes a sort of coevolution -- the more common a language becomes, the more incentive there is for default learners to evolve, since they are more effective; the more default learners there are, the less selective advantage there will be for languages they are not adapted to learn.

9.10

In addition to the possibility of being set as defaults (which can be changed at least once), parameters may also be set as absolutes (which cannot be changed). Absolute settings would be theoretically advantageous if the languages of a population did not change at all; then the most effective learners would be the ones who were essentially "born" knowing the language. However, in situations marked by linguistic change, default parameters or unset parameters would be most adaptive, since an "absolute" learner might become unable to learn a language that had changed over time. Interestingly, results indicated that when linguistic change was relatively slow (which was modeled by not allowing migrations of other speakers), language learners evolved that replaced unset parameters with default ones compatible with the dominant language. "Absolute" parameters, though still somewhat common, were not nearly as popular (making up about 15% of all parameter types, compared with about 70% default and 15-20% unset).

9.11

In cases with much more rapid linguistic variation (modeled by a large amount of migration), LADs still evolved. However, there was an even greater tendency to replace unset parameters with absolute parameters than there was in cases of little linguistic variation. (Approximately 60% of parameters were set to default, while as many as 30% were absolute). This may seem counterintuitive, but Briscoe theorizes that the migration mechanism -- which introduces adults with identical p-settings to the existing majority -- may favor absolute principles that have spread through a majority of the population. In general, genetic assimilation of a language (seen in the evolution of an LAD) can be explained by recognizing that the space of possible grammars is much larger than the number of grammars that can be sampled in the time it takes for a default parameter to get "fixed" into a population. In other words, approximately 95% of the selection pressure for genetic assimilation of any given grammatical feature is constant at any one time. Thus, unless linguistic change is inconceivably and unrealistically rapid, there will be some incentive for genetic assimilation, even though it may not be strictly necessary for acquisition.

9.12

When the LAD incorporated a Bayesian learning mechanism (Briscoe [1999a](#), [1999b](#)), the general trends were similar, with a large proportion of default parameters being set (e.g. 40-45% default parameters, 35-40% unset parameters, and 20% absolute parameters). This is a clear indication that a minimal LAD incorporating a Bayesian learning procedure could evolve the prior probabilities necessary to make language acquisition more robust and innate.

9.13

Having seen the pressures driving the evolution of the individual to respond to language, what can we say about the pressure driving evolution of the language itself? Briscoe identifies three: learnability, expressivity, and interpretability. They typically conflict with each other. Learnability is reflected by the number of parameters that need to be set to acquire a target grammar (the lower the number, the more learnable it is). Expressivity is reflected roughly by the number of trigger types necessary to converge on the language, and interpretability by parsing cost (in terms of working memory load). These three pressures can interact in complex ways with each other; for instance, a language that is ideally learnable will typically be quite unexpressive.

9.14

In general, agents tended to converge on subset languages (which are less expressive than full languages but more learnable) unless expressivity was a direct component of selection (i.e. built into the fitness function). When it was, agents did not learn subset languages even when there was a highly variable linguistic environment due to frequent migrations.

Discussion and Comments

9.15

Briscoe's work is noteworthy and revealing because it demonstrates that it is possible to acquire languages that are an important kernel of UG using a parameter-setting model. In addition, his work is valuable by suggesting how it is possible for a LAD and a language to coevolve; this notion suggests, perhaps, that the answer to the eternal debate about whether language is "innate" or not probably lies somewhere in the middle ground. Finally, this work is important by providing a paradigm in which to model and discuss the evolution of the languages themselves: according to the conflicting constraints of expressivity, learnability, and interpretability.

9.16

Nevertheless, there are definitely shortcomings/caveats that it is important to keep in mind regarding much of this work, at least as it applies to our purposes. First, as a model of actual human language evolution, it is unrealistic in a variety of ways. The account presupposes the pre-existence of a parser, language acquisition device composed of certain parameters (even if those parameters are initially unset), and an algorithm to set those parameters based on linguistic input. There is no room in the account to explain how these things might have gotten there in the first place. Similarly, the agents require language input from the beginning; thus, there is no potential explanation for how the original language may have originally come about. Since it's not obvious

that Briscoe intended this to be a perfect model of actual human language evolution (but rather an experimental illustration of the possibilities), this observation is less an objection than it is something to keep in mind for those researchers who are interested in models of actual human language evolution.

9.17

There are clear shortcomings even as a model of the coevolution of the LAD and human languages. The primary problem is that there is very little difference in the simulation between the time required for linguistic change and the time required for genetic change. Slightly less time is required for linguistic change (it is on the order of 10 times as fast), but it is not clear that the same relative rate is applicable to actual human populations. At least, it is intuitively quite likely that actual genetic change occurs much more slowly, relatively speaking (since entire languages can be created in the space of a few generations, as in creolization, but it may take many millenia or more for even slight genetic variation to be noticeable on the level of a population). (e.g. Ruhlen [1994](#))

9.18

A final concern with the research presented here lies in the discussion comparing learnability, expressivity, and interpretability. These represent a key confluence of pressures, but it is somewhat unrealistic that the simulation required some of them (e.g. expressivity) to be explicitly programmed in as components of the fitness measure. If one wanted to apply these results to human evolution, one would need to account for how the need for expressivity might arise out of the function of language -- an intuitively clear but pragmatically very demanding task.

Models of the Induction of Syntax

9.19

Simulations modeling the evolution of syntactic properties using induction algorithms specifically claim that these properties arise automatically out of the natural process of language evolution. (Kirby [1998](#), [1999a](#), [1999b](#)) In this research, agents are equipped with a set of meanings, ability to represent grammars (but no specific grammar), and an induction algorithm. After many generations of interacting with each other, they are found to evolve languages displaying interesting linguistic properties such as compositionality and recursion.

9.20

Many of the conclusions suggested by Kirby in this research are crucially dependent upon the methods and parameters of his models.

Method

9.21

Kirby makes a special distinction between I-Language (the internal language represented in the brains of a population) and E- Language (the external language existing as utterances when used).

The computational simulation he uses incorporates this distinction as follows: individuals come equipped with a set of meanings to express (the I-Language). These meanings are chosen randomly from some predefined set, and consist of atomic concepts (such as *{john, knows, tiger, sees}*) that can be combined into simple propositions (e.g. *sees(john,tiger)* or *knows(tiger,(sees(john,tiger)))*).

9.22

Individuals also come equipped with the ability to have a grammatical representation of a language. This representation is modeled as a type of context-free grammar, but does not build compositionality or recursivity in. For instance, both of the following grammars are completely acceptable as representations of a learner's potential I-Language. (Kirby [1999a](#))

1. *Grammar 1*: $S / \text{eats}(\text{tiger}, \text{john}) ? \text{tigereatsjohn}$

2. *Grammar 2*: $S / p(x,y) \rightarrow N/x V/p N/y$

- $V / \text{eats} \rightarrow \text{eats}$
- $N / \text{tiger} \rightarrow \text{tiger}$
- $N / \text{john} \rightarrow \text{john}$

9.23

The population dynamic differs slightly among the different studies discussed here, but there are some key features that apply to all three. In all of them, individuals of a population initially have no linguistic knowledge at all. Individuals are either "listeners" or "speakers," though each individual plays both roles at different times. As speakers, individuals must generate words corresponding to the meanings contained in their part of I-Language. If the individual has a clear mapping between a given meaning and a string (based on its grammatical representation), then it produces that string. If not, it produces the closest string it can, based on the mappings it does have. For instance, if an agent wished to produce a string for the meaning *sees(john, tiger)* but only has represented strings for the meaning *sees(john,mary)* it would retain the part of the string corresponding to the part that was the same, and replace the other part with a random sequence of characters. In this way speakers will always generate strings for any given meaning.

9.24

The job of the listener is to "hear" the string generated by speakers and attempt to match that to a meaning. Communicative success is based on to what extent the meaning/string mapping is the same between speaker and listener. Crucially, then, everything depends upon the induction algorithm by which listeners abstract from a speaker's string to the corresponding I-Language meaning. The core of the algorithm relies on two basic operations. The first is incorporation, in which each sentence in the input is added to the grammar by a trivial process of assigning it to a string. For instance, given the meaning pair (*johnseestiger,sees(john,tiger)*) the rule for induction would be $S/\text{sees}(\text{john}, \text{tiger}) \text{johnseestiger}$. The second operation is duplicate deletion, in which one of a set of duplicate rules is deleted from the grammar whenever it is encountered.

9.25

In order to give the induction algorithm the power to generalize, an additional operation exists. This operation basically takes pairs of rules and looks for the most specific generalization that might be made that still subsumes them within certain prespecified constraints. For instance, given the two rules $S/sees(john,tiger) \rightarrow johnseestiger$ and $S/sees(john,mary) \rightarrow johnseesmary$ this can be subsumed into the general rule $S/sees(john,x) \rightarrow johnsees N/x$ and the two other new rules $N/tiger \rightarrow tiger$ and $N/mary \rightarrow mary$.

Results**9.26**

Agents did indeed develop "languages" that were compositional and recursive in structure. In his analysis, Kirby found that development proceeded along three basic stages. In Stage I, grammars are basically vocabulary lists formed when an agent that does not have a string coinciding to a given meaning invents one. The induction algorithm then adds it to the grammar. After a certain point, there is a sudden shift: the number of meanings covered becomes larger than the number of rules in the grammar. This can only reflect the fact that the language is no longer merely a list of rules, but has begun to have syntactic categories intermediate between the sentence level and the level of individual symbols. This is what Kirby designates as Stage II. Stage II typically ends with another abrupt change into Stage III, which is apparently completely stable. In Stage III, the number of meanings that can be expressed has reached the maximum size, and the size of the grammar is relatively small. The grammar has become recursive and compositional, enabling agents to express all 100 possible meanings even though there are many fewer rules than that.

9.27

Interestingly, agents even encoded syntactic distinctions in the lexicon: that is, all the objects were coded under one category, and all the actions under a separate category. (Kirby [1999a](#), [1999b](#)) This may indicate that agents are capable of creating syntactic categories using their E-Language that correspond in some sense to the meaning-structure of their I-Language.

Discussion and Comments**9.28**

Kirby suggests that the emergence of compositionality and recursion can be explained by conceptualizing I-Language as being built up of replicators that are competing with each other to persist over time. That is, whether an I-Language is successful over time is dependent up on whether the replicators that make it up are successful. For every exposure to a meaning (say johnseesmary), the learner can only infer one rule (I-Language replicator) for how to say it. Thus, rules and replicators that are most general -- those that can be used to create multiple meanings -- are the ones that will be most prolific and therefore most likely to survive into succeeding generations. In this way, I- Languages made up of general rules that subsume other categories will be ultimately the most successful.

9.29

This is an intriguing analysis, since it paints a picture of the language adapting and evolving, forming a coevolutionary relationship with the actual individuals. That is, certain languages will be more adaptive and therefore more selected for, indicating a language/agent coevolutionary process. Nevertheless, it is difficult to conclude (as Kirby does) that compositionality and recursive syntax emerges inevitably out of the process of linguistic transmission and adaptation. His induction algorithm, in fact, heavily favors a grammar that is compositional and recursive. This is due to the second step, which attempts to merge pairs of rules under the most specific generalization that can subsume them both. By specifically looking for -- and making -- every generalization that can be made, this algorithm automatically creates compositionality whenever a grammar grows rich enough to have vocabulary items with similar content.

9.30

Even the algorithm used to create new string/meaning pairs implicitly favors compositionality and recursiveness. Recall that when given a meaning that has not itself been seen but that is similar to something that has been seen, the algorithm retains the parts of the string that are similar and randomly replaces that parts that are not. In doing so, it essentially creates new strings that already have begun to generalize over category or word.

9.31

The theorizing about the success of replicators in an I-Language is both fascinating and possibly applicable. However, it must be at least considered that the final grammar is compositional and recursive merely because the algorithm heavily favors compositionality and recursivity.

9.32

Kirby uses his results to suggest that the "uniquely human compositional system of communication" need not be either genetically encoded or arise from an intrinsic language acquisition device. As we have seen, his position that syntax is an inevitable outcome of the dynamics of communication systems is not supported by the experiments detailed above. If one were to try to draw the analogy between Kirby's agents and early humans, the induction algorithm could be seen as either an LAD or as a more general-purpose cognitive mechanism that had been recruited for the purpose of language processing. Each of these alternatives is completely distinct from one another and both are still a valid possibility based upon what we have seen so far. However, the difference between these possibilities needs to be further elaborated. Additionally, it would be useful to further discuss how well each alternative accords with the viewpoint that our system of communication is a natural outcome of the process of communication in general.

9.33

One final issue is less of a problem with Kirby's model per se than an observation of how it fails to meet our purposes here. Specifically, it makes an enormous amount of assumptions about the basic structure of meaning representation: all meanings are already located in an agent's "brain", and all are already stored in an orderly -- if not hierarchical and compositional -- form. Thus, the most

shown by Kirby is that, given this sort of meaning representation, compositional and recursive speech can evolve. The question which I am most interested in is: to what extent is this result dependent upon the structure of meaning representation? How does meaning representation itself evolve? How might language be different if the underlying meaning structure were otherwise? Kirby's model, as valuable as it is in other domains, doesn't attempt to answer these questions.

A Neural Network Model

9.34

Both of the approaches detailed above relied on genetic programming in a broad sense, but a few hardy researchers have explored issues in the evolution of syntax using models based on neural networks. In the work discussed here (Batali [1998](#)), agents containing neural networks alternate between sending and receiving messages to one another, updating their networks as they do so. This is considered one "episode" of communication. After multiple episodes, the agents have developed highly coordinated communication systems, often containing structural, syntax-like regularities.

Method

9.35

The communicative agents in this model contain a "meaning vector" made of ten real numbers (between 0.0 and 1.0). In any given episode of communication, each value of the meaning vector is set to 0.0 or 1.0, depending on what meaning is to be conveyed. The agents also contain a recurrent neural network that is responsible for sending and receiving characters from and to the other agents in the population. The neural networks have three layers of units (one input unit for each character, thirty context input units, a 30-unit hidden layer, and ten output units corresponding to meaning vectors).

9.36

The sequence of characters sent in any given situation is determined from the values in the speaker's meaning vector. Speakers are self-reflective; that is, they decide which character to send at each point in the sequence by examining which character would make its own meaning vector closest to the meaning it is trying to convey. This is quite similar to the approach discussed in Hurford ([1989](#)) as well as others reviewed here (e.g. Oliphant & Batali [1997](#)). Hurford found that when an agent uses its own potential responses to determine what to send, highly coordinated and complex communication systems may develop. Thus, an implicit assumption of this model is that agents will use their own response in order to predict other's response to them.

9.37

Listeners have the difficult task of setting their meaning vectors appropriately upon "hearing" a certain sequence of characters. Classification of these sequences is determined by examining the agent's meaning vector after hearing the sequence. Values are considered to have been classified

``correctly" if they are within 0.5 of the corresponding position in the hearer's meaning vector. Networks are trained using backpropagation after each character in the sequence is processed.

9.38

Meanings themselves correspond to patterns of binary digits, ten different predicates and ten different referents. The predicates are encoded using six bits (for instance, excited = 110001 and hungry = 100110). Referents are encoded using the remaining four (e.g. me = 1000 or yall = 0101). Thus, there are 100 possible meaning combinations that can be represented. The vectors for the predicates are randomly chosen, but each bit of the referent encodes an element of meaning. For example, the first position indicates whether the speaker is included in the set or not, and the second position represents whether the hearer is included. Agents were completely unaware initially of this structure as well as the distinction between predicate and referent.

9.39

In each round of the simulation, agents alternate between being speakers and listeners. When designated a listener, agents are trained to correctly distinguish the sequences sent by a randomly selected speaker, then both are returned to the population.

Results

9.40

In initial rounds of the simulation, agents are incorrect nearly all the time, not surprisingly. Even after 300 rounds speakers are sending many different sequences for each meaning, and listeners are not very accurate in interpreting them. However, there are naturally slight statistical fluctuations that increase the likelihood of certain sequences being sent for a certain meaning. These are capitalized on, and gradually agents are exposed to less contradictory input, enabling them to achieve a high degree of communicative accuracy by round 15000. By the end, over 97% of meanings are interpreted correctly, and sequences are generally much shorter than they were originally.

9.41

The sequences in the communication system that developed exhibit some regularity, although the syntax is not completely systematic. Each sequence can be analyzed as a root expressing the predicate, plus some modification to the root expressing the referent. (Batali [1998](#)) For some meanings, these sequences are perfectly regular, although for some there are significant deviations.

9.42

In addition to the basic simulation, agents were trained on input from which 10 meanings were systematically omitted. Following successful creation of a communication system, one agent was used as a speaker to generate sequences for each omitted meaning, and another was used as a listener to classify the sequences. They did so with considerable accuracy, suggesting that they made use of their similar mappings from sequences to output vectors to convey novel meaning

combinations.

Discussion and Comments

9.43

Although this simulation involves agents that create coordinated communication systems with structural regularities, it is difficult to generalize these results beyond this specific situation. This is because the neural network model involved may, like Kirby's induction algorithm, be implicitly biased towards detecting and creating regularities.

9.44

Why? The algorithm used is back propagation, which by definition attempts to assign "responsibility" to which input units were responsible for a given output. As Batali himself recognized, the most plausible explanation for the success of the simulation is that characters and short sequences of characters were effective because they encoded trajectories through the vector space of network activation values. This encoding probably also occurred as a by-product of the fact that neural nets were updated in the same temporal sequence as the characters were received, with two probable outcomes.

9.45

First of all, characters that were in close proximity together therefore naturally tended to have more influence on the outcome (together) than if they were widely separated. This itself may have driven the algorithm to "clump" characters into sequences approximating words. Secondly, characters that came first (the predicate) therefore were more important in driving the trajectory than were later characters (in the same sense that the direction one takes at the beginning of a long trip is most important in getting close to the final destination). Given this fact, it is not surprising that predicates tended to be analyzed as roots while referents were only modifications to that root. Probably if the referent were to be first in the meaning vector (or greater than four bits long), the results would be opposite.

9.46

Overall, it is difficult to apply the results discussed here to a more general picture of communication because it is difficult to tell what assumptions are necessary in order to get the results described. In addition to the implicit bias of the back propagation algorithm and neural network update process, there are apparently arbitrary characteristics of the model. For instance, why are predicates six bits long and referents only four? Why is the neural network updated after each character? How were the sizes and settings of the layers of the neural network arrived at? How plausible is the assumption that pre- linguistic individuals have enough theory of mind capabilities to use their own responses in predicting those of others?

9.47

These questions pose a difficulty because it is unclear how much the success of the communication strategy may have resulted from one of these seemingly arbitrary decisions. Batali

himself confesses that the ``model used in the simulations was arrived at after a number of different approaches...failed to achieve anything like the results described above." (1998) What were the reasons for their failure? What assumptions were made here that caused this model to avoid this failure?

9.48

Until we know the answer to these questions, we cannot generalize the results or draw solid conclusions about what they might mean regarding human language evolution and/or acquisition. This approach has potential, once these questions are answered, but until then we must wonder.

General Discussion of Evolution of Syntax

9.49

We have seen a variety of approaches attempting to simulate the evolution of syntax. Though there are definitely characteristics of these studies that have potentially fascinating repercussions for our understanding of the topic, it is also unclear how well any of them generalize to human language evolution.

9.50

The most obvious drawback is that all three models make a large number of assumptions about the characteristics of agents and their learning algorithms. Briscoe assumed that agents came equipped with the ability to set parameters (even if they were initially unset), in addition to the ability to parse sentences, an algorithm for setting parameters, and a mental grammar already fully capable of representing context-sensitive languages. Kirby assumed that agents came equipped with mental representations of meanings that were already compositional and hierarchical in nature, and his induction and invention algorithms were strongly biased towards creating and seeing compositional regularities in the input. And Batali's algorithm, based on time-locked backpropagation on the agents' neural networks, almost certainly biased the agents toward detecting and creating regularities in their speech.

9.51

In addition to these assumptions, all the researchers included more fundamental and basic ones. All the studies we have examined so far have automatically created a conversational structure for the agents to follow -- that is, agents did not need to learn the dynamics of conversation on any level. All agents were motivated to communicate with each other. In almost every case, fitness was based on the direct correspondence between speakers' and listeners' internal meaning representations.

9.52

Why is this a problem, you may ask? Insofar as we examine these studies on their own, it is not. But in the evolution and acquisition of human language, we must account for where the motivation for communication came from (especially given the potential costs associated with making noise

and drawing attention to oneself). We must account for the emergence of conversational structure. We must account for the fact that, in ``real life'', fitness is never based on a direct correspondence between two individual's internal meanings; it is based on how that correspondence translates into fit or unfit behaviors. And we must not assume that humans somehow ``came equipped'' with key tools such as parameter settings, parsers, appropriate induction and revision algorithms, or meaning representations. Otherwise, we are still left with the largest chicken-and-egg problem left unanswered: where did those come from?

Evolution of Communication

10.1

The questions above are key to our eventual understanding of human language evolution, as well as to determining how far we can generalize the results from these simulations of the evolution of syntax. Because answers to these questions are so important, computational work has been done in an effort to find them. In this section we will review some of the most promising work in the field.

Evolution of a Learning Procedure

10.2

The most prevalent assumptions in the work reviewed in the last section were the assumptions stemming from the nature of the learning procedure used in the simulation. Quite often, we found, the learning procedure itself was implicitly biased towards developing syntax or other language-like properties. However, the problem is not the existence of a biased learning procedure per se -- the problem is only that no explanations are made for how one might evolve. The first research we will discuss here examines this very topic, asking how coordinated communication might emerge in the first place among animals capable of producing and responding to simple signals. Clearly this question is more basic than the ones analyzed in the last section; thus, satisfactory answers to it may serve as a solid stepping-stone toward our larger goals.

Method

10.3

In order to benefit from linguistic ability, animals must first have the ability to coordinate their communicative behavior such that when one animal sends a signal, others are likely to listen and respond appropriately. Oliphant and Batali ([1997](#)) investigate how such coordination may have evolved.

10.4

Their analysis revolves around what they term a ``communicative episode.'' In such an episode, one member of a population produces a signal upon noticing a certain type of event. The other animals recognize the signal and respond to it. It is a successful episode if the response is appropriate to the situation. Any given individual's behavioral dispositions to send or receive

(appropriately recognize) signals is characterized with two probability functions, aptly titled "send" and "receive." For instance, imagine that a leopard is stalking one of our agents. Then the meaning it wishes to impart is leopard. It has a variety of signals it can use to send this: barking, coughing, chattering, etc. The probability function encodes the probability that any of those methods will be the one chosen: an example probability set might be: [bark = 0.7, cough = 0.2, chatter = 0.1]. Communicative accuracy, under this paradigm, is defined as the probability that signals sent by an individual using its "send" function will be correctly interpreted by another individual using its "receive" function.

10.5

The key concern of this research is to determine how individuals might learn to communicate, and thus the bulk of Oliphant and Batali's paper is devoted to an analysis of different learning procedures. The simplest learning procedure that might theoretically have a chance of success is dubbed Imitate-Choose. Using this procedure, learners will send the signal most often sent for any given meaning and will interpret each signal in the way most of the population does.

10.6

The other learning procedure, called the Obverter, is based on the premise that if one wants one's signal to be interpreted correctly, one should not send the signal most often sent for that meaning but instead the signal most often interpreted for that meaning. Since it is implausible to assume that a learner actually has access to the population send and receive functions, they are in all cases restricted to only approximations based on a finite set of observations of each.

Results

10.7

The Imitate-Choose strategy exaggerates the communicative dispositions in the population. In other words, if the system is highly coordinated to begin with, the strategy will maintain this coordination and prevent degradation. However, if it is initially non-optimal, it will do nothing to make it more coordinated; it may even become further degraded over time.

10.8

In contrast, the Obverter procedure is quite effective: communication accuracy reaches 99% after only 600 rounds of the simulation. Even approximations to the Obverter -- which are more realistic by relying on a limited set of observations of communicative episodes - achieve excellent accuracy (98% after 1200 rounds for the Obs-25 (the one based on 25 observations)). As the number of observations declines, accuracy naturally goes down. However, even with Obs-10 (based on 10 observations), learning occurs; accuracy for that procedure eventually asymptotes at approximately 80%.

Discussion and Comments

10.9

Oliphant and Batali interpret the success of the Obverter learning procedure to indicate that what is important for an agent to pay attention to is not other's transmission behavior, but instead its reception behavior. On one level, this makes a great deal of sense; on the other hand, it is quite doubtful that this process accurately describes human language acquisition. First of all, it is well-established that young children's utterances are exceedingly well-coordinated with the frequency and type of words in their input - - that is, the transmission behavior of the people around them. (e.g. Akhtar [1999](#); Lieven et al [1997](#); De Villiers [1985](#)) Secondly, it is implausible to suggest that children keep statistical track of the average reception behavior of other people as they are learning language; indeed, children seem not to tune into language not geared specifically for their ears.

10.10

Another issue with Oliphant's and Batali's research is that, contrary to their claims, it does not explain how coordinated communication might emerge. It does suggest a learning algorithm by which agents who initially do not coordinate their communication might eventually do so. But it provides no justification for the evolution of the Obverter in the first place, either as a language-specific algorithm or as a general cognitive function that has been coopted for the use of language. Lacking such a justification, we are nearly in the same place we began: with no solid link between a pre-linguistic human ancestor and who we are today.

10.11

Finally, as before, this work makes certain fundamental assumptions that are still unanswered. For instance, the agents here are automatically provided with a set of meanings, as if they sprang full-blown into their heads. Although no special assumptions were made about the structure of those meanings, we are still left wondering where they came from in the first place. As with other work covered here, this is not an objection to their work itself, only to how it fills our needs. Oliphant and Batali were not seeking to eliminate all basic assumptions and start from scratch, so the fact that they didn't is not their problem. Nevertheless, since we are ultimately interested in this, it makes the research reported here less valuable to our purposes than it might otherwise be.

10.12

Our questions about what might have caused coordinated communication to emerge in the first place have not been answered to satisfaction so far. Let us move on to two other pieces of research investigating that very topic.

Evolution of Coordination

10.13

In order for a successful communication system to evolve, there must be some selective advantage to both speakers and listeners of that language. This poses a difficulty, because it is difficult to see what the advantage to a speaker might be in the simplest of situations. For a listener, it is obvious; those individuals better able to understand and react appropriately to warnings about predators,

information about food, etc, are more likely to survive into the next generation. Yet what motivation does an animal have for communicating a warning when making noise might make it more obvious to a predator? Why should an animal tell others where the food is when keeping quiet would allow him to eat it for himself?

10.14

These questions are definitely reminiscent of work on the so-called Prisoner's Dilemma and the difficulty coming up with an evolutionary explanation for altruism. The two studies we will examine here both take on these questions, albeit from slightly different angles. (Oliphant [1996](#); Batali [1995](#))

Method

10.15

Both studies involve agents who can be either listeners or speakers, and both analyze the parameters necessary for coordinated communication to evolve. In Batali ([1995](#)), agents contain a signaling system made up of two maps: a "send" map mapping from a finite set of meanings to a set of signals, and "receive" map mapping in just the opposite direction. All members of the population are assumed to have a signaling system with the same sets of meanings and symbols, though not necessarily the same mappings. During communicative episodes, one animal begins with a specific meaning and produces the meaning corresponding to it according to its "send" map. A second animal, overhearing the signal, attempts to determine what meaning it may be mapped onto by using its "receive" map. A conversation is a success if the animals have made the same meaning/signal mapping.

10.16

Each individuals' receipt coordination is defined as the average (over all members of the population) of the fraction of their signals that the individual can provide the correct mapping for. Because, evolutionarily speaking, there may be little advantage to speaking but large fitness advantage to listening, only receipt coordination is important for fitness; success at sending messages is irrelevant. The question is whether the signaling coordination of a population (the average of the values of receipt coordination of each individual) converges to a high value. In other words, do populations that only reward listeners, but not speakers, ever generate coordinated communication?

10.17

Michael Oliphant ([1996](#)) asks the exact same question, but his agents are bit-strings using genetic algorithms that are made up of a two-bit transmission system and a two-bit reception system. The transmission system produces a one-bit symbol based upon a one-bit environmental state (so the system '01' might produce a 1 when in environmental state 0). Similarly, the reception system produces a one-bit response based upon the one-bit symbol sent by the speaker. As in Batali's work, the fitness function discriminates between transmission and reception systems: fitness is

based upon only the receiver's average communicative success. In other words, if a speaker and listener communicate successfully, the receiver gets rewarded; otherwise, it gets punished. Nothing happens to the speaker either way. Again, this is done in order to simulate the perceived lack of reward for speaking in the real world.

Results

10.18

In both studies, simple reward of only receiver's actions does not result in a completely coordinated, stable system. However, it does result in a "bi-stable" system in which the particular communication system that emerges to be dominant at any one time does not stay dominant, but transitions sharply back-and-forth between another communication system. In this way, two communication systems flip-flop back and forth indefinitely.

10.19

This bi-stable equilibrium can be explained by recognizing that since reception is the only behavior that contributes to fitness, it is profitable to agents to converge on a system so that reception improves. However, it is also profitable for speakers to not speak according to that system (but hope that all other agents do) in order to maximize reception-based fitness relative to everyone else. In this way, systems of communication will emerge and become dominant, only to suddenly make a sharp transition to another system as soon as enough "renegade" mutants form. This result is clearly a robust one, since we see similar behavior in both studies, even though their implementations are significantly different.

10.20

The parallels between this situation and the Prisoner's Dilemma are striking, so Oliphant ([1996](#)) pursued the analogy further by simulating variants of the scenario that are analogous to strategies successful in promoting altruistic behavior in the typical Prisoner's Dilemma. In one such variant, individuals are given a three-round history allowing them to document the actions of themselves and their opponents so that they know who is trustworthy. They are also given a means by which to alter their behavior based on the past behavior of the opponent. The idea, of course, is that individuals who constantly renege by speaking a language that is not the common one will shortly find themselves being spoken to in an unpredictable language as well.

10.21

This is indeed the case. Individuals eventually evolve a "nice" communication system that is primary and unchanging over time (and therefore predictable for receivers) as well as a "nasty" one that is unstable and therefore unpredictable. The most successful agents are those that begin with the most stable system, but punish those who have given them incorrect information by switching to the secondary system. It should be noted that this system is not completely stable, since after multiple rounds all individuals are consistently using the primary system, and there is no longer selection pressure on the secondary system. It hence begins to "drift" towards being accurate, and a few non-cooperators begin to infiltrate the system. Eventually a slightly more

careful strategy emerges.

10.22

In addition to this explanation of altruism (which is strongly reminiscent of Axelrod's 1984 Tit-for-Tat approach), many theorists have suggested that altruism may evolve through some process of kin selection. In other words, an agent will tend to be "nice" to others -- even if there is potential harm to itself -- in proportion to the degree that those others are related. That way, even though it might die, its genetic material is more likely to survive than if it didn't. Oliphant applies this approach to explaining the emergence of communication systems, suggesting that it is in an individual's interests to communicate clearly with kin, and hence stable systems can evolve.

10.23

This is simulated by creating spatially organized populations in which agents are more likely to mate with individuals close to them, and their offspring end up nearby as well. The result is a space where individuals are more related to those nearer to them. After 100 generations or so, there is indeed a stable communication system dominating the entire population. The more individuals communicate and mate only with those very close to them, the more pronounced the effect is; as distance increases, the general pattern remains, but much less stably.

Discussion and Comments

10.24

This research, unlike all the rest that we have discussed so far, genuinely gets at the heart of the question of how coordinated communication might evolve, given the selection pressures that are always acting against it. While clearly the domain is highly simplified and idealized, it takes no huge liberties with the essence of early human communication.

10.25

In general, these simulations give a plausible explanation for how agents in a population might converge on the same language system, even when they only personally gain by having good reception behavior. It should be noted that this result has only been found to hold for systems that have a relatively small number (less than 10 or so) of distinct signals to be sent. Thus, while analogous results might begin to explain the emergence of coordinated systems of communication such as those seen today among animals such as vervet monkeys (Cheney & Seyfarth [1990](#)), it is not clear that they can be extended towards explaining how more complex systems, like human communication, might evolve on top of that.

10.26

As always, there are a few assumptions made: for instance, one is that agents already have the ability to voluntarily send and receive signals. Another is that all agents have the same set of meanings and signals. And still another is that selection pressure is directly for communicative success (even if in this case it is solely receptive success). As we have already noted, such a

directed fitness function -- though it definitely simplifies the creation of the model - is implausible in an actual evolutionary context. Agents are never rewarded directly for their success in communication, only for the greater ability to handle their environment that successful communication bestows.

10.27

In the following section, we shall review a work that rectifies this shortcoming by assigning fitness scores based on success in a task that itself relies on communicative success.

Evolution of Communication Among Artificial Life

10.28

The basic idea behind this research is to simulate environments that themselves exert some pressure for agents to communicate. (Werner & Dyer [1992](#)) In this way, animal-like communication systems may evolve. Theoretically, as the environment gets more complex, progressively more and more interesting communication systems result, providing a possible explanation for the emergence of human language.

Method

10.29

In (Werner & Dyer [1992](#)), simulated animals are placed in a toroidal grid, occupying about 4% of the approximately 40,000 possible locations in the environment. Individuals designated ``females'' are the speakers: they have the ability to see the males and emit sounds. Males, on the other hand, are blind, but can hear the signals sent out by females. It is the job of the female to get the blind male to come near her so they can mate and create offspring. Thus, only those pairs who are successful at communication will consistently find mates and reproduce two offspring (one male and one female), ensuring that their genetic material exists in future generations.

10.30

Both males and females have a distinct genome that is interpreted to produce a neural network that governs its actions. Thus, this is a GA application in which each gene in the genome contains an 8-bit integer value corresponding to the connection strength of each unit in the neural network of the animal. This network is a recurrent network in which all hidden nodes are completely interconnected and can feedback to themselves. All individuals have coding in the genome to be both male and female, and the sex of the animal determines which part of the genome is interpreted to create the neural network (different for females and males).

10.31

What happens in a simulation is this: a female ``spots'' a male using an eye that can sense the location and orientation of nearby animals. This creates activation of her neural net and produces a pattern of activation of her output units, which is translated as a sound by any males that might

overhear her. This sound serves as input to the male, activates his neural net, and results in outputs that are interpreted as moves. In this simulation, females have three-bit outputs (hence 8 different possible sounds).

Results

10.32

In the initial phases of the run, males and females behaved randomly: females emitted random "sounds" for the males to hear, and males moved in random directions. Over time, the males started demonstrating strategies: agents who stood still were selected against, and those that continuously walked in straight lines, maximizing the area they covered, were selected for. At this point, there was no effect of the signals females sent; indeed, males who paid attention to those were usually selected against, since there was no consistent communication system between females. Thus, what worked for one was unlikely to work upon encountering another one.

10.33

After enough males began incorporating this straight-line strategy, more and more males began to pay attention to the females. This is almost certainly because, given that all of them were incorporating an optimal non-communicative strategy, increased fitness could only be possible by endeavoring to communicate. As more males paid attention to females, there was pressure on females to send signals that were likely to be interpreted correctly. Thus, over time, a stable system of communication began to emerge.

10.34

Interestingly, the best males were essentially "bilingual", using some of their bits to respond to signals that one dominant subpopulation of females sent, and using the rest to respond to signals from another dominant subpopulation. After a very long time, even these subpopulations converged into one single communication system.

10.35

In addition to this basic scenario, Werner and Dyer modeled the creation of dialects by separating agents by the means of barriers. They found that when barriers let approximately 80% of the individuals change sides, dialects tended not to form over the long term; there was enough exchange of genetic and linguistic material to create a single communication system (though it took longer). When barriers were more impermeable, separate dialects did indeed form, with individuals on either side of the barrier converging on their own separate languages.

Discussion and Comments

10.36

Though in many ways this study is not a realistic and applicable model of human language evolution, in one respect it is the best of all the ones discussed so far. Unlike the others,

communicative fitness is measured only insofar as effective communication helps individuals to succeed at some other task. This type of fitness is probably much more reflective of the effects of selective pressure in the real world.

10.37

Another forte of the simulation, in comparison with the other research discussed here, is that it does not pre-equip agents with too many capabilities. At no point is a conversational structure programmed in, except insofar as females emit sounds and males hear them. In other words, there is nothing compelling a back-and-forth exchange reminiscent of dialogue, nor even compelling males to act on the signals emitted by females. Indeed, in the beginning of the simulation they do not act on them. Additionally, unlike all other research reviewed here, agents do not come pre-equipped with sets of meanings. Instead, any meaning that exists in the scenario only results as an emergent property of the task of agents.

10.38

However, one potential issue is the nature of the environment that is modeled. Though separating the functions of female and male is an excellent first step in simplifying the conditions of communication, it is highly unrealistic in the real world. One of the difficulties in managing communication among humans, in fact, involves how to manage the alternation between listener and speaker.

10.39

In addition, this simulation suffers from some of the same difficulties in generalization as does the research covered earlier (Oliphant & Batali [1997](#); Oliphant [1996](#)). That is, it essentially stops at the linguistic stage of certain animals like vervet monkeys. It has demonstrated a plausibly and highly simplified -- but realistic -- account of how basic communication systems like those shared by monkeys might have evolved. Nevertheless, there are huge gaps between the communication systems of other animals and the communication system of humans: gaps not only in degree, but probably in kind as well. Human language, as we have discussed, employs compositionality (and many other grammatical strategies) to convey a potentially infinite number of meanings with a small grammar. Even more basically, while animals can communicate a small number of meanings, this communication is usually ritualized, involuntary, and limited to only that set: there is very little production of new meanings among animals, except possibly over the span of generations.

10.40

That said, the paradigm used by Werner and Dyer may be able to be elaborated to incorporate more complexity and require more of the agents in the scenario. For instance, the "ears" used by the males can be improved, allowing them to hear multiple females at once. This would require them to develop the ability to screen out which calls were most important (i.e. which females were closer). As more complexity is added to the scenario, more complex language-like behavior could potentially emerge.

A Synthetic Ethology Approach

10.41

One objection to the research by Werner and Dyer is that -- even though it is highly simplified in comparison with other work discussed in this chapter -- it still unavoidably contains multiple assumptions about the agents in the environment. A piece of research by Bruce MacLennan and Gordon Burghardt (1995) attempts to rectify this problem by simplifying even further. The investigators created a population of individuals whose fitness was a measure of the degree of cooperation between them. The organization of the signals used by the population as well as average fitness was compared under three conditions: when communication was suppressed, when communication was permitted, and when communication as well as learning were permitted. When communication was allowed, cooperative behavior evolved, while when it was suppressed, cooperation rarely arose above chance levels. And when learning was also permitted, evolution proceeded significantly faster still.

Method

10.42

MacLennan and Burghardt began with moderately sized populations (around 100 organisms) of finite-state machines coded as genomes incorporated into a genetic algorithm. Each finite state machine is determined by a number of condition/effect rules of the form $(\Sigma \gamma, \lambda) \rightarrow (\Sigma', R)$. Σ is a value representing the internal state of the organism, γ represents the global state of the simulation, λ represents the local state of the organism, and R is a response. Essentially, organisms base their "behavior" on the state of the world, their own internal state, and something they know that is also inaccessible to the other organisms (the local state λ).

10.43

Each organism is a finite-state machine consisting of a transition table for all possible states; thus, it is completely determined. The transition tables are represented as genetic strings based on the idea that each state can be represented by a finite number of integers. For example, the global environment states can be represented by integers $(1, 2, 3, \dots, G)$, local environment states by $(1, 2, 3, \dots, L)$, and internal states by $(1, 2, 3, \dots, I)$. A transition table will therefore have IGL entries in a fully-defined organism.

10.44

An organism's responses may fall into one of two categories: either an emission or an action. An emission has the form *emit*(γ) and puts the global environment into state γ' . An action has the form *act*(λ') and represents an attempt to communicate with an individual with local state λ' . Thus, *act*(λ') does nothing besides comparing λ' to the local environment of the last organism; if they match, the organisms are considered to have cooperated. In order for successful communication to occur, organisms need to make use of both responses at some point: emissions are necessary to transfer information about local environment into the global environment, where it is accessible to

other organisms. And actions are necessary to base a behavior on the content of another organisms' local environments - in other words, cooperate.

10.45

What is the importance of cooperation in this simulation? Quite simply, fitness is directly calculated from the number of times an organism has cooperated with another. Thus, it is essentially measuring the number of times the organism has acted based on another individual's local environment. Since local environment itself is unavailable except through communication via the global environment, measures of cooperation are a direct measure of communication. If a group of organisms cooperates significantly more often than they would by chance, we can say they are *communicating* in an elemental sense with each other.

10.46

The difference between cooperation and cooperation plus learning is also explored here. When learning is enabled, organisms that "make a mistake" by acting noncooperatively can change the rule matching the current state so that it would have acted correctly. For example, if the rule that matches the current state is $(\Sigma, \gamma, \lambda) \rightarrow (\Sigma', \text{act}(\lambda))$ but the local environment of the last emitter is in state λ'' , which is not equal to λ' , then cooperation fails. In that case, its rule would be changed to $(\Sigma, \gamma, \lambda) \rightarrow (\Sigma', \text{act}(\lambda''))$. Note that this is the simplest possible form of learning, since it is only based on a single case, and it is not necessarily true that the next time these conditions recur, that will be the correct action. This does not represent Lamarckian learning, however, since the "genotype" -- the GA corresponding to the transition table -- is never modified during the course of the organism's life, even if learning takes place.

10.47

The number of global environmental states G of each organism precisely matches the number of local environmental states L possible, ensuring that there are just enough "sounds" to match the possible "situations." The machines have no internal memory, so there is just one internal state.

10.48

Overall, experiments are run for an average of 5000 breeding cycles, although some are run an order of magnitude longer. Each breeding cycle consists of environmental cycles, each of which is made up of several action cycles. In an action cycle each organism reacts to its environment as determined by its transition table. After five action cycles, the local environments are randomly changed and five more action cycles occur, making one environmental cycle. After ten environmental cycles, breeding occurs. Thus, each breeding cycle consists of 100 action cycles and 100 opportunities for cooperation.

Results

10.49

Not surprisingly, the condition in which communication is suppressed by adding a large amount of

``noise" to the global environment results in levels of cooperation no different from chance. However, when this constraint is removed, cooperation is significant. By the end of 5000 breeding cycles, populations achieve 10.28 cooperations per cycle - a number 65% above the chance level. Linear regressions indicate that fitness increases 26 times as fast as when there is no communication. Thus, there is a clear indication that communication is having an effect.

10.50

When learning is enabled, fitness is dramatically increased. There are now 59.84 cooperations per breeding cycle, which is 857% above chance, increasing at 100 times the rate when communication was suppressed. We can see evidence of communicative activity when we examine the denotation matrix representing the collective communication acts of the entire population. By the end of the run, some symbols have come to denote a unique situation, and certain situations have symbols that typically denote them. The entropy of the denotation matrices is much smaller when communication is enabled ($H = 3.95$) and when communication and learning are enabled ($H = 3.47$) than when neither is ($H = 5.66$ -- almost the maximum level of 6). In this way it is possible to tell that the strings emitted by the agents are in some way contentful.

10.51

Possibly of most interest to those interested in the next step - the development of syntax -- are the experiments done where there are fewer global environmental states than local environmental states. Thus, an adequate description of a local environment situation would require two or more symbols, and possibly push towards a rudimentary ``syntax." In this situation, as before, organisms can only emit one symbol per action cycle; however, they now have the theoretical ability to remember the last symbol they emitted, making them capable of emitting coordinated pairs of symbols. Evolution runs for longer, but results in successful communication: entropy drops from a maximum of 7 to a level of 4.62.

10.52

Most interesting are the characteristics of the ``language" that evolves. For the most part, there is an extensive reliance on the second (most recent) symbol of a pair -- not surprising, since that doesn't require the organism to remember the first. However, there are occasional forms where both symbols were used, though they are not prevalent. This seems to indicate that, while they aren't completely ineffective, the machines don't evolve to make full use of the communicative resources at their disposal by developing multiple-symbol ``syntax." MacLennan and Burghardt suggest that this indicates that this step is evolutionarily hard, especially since it doesn't seem to improve as the organisms are given more time to evolve -- rather, they plateau at a certain point and never improve after that. Nevertheless, even under circumstances where a multiple-symbol language would have resulted in improved communication, organisms were capable of developing something.

Discussion and Comments

10.53

As with the Artificial Life task, this is noteworthy because it represents an attempt to evolve communication by selecting for performance on another task, namely cooperation. Nevertheless, it is worth pointing out that it has only limited applicability to actual evolutionary scenarios, since fitness is a direct measure of cooperation, which itself is a direct measure of communication. That is, there are no other ways for an organism to cooperate except through communication. Thus, essentially, the fitness function is a direct measure of communication. There is nothing wrong with this per se -- however, if one's goal is to see how communication evolves when there is no direct pressure for it (as probably happened on the evolutionary level) then this is not applicable.

10.54

It is also somewhat interesting how long it takes the organisms here to achieve successful communication, at least in relation to the other simulations reviewed here. The best population achieved 59% accuracy after 5000 breeding cycles -- and while this is far above chance performance and reveals significant communication, it is also far below a level that our ancestors presumably attained (or even the accuracy that vervet monkeys attain now). What might be an explanation for this, especially compared to other, more successful simulations?

10.55

A final question stemming from this work is the extent to which it may be used to achieve valid insights regarding the evolution of syntax. MacLennan and Burghardt indicate that their organisms' failure to fully make use of the multiple-character "syntax" means that, as an evolutionary step, that is difficult. Yet they do not rule out the possibility that this failure stems only from the difficulty of the scenario they have set up -- a scenario that does not correspond plausibly to this stage in evolutionary time. For instance, their agents do not have the ability to remember more than one digit at a time, and can only remember a maximum of two. This is highly implausible, as experiments -- and common sense -- have demonstrated that even animals like dogs can remember multiple-word commands.

10.56

Furthermore, even if the organisms in this scenario had developed the ability to use multiple symbols, this does not necessarily serve as the initial stages -- or even the logical precursor -- of the development of syntax. It might have, if the first symbol and the second symbol were related in a way that was more than just the sum of the parts. But it is more likely that they would just be combined in such a way that there become 8 different combinations for each of the 8 different global states. To truly force syntax, one might need to create an environment where there is truly no way to communicate something in the amount of space given, necessitating the development of marking certain structures and coding others appropriately.

Comments on Evolution of Communication

10.57

The research about the evolution of communication reviewed here definitely covers an earlier evolutionary time frame than the research about evolution of syntax, and is valuable in that it

provides a stepping-stone by which to account for some of the assumptions made by the latter. As we have seen, there are some fundamental issues encountered by researchers wishing to account for how stable communication systems might evolve. How does stability arise in communication, given that there is selective pressure for listeners to improve their skills, but -- because speakers probably do not get the same direct benefit from communication as do listeners -- no equivalent pressure on speakers? How might selective pressure for communication be modeled in a way that does not involve a fitness function that directly selects for communication? How might learning procedures capable of analyzing language evolve out of an initially non-linguistic state?

10.58

The research we have discussed begins to shed light on some of these issues. It has demonstrated that, due to kin selection and evolution of altruism, it is at least possible for stable systems of communication to emerge even if there is no selection for effective speakers. It has also demonstrated that in a scenario in which fitness is based only indirectly on communication, the evolution of stable systems is still possible.



General Discussion

11.1

Most of the work on computational simulations of language evolution, as we have seen, can be classified into work on the emergence of syntax, work on the emergence of coordinated communication, or work attempting to approach the larger issues regarding the innateness of language. This work is valuable and impressive for a variety of reasons, including its ingenuity, many of the findings and their implications, and their structure and methodology.

11.2

However, while an enormous amount has been learned from both approaches, there is still much to be done. One of the largest shortcomings in the research discussed here is the enormous gap between the evolution of communication and the evolution of syntax. Studies on the emergence of coordinated communication often begin by addressing the most basic issues of human language evolution: how do stable systems of communication arise, given constraints on what types of selection plausibly affect agents and the environments they are in? How may we account for simple features of communication systems, like the fact that they are shared by all members or a population, or the fact that in order to use them, individuals must use appropriate listening and speaking behavior at the appropriate times? Thus, while these simulations can suggest how the initial steps of language evolution might have occurred, they generally fall far short of making any claims about human (as opposed to animal) language. This is not surprising, given that the intent of most of these studies was not to make strong claims about human language per se; nevertheless, this latter step is one we would like to ultimately make.

11.3

By contrast, studies on the emergence of syntax typically include a vast number of assumptions

about these more fundamental questions. Typically, agents come equipped with meanings already represented (often in a structured manner) in their ``brains"; learning algorithms, grammar structures, and parsing algorithms are also usually specified. In all the work we have studied, dialogue structure and turn-taking has been explicitly programmed in. Therefore, although the question of how coordination evolves is as yet unanswered, these assumptions have made it such that we are not much closer to an answer than we were before. While the simulations we have reviewed can and do tell us a great deal about how various initial assumptions may account for the evolution of syntax, they tell us very little about how valid those assumptions are in the first place. Again, this is natural given that most of these studies explicitly recognized that they were incorporating many assumptions and were not seeking to fully eliminate them. In order to create a complete theory of human language evolution, however, we must work to begin challenging them.

11.4

There are few attempts to bridge the gap between work on the evolution of syntax on one hand, and work on the emergence of communication on the other. Those that do, while valuable for other reasons, often fail to provide a convincing explanation that can be easily and appropriately generalized to the case of human language. (MacLennan & Burghardt [1995](#)) Research that links the two approaches by avoiding the assumptions of the first while extending the implications of the second would be incredibly valuable. Additionally, insights from studying the innateness of language could be used to shed light on to what extent nativist assumptions -- both about meaning and about the emergence of dialogue structure -- might be validly utilized. Most interesting of all would be the development of an simulation that did this while incorporating the strengths of the various studies reported here -- for instance, a reliance on a fitness function that does not directly measure communication, while still having complex enough input to allow for the development of syntax-like constructions.



References

- AKHTAR, N (1999) Acquiring basic word order: evidence for data-driven learning of syntactic structure. *Journal of Child Language*, Volume 26. Cambridge University Press: 339-356.
- BATALI, J. (1994) Innate Biases and Critical Periods: Combining Evolution and Learning in the Acquisition of Syntax. *Artificial Life: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Organisms*. eds. R. Brooks and P Maes. MIT Press: Cambridge, MA.
- BATALI, J. (1995) Small signaling systems can evolve in the absence of benefit to the information sender. Draft.
- BATALI, J. (1998) Computational simulations of the emergence of grammar. *Approaches to the*

Evolution of Language: Social and Cognitive Bases. eds J. Hurford, M. Studdert-Kennedy, C. Knight. Cambridge University Press.

BELLUGI, U; MARKS, S; BIRHLE, A; SABO, H. (1991) Dissociation between language and cognitive function in Williams Syndrome. *Language development in exceptional circumstances*, eds D. Bishop and K. Mogford. Hillsdale, NJ:Lawrence Erlbaum.

BICKERTON, D. (1981) *Roots of Language*. Ann Arbor, MI: Karoma.

BICKERTON, D. (1984) The Language Bioprogram Hypothesis. *Behavioral and Brain Sciences*, Vol 7.2: 173-222.

BICKERTON, D. (1990) *Language and Species*. University of Chicago Press, Chicago, IL.

BRISCOE, E.J. (1998) Language as a complex adaptive system: Coevolution of Language and of the language acquisition device. Proceedings of the 8th Meeting of Computational Linguistics in the Netherlands. Rodopi, Amsterdam:3-40.

BRISCOE, E.J. (1999a) The acquisition of grammar in an evolving population of language agents. Proceedings of the Machine Intelligence, Volume 16. Oxford University Press, Oxford.

BRISCOE, E.J. (1999b) Grammatical acquisition and linguistic selection. draft for Linguistic evolution through language acquisition: formal and computational models, (ed.) Briscoe, E.J., CUP, in prep.

BROWN, R. (1958) *Words and things*. New York:Free press, MacMillan.

CHENEY, D and SEYFARTH, M. (1990) *How monkeys see the world: inside the mind of another species*. University of Chicago Press.

CHOMSKY, N. (1981a) *Government and Binding*. Foris, Dordrecht.

COPPOLA, M; SENGHAS, A; NEWPORT, E; SUPALLA, T. (1998) *Evidence for verb agreement in the gesture systems of older Nicaraguan home signers*. Boston University Conference on Language Development. Boston, MA.

CURTISS, S. (1977) *Genie: A linguistic study of a modern-day "wild child"*. New York:Academic Press.

DEACON, T. (1997) *The Symbolic Species*. Penguin: London.

DEMETRAS, M.; POST, K.; and SNOW, C. (1986) Feedback to first language learners: the role

of repetitions and clarification questions. *Journal of Child Language*, 13:275-292.

DE VILLIERS. J. (1985) Learning how to use verbs: lexical coding and the influence of the input. *Journal of Child Language*, Volume 12. Cambridge University Press: 587- 595.

EIMAS, P.; SIQUELAND, E.; JUSCZYK, P; VIGORITO, J. (1971) Speech perception in infants. *Science*, 171:303-306.

FERNALD, A and SIMON, T. (1984) Expanded intonation contours in mother's speech to newborns. *Developmental Psychology*, 20:104-113.

FERNALD, A; TAESCHNER, T; DUNN, J; PAPOUSEK, M; DE DOYSSON-BARIES, B; FUKUI, I. (1989) A cross-linguistic study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16(3):477-502.

FOUTS, R. (1972) Use of guidance in teaching sign language to a chimpanzee (Pan troglodytes). *Journal of Comparative and Physiological Psychology*, 80:515-522.

FROMKIN, V; KRASHEN, S; CURTISS, S; RIGLER, D; RIGLER, M. (1974) The development of language in Genie: A case of language acquisition beyond the "critical period." *Brain and Language*, 1:81-107.

GIBSON, E; WEXLER, K. (1994) Triggers. *Linguistic Inquiry*, Volume 25.3: 407-454.

GOLDBERG, D. (1989) Genetic Algorithms in Search, Optimization, and Machine Learning. Addison- Wesley Publishing Company: Reading, MA.

GOLDSTEIN, E. (1989) Perceptual development. *Sensation and Perception*, 3rd ed. Belmont: Wadsworth Publishing Company 327-353.

GOPNIK, J; CRAGO, M. (1991) Familial aggregation of a developmental language disorder. *Cognition*, 39:1-50.

GOSCH, A; STADING, G.; PANKAU, R. (1994) Linguistic abilities in children with Williams-Beuren Syndrom. *American Journal of Medical Genetics*, 52:291-296.

HOFFMAN, B. (1996) The formal properties of synchronous CCGs. Proceedings of the ESSLLI Formal Grammar Conference, Prague.

HURFORD, J. (1989) Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua*, Volume 77: 187-222.

JOHANSON, D and EDGAR, B. (1996) *From Lucy to Language*. Nevraumont Publishing Co: New York, NY.

KIRBY, S. and HURFORD, J. (1997) Learning, Culture, and Evolution in the Origin of Linguistic Constraints. In Fourth European Conference on Artificial Life, eds. P. Husbands and I. Harvey. MIT Press, Cambridge, MA: 493-502.

KIRBY, S. (1998) Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. *Approaches to the evolution of language: The emergence of phonology and syntax*. Eds C. Knight, M. Studdert-Kennedy, and J. Hurford. Cambridge University Press.

KIRBY, S. (1999a) Learning, bottlenecks, and the evolution of recursive syntax. *Linguistic evolution through language acquisition: formal and computational models*. ed E.J Briscoe. Cambridge University Press.

KIRBY, S. (1999b) Syntax out of learning: the cultural evolution of structured communication in a population of induction algorithms. *Advances in Artificial Life*. eds. D. Floreano, J.-D. Nicoud, and F. Mondada. Lecture Notes in Computer Science 1674. Springer.

KOZA, J. (1992) *Genetic Programming: On the programming of computers by means of natural selection*. Cambridge, MA: MIT Press.

KOZA, J. (2000) *CS 426 Course Reader: Genetic Algorithms and Genetic Programming*. Stanford Bookstore Custom Publishing Department: Stanford, CA.

LENNEBERG, E. (1967) *Biological foundations of language*. New York: Wiley.

LEWIN, R. (1993). *Human Evolution*. 3rd ed. Blackwell Scientific Publications, Inc.

LIEBERMAN, P. (1975) *On the Origins of Language*. MacMillan Publishing Co: New York.

LIEBERMAN, P. (1992) On the evolution of human language. *The Evolution of Human Languages*. eds J. Hawkins, and M. Gell-Mann. Proceedings in the Santa Fe Institute Studies in the Sciences of Complexity, Proceedings Volume X. Addison-Wesley Publishing Co.: 21- 48.

LIEVEN, E; PINE, J; BALDWIN, G. (1997) Lexically-based learning and early grammatical development. *Journal of Child Language*, Vol 24. Cambridge University Press: 187-219.

MACLENNON, B. and BURGHARDT, G. (1995) Synthetic etiology and the evolution of cooperative communication. *Adaptive Behavior*.

MARCUS, G. (1993) *Negative evidence in language acquisition*. *Cognition*, 46:53-58.

MORGAN, J and TRAVIS, L. (1989) Limits on negative information in language input. *Journal of Child Language*, 16(3):531-52.

NINIO, A. (1999) Model learning in syntactic development: Intransitive verbs. *International Journal of Bilingualism*, Volume 3.

OLIPHANT, M. (1996) The development of Saussurean communication. *BioSystems*, Volume 37:31- 38.

OLIPHANT, M; BATALI, J. (1997) Learning and the emergence of coordinated communication. [unpublished MS?]

PINKER, S. and BLOOM, P. (1990) *Natural language and natural selection*. *Behavioral and Brain Sciences*, Volume 13: 707-784.

PINKER, S. (1994) *The Language Instinct*. William Morrow and Company.

PINKER, S. (1995) Why the child holds the baby rabbits: A case study in language acquisition. *Language: An invitation to Cognitive Science*, 2nd ed, vol. 1. L. Gleitman and M Liberman, eds. Cambridge, MA: MIT Press. 107-133.

RUHLEN, M. (1994) *The Origin of Language*. John Wiley & Sons, Inc: New York.

SAFFRAN, J; ASLIN, D; NEWPORT, E. (1997). Statistical learning by 8 month old infants. *Science*, 274:1926-28.

SAVAGE-RUMBAUGH, E; RUMBAUGH, D; SMITH, S; LAWSON, J. (1980) Reference: The linguistic essential. *Science*, 210:922-925.

SAVAGE-RUMBAUGH, S. (1987) A new look at ape language: comprehension of vocal speech and syntax. *Nebraska Symposium on Motivation*, 35:201-255.

SENGHAS, A; COPPOLA, M; NEWPORT, E; SUPALLA, T. (1997) Argument structure in Nicaraguan sign language: The emergence of grammatical devices. *Proceedings of the Boston University Conference on Language Development*, 21. Boston, MA: Cascadilla Press.

SHIPLEY, E. and KUHN I. (1983) A constraint on comparisons: equally detailed alternatives. *Journal of Experimental Child Psychology*, 35:195-222.

STROMSWOLD, K. (1995) The cognitive and neural bases of language acquisition. *The cognitive neurosciences*, ed M. Gazzaniga. Cambridge, MA: MIT Press.

TALLAL, P; ROSS, R; CURTISS, S. (1989) Familiar aggregation in Specific Language Impairment. *Journal of Speech and Hearing Disorders*, 54:167-171.

TOMASELLO, M. (1992). *First words: A case study of early grammatical development*. Cambridge: Cambridge University Press.

TOMASELLO, M. (1995) Language is Not an Instinct. *Cognitive Development*, 10:131-156.

WERKER, J. and TEES, R. (1984) Cross-language speech perception -- evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7:49-63.

WERNER, G; DYER, M. (1992) Evolution of communication in artificial organisms. *Artificial Life II*. eds C. Langton, C. Taylor, J. Farmer, and S. Rasmussen. Addison-Wesley Publishing Company: Redwood City, CA. 659-687.

WRIGHT, B.A., LOMBARDINO, L.J., KING, W.M., PURANIK, C.S., LEONARD, C.M., and MERZENICH, M.M. (1997). Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature*, 387, 176-178.

[Return to Contents of this issue](#)



© [Copyright Journal of Artificial Societies and Social Simulation, \[2002\]](#)