SIMULATED EVOLUTION OF COMMUNICATION: THE EMERGENCE OF MEANING

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF LINGUISTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

Amy Perfors
July 2000

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Master of Arts in Linguistics.

David Beaver
(Principal Adviser)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Master of Arts in Linguistics.

John Koza

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Master of Arts in Linguistics.

Approved for the University Committee on Graduate Studies:

Acknowledgements

For the second time in two years I find myself sitting down to write acknowledgements for a thesis, and for the second time in two years I am overwhelmed when I consider all the people who deserve to be mentioned on this page. My primary thanks, above all else, must go to my advisor, David Beaver. He took the early bloom of the idea I came to him with in September, translated it to incorporate many of his own ideas, and worked with me ceaselessly to cause it to blossom into the completed project reported on here. He has been unfailing in his encouragement when I was completely disillusioned about finding yet *another* bug or convinced that I would *never* get interesting results. Over the months his thoughtful analysis of the issues and probing questions have given me important intellectual guidance, and his continued support has enabled me to make it over some rough spots, both personally and professionally. This thesis would not exist without David Beaver.

Tremendous kudos also belong to the Secow (Simulated Evolution of Communication) Group that allowed me first to crash their meetings, and then to steer them in my own direction as the project took off. It was an incredible bonus for me to have a group of intelligent, insightful people to knock ideas around with and bounce insights off of. Without the assistance of the Secow members – Gary Gongwer, Richard Pocklingtion, Fritz Schlerith, Kathryn Campbell-Kibler, James Warren, Dominic Hughes, and Roger Levy – the thesis would not be in its current form.

Special thanks must also go to Professor John Koza, my second reader, who was willing to dive into a linguistics thesis on very short notice and an even smaller amount of preparation. He came through with excellent insight and strong suggestions, and his conversation and questions provided a guiding light for much of the analysis here.

Not to mention that his work on Genetic Programming is the underpinning upon which the entire methodology of this research is based!

On a more personal note, I would also like to thank my friends Sara Fisher and "Foxy" Roxy Barbulescu, for showing me that real friends stay that way no matter how far away they may be, and that I have at least two people I can count on no matter what. I'd also like to acknowledge my other friends, my rugby team, and my fellow Human Biology course assistants, for doing their best to make me have fun even when I seemed determined to spend my life in front of the computer. They made this year memorable.

Finally, I would like to thank my siblings: Julie, Tracy, David, and Steve, for their unwavering support during what has been, in many ways, a difficult year for me. Many times the only thing that got me through the rough parts was the knowledge that they were there. Thank you, all of you, and a special thank you to mom and dad for giving me the encouragement to reach for my dreams.

Contents

List of Tables

List of Figures

Chapter 1

The Issues

"Language is the light of the mind." - John Stuart Mill

1.1 Introduction

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. Is it our habit of walking on two legs rather than four, which frees our hands and enables us to spread far and wide? Is it our general intelligence, without which we would not be able to adapt and flourish in the wide range of environments on the earth? Or is it our language, which allows us to communicate with one another so subtly and efficiently?

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.

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) In other words, chimps like Washoe can be taught (not easily, but reliably) to have vocabularies of up to hundreds of words. However, 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.[1] 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. In 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 (Cheyney, 1990).

The precise *extent* to which our closest relatives may be said to possess or be able to learn language is a matter of some debate. It is fairly unimportant for my purposes; it is enough to agree that there *is* a significant difference between the language abilities of humans and those of any other animal. Of this there can be little doubt.

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? The two are, after all, both indisputably things that contribute to the gulf between humans and animals. Should they really be separated from each other?

This question can be rephrased as whether language ability is somehow "prewired" 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

[1]Claims about animals mastering some basic elements of syntax – e.g. the widely-circulated anectode of Washoe's phrase *water bird* to mean swan – are both difficult to substantiate and, even if true, of arguable importance. *Water bird*, for example, can easily reflect the juxtaposition of two words corresponding to things in the same place at the same time with no grammatical intent involved. All claims of syntactic knowledge are defeated by rigorous observation indicating that the primates involved do not use such language productively, communicatively, or spontaneously. (e.g. Savage-Rumbaugh, 1987)

the linguistic and psychological research and debate of the latter half of this century has been focused on analyzing the truth of such a claim. Naturally, the debate has 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) The truth undoubtedly lies somewhere in the nebulous overlapping region between these two extremes. The big question is, where?

The controversy over the innateness of language touches on another of the largely unexplored and most 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. This is only one of the issues confronting scientists seeking to describe the emergence of language. They face many of the same difficulties faced by any evolutionary biologist, tripled – how do you explain processes that happened thousands or millions of years ago?

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 mysterious and powerful characteristics of language. Additionally, syntactic ability is apparently unique to humans, as we have seen. 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?

In this work I explore an approach that primarily sheds light on the question of how language evolved, and secondarily applies to the issue of how innate it is. In order to properly ground the research in theory, I will devote the first chapter to exploring the theoretical background to date for both questions, highlighting the important issues as I go. In the second chapter, I will discuss recent efforts to create computational simulations addressing these questions. Such simulations are important since they provide a rigorous and clear method for evaluating the plausibility and effectiveness of the theories that currently exist. Finally, I will present a new approach using a computer simulation that rectifies some of the weaknesses and incorporates the strengths of previous work. In so doing, it illuminates the larger issues that have confronted linguists and philosophers since time immemorial.

1.2 The Innateness of Language

Possibly the most fundamental issue debated among linguists is the extent to which our ability to communicate is a result of a language-specific ability in our brains. This controversy is as old as modern lingusitics – both the controversy and the approach were started by Noam Chomsky in the middle of the twentieth century when he published his earth-shattering 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. In this section, I shall consider the strongest arguments for each position, beginning with the nativist viewpoint and concluding with the non-nativist approach.

1.2.1 The Nativist Viewpoint

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 ot the nativist view, however, is the idea that language ability is in some sense separable from the more general cognitive capacities of the human mind. In other words, we learn language as a result of having a specific biological adaptation to do so, and this adaptation is distinct from our other mental abilities such as mathematical skill.

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)

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.

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 critical 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.

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) To some extent it is difficult to draw too hasty of 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)

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 interestingly 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 wherever they hear the lack of it. 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)

The final bit of strong evidence that language is innate comes from genetics and biology. After all, if language were truly innate and separable from other cognitive skills one would expect to be able to find instances of having normal intelligence but extremely poor grammatical skills, and vice versa. And indeed, both of these cases exist. Specific Language Impairment (SLI) is an example of the former; individuals diagnosed with this 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, 1993) 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)

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.

1.2.2 The Non-Nativist View

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 (aside from computer simulations, which we will consider in Chapter Two) is therefore actually evidence *against* the nativist view.

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.

More strikingly, children are incredibly 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 combined facts together suggest that a non-nativist strategy may account for much of language acquisition just as well as theories that make far stronger claims. In other words, children who are incredibly 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". Thus, there is no need to

make that additional assumption.

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.

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.

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.

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 nativist perspective of language development because they do not show how language is separable from general intelligence.

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, 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 no 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.

The existence of genetic impairments like Specific Language Impairment seem to be incontrovertible proof that language ability must be innate, 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 in innate to the notion that syntax is. [2]

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?

[2]Finally, a word should be said about the process of creolization, which is often taken as evidence that children must have an innate component of language ability. I am unfamiliar with much of the literature on creoles and creolization, but there is definitely debate about many of the claims originally made by Bickerton in the early 1980s. It is not obvious that pidgins and creoles are as different from each other as originally portrayed, nor it is obvious that creoles from different parts of the world are linguistically similar (which they would have to be, if they were created out of innate and universal mechanisms). I regret that due to time constraints I could not delve more deeply into this topic, but it is enough for our purposes to be aware that it is still a debated issue.

1.3 The Evolution of Language

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 accordance with the growth of our general cognitive capacities, that would be evidence that it was *not* innate.

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 question of the "Evolution of Communication" and the "Evolution of Syntax," respectively.

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.

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 the nativist perspective, while others argue against it.

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 chapter with a commentary tying together research to date about both the innateness and the evolution of language, leading so suggestions of where to go from here. [3]

1.3.1 Bickerton: Incorporating the fossil record

The Theory

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 in the brain.

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." (1990) The question is, what are the properties of that representation system?

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, daschund,

[3] There are naturally far more issues in the evolution of language than can be easily classified into either of these two categories. For example, how does one explain the evolution of the human articulatory tract? Or the evolution of the ability to parse a speech stream, both phonetically and syntactically? These are important questions in their own right; however, for this work they are less important *for themselves* than for what evidence they might contribute to our understanding of the two basic stages of language evolution, and so will be of less consideration.

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.

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

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 of our PRS.

At this point one is inclined to object that humans *are* perfectly able to represent even things that are not necessary 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.

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. [4] This creates a paradox: such a vocal tract is evolutionarily disadvantageous unless it is 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.

The other reason for believing that full language did not exist until relatively recently is that there is little evidence in the fossil record for the sorts of behavior facilitated by full language until 100,000 to 40,000 years ago (the beginning of the Upper Paleolithic). (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

[4]Gestural language may have evolved to fill this gap, but if it did, the mechanism and motivation for the transformation into spoken language is entirely unexplained. The question of gestural language as a possible intermediary is fascinating, but only indirectly to the issues discussed here. Nevertheless I regret that space constraints make me unable to consider it in more depth.

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.

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 of 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.

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 more gradual development of syntax. Additionally, Bickerton has drawn strong parallels between protolanguage and pidgens or the language of 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.

Discussion

The initial part of the 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.[5]

However, there is one glaring drawback to Bickerton's theory. The problem with an explanation relying on a sudden genetic mutation 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.

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 recaptulates 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.

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.

[5]Many argue that although language skill would have both facilitated these behaviors and correlates nicely with an increase in cranial capacity, this is only circumstantial evidence (at best) for language use. At worst, it is merely a Just-So story. This objection is valid. Nevertheless a theory (like Bickerton's) that attempts to take the fossil record into account is stronger than one that ignores or contradicts it (as some do).

1.3.2 Pinker and Bloom: On Natural Selection

The Theory

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.

The first criteria, 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! More importantly, and 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 potentialy 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.

The second criteria, 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.

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.

Exaptation is a bit more difficult to explain away. It refers to the process of coopting parts (usually termed spandrels) 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. (Deacon, 1992) 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. (Deacon, 1992; Lieberman, 1975)

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 an 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 brain 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.

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.

If the mind is indeed a multipurposive learning device then Pinker and Bloom suggest that it was certainly *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.

Additionally, scientists have thus far been unable to develop any psychologically realistic computational inference mechanism that is general purpose 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.

Discussion

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 in 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)

This is unlikely for a few reasons. First of all, no neural substrates corresponding to grammatical rules *themselves* 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?

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.

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 as a largely unmodified spandrel to serve the purpose of language.

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.

1.3.3 Deacon: The Natural Selection of Language Itself

One of the most plausible arguments to the viewpoint that language is the product of biologically-based natural selection is the viewpoint that rather than the *brain* adapting over time, language *itself* adapted. (Deacon 1992, 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 have 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)

Discussion

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.

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.

Nevertheless, there are drawbacks to Deacon's idea, most obviously with the learning procedure he suggests. Though he gives reasons for believing that such a procedure would produce at least some of the features of natural language, he demonstrates no convincing evidence for accepting that such a learning procedure actually exists. For instance, a mildly context-sensitive language with cross-serial grammatical dependencies like those found in German can be parsed by his learning procedure. (Briscoe

1998) However, a construction in which *any* ordering of *a*s, *b*s, and *c*s is grammatical cannot be parsed. In general, no formal or computational demonstration of learnability of mildly context-sensitive languages – which are possibly most like Natural Languages – has been found. (Joshi et al., 1991) As before, the fact that one hasn't been found doesn't necessarily mean one *won't* be found. However, it is definitely problematic that the very learning procedure Deacon proposed has been unable to parse formal systems with properties like those found in Natural Language.

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.

1.4 Comments and Discussion

The issues in both language acquisition and language evolution overlap 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 some combination of the two extremes under debate.

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 may 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.

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 is 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 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?

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.

Chapter 2

Computational Simulations

2.1 Introduction

As we have seen, there is a great deal of interest in many of the most intractable issues confronted by modern linguistics. Questions about the evolutionary history of language are by nature difficult to answer; not only is human language a unique phenomenon, but it emerged long ago. Nevertheless, these questions are strikingly important. What evolutionary forces propelled language evolution? By what mechanism did language eventually develop syntactic characteristics such as compositionality and recursion? What evolutionary adaptation did the acquisition of communication serve?

There are equally difficult but related questions in the field of language acquisition. Are the language abilities of human beings due to an innate, biologically-based (and therefore biologically evolved) language acquisition device? Or are they reflective of either more general cognitive abilities of the human brain, language evolution, or both? The answers to these queries will naturally have some bearing on the answers to our first set of questions, and vice-versa.

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. 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.

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.

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 rest of this chapter, 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.

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. Special attention will be paid to genetic programming, since it is the basis of the work upon which this thesis is based.

2.2 GA, GP, and A-Life

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.

This is most clear with an (albeit 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 equally effective, 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.

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.

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.

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.

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

2.3 The Nativist vs. Non-Nativist Dilemma

As we saw in the last chapter, 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 a some of the research involving computational simulations that sheds some light on this issue, as well as discuss the implications of simulations we have already reviewed. The sections following will concentrate on the specific issues, raised in the last chapter, of the Evolution of Communication and the Evolution of Syntax. What can computational simulations do to shed light on these unanswered issues?

2.3.1 The Role of Learning

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. In other words, 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

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.

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.

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 uninformative 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 analysed 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 analysed 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.

Basically, then, the learning algorithm only takes up a new parameter if the input cannot be analysed 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.

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. Before that point, 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 analysable, and on how many utterances that were heard were successfully analysed.

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 is 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

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 being 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. In order to test the non-nativist theory, the percentage of *triggers* whose parsability is scored by the learning algorithm is zero but the percentage of utterances is 10. In this way, there is no way that language itself can evolve.

Results indicated that evolution completely fails to respond to the functional pressure on a purely cultural-evolutionary 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...].

Next, linguistic selection was enabled by scoring the parsability of triggers 10 percent of the time, and parsability of utterances 10 percent as well. After only 200 generations, two optimally parsable languages predominate: this is Deacon's cultural adaptation of languages in action. Interestingly, natural selection seems to follow the linguistic selection here. After 57 generations – at which point there are already only a few highly parsable languages in existence – this has been formalized as only one parameter in the LAD. Yet as time goes on and linguistic selection has already occurred, LADs eventually seem to evolve to partially constrain learners to learn functional languages. The interesting thing about this is that "biological" emerges only *after* substantial cultural evolution, rather than vice-versa as predicted by most nativists.

Comments and Discussion

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.

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.

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.

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.

2.3.2 Learning of Context-Free Languages

John Batali (1994) did a similar simulation as Kirby and Batali, 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

Batali used a combination of genetic algorithms and neural networks in a structure very similar to the work discussed above (CITE here). 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 3 input units, 10 hidden units, and 7 recurrent connections from hidden units to the input layer.

In each generation, networks were trained on a set of inputs. After this, fitness 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.

Training occurred with the context-free language $a^n b^n$. 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 negative evidence in child acquisition was simulated here as well.

Results

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* switched 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 is 0.244.

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 are the ones that gradually develop an innate bias toward learning the languages in that class.

Comments and Discussion

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.

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.

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 just one. If it did that, we would still not know whether to ascribe this to general-purpose learning or language-specific inference, however.

2.3.3 Discussion of Language Innateness

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 chapter) 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.

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 Chapter One. 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.

2.4 Evolution of Syntax

2.4.1 Parameter Setting Models

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.

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

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. While this 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 1995, 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 a growth in complexity of the LAD.

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).

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 inheritence 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

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)

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.

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 Chapter One.

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.

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).

In cases with much more rapid linguistic variation (modelled by 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.

When the LAD incorporated a Bayesian learning mechanism (Briscoe 1999a, 1999b), the general trends were similar, with an large proportion of default paramaters 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.

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.

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

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.

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 (like ourselves) who *are* interested in models of actual human language evolution.

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). (Ruhlen, 1994)

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.

2.4.2 Models of the Induction of Syntax

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.

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

Method

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)))*).

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)

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

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

1. $V/\text{eats} \rightarrow \text{eats}$
2. $N/\text{tiger} \rightarrow \text{tiger}$
3. $N/\text{john} \rightarrow \text{john}$

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 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.

The job of the listener is to "hear" the string generated by speakers and attempt to match that to a meaning. Fitness is based on to what extent the meaning/string mapping is the same between speaker and listener. Naturally, at the beginning of evolution, speaker and listener will not coincide between mappings except by chance. However, as evolution progresses, those individuals who share the same string/meaning mapping will be preferentially selected for while those who do not will tend to die out.

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/sees(john,tiger)* $\rightarrow$ *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.

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. (Kirby, A) 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

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.

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 section category. (Kirby, 1999a, 1999b) This may indicate that agents are capable of creating categories syntactically in their E-Language that correspond in some sense to the meaning-structure of their I-Language.

Discussion and Comments

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.

This is an intriguing analysis, since it paints a picture of the *language* actually 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.

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.

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.

Kirby uses his results to suggests 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.

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.

2.4.3 A Neural Network Model

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

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).

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.

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.

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 contains syntactic 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.

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

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.

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.

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 to 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

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.

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.

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.

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?

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? 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.

2.4.4 General Discussion of Evolution of Syntax

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.

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.

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* correspondance

between speakers' and listeners' internal meaning representations.

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 correspondance between two individual's internal meanings; it is based on how that correspondance 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?

2.5 Evolution of Communication

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.

2.5.1 Evolution of a Learning Procedure

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

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.

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, chuttering, 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, chutter=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.

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 that any given meaning and will interpret each signal in the way most of the population does.

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

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.

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 only a finite set of observations of communicative episodes – achieve excellent accuracy (98% after 1200 rounds for the Obs-2 (the one based on 25 observations)). As the number of observations declines, accuracy naturally goes down. However, even with Obs-10, learning occurs; accuracy for that procedure eventually asymptotes at approximately 80%.

Discussion and Comments

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; Lievel 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.

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.

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 abjection 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.

There are certain other fundamental assumptions about the individual's cognitive capacities that we should be aware of. Implicit in the model is the thought that agents must be capable of classifying different situations, voluntarily producing several possible calls in these situations, and recognizing that other animals are performing them. There is some doubt that even these simple skills can exist without language. For instance, few if any non-human primates have voluntary control over their vocal apparatus. (Lieberman, 1992) Therefore, it is unlikely that our common ancestor could voluntarily produce several possible calls in varying situations. This is especially true when one considers that there should be no reason to develop the ability to voluntarily control the voice, *other* than to use language. But that would imply that this ability came about *after* language, a suggestion that directly contradicts the assumptions made here.

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.

2.5.2 Evolution of Coordination

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?

These questions are definitely reminiscent of work on the so-called Prisoner's Dillema 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)

Methods

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.

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?

Michael Oliphant (1996) asks the exact same question, but his agents are genetic algorithms 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

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.

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.

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 analagous to strategies successful in promoting altruistic behavior in the typical Prisoner's Dilemma. In one such varient, 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.

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.

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, it's 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.

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

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.

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 analagous 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.

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.

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.

2.5.3 Evolution of Communication Among Artificial Life

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

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.

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).

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

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.

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.

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.

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 about 80% or less of the individuals on either side, 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

In one respect (as a realistic and applicable model of human language evolution), this study 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.

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 my females. Indeed, in the beginning of the simulation they do not act on them. Additionally, unlike all other research reviewed here, agents do not coem 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.

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.

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.

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.

2.5.4 A Synthetic Etiology Approach

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

MacLennan and Burghardt began with moderately sized populations (around 100 organisms) of finite-state machines implemented as genetic algorithms. Each finite state machine is determined by a number of condition/effect rules of the form (Σ,γ,λ) $\rightarrow$ (Σ', R). Σ is an 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).

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,...G*), local environment states by (*1,2,3...L*), and internal states by (*1,2,3,...I*). A transition table will therefore have *IGL* entries in a fully-defined organism.

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.

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.

The difference between cooperation and cooperation plus learning is also explored here. When learning is enabled, organisms that "make a mistake" by acting non-cooperatively 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.

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.

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

Not surprisingly, the condition in which communication is supressed 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.

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 matrixes 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.

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.

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

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.

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?

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.

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 the 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.

2.5.5 Comments on Evolution of Communication

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 initiall

non-linguistic state?

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.

2.6 General Discussion

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. 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.

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. Thus, while these simulations can 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.

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 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. The approach outlined in the next chapter attempts to do just this.

Chapter 3

Method

3.1 Introduction

As we have seen, the work done so far investigating the evolution of language using a computational approach is evocative but still incomplete. While there has been significant work on how syntax might evolve, it all incorporates strong assumptions about the learning algorithms available to agents as well as the underlying representation of meanings and/or syntactic structures. Research into evolutionarily "earlier" states of language evolution – such as the development of communication itself – attempts to bridge this gap but still falls short. It is often difficult to generalize to claims about cognitive mechanisms found in humans (e.g. Werner & Dyer 1992; MacLennan & Burghardt 1995), and often unrealistically rewards communicative behavior *itself* (e.g. Oliphant 1996; Batali 1995). Additionally – and most problematically – this work makes the same mistake that plagues research on the Evolution of Syntax: that is, it incorporates specific assumptions about the nature and structure of meaning representation in the brain.

Why is this a problem? After all, meanings must be represented *somehow*. Why not make plausible assumptions and study what results follow from those assumptions?

There are two main difficulties with this "let's see what happens" type of approach. Although it is perfectly valid and even necessary when one is determining the space

of parameters involved in a problem or surveying the landscape of possibilities, it is inadequate if one wants to make evolutionary claims. If the latter is our goal – which it is – then we must ground our assumptions in plausible evolutionary scenarios. Otherwise, we just "push back" what we don't know another step, without truly solving anything. Most of the theories reviewed here fail to provide the necessary grounding (e.g. Pinker & Bloom, 1990; Deacon, 1992, 1997), making it particularly essential to develop an account of early meaning representation and how it arose.

The second main difficulty with starting with assumptions about meaning/syntactic representation is that it is one domain in which there are no intuitively "correct" answers to use as defaults. There are a wide number of competing theories about meaning representation, from images (e.g. Cooper & Shepard, 1973; Finke et. al. 1992), to heirarchies (e.g. Bickerton, 1990; Collins & Quillian, 1969) to connectionist models (e.g. McClelland & Rumelhart, 1986). The differences between these theories of representation may result in significant differences in results, since outcomes are often crucially dependent on the underlying representation. (see discussion, Chapter 2; Briscoe, 1998; Kirby, 1998, 1999a, 1999b; Bickerton, 1990)

The work covered here seeks to shed light on the evolution of meaning – and, hence, the first stages of communication – by suggesting that meaning itself naturally emerges in situations where fitness would be increased through successful communication. In other words, meaning and communication are emergent properties of optimal performance on certain types of joint activities. (Clark, 1996) The structure of meaning representation, then, is determined by properties of the world as well as the dynamics of the joint activity in question.

The method discussed here has three properties that are integral to its success. First of all, it incorporates few if any assumptions about the type (or even existence) of either meaning representation or dialogue/communication structure.[1] This ensures that any communication that *does* emerge was not a result of those assumptions. Additionally, by examining the nature of any emergent communication, we might be able to better understand which assumptions are indeed most warranted.

[1]While no assumptions are explicit in this paradigm, it might be argued that the program trees themselves might be interpretable as the meaning representation of the agent.

The second property lies in the fitness function: the fitness of agents in this scenario is not based directly on their ability to communicate with each other. Rather, fitness is based on success in completing a joint activity that can be performed adequately without communication, but optimally only with it. This is advantageous in two ways: it is by far the most evolutionarily plausible approach, and it is consistent with theories of meaning (e.g. Clark, 1996; Wittgenstein, discussed in Kripke, 1982) that suggest that meaning is to be found only in the context of the activity (or "game") in which it occurs.

Finally, the third property lies in the generalizability of the approach covered here. It is designed to be applicable (with few modifications) to the richer questions involved in the evolution of language, such as the emergence of syntax. This is a desired result not only because it contributes to the ease and flexibility of *practical* use, but because it is explanatorily most powerful. If a theory postulates qualitatively similar mechanisms that can be used to explain seemingly distinct areas of language evolution (as is the hope here), it is a stronger theory than one requiring multiple distinct mechanisms.

In the following sections I will describe the setup of the computational approach used here, beginning with a general illustration and moving quickly to specific details.

3.1.1 Setup

This research utilizes a genetic programming approach involving populations of agents that have the opportunity to interact with one another. The original form of this database coordination task was outlined in Beaver (1999), which served as the starting-point for the work done here. Implementation incorporated the use of the genetic programming software lil-gp (v. 1.1), written by Douglas Zonker and Dr. Bill Punch of Michigan State University. Both lil-gp and the original task outlined by Beaver have been altered significantly during the course of this research.

In this setup, each agent in a population is equipped with its own "database" – a knowledge structure containing information (represented by the digits 0 to 8) and unknowns (represented by the digit 9). Agents are also equipped with their own

"blackboards" – essentially, structures to which they can write digits. While agents do not have the ability to read or write to another agent's database, they can read from another agent's blackboard. It is therefore possible for communication to occur - one agent can write a sequence of digits to the blackboard and another agent can read them and act appropriately.

What constitutes appropriate action? The key element of this scenario is that the databases given to each agent are incomplete – that is, many of the digits are unknown (9s). Yet together, the agents have knowledge of the full database. Fitness of an agent is determined by to what extent it has replaced the unknowns in its blackboard with correct information. The databases are constructed in such a way that the probability of evolving a "blind" strategy that consistently fills them in correctly, without communication, is very low. Thus, for an agent to perform optimally, true "communication" must occur. In this way, communication emerges naturally out of the joint activity of attempting to match individual

Agents themselves consist of program trees constructed from a basic function set (see Table 3.1). Included in this function set are functions that allow agents to read and write to their own database as well as functions that let agents write to their own blackboard or read from the other agent's blackboard. Write-Bit-To-DB-At-Location takes three arguments, the bit to be written and the location (row and column) in the database to which to write it. Upon writing the digit – which it only does if there is currently an 'unknown' in the database at that location – it advances to the next location in the database and returns a 9. Write-Bit-To-BB-At-Location operates in much the same way except it takes only two arguments (since the blackboard is one-dimensional) and will write at that location regardless of whether or not there is an 'unknown' currently there. The other two "write" functions are essentially identical, except they take only one argument (the bit to be written) and automatically occur at the current location of the database or blackboard. As before, symbols that are *not* unknown will not be written over in the database but will be written over in the blackboard, and both functions will return a unknown (9).

Read-Bit-From-DB-At-Location and Read-Bit-From-BB-At-Location take two and one arguments, respectively, referring to the locations at which bits are read from.

They return the bit on the specified location and automatically advance the location one step as they do so. All "read" functions do *not* occur on the agent's own blackboard, since agents may only read *other* agents' blackboards. Read-Bit-From-DB and Read-Bit-From-BB are identical to the other read functions except that they each automatically read from the current location, and therefore have zero arguments.

The other functions in the function set correspond to the operators in first-order predicate logic. The function Or, for instance, evaluates two subtrees, returning 1 if at least one of them is true and 0 if not. The function Not evaluates one subtree, and returns its opposite value (0 if it was 1, 1 if it was 0, and 9 otherwise). The function And evaluates two subtrees, returning 1 only if both subtrees evaluate as 1. The function Equal evaluates two subtrees, returning 1 if they return the same value and 0 otherwise. It may be used in the function If (either by itself or in conjunction with other logical operators), which evaluates one subtree, taking one action if it returns 1 and another action if it returns 0. And finally, there are three connector functions (Connector2, Connector3, Connector4) which consecutively evaluate 2, 3, and 4 subtrees respectively and automatically return a 9 (unknown) at the end.

Databases are two-dimensional matrices containing the integers from 0 to 9. Integers from 0 to 8 are considered "information", while 9s represent "unknown" material. At the beginning of each generation, a master database for the population is created – this database contains no unknowns and therefore represents "perfect knowledge." Individual agents are given databases derived from this master – the information contained in individual databases is correct, but much of it has been replaced with 9s. All individuals in these simulations begin with 50% of their databases converted to unknowns and 50% untouched, although *which* 50% differs from individual to individual. Agents may both write and read to their own database, although write capabilities are limited in that they cannot write over "known" information (i.e. numbers 0 to 8).

Blackboards are one-dimensional matrixes of specified length that – unlike databases – are visible to other individuals in a population. All blackboards are initially "erased" by putting 9s in each location. Agents can automatically write over any information (whether known or unknown) contained on their own blackboard, but do not have

Table 3.1: Function Set of Agents

Function	Args	Value returned
Get-Integer	0	Between 0 and max length of db/bb
Write-Bit-To-DB-At-Location	3	9
Write-Bit-To-DB	1	9
Write-Bit-To-BB-At-Location	2	9
Write-Bit-To-BB	1	9
Read-Bit-From-DB-At-Location	2	Bit on DB at specified location
Read-Bit-From-DB	0	Bit on DB at specified location
Read-Bit-On-BB-At-Location	1	Bit on BB at specified location
Read-Bit-On-BB	0	Bit on BB at specified location
Equal	2	1 if equal; 0 if not
If	3	1 if comparison true; 0 if not
And	2	1 if both subtrees true; 0 if not
Not	1	Opposite value of subtree
Or	2	1 if at least one subtree true; 0 if not
Connector2	2	9
Connector3	3	9
Connector4	4	9

similar write capabilities for any other agent's blackboard.

Since this research is exploratory, I sought to simplify the situation as much as possible for the initial exploration. Therefore, agents do not have the ability to "talk" to all other members of the population at once. Instead, agents are paired at the beginning of each generation with another agent. The two agents (call them Bert and Ernie) interact only with each other over the course of the generation, and they earn the same fitness score. In order to speed up evolution, Bert and Ernie are made up of the same program tree: in other words, they are identical except for the database that each contains. By analogy, this is the same as saying that Bert and Ernie are identical twins with different knowledge of the world.

Diagram 3.1 illustrates the make-up of a typical pairing of agents. Each "pair" makes up one individual in the population; thus a population of size 500 would contain 500 *pairs*, each consisting of one Bert and one Ernie. This "identical twin" setup was designed in order to speed up evolution and simplify fitness assignment only; it was not motivated by theory or designed to be particularly evolutionarily realistic. Thus,

one direction of further research lies in exploring what effect altering this structure has on the final results.

diagram 3.1 goes here

Each generation, as discussed earlier, begins with the creation of a master database and the derivation of each individual's database from that. Bert and Ernie each have 50% of their databases filled with 9s, in a fashion perfectly complementary to each other. Thus, in theory, they contain the same information *together* as the master database does. After all individuals in a population are created, each pair interacts among themselves.

This interaction is characterized as follows: first Bert's program tree is run, and then Ernie's program tree is run. This is repeated either until both agents have completely replaced the unknowns in their databases with integer values, or until some specified maximum number of conversation steps has been reached (default is 100). It is important to note that although this setup runs Bert and Ernie in a specified manner, it is *not* making the implicit assumption that dialogue structure requires one agent to "speak", followed by another. This is because each agent has almost unlimited freedom for what to do while being run – in other words, Bert could

easily not write *anything* to his blackboard (i.e. he could remain silent) as fill and refill his blackboard (i.e. essentially give a monologue) or do anything in between. Thus, aside from giving each agent at some point the *opportunity* to speak or listen as they choose, this structure makes no implicit assumptions about how dialogues should be set up.

Once the interaction is terminated, fitness for each individual is measured. It is made up of three equivalent subcomponents: the percentage of Bert's database that matches the master database, the percentage of Ernie's database that matches the master, and the percentage of Bert's database and Ernie's database that matches each other. Thus, optimal fitness for any interaction is 1 + 1 + 1 = 3. Fitness for agents that do not change their databases in any way, by contrast, is 0.5 + 0.5 + 0 = 1; 0.5 of each agent's database matches the master, and there is 0 overlap between the two. Fitness for an optimal non-communicative strategy (which would work by overwriting unknowns with random integers from 0 to 8) is approximately 0.55 + 0.55 + 0.11 = 1.21; each digit has approximately a 1 in 9 chance of being correct, resulting in 61% matches with the master on average, and 11% matches between each other.

Once fitness is calculated, the master database for the population is changed, as are the databases for each individual in the population. Then each Bert and Ernie pair interacts once more, as detailed above. Fitness is again calculated and added to the fitness values from prior interactions. This repetition occurs multiple times (default is set at eight, resulting in an optimal fitness value of 24). The reason for this repetition is to make it ever-more unlikely for an agent to achieve perfect fitness merely by "lucking out" with a random strategy, and hence without using communication. In this way, fitness is very dependent on communicative ability while only *directly* measuring performance on the joint activity performed by Bert and Ernie.

After all interactions have been repeated the correct number of times, reproduction of individuals into the next generation occurs. 90% of the population is selected for crossover with one another, using tournament selection of size 5. (see Koza, 1992 for details) In this way, the best individuals in a population are likely to reproduce more copies into the next generation than are the worst individuals, in a way that parallels competition in nature among individuals for the right to mate. The remaining 10%

of the population is reproduced straightforwardly into the next generation, also using the same type of tournament selection. Program trees are restricted to depths of less than 17, but otherwise there are no limits on which trees may be crossed over with one another.

3.2 Parameters of Variation

The simulation described here is simple in the sense that it makes few basic assumptions, but complicated in the sense that there are multiple parameters whose variation might have a significant effect on the final outcome. In this section, we will consider some of the basic ones and outline why they are (or may be) important. Pilot studies occurring before this research were used to give initial insight into which ones might be of most importance, as well as in establishing the details of the setup described earlier.

The most important finding of these pilot studies showed that using binary databases with very few repetitions of conversation has the result of dramatically increasing the probability of "random" strategies achieving optimal or near-optimal fitness without actually using *communication*. In other words, when the master database was "recycled" very few times per generation, the fitness landscape has such a large maximum for non-communicative strategies that no communicative strategies evolved. As a result, the defaults for the experiments reported here use decimal-numbered databases as well as multiple database exchanges per generation. Instances where the database was binary as well as where there were fewer databases per generations will be analyzed and reported on in the next chapter as well.

Much of the experimentation reported on in the next chapter consists of varying the following parameters in a principled way in an effort to determine the effect of each. All of the parameters suggested below have been manipulated in some ways, although not exhaustively; the intent of this research, after all, is not to exhaustively determine the extent of the importance of all variables. It is to determine whether – and in what circumstances – effective (and/or optimal) communication could evolve, given a fitness metric that does not directly favor it. Thus, the parameters have been

manipulated in principled and motivated ways, but all permutations and combinations of parameters (which would be an extraordinarily large number) have not been considered.

Population Size and Generation Size. These are definitely parameters that make a difference for the success of a simulation, since greater population and generation sizes increase the probability of creating perfectly fit individuals. Default for this is a population size of 500 and run lengths of 50 generations (again following standards set by Koza 1996) but variations for more complicated runs occur.

Information in the Database. The information in the database has so far been set to be made up of a random assortment of the integers 0 to 8, with 9s standing for "unknown." It is possible to alter this in order to create more information-rich or information-poor databases – for example, making them binary would result in databases with less entropy (i.e. more order).[2] Since this has an effect on how effective non-communicative random strategies are – and thus changes the fitness landscape – it might be instructive to alter this systematically.

Another possibility regarding modifying the information in the database is to make it more systematic or structured. In the current implementation, the integers 0 to 8 are assigned randomly, so each integer has an equal probability of occuring in each location. However, if there was structure in the assignment of database information – say, 4s always followed 2s, or certain sequences of integers only occurred in certain places – then strategies taking advantage of these sequences might occur. Most interestingly, *communicative* strategies referencing these underlying regularities might evolve, creating a rudimentary "syntax" of sorts.

Database Richness. Defaults of the amount of a database each agent has access to are set at 50% – half of each agent's database consists of unknowns. This is easily modifiable, however. Does decreasing the number of unknowns lessen the probability of communication evolving? If so, is there a threshold amount? How about situations

[2]Order is a measure of how much information is contained in the database. The more order there is, the shorter the description of the information needs to be in order to get it all – a perfectly random database has the least order, since the shortest description of that information would just *be* to print out that information. A binary database of length four has more order than a decimal database of equivalent length, since the shortest description of the binary database could fit into two base-ten digits while the shortest description of the decimal database would require four.

when agents know less than half of the database?

Database Write Capability. Agents currently do not have the ability to replace "known" information in their databases – in other words, they can only write over 9s. This is motivated by analogy: in the real world, individuals do not "forget" known information or replace it with falsehoods as a matter of course. Nevertheless, we *do* occasionally do so – thus, it would be instructive to determine to what extent successful communication depends on leaving known information alone.

Database Size. It is also possible to alter the size of databases. At the very least, increased size of database might result in longer conversations and more dependency on formalisms and regularities in interaction between agents, since there would be more chance of becoming uncoordinated over the course of conversation.

Blackboard Size. Blackboards are the sole means of communication between agents, so size of blackboard creates a natural bottleneck to that communication. At its most extreme, blackboards of only one digit long might limit effective dialogues to short bursts and may dramatically curtail the ability of agents to coordinate. Very long blackboards, on the other hand, might provide agents with more flexibility and communicative tools, but also increase the chance of Bert and Ernie "getting confused" about what the other might be trying to say.

Change in Function Set. The functions chosen here were deliberately chosen to make as few assumptions about an agent's necessary capabilities as possible. Nevertheless, as activities grow more complicated (say, by increasing database size or richness), it might be valuable for agents to have more ability to manipulate the database and blackboard. For instance, additional functions might give them the ability to reposition themselves at the beginning of a database or blackboard; erase a blackboard; or look at specific locations on a database or blackboard. Such a change might either make it easier for communication to evolve, or else more difficult since coordination would be less assured.

Change in Conversation Length. Conversation length is currently set to a default of 100 permissible read and writes, but there is no principled reason for this. As databases get longer and more complicated, lengthier conversations might be necessary in order to achieve optimum performance.

Change in Repetitions of Interaction. Currently, pairs of agents change master databases eight times per generation. As with conversation length, however, this is an arbitrary value. Presumably more repetitions results in even lower probability of a random strategy taking effect, while fewer is faster but results in more likelihood of effective random strategies. Thus, principled manipulation of this variable might serve to shed light on how divergent random and communicative fitness needs to be in order to create agents that do communicate.

Change in Fitness Function. The current fitness function depends only on the degree of overlap between the databases of Bert, Ernie, and the master. Yet there are certainly theoretical reasons to believe that other factors might be relevant. For instance, one could build in a "bottleneck" on linguistic output by explicitly rewarding agents for shorter conversations or fewer database and blackboard writes. This is theoretically motivated in part by Kirby (1998), who suggests that linguistic bottlenecks were part of what provided the impetus for communication (and especially syntax) to emerge. In addition, we see this in the real world; messages which are shorter and easier to say tend to be heard – and be more effective – more often than those which are long or complicated.

In the next chapter, we will examine basic results using the default values of each of these parameters. In an effort to determine which parameters are most crucial to the development of effective communication, we will further consider what happens when some of the parameters are varied and analyze the "conversations" that result from each case. Finally, we will discuss implications of these results and consider future directions of research.

Chapter 4

Results

4.1 Overview

Since this work is primarily exploratory research into a new paradigm, most of the experiment involved varying many of the relevant parameters to determine which are most important in achieving the best performance. It is important to note that the intent of this work was *not* to make an exhaustive survey of the parameters of variation; rather, it was to aim for an interesting "existence proof" showing that effective, even optimal, communication can evolve in a generalizable task that does not explicitly favor communication in the fitness function itself. An exhaustive survey would not only be impractical given the number of parameters, but – more importantly – the dynamics of evolutionary algorithms in this paradigm are not what we are principally interested in. Rather, we are most interested in the linguistic implications of this research. Thus, the 41 experiments that were run involved principled variations in the parameters according to the implications the data gave as to what was significant. (The table at the end of this chapter contains the raw data for the runs; column headers in that table correspond to the italicized words in List 4.1 below). The parameters considered during these experiments were as follows:

Sections 2 and 3 will be devoted to a thorough exploration of the two major findings of this study. The first and most exciting finding is that optimal performance on a variant of the database task *did* indeed emerge. Agents developed "communication"

List 4.1: Parameters	
Num	Parameter
1	*Pop:* Population Size
2	*Gen:* Generation Size
3	*Info:* Information in the Database (binary vs. decimal numbers)
4	*Write:* Database Write Capability
5	*DB:* Database Size
6	*BB:* Blackboard Size
7	*Function:* Change in Function Set
8	*Conv:* Conversation Length
9	*Reps:* Repetitions of conversation
10	*ShortFit:* Fitness reward for shorter conversations

– and hence, arguably, a notion of *meaning* – as an emergent property of attempting to accomplish the joint activity of database matching. This performance was crucially and strongly dependent on the size of the databases in question. In other words, of all the parameters of variation, database size was found to be – by far – the most important factor in evolving effective communication. A great deal of the analysis in Section 3 will be devoted to exploring why database size is so important, and what may be done to improve performance as size increases.

Although database size is by far the most important factor influencing performance in the database communication task, multiple parameters have an effect on the outcome of a run. In Sections 4 through 8 we will examine some of the most interesting parameters in more detail, exploring precisely what effect they have on both performance and the nature of the "conversations" between the agents. Although not all parameters that *might* effect the database communication task have been considered, a great deal of them have, in accordance with the purpose of this exploratory study.

4.2 The Successful Emergence of Communication

The primary purpose of this work was to shed light on the evolution of meaning, based on the premise that meaning itself is an emergent property of optimal performance

in certain joint activities. In this case, that joint activity is the task of matching the individual agent's database to the 'master' database that neither agent has direct access to. Given this purpose, the fundamental question is: did such communication emerge?

The short answer is *yes, it did*. The long answer requires a more detailed explanation.

4.2.1 Basic Results

The nature of the fitness function is such that it is statistically so unlikely to get optimum fitness through chance[1] – i.e. without communication – that evidence of consistently optimum individuals over multiple generations in a population is strong indication that communication is occurring. By that definition, the goal of developing emergent communication has clearly been met. Multiple runs of the simulation produce a number of optimum individuals, even when some parameters are varied. Further evidence that the optimal fitness achieved indicates true communication is that, when runs are identical except that the blackboard is disabled so that agents cannot "write" to it, fitness drops to chance levels. Since the blackboard is useful *only* for communicative purposes, this drop is another strong indication that true communication has emerged.

In all experiments, each parameter has certain defaults; unless explicitly mentioned otherwise, these defaults are the standard. For easy reference, they are as follows:

There are three interesting measures produced by each run of the simulation; these

[1] We can estimate the statistical likelihood of the success of a strategy that randomly fills the 'unknowns' in a database with digits. In a database of length L and base B, with 50% of the digits filled with unknowns, there is a 1 in B probability for each digit that a random strategy will replace it correctly. Since each individual Bert/Ernie pair has L unknowns (0.5L each), the probability of a pair correctly filling in *every* digit is $1/B^L$. If the database is exchanged N times per generation, the probability of a pair correctly filling in all digits for all databases in one generation – and thus achieving optimal fitness – is $1/B^{LN}$. For our default values (B=10, L=4, N=8) this translates to a $2.9 * 10^{-31}$ probability of randomly achieving optimal fitness. In other words, at populations of 500 individuals per generation, $1.7 * 10^{33}$ generations would be necessary in order to randomly get optimally fit individuals.

List 4.2: Parameter Defaults	
Parameter	Default Value
Population Size	500
Generation Size	50
Information in the Database	Decimal numbers
Database Write Capability	cannot write over "known" material
Database Size	4 digits per entry, 1 entry
Database Structure	digits unpatterned, randomly assigned to database
Blackboard Size	8 digits
Function Set	see Table 3.1
Conversation Length	100 turns
Repetitions of conversation	8
Fitness	No reward for shorter conversations

measures are used to measure various aspects of the "success" of each run. Most importantly is the measure AVGHI, which is the average fitness of the highest-fitness individual for the final 10 generations of each run, expressed as a percentage of total fitness. An AVGHI of 100 means that each of the highest-fitness individuals of the last ten generations was perfectly optimal. Since genetic programming is always primarily concerned with the best-fitness individuals, AVGHI is thus a direct reflection of the success of the run. And since optimal fitness is almost certainly a reflection of successful communication, AVGHI is a measure of the degree of successful communication between agents of a given run. Average AVGHI for the "basic" runs considered here (the runs that do not stray significantly from the default paramaters) is 98.4, arguably indicating the existence of true communication. By contrast, AVGHI of a run that is identical in every way except that writing to the blackboard is disabled is 48.5 – not much above chance values of 40.33.[2] This also strongly implies that communication (via the blackboard) was essential to the success of the default runs.

A similar measure, called HIFIT, reflects the highest-fitness individual created during the run, expressed as a percentage of total possible fitness. Thus, if a run created a completely optimal individual, it would have an associated HIFIT value of

[2]AVGHI is expected to be slightly above chance values here, since it represents the average of the *highest* scoring individuals of a run. A normal distribution around the mean would result in an AVGHI of slightly above chance.

100. If the best individual had only a fitness of, say, 20 (out of a total possible 24) then the HIFIT for that run would be 83.3. AVGHI and HIFIT measure aspects of the same thing – it is impossible to have an AVGHI of 100 without also having a HIFIT of 100 – but they also reflect slightly different characteristics. AVGHI is much more of a reflection of the *stability* and general nature of the population, since it reflects the 10 final generations; HIFIT merely reflects the best performance generated. The HIFIT of all the "basic" runs considered here is 100, indicating that in all of them at least one optimally fit individual was created. In contrast, the HIFIT of the blackboard-disabled runs is only 66.7.

The final measure, CONVERGE, attempts to quantify speed of convergence. It measures the number of generations necessary before the highest-fitness individual was evolved. Thus, a low CONVERGE value usually indicates quick convergence. This variable is occasionally problematic, in that it is possible for a run to generate an early high-fitness individual far before reaching population convergence; nevertheless, it provides a rough measure of the "speed" of evolution. The average CONVERGE for the runs considered in this section is 24.2, indicating that the optimum-fitness individuals are usually evolved after slightly more than 20 generations.

It is quite clear from the length of time necessary for optimal communication to arise that the communication does not evolve as the result of a blind random search. The highest fitness value in the first generation on all the basic runs considered here was 62.5, which is above chance – itself not surprising in a population of 500 – but still indicates a quite high error rate. The *average* fitness value of the first generation – which gives a measure of the value of pure random search – is 36.95, approximately chance.[3] The "conversations," then, between agents in the first generation are always marked by the lack of a stable convention of communication – in other words, the lack of a shared system of meanings between the agents. Table 4.1 gives a randomly chosen example of a fragment of conversation occurring between individuals during the first generation. In each of the tables considered in this chapter, the "write"

[3]Note: "chance" for a strategy that does not change the agent's databases at all is 33.33, but "chance" for a strategy that randomly replaces unknowns with other digits is 40.33. The value of 36.95 most likely indicates that while some agents are replacing at least *some* unknowns, many are not even taking that step.

Table 4.1: Conversation 1	
Master Db: 6881	Bert's: 9981 Ernie's: 6899
Action	Result
Bert: Db (0,1): 3	9 3 8 1
Bert: Bb (0): 1	19999999
Ernie: Bb (0): 1	19999999
Ernie: Db (0,2): 9	6 8 9 9
Ernie: Db (0,3): 9	6 8 9 9
Bert: Bb (3): 3	19939999
Ernie: Bb (6): 1	19999919
Ernie: Db (0,2): 9	6 8 9 9
Ernie: Db (0,3): 9	6 8 9 9
Bert: Bb (3): 3	19939999
Ernie: Bb (6): 1	19999919
Ernie: Db (0,2): 1	6 8 1 9
...	

actions of each agent are portrayed, including the location of the "write" as well as what digit was replaced by what value and the final outcome. Therefore, the line "Bert: db(0,1) 2: 8729" can be read to mean "Bert writes a 2 at location (0,1) of the database, and now the database contains 8729."

The conversation depicted here in Table 4.1 is actually in some respects a "cut above" many of the first-generation conversations, which were not depicted in a table because they would be too boring. In many conversations neither Bert nor Ernie "says" a word; they write nothing to their database or blackboard (or both). In others, Bert and Ernie immediately write digits to their database, completely regardless of what the other has written on the blackboard. In still others, Bert and Ernie write the same digit continuously on their blackboard and/or database regardless of anything else.

In *this* conversation, we at least see Bert and Ernie writing to both database and blackboard. They are clearly not communicating fully effectively with each other, as evidenced by the fact that Bert writes a digit to his database before Ernie even has a chance to write anything to his blackboard. When they *do* write to the blackboard, the digit is sometimes entirely different from any digits on the database. And we

can see that they are having trouble coordinating, since Bert and Ernie are clearly writing to different sections of their blackboard, which we do not see in conversations of optimal agents.[4] In the next section we shall look in more detail at conversations between optimally fit agents in order to better highlight the differences that evolution brings.

4.2.2 Analysis of the Conversation

It is clear solely from the values of the three variables considered above that optimum fitness, and hence communication, evolves in the default case. This finding applies even for runs involving different population and generation sizes. However, knowing that optimal fitness is reached does not tell us much about the *nature* of the communication that is created. To what extent can it be understood? Are there clear generalizations that can be made about the types of conventions evolved by the agents? Do agents of one run evolve similar strategies (i.e. a common "language" of sorts)? In what ways are agents from different runs similar? Different?

These questions, by nature, are difficult to answer in a quantitative manner. Nevertheless, an analysis of the issues they suggest is important to achieve a full understanding of the emergence of communication in general, especially in reference to the question of how similar this communication is to *human* conventions. In order to treat the differences in conversations somewhat quantitatively, we defined three variables that reflected differences in conversation: BBWrites (number of writes per "turn" to the blackboard by Bert or Ernie), BBSpace (number of locations on the blackboard actually used by the agents), and Order (whether the first unknown in the database was overwritten by both agents first, the second was, or one agent did one and one agent did another).

Data from three runs using default values on all parameters were analyzed. Segments of conversations are presented in the tables below. Though there is no such

[4]This is especially interesting since it is theoretically possible for fully coordinated agents to write to different portions of their blackboards than each other. They would just have to "know" where to look. Nevertheless, we have not seen this in any of the optimally-fit conversations analyzed so far.

Table 4.2: Conversation 2	
Master Db: 7630	Bert's: 9930 Ernie's: 7699
Action	Result
...	
Ernie: Db(0,3): 9	7699
Ernie: Bb(0)-(2): 966	96699999
Bert: Db(0,1): 6	9630
Bert: Bb(0)-(2): 900	90099999
Ernie: Db(0,3): 0	7690
Ernie: Bb(0)-(2): 966	96699999
...	
Ernie: Db(0,2): 9	7690
Ernie: Bb(0)-(2): 977	97799999
Bert: Db(0,0): 7	7630
Bert: Bb(0)-(2): 933	93399999
Ernie: Db(0,2): 3	7630
...	

thing as a "baseline" conversation, the examples in the tables are typical of their runs in terms of the variables BBWrites, BBSpace, and Order.

Conversational analysis indicates two main things. First, and most importantly, conversations are marked by the use of conventions that are shared by both participants. In other words, communication is achieved through the use of shared, coordinated mechanisms – for instance, both agents will write in the same location on the blackboard (BBSpace) and use the same number of writes in order to do so (BBWrites). These mechanisms undoubtedly arise from the shared 'genetic' structure of the agents, but can be understood (metaphorically, at least) as the "language" shared by the agents. Due to their simplicity, they are a language only in the most general possible sense, that of being a shared convention enabling successful communication.

What is the nature of these conventions? Consider two segments of typical conversations from perfectly fit individuals from different runs (see tables 4.2 and 4.3 below). These conversations illustrate some of the factors that seem to characterize typical interactions in the database communication task. Ellipses (...) indicate sections where lines of conversation have been omitted.

The first thing that one notices about these conversations is that they rely on

Table 4.3: Conversation 3	
Master Db: 2874	Bert's: 2994 Ernie's: 9879
Action	Result
...	
Ernie: Db(0,0): 9	9879
Ernie: Bb(0): 8	89999999
Ernie: Bb(0): 7	79999999
Bert: Db(0,2): 7	2974
Bert: Bb(0): 4	49999999
Bert: Bb(0): 2	29999999
Ernie: Db(0,0): 2	2879
...	
Ernie: Db(0,3): 9	2879
Ernie: Bb(0): 2	29999999
Ernie: Bb(0): 8	89999999
Bert: Db(0,1): 8	2874
Bert: Bb(0): 7	79999999
Bert: Bb(0): 4	49999999
Ernie: Db(0,3): 4	2874
...	

a shared underlying convention in order to have effective communication. In Table 4.2, we see that agents both end up first "speaking" the second unknown that occurs in their databases, and when that has been successfully communicated, they begin "speaking" the first unknown. Moreover, the digits are always in the same location on the blackboard and prefaced by a 9. In Table 4.3, we see another (equally effective) convention. In this conversation, agents only use location (0) on the blackboard, and consistently write *both* digits of the database onto it. Naturally, since the second write action covers the effects of the first, this is the same as if the agents had merely written the second known digit onto the board. Again, the other agent somehow realizes that this is the second known digit, and fills it into the correct spot on the blackboard. These conversational differences are reflected in the variables BBSpace (which is 4 in Table 4.2 and 1 in Table 4.3) and Order (which is "second" in Table 4.2 and "both" in Table 4.3).

One thing is unclear from the conversations evident in Tables 4.2 and 4.3. Each

agent had to realize when the other one successfully filled in the database with the first digit so that they could begin "speaking" the second digit. It is not clear from an analysis of either conversation (during the ellipses) exactly how this was accomplished, but it seems to be different in each conversation.

The fact that both agents here happen to focus first on the second known digit before moving to the first is coincidental; there are conversations in which they begin with the first, and even occasions when one focuses on the first and one focuses on the second. Since genetic programming is an inherently *probabilistic* algorithm, this variation is not unexpected. What seems to be fundamental is that in all cases each agent is aware of which digit the other one is referring to. This awareness might spring from two things: a convention based in the 'genetics' – the program structure – of the agent, or else an explicit communication about location on the blackboard in addition to the direct communication of the digit itself. Most conversations seem to rely on the former, but the latter possibility is certainly conceivable.

The importance of the "genetic" programming code of the individual agents is nowhere more evident than in comparing conversations from different generations within the same run. The question is whether individuals sharing a run (and hence, possibly, more genetic similarity) are likely to share or converge upon the same conversation strategies involved in communication. In other words, are the mechanisms involved in conversation more likely to be qualititatively similar for agents that belong to the same population run, or is the genetic distance caused by crossover and evolution large enough that no such similarities can be found? Do agents have similar "languages" based on their genetics, or not?

We can get a partial answer to this question by analyzing the conversation structure of multiple agents in the same run. Results of analyzing multiple conversations are illustrated well by the examples of Table 4.2, 4.4, and 4.5. All three represent conversations from the same run – Table 4.2 is a conversation from the first optimally fit individual (generation 21), Table 4.4 a conversation from a random generation 32 individual, and Table 4.5 a conversation from a random generation 41 individual.

The most striking conclusion from analyzing Tables 4.4 and 4.5 is that the conversation structure of 4.4 is almost identical to that of 4.2, even though the two agents

Table 4.4: Conversation 4	
Master Db: 1781	Bert's: 9981 Ernie's: 1799
Action	Result
...	
Ernie: Db(0,3): 9	1799
Ernie: Bb(3): 9	99991199
Ernie: Bb(4)-(5): 7	99997799
Bert: Db(0,1): 7	9781
Bert: Bb(3): 9	99991199
Bert: Bb(4)-(5): 1	99991199
Ernie: Db(0,3): 1	1791
Ernie: Bb(3): 9	99997799
Ernie: Bb(4)-(5): 7	99997799
...	
Ernie: Db(0,2): 9	1791
Ernie: Bb(3): 9	99991199
Ernie: Bb(4)-(5): 1	91199999
Bert: Db(0,0): 1	1781
Bert: Bb(3): 9	99998899
Bert: Bb(4)-(5): 8	99998899
Ernie: Db(0,2): 8	1781
...	

Table 4.5: Conversation 5	
Master Db: 1781	Bert's: 1991 Ernie's: 9789
Action	Result
...	
Ernie: Bb(0)-(7)	98899988
Bert: Db(0,2): 8	1981
Bert: Bb(0)-(7)	98199911
Ernie: Db(0,0): 1	1789
Ernie: Bb(0)-(7)	98899988
Bert: Bb(0)-(7)	98199911
Ernie: Bb(0)-(7)	98799988
Bert: Db(0,1): 9	1981
Bert: Bb(0)-(7)	19819991
Ernie: Db(0,3): 1	1781
Ernie: Bb(0)-(7)	98799977
Bert: Db(0,1): 7	1781
...	

are communicating over different databases and are separated by 11 generations. Other than the fact that digits are written to different locations on the blackboard, the conversation structure is exactly the same. More generally, qualitatively similar conversations are highly likely to be found throughout individuals in all parts of a run. For instance, Table 4.3 came from a run, *all of whose* analyzed optimally-fit members (23 out of 50+) had BBSpace values of 1 and BBWrite values of 2. That is, they did two writes to the blackboard per turn, and used up only one location on the blackboard throughout. The run that Table 4.2, 4.4, and 4.5 came from was a bit more variable, but in all of its analyzed conversations a large portion of the blackboard was used (3 or 4 digits in the beginning of the run, up to 8 or 9 by the end) and agents almost exclusively had an Order value of "both", with Bert first writing over the second unknown in the database and Ernie writing over the first. The similarity within runs is not too surprising given that agents of a run are likely to share a high degree of genetic similarity, but it is nevertheless provocative, as it demonstrates the possible existence of a common, genetically-based language even though members of the population never speak to one another.

Keeping that in mind, it is important to realize that Table 4.5 suggests that *all* conversations in one run need not be qualitatively similar. The conversation in Table 4.5 is closer (in terms of generations) to Table 4.3 than it is to Table 4.4 from 4.2, yet Table 4.4 and 4.2 are indisputably much more qualitatively well-matched. Thus, while there definitely seem to be forces driving the creation of a "common language" over generations, there are countervailing forces (e.g. during crossover) that ensure that conversation structure is potentially changeable.

In terms of conversation structure itself, it is difficult to analyze exactly *what* is going on in Table 4.5. Again, it is clearly communication – that is, the agents are clearly employing a consistent convention in order to arrive at perfect fitness – but, as is often typical in genetic programming applications, the exact analysis of that mechanism is opaque. At the very least, we *can* say with some degree of certainty that the mechanism involved is probably quite different than one which a human would come up with.[5] Implications and reasons for this will be considered briefly in

[5]We cannot make certain claims about what a typical, optimal *human* strategy would be, but

the next chapter.

Dialogue structure is another area in which this research may have interesting implications. The agents in this paradigm were not programmed with *any* implicit notions of "turn-taking" or other elements of dialogue structure, and as we saw in the analysis of first-generation individuals, there are multiple agent interactions that do not incorporate any of these notions at all. Agents will indulge in monologues, or refuse to speak entirely, or write over the entire blackboard multiple times before the other agent has a turn. In contrast, the communicating agents *usually* displayed none of these behaviors. A typical turn consisted of writing digits in some specified sequence to all or part of the blackboard and writing a digit or two to the database. In that sense, then, turn-taking seemed to emerge. That said, it is important to remember that while dialogue structure is not hard-wired into agents, the turn-taking inherent in the simulation may have introduced a *proclivity* to turn-taking in conversation. It is difficult to tell to what degree this may have been a factor.

While some form of turn-taking similar to human conversations emerged, dialogue structure was quite unusual in another sense. Conversations generally took a great deal *longer* than the minimum time necessary, and much of that time was spent writing 9s to the blackboard or the database. As a result, agents might write a digit (say, 7) onto the same place in the blackboard – and write a 9 to the same place in the database – for multiple turns in a row before changing one or both actions. This behavior would be quite odd to find in conversations between humans, and it is still unclear what its cause is.

4.3 The Effect of Database Size

The discovery that true communication does indeed evolve as an emergent property of the joint activity of database matching is the primary finding of this research,

our intuitions here can be a solid guide. Given a database of length four with 50% of the locations unknown, a human would probably print out the length of their database – verbatim – on their blackboard, read the other agent's blackboard, and revise accordingly. If agents knew that the other individuals "didn't know" the very things that they themselves *did* know, they might only bother to print out the "known" information. In any case, it *seems* intuitively unlikely that humans would go through all of the manipulations typical of our evolved agents here.

but it is only a starting point. Given this realization, it is especially important to determine what parameters this result rests upon. Doing so will help to define not only which assumptions are necessary for communication to emerge, but also determine the extent to which the results can be generalized to cover other paradigms and situations.

To that end, the data from the 41 different runs was analyzed. Specifically of interest was the degree of correlation between each parameter and the dependent variables measuring communicative success, AVGHI and HIFIT. Both of these dependent variables varied widely across the 41 different runs: AVGHI had a mean value of 84.54 and a standard deviation of 18.15, while HIFIT had a mean value of 88.13 and a standard deviation of 20. Thus, there was enough variation in the dependent (outcome) measures to make them potentially dependent on a variety of independent parameters.

Results strongly indicated that the only parameter with a statistically significant correlation to either HIFIT ($r = -0.708$, $p < 0.0001$) or AVGHI ($r = -0.674$, $p < 0.0001$) was the parameter of *Database Size*. In other words, longer database sizes were strongly predictive of increasingly poor performance (i.e. increasingly unsuccessful communication). Specifically, databases as small as length 8 could not be made to evolve optimally performing individuals, although fitness was still above chance (AVGHI = 74.5, HIFIT = 78.2). Databases of length 12 did even worse, in general, garnering a mean AVGHI of only 62.3 and a mean HIFIT of only 65.4 – still above chance, but nowhere near optimal performance.

In addition to correlating with measures of performance, database size also was a predictor of CONVERGE, although less strongly ($r=0.371$, $p < 0.0210$). In other words, smaller databases were likely to converge more quickly on the highest fitness of the run than were large databases. This finding is not surprising, given the increased complexity of larger databases; however, it is interesting that the single variable of database size had enough impact to be so highly statistically significant. The only other statistically significant predictor of CONVERGE is blackboard size, and that one only barely ($r=0.323$, $p < 0.0478$).

It appears, then, that database size is the key variable in creating populations

that evolve emergent communication. Yet under what conditions does fitness within these limits vary? In other words, to what extent is it possible to tease more and more optimal performance out of longer and longer databases? What are the limiting factors preventing fully optimal communication from developing?

4.3.1 The Effect of Conversation Length on Large Databases

One main parameter has the most effect on increasing performance for larger databases: conversation length. The default conversation length is set at 100, meaning that each individual Bert and Ernie are "run" 100 times during each conversation, and thus have 100 "turns" of conversation. Although this is far above the minimum number of turns that is strictly necessary for perfect communication in a run for a database of length 12, it is clear that longer databases would have more of a need for multiple turns than smaller ones. Thus, additional simulations were run that varied only the conversation size parameter: one set it at 500, and another to 1000.

For length 12 databases, increasing conversation size to 500 resulted in an increase of AVGHI from 62.3 to 75.3 and an increase of HIFIT from 65.4 to 79.2 – a substantial improvement, although still not the type of conversation necessary for optimal performance. By contrast, the same run implemented with a disabled blackboard resulted in an AVGHI of 47.9 and a HIFIT of 50.7 – approximately chance levels for both variables. The discrepancy between these two runs is indication that the success of the former was dependent on some communication, even though it clearly wasn't optimal.

How might optimal communication for length 12 databases evolve? On the theory that if some is good, more is even better, conversation length was increased to 1000. This, however, did not increase either HIFIT or AVGHI at all (garnering values of 74.6 and 77.8 respectively). This strongly indicates that while conversation length is important, it is only effective up to a certain threshold of performance, and not beyond.

The same influence of conversation length is evident in runs using smaller databases in which there is no emergence of optimal performers. In databases of length 8,

AVGHI is 69.5 when all parameters are at the standard defaults; however, when conversation length is set at 400, it increases only to 79.5. While this is the best AVGHI value of all databases *larger* than size four, it is still far from the optimal performance achieved for size four databases.

4.3.2 The Effect of Blackboard Size

Blackboard size, as well, appears to have a non-negligible impact on successful communication up to a certain threshold. Limitations on blackboard size create a bottleneck regulating how much information can be communicated during one turn. Thus, one might expect for lower blackboard sizes to hinder communication for larger databases. In fact, this speculation is substantiated by the data: agents with databases of length 12 and blackboards of size 1 (rather than the default 8) have an AVGHI of only 54.3, and achieve a HIFIT of merely 56.3. In other words, when a bottleneck is created by allowing agents to only print one digit per turn, performance decreases substantially for larger databases.

Although performance seems to go *down* when very small blackboards are used, performance does not similarly go *up* when very large blackboards are used. In a run whose parameters were identical to the last example except that the blackboard size was set to 25, AVGHI was found to be only 49.6, and HIFIT only 52.8. This seems to indicate that, as with conversation length, there is a threshold level before which performance is affected, but above which differences seem to have no effect on performance. This makes sense: bottlenecks on the output (blackboard) will have the most effect when they severely limit the ability of the individual to communicate. They will have very little if the blackboard increases past the point that agents are utilizing it in the first place (and as we have seen, agents don't always use the full 8-length blackboard anyway).

Can the effects of smaller blackboards (i.e. bottlenecks on the output) be mitigated by removing the *other* bottlenecks on output that exist, namely conversation length? In other words, will an agent with a very small blackboard but very long conversation length perform as well as an agent with the defaults of both? Which "bottleneck"

has the most effect on performance?

In a run with default values of everything except for large conversation lengths (500) and small blackboard lengths (1), performance was quite good (AVGHI = 84.8, HIFIT = 89.5). This contrasts with the finding covered earlier showing that runs with size 8 blackboards and short conversation lengths have significantly diminished performance. These two findings, taken together, indicate that the bottleneck created by conversation length has more of a detrimental effect on performance than the bottleneck created by limitations on blackboard size. In Section 7 we will consider the effects of blackboard size in more detail.

4.3.3 Effect of Internal Database Structure

One other factor that seems to have an effect on performance for larger length databases is internal database structure. Usually the databases were randomly filled with the digits 0 through 8; that is, there was no *internal* database structure. However, the presence of internal database structure has the ability to change the nature of the task in two potential ways. First, the strategy used by agents could become less communicative and more strategic, taking advantage of the knowledge of the structure in order to "infer" what goes where without having to ask the other agent. Secondly, the communication could become shorter and clearer while communicating the same thing. For example, if all information in the database was always packaged in clumps of four identical digits (making '4444' a possible entry but not '4328') and both agents "knew" this, then agents could get away with writing only one digit (4) on the blackboard, rather than all four digits. Either way, increased database structure would increase the probability of achieving optimal performance, especially on larger databases.

Results of runs implementing this change indicate that performance *does* increase when database structure is added (and conversation length is set at 500). In one run, data was always packaged in clumps of four identical digits, resulting in databases like '333377774444' but not something like '287456231158'. In that case, AVGHI was 83.3, the highest of all length 12 databases, and HIFIT was all the way up to 86.1.

While this definitely reflects improved performance, it is worth noting that it *still* did not result in optimal fitness, even though the amount of information in the databases was no more than that found in databases of size 3 (which *do* achieve optimal performance). This may indicate that the problem with larger databases has less to do with the amount of information needed to be communicated than the difficulty in coordinating information about all possible database locations. We will explore the implications of this finding in further detail in Chapter 5.

In another run, data was packaged in clumps of four such that the first and third digit of each was identical, as was the second and fourth. Thus, the database '343482820505' would be a possibility, but '762434518724' would not be. In this case – which is a bit more complicated than the first – performance goes down dramatically; AVGHI is 67 and HIFIT 72.9, scarcely different from the performance for non-binary databases. Thus, it seems agents can recognize or take advantage of only the simplest of structure in the database.[6]

While this definitely reflects improved performance, it is worth noting that it *still* did not result in optimal fitness, even though the amount of information in the databases was no more than that found in databases of size 3 (which *do* achieve optimal performance). This may indicate that the problem with larger databases has less to do with the amount of information needed to be communicated than with the difficulty in coordinating information about all possible database locations.

[6]Note: It is important to realize that both types of structure considered here are types of structure that might be utilized by *non-communicative* agents. In order to avoid the question of whether any improved performance during these runs is due to effective use of non-communicative strategies or to linguistic, communicative strategies making use of this structure, it would be ideal to incorporate structure that did not allow non-communicative strategies to be effective. However, the distinction between structure allowing effective non-communicative strategies and structure preferentially selecting communicative strategies is a very fine one at best. *Any* structure, by nature, includes within it the possibility of a non-communicative strategy making use of that structure. As the structure gets more complicated it may be more and more *unlikely* that a non-communicative strategy that can effectively use it will evolve – but it is never impossible. Indeed, as structure gets more and more complicated it is probably more unlikely that a *linguistic* strategy could evolve to use it. This difference is key in any ultimate attempts to evolve structured, syntactic language using this paradigm, and these issues must be considered at far greater length before attempting to do so. Since the Evolution of Syntax *per se* is not our concern in this piece of work, our consideration of these issues shall be limited to this footnote.

4.4 Database Richness

Variations in the structure of the information in the database translate, at root, to variations in the *amount* of information in the database. That is, highly structured information has lower entropy and more order than completely random information. Thus, increased performance when the databases contain structured information could result either from the fact that agents actually need to communicate less information when the database is structured, *or* from the fact that agents are taking advantage of the structure to coordinate conversations and communicate more simply.

We can piece apart these two possibilities by analyzing the case when the database is filled with binary information rather than decimal. In the default case, information in the database is made up of the digits 0 through 8 (with 9 being unknown). Changing this so that information in the database is binary (digits 0 and 1) has the effect of decreasing the amount of *information* in each database but keeping size constant. In other words, a database of size 12 that is binary will contain less information than an equivalent database of size 12 that is decimal. Thus, by comparing performances, it is possible to begin to distinguish whether it is the *information content* or the *size* that makes databases of length 12 not reach optimal performance.

One primary characteristic of binary databases is that not only do they contain less information in equivalently sized databases than do decimal, but they make it more probable that a non-communicative strategy will achieve high levels of fitness. For every digit, a random strategy in a binary database will have a 50% probability of filling in the "correct" integer (either 0 or 1) while a strategy in a decimal database will only be correct with 1 in 9 (11%) probability.

We can see this result clearly when we compare performance on binary databases whose databases are repeated different numbers of times. Recall that by default, the master database is exchanged 8 times during each generation; by changing this number, we can affect the probability of a non-communicative strategy achieving high or even optimal performance. Indeed, there *is* a difference between runs in which databases are exchanged 8 times, and when they are exchanged fewer (5 times).

Runs in which binary databases are exchanged 8 times have the default values for all parameters except for database richness. In fact, performance is not significantly different than the default case: individuals reach optimal fitness by generation 21, and the AVGHI is 100, indicating consistent attainment of optimal performance. These figures are quite comparable – indeed, almost identical – to the results of the general, default case.

We can compare this result to the case in which binary databases are exchanged 5 times. Performance is still consistently high and even, often, optimal; however, there is not evidence of the same *consistently* optimal performance found in the former case. AVGHI is 93, which is not poor by any means. But many generations – even many *later* generations – do not contain optimal performers, which is in stark contrast to the exchange-8 case. A possible explanation for this is that optimal performance here is often a by-product of "lucky guessing" rather than actual communication strategies, and thus cannot be consistently relied upon.

In order to further test this hypothesis, it is necessary to examine the conversations to see if they indicate true communication or, instead, illustrate the usage of non-communicative strategies. The result of such analysis is a mixed bag. Some conversational segments indicate communication between the agents, but some clearly don't. For instance, some agents will write digits occasionally to a database, even if those digits were not communicated beforehand via the other agent's blackboard. And generations that happen to, by chance, have databases that are all similar (e.g. 0000, 0100, 1000) tend to have more optimally performing individuals, which can only be explained by the fact that they make it more probable for a random strategy to be effective.

It is therefore clear that if the task is simple enough to be solvable using a non-communicative strategy, such a strategy will be employed. In other words, non-communicative strategies form local maxima on the fitness landscape. In the default situation, when databases are more complex than binary ones, these local maxima are relatively small in comparison to the global maximum of optimal communicative performance, and therefore agents do not fall into the trap of using non-communication

strategies. Thus, one of the keys to evolving successful communication lies in creating a fitness landscape with a relatively small probability that non-communicative strategies will produce high fitness.

This is an important insight, but we still have not answered the question we originally set out to answer – does the difficulty with large database sizes lie in the quantity of *information* needed to be communicated, or in the size of the database itself? In order to answer it, we considered a run with database of size 12, standard blackboard size, large conversation length (500), and binary rather than decimal database. If the problem is information rather than sheer size of the database, then one would expect to see high, even optimal, performance for large binary databases.

Results indicate that it is probably a combination of information and sheer size: while fitness never reaches optimal levels, it is certainly higher than any score attained on any other database of size 12. AVGHI is 88.9 and HIFIT as much as 91.7. A possible analysis of this is that much of the problem that agents have with larger databases is writing the wrong digit to certain locations – with fewer choices of digits, there is less opportunity for a mistake in *communication* to result in a mistake of writing to the database. Therefore, this higher fitness score probably does not reflect improved communication for binary databases, only increased reliance on non-communicative strategies. The underlying problem, then, is that agents can't seem to handle the actual size of the databases – coordinating what is written on the blackboard with what should be written to the database just becomes too complicated. The implications of this will be explored more thoroughly in Chapter 5.

4.5 Altering the Fitness Function

There are two important things that we have touched on so far but not explored completely. First is the importance of conversation length as a bottleneck encouraging (or limiting) the evolution of successful communication, specifically for medium-sized or small databases. Second is the fact that the nature of the fitness function itself can have a large effect on the nature of what evolves. Up to this point we have only considered bottlenecks on output (i.e. conversation length) as *implicit* limitations

on fitness, not direct ones. How would our results change if conversation length was explicitly factored into the fitness function? Would the existence of a bottleneck prevent the evolution of an optimal communicative system? Or would it merely result in a system what was shorter and more compact?

In order to answer these questions, the fitness function was changed to accomodate information about conversation length. Because we are interested in bottlenecks on output, agents were penalized for each time they wrote to either a database or blackboard, with an upper limit of 300 'writes' per conversation. Fitness was assessed by adding the number of writes for each of Bert and Ernie, dividing the total by the number of total writes allowed (300 + 300 = 600), and subtracting the whole thing from one. Thus, an individual who didn't write *anything* would have this component of the fitness score be 1. An individual whose Bert and Ernies both took the full 300 writes allocated to them would have this component be 0.

The rest of the fitness function here remains exactly the same, making "optimal" performance now 32 rather than 24 (since for each database the maximum fitness is 4, and there are 8 databases exchanged per generation). However, this situation is unlike all previous situations, since now the fitness function is made up of components that are mutually exclusive to one another. Thus, reaching "optimal" fitness is a sheer impossibility – there is no way that an agent could correctly fill in its database while not writing anything to the database, much less its own blackboard.

Further analysis, in fact, reveals that incorporating this component into the fitness function results in agents that do not evolve completely successful communicative systems. The mean AVGHI value of runs using this fitness function and default values for everything else is 65.8, meaning that the average fitness score is 65.8% of 32, or an absolute value of 21.1. This would be sub-optimal performance even if fitness were measured out of 24 rather than 32, and represents quite poor performance when the additional fitness component is taken into account. In trying to create a streamlined, compact system of communication, we have evolved one that is neither streamlined nor even fully accurate.

Is the problem, perhaps, in the weight being put on the component of fitness

devoted to output? In other words, perhaps the problem isn't with an explicit bottleneck on output *per se* – perhaps the problem is that such a bottleneck needs to be less extreme. In order to test this possibility, the output component of fitness was divided by four, making the fully optimal fitness 26 (24 + (8*0.25)).

This option creates a different result, but one that is hardly distinct from the default case when there is *no* explicit bottleneck on fitness. The mean AVGHI becomes 92.2, which translates to an AVGHI of 99.8 (considered out of a total of 24). This is highly indicative of near-perfect if not perfect conversational accuracy (recall that the "optimal" goal of 26 is theoretically impossible). Unfortunately, it does not result in significantly shorter conversations. The 10 highest-fitness individuals of any run (those handful with fitness scores of 24 or above) include individuals scoring nearly '0' on the output component, but generating perfect communication. It seems as if the component of fitness devoted to output is simply *not strong enough* to make a difference.

In order to remedy that, another run was done in which the output component of fitness was divided by two rather than four, making fully optimal agents have a score of 28. This action, not too surprisingly, had the effect of evolving a population midway between the first two. There is occasionally perfect and near-perfect accurate communication between Bert and Ernie. And there are fewer high-scoring individuals that score high merely by virtue of communicative accuracy, without taking output considerations into account.

That said, the effect of this component of the fitness function on output is clearly negligible in comparison with the effect of the other components. The mean AVGHI for runs using this fitness function is 86.1, which translates to a mean *absolute* fitness of 24.1. Mean HIFIT is 87.5, which translates to an absolute of 24.45. Thus, although there are a few agents that manage to both communicate without error *and* have slightly fewer writes than the maximum allowed, there are none that do so strongly or consistently. And there are far fewer agents, total, that have optimal accuracy of communication since many of the ones with the highest scores do so by virtue of having short (but not necessarily fully accurate) conversations.

Thus, it seems that figuring a bottleneck on the output explicitly into the fitness

function produces mixed results at best. There is no emergence of a population with optimal or near-optimal communication as well as relatively few writes. In the best case there are some optimally communicating agents with slightly shorter conversations. However, there is less effective communication in the population taken as a whole. In general, the modified fitness function results only in a population with less true communication and not much recompense in the form of shorter conversation.

4.5.1 Change in the Function Set

So far we have considered many of the paramaters of variation first discussed in Chapter 3, but one that we have not yet discussed is the function set making up the individuals themselves. One of the key limitations of genetic programming is its fundamental reliance on the function set to begin with. The default function set was originally chosen on the basis of its simplicity, since it contained relatively few functions and the ones that it did contain were among the most fundamental. Yet, especially given the difficulties with larger databases, it is always possible that using more functions – and thus potentially increasing the abilities of agents to manipulate the database and blackboard – would result in better performance. To what extent are the results we have so far obtained dependent upon the function set used?

In order to answer this question, several modifications were made to the function set. First of all, the functions taking the location of the blackboard and database as a variable were removed.[7] The reasoning behind this was that it is still theoretically possible to have a perfectly adequate conversation without this ability; thus, if one wishes to make the fewest assumptions possible, these functions should be eliminated.

Results indicate that removing these "location" functions has little effect on performance. As we have seen, the default AVGHI for the default function set was 98.4, and the HIFIT was 100. By comparison, the AVGHI for the modified function set – without the "location" functions – was 95, and the HIFIT was 100. Rates of convergence for each function set were approximately similar as well.

While removing the "location" functions is the most intuitive and obvious way

[7]These are functions Write-Bit-To-DB-At-Location and Write-Bit-To-BB-At-Location.

Table 4.6: Modified Function Set		
Function	Num. Arguments	Value returned
Move-Back-On-BB	0	9
Move-Back-On-DB	0	9
Move-To-Beginning-Of-BB	0	9
Move-To-Beginning-Of-DB	0	9
Move-Ahead-On-BB	0	9
Move-Ahead-On-DB	0	9
Erase-BB	0	9

to simplify the function set, another method would be to remove the *other* type of functions, leaving *only* the location ones. Results indicate that the "non-location" functions are more important to achieving optimal performance – AVGHI of the run was only 81, and no perfectly communicating individuals evolved at all; HIFIT was only 83.3. Apparently the simpler functions were necessary in order to better coordinate the agents with one another.

Given that in at least some instances having more functions is better than having fewer, it makes sense to ask what would happen if individuals had access to even greater functionality. Would these added abilities make performance increase? To answer these questions, the following functions were added to the function set. Move-Back-On-BB and Move-Back-On-DB gave agents the ability to move backwards one location on the blackboard and database without writing to either, respectively; Move-Ahead-On-BB and Move-Ahead-On-DB, unsurprisingly, gave agents the ability to move ahead on digit on each, also without writing to either. Move-To-The-Beginning-Of-BB and Move-To-The-Beginning-Of-DB moved agents to the beginning of the blackboard (location [0]) and the database (location [0,0]). Erase-BB replaced all digits on the blackboard with unknowns, and moved the agent to location [0].

Results were catastrophic; the presence of so many additional functions seemed to cause agents to be completely unable to coordinate with one another. AVGHI reached only 41.6, and HIFIT was a meager 44.5 – approximately chance levels. With the additional functions, individuals became unable to coordinate and therefore could not successfully communicate with each other at all.[8]

[8]This result is fairly consistent with general properties of Genetic Programming, which indicate

It is evident that function set does play a key role in final performance. Contrary to expectation, simpler function sets may, in fact, be better, since they better allow agents to coordinate with one another. Thus, the route to hopefully achieving optimal performance on even larger databases probably does not fall by way of enlarging the function set, unless it would be possible to do so without potentially sacrificing coordination between the agents.

Given that coordination is so important, a key question is whether this paradigm been building in certain assumptions that automatically make coordination easier. For instance, we have so far been assuming that individuals have the ability to write over "unknown" information but not "known" information. To some extent, this is intuitively a justified goal; after all, most people don't routinely *replace* remembered information in their head with things they have just learned. Nevertheless, people *do* forget things, and there is no biological or logical reason to assume from the beginning that agents should be "born" with the ability to distinguish between "knowns" and "unknowns."

Therefore, we did a run with default values of everything except that agents had the ability to write over all digits in the database, including known values. Fitness plummeted dramatically: AVGHI was only 45.82 and HIFIT merely 55.95. Furthermore, an analysis of conversations revealed that agents were having a difficult time achieving any degree of success because of the strong tendency to write over "knowledge", making it impossible to reclaim later. Thus, it seems as if the ability to write over all types of information in the databases is actually hostile to the ultimate development of communication.

---

that enlarging a function set may or may not significantly change the results, depending upon which functions are added. (Koza, 1992) Basically, if the additional functions do not add functionality (i.e. they could be formed with the already-existing functions) then adding them will usually result only in differences in how long it takes optimal solutions to appear, without significantly changing *whether* they appear or not. If the added functions *do* add functionality – as these do – then they can have strong effects in either direction.

4.6 Summary of Findings

The results considered here have answered many of the questions we began with while raising more queries as well as crystallizing the important issues. The key results gleaned from this work tell us about the power as well as the limitations of the paradigm discussed here. The fact that it is *possible* for communication – and hence meaning – to emerge out of the performance of a joint activity has been demonstrated by the emergence of optimally communicating individuals. However, our joy at this result must be tempered by the knowledge that this successful communication is apparently severely limited by the size of the database needing to be communicated *about*. While it was possible to twiddle with the parameters of variation far enough to get a performance level significantly above chance, nothing was sufficient to create optimally communicating individuals for databases larger than size 4.

Systematically altering parameters that created bottlenecks on the output served to demonstrate that some types of limitations were more severe than others. For instance, blackboard size seemed to have relatively little impact except for extreme cases – and even in those cases, the diminished performance could be remedied by reducing the bottleneck due to conversation length. By contrast, conversation length was a more fixed and effective bottleneck. Larger databases did significantly better with longer conversations, and when conversations were short, decreasing the bottleneck due to blackboard size (by lengthening the blackboard) had little if any effect.

These *implicit* bottlenecks, however, had a much more interesting effect on the populations of agents than did an explicit modification of the fitness function to create a bottleneck. Although a few individuals did evolve that still managed to communicate adequately, the goal of achieving consistent perfect communication while *still* having short conversations (or otherwise limiting output somehow) was not met.

Much of the variation in the parameters making up the experiment indicated that coordination between agents was of primary importance in establishing successful communication. In other words, agents needed to "agree" on a system for telling which digits written to the blackboard corresponded to which digits should be written to the database (and *how* they should correspond). With shorter databases, this was

possible, but as they got longer, agents tended to lose track of the vastly larger number of digits that needed to be communicated.

It was thought that increasing the structure – decreasing the entropy – of the information in the database would aid in coordination and simplification of conversation. When databases had a very simple, very clear structure there was indeed an increase in performance; but this did not hold for databases with even slightly more entropy. Additionally, the increase in performance during low-entropy cases is due to the increased reliance on non-communicative strategies rather than modification of the "language" of the agents. Thus, this performance did not derive from the help in coordination that database structure could have added.

The need for coordination was also evident in examining performance for different function sets. Adding functions to the basic function set, while increasing the agents' ability to manipulate their databases and blackboards, also made it more and more difficult for the agents to coordinate with each other. This resulted in dramatically declining performance. Similarly, giving agents the ability to write over "known" information as well as unknown information further inhibited the coordination capabilities of agents, since it became much more difficult to glean what locations the other agents were at. This ability also hurt performance by writing over the very information that needed to be communicated, before it could be.

In the next chapter we will examine the implications of these results as they apply to two things. First of all, how might we modify or change the experimental paradigm in order to make agents with larger and more complex databases achieve optimal performance? And secondly, to what extent can we generalize the results on this specific research to apply to the more theoretical questions we have considered regarding the nature of language evolution and acquisition as well as notions of meaning? Answering these questions is essential in pinpointing the value of this research as well as lighting the way for research to come.

Chapter 5

Conclusions

As an exploratory survey of this new paradigm in computational simulations of the evolution of language, this research had two primary aims. First, it was intended to determine if – and under what conditions – it is possible to evolve communicating agents in the database coordination paradigm. Secondly, it was intended to provide the foundations and "map the territory" of possibilities of the paradigm in order to determine to what extent it can be generalized to consider other issues in evolutionary linguistics, such as the Evolution of Syntax. In this chapter we will consider to what extent we have succeeded in these goals. Most importantly, we will begin to discuss the implications of our findings and look ahead to their impact on future research.

5.1 Relation to Other Work

The most general, overarching goal of this research was to provide a plausible account for the emergence of coordinated communication in humans by implementing a simulation of it on the computer. This goal rests on the implicit assumption that evolution in genetic programming scenarios and the evolution of human langauge have key similarities. This assumption seems plausible, if not downright probable – much of genetic programming is, after all, based on the parameters of evolution in the natural world. However, it is very important to note that the assumption of similarity between GP and real-world evolution is not unique to the research reported here; it is

a fundamental assumption of any computational simulation of evolution that claims to shed light on evolutionary phenomena in the real world. All the research considered in Chapter 2, then, is equally dependent upon this assumption, and our results should most appropriately be considered in comparison to that research.

As we saw in Chapter 2, there is as yet very little computational work done on the emergence of communication (and therefore, possibly, meaning). Theorists of meaning such as Clark and Wittgenstein have suggested that meaning is not, even in principle, separable from the context and process of its use. (Clark, 1996; Wittgenstein, discussed in Kripke, 1984)[1] In other words, meaning is *the action itself* in a conversation. Thus, in a simulation like ours, one can interpret the meaning of, say, the fourth line of table 4.3 to be the "what it is that makes an agent, upon seeing a '7' in the first location of the other agent's blackboard, write a '7' into a certain location of its own database." The setup of the paradigm as is included an implicit link between the symbols in the database (the "mentalese" of the agents, if you will) and the symbols on the blackboard (the "phonemes" or "words" of the agents).

The theories provided by Clark and Wittgenstein give us a framework in order to look at the emergence of communication in this simulation. According to this concept of meaning, we can say with a fair degree of justification that *meaning* has been shown to be an emergent property of an evolutionary process that does not directly select for communication at all. Most of the research we discussed in Chapter 2 does not have this property. However, some could be considered to show the emergence of meaning, as in Werner & Dyer's A-Life scenario (section 5.3) or even MacLennan & Burghardt's synthetic etiology approach (section 5.4). What makes the approach detailed here different than this prior research?

The primary difference and advantage to our approach lies the generalizeability of both its methods and its results. Both MacLelland & Burghardt and Werner & Dyer's A-Life scenarios limited their agents to acquiring language at the level of vervet monkeys – developing more complicated interactions, possibly even syntax, is impossible in that scenario because the agents do not come equipped with the necessary

[1]Note: By lumping Clark and Wittgenstein together here, I am not trying to imply that their theories are alike in any way *other* than insofar as they both suggest that meaning cannot be isolated from the context and process of the activity in which language is used.

structure. In contrast, this simulation equips agents with an internal representation of the world (the database), the ability to "speak" and "listen" to other agents, and the rudimentary processing capabilities (operations in first-order logic) necessary to – in theory, at least – be able to handle far more complicated systems. And it does this while making only minimal assumptions about the *nature* of meaning representations, dialogue structure, and agent interaction.

Although we have plausibly demonstrated the emergence of meaning, the setup so far has not encouraged principled exploration of the emergence and nature of reference. The alphabet of the "output" language of our agents corresponds exactly to the alphabet making up the internal representation of the agents – both consist of the numbers 0 through 9, with 9 being "unknown." This identity is implausible when applied to human agents (since our internal representations of words, without much doubt, do not consist of sound waves corresponding to phonemes). More importantly, as long as the internal and external "languages" are the same, we cannot draw any conclusions about the emergence of reference. A possible direction for future reseach would be, therefore, to construct a scenario in which the two alphabets are distinct; where, for instance, the language written to the blackboard consists of the letters A through J and the language inside the databases consists of the numbers 0 through 9. Effective communication in such a setting would be provocative evidence that reference as well can emerge in this paradigm.

Reference is one property of human language displays and that, therefore, evolutionary simulations of language should endeavor to account for. The consideration of it and other issues of meaning, however, point to a larger issue that we must discuss if we are to appropriately analyze the results of this simulation. To what extent is it possible to generalize from our findings here to what we know about how humans actually communicate and how language actually evolved? We begin answering this question by considering what we have learned about the importance of coordination.

5.2 The Importance of Coordination

The main thing we have learned about coordination is that it is, not surprisingly, key in developing effective communication strategies. Much of the difficulty in dealing with larger size databases did not consist directly in the amount of information in them, but rather in the increased difficulty in coordinating actions between the agents. In other words, agents would "speak" certain digits to their own blackboard and write digits to their database based on what was on the other agent's blackboard – but they would get messed up in keeping track of what digits went where, and perform far below optimum. They could not coordinate appropriately.

In some sense, this lack of coordination *is* the lack of a fully developed and complete system of meanings – after all, if meaning consists in the actual actions during an activity, then poor performance on the joint activity reflects poor or inadequate meaning characterization. Thus, to say that the agents could not communicate because they could not coordinate is, in some sense, tautological. This criticism has an element of truth in it, but the finding is still valuable because coordination is only a part of the action of the agents, and therefore only part of the meaning. Action also consists in knowing what is important to speak, knowing to look at the other agent's blackboard for information, taking "turns" in a rough sort of dialogue, and using an implicit notion of reference between the symbols on the blackboard and the database. Our agents adequately do all of these things, even for large databases. The hard part, apparently, is coordination.

To what extent can these coordination difficulties be translated into human terms? After all, humans do not speak digit by digit, and the sound stream cannot be "jumped around on" in the same way that our agents can move around and manipulate the blackboard. Furthermore, it is fairly safe to say that the information in our heads probably does not appear as a matrix of digits in a database.

However, neither of these issues inhibit the generalizeability of our findings because they are probably not fundamental assumptions upon which the results were based. The results here were not results about the *nature* of what was communicated; for the most part conclusions were based on concepts that can likely be generalized

relatively easily – size of database (which could correspond to complexity of the world or mental representation of the agent, among other possibilities), limitations imposed by bottlenecks on the output, etc.

Additionally, many of our findings may be useful in suggesting precisely *which* characteristics of humans were essential in allowing us to begin developing communication. For instance, the inability to "jump around" the sound stream – in other words, the fact that the sound stream is continuous and linear – is probably not strongly important in the evolution of communication. (Recall that agents with function sets containing "location" functions did approximately equally as well as agents without them). Nevertheless, if agents could *only* jump around, or could jump around freely and often, communication and coordination would at the very least be far more difficult.

Another implication of the results regarding coordination lies in the distinction we saw between *information* and *size* of the database. Although both characteristics seemed to have some effect on whether coordination could emerge, size seemed to be a much more important parameter. What does this mean? One interpretation is to see it as a reflection of the age-old balancing act between expressivity and interpretability/simplicity. In other words, high database information encourages the development of more expressive but less interpretable (because longer) language (blackboard writes). Low database information and low database size encourages more interpretable but less expressive language.

The middle case – low database information but high database size – encourages less interpretable *and* less expressive language. Less interpretable, because longer database sizes favor more 'writes' to the blackboard, and hence more possibility of confusion; less expressive because there is less information to express. The low interpretability is not a necessary by-product, however: agents *could* evolve systems that take advantage of the database structure to increase their interpretability by settling on conventions for referring to it. If agents actually do that, it may be the beginnings of a plausible account for the beginning of the evolution of syntax. We saw in examining other computational work (e.g. Kirby 1998, 1999a, 1999b; Batali 1998) that

syntax may emerge in response to pressure for more interpretable but equally expressive output. Taking advantage of database structure to increase interpretability in this scenario would be one means of responding to this pressure, and hence possibly a means by which syntax might begin to develop.

What do our results suggest? Essentially, our agents managed to take some advantage of the structure inherent in the database when it was *very* simple. As it become even slightly more complicated, the agents could not "clue in" to this structure over the evolution of the population, and thus did not take advantage of it to increase interpretability. This is not a conclusive result by any means, however. Only two types of – very simple – database structure were considered; it would be informative to experiment with more and less complicated internal structure. And if optimal communication could be established for larger database sizes, a great deal of variation could be analyzed in a principled manner.

This finding does imply that the non-linguistic skills of agents are very important in developing ultimately *linguistic* skills. In other words, it is necessary for the agent to have developed strong memory and pattern recognition abilities, otherwise rudimentary syntax cannot develop. Our agents have very little in the way of either memory *or* pattern recognition capabilities. They have no memory: all they can see is the most recent writes to the other agent's blackboard.

Furthermore, pattern recognition capabilities are limited by the functions of first order logic (FOL) *sans* quantifiers. While FOL can in theory be an extremely flexible and powerful representation tool, building up complicated representations with only the simplest operators (Or, If, And, and Not) is in practice quite difficult. Or, rather, it results in extremely long, unwieldy representations that can be much more easily expressed – and therefore are more likely to evolve – with complicated function sets. Thus, *pragmatically speaking*, the pattern recognition capabilities of our agents are arguably not very large either. This implies that much of the difficulty our agents had with the larger size databases resulted from the inability to take advantage of the inherent structure in the database by generating a syntax. Increasing the complexity of our agents – either by adding in functionality that aids in pattern recognition, or adding a memory buffer, or increasing the ability to communicate large blocks of

information at once – might therefore remove the limitation on database size that we are currently laboring under.

If this turns out to be true, it has strong implications for the actual evolution of human language. We have explored many theories that suggest that it involved a coevolution between a specific language acquisition device and more generalized intellectual abilities – this finding is consistent with them. (Deacon, 1992, 1997; Kirby & Hurford, 1997) It suggests that a reservoir of general intellectual abilities is necessary in order for syntax to first develop, and provides a mechanism for understanding how general intellectual abilities might "leapfrog" off of syntax and become more powerful (through communication of more information, etc).

Furthermore, as we have seen, the fossil record is highly suggestive of the view that language did not emerge until after a certain cognitive complexity was reached. Recall that evidence of fully modern language is not conclusive until about 100,000 to 40,000 years ago. And indirect evidence of *any* sort of language does not emerge until approximately 2 million years ago, with the development of stone tools and the beginning of alteration in the voice box. (Johanson & Edgar, 1996) Coincidentally, this comes immediately after an enormous increase in cognitive complexity as we evolved from *Australopithecus* to *H. Habilis*. This is of course an area subject to immense debate, but it is suggestive that our finding is consistent with the reigning view on human language evolution.

5.3 The Role of Bottlenecks

An important factor of human language evolution that we have so far not discussed completely is the role of bottlenecks – specifically, bottlenecks on output. In the real world, there are many sources of bottlenecks on output. If agents are ever in a hurry, bottlenecks in output inevitably emerge, since fitness would favor individuals who manage to communicate more information in a shorter period of time. The fact that language is verbal (auditory) implies that there are bottlenecks caused by the limitations of the human vocal tract and the speed at which phonemes can be produced. Most bottlenecks relate to the inherent time constraints possible and likely

to occur in the real world.

The bottlenecks modeled in this work were all bottlencks on time constraints in different ways – either bottlenecks on actual length of conversations or bottlenecks on the actions agents took, either implicit in the setup, as an absolute barrier, or explicit in the function set. In most cases, the bottlenecks acted as a threshold – that is, there were values for each characteristic above which the nature of the characteristic made no difference but below which there was a strong impact. So blackboard size was an important bottleneck when blackboard size was *very* small, but as it increased it rapidly became unimportant, there being essentially no difference between blackboards of size 8 and those of size 25. We saw a similar phenomenon in considering conversation length.

This finding generalizes to the interesting – and sensible – conclusion that most of the bottlenecks on human linguistic output are only significant below some threshold. This is not an inevitable inference from the finding by any means, but it *is* reasonable and a good place to start. It is logical from the evolutionary perpective – the bottlenecks we do observe seem to have this "threshold" quality. Humans can only recognize and speak a limited number of phonemes, probably due to our limitations in effectively processing hundreds of them. Similarly, sentences are generally limited in complexity and length, since we do not have the memory or processing resources to handle lengthy, heavily nested monologues. Furthermore, humans are typically quite time-pressured in their conversations – for instance, a communication of "Help!" or "Watch out! A lion's coming!" is far more likely to be effective if it takes less than a second to speak than if it takes half a minute. It can be argued quite reasonably that even everyday conversation takes place under some form of time-pressure, given the multiplicity of other factors vying for an individual's attention. Thus, it is possible that bottlenecks on output served the same function in early human evolution that they serve in our simulation here; namely, helping to streamline conversation and (potentially, at least) provide a pressure for the ultimate development of syntax.

With the consideration of output bottlenecks, we should also examine the effect of *other* bottlenecks and obstructions – in other words, noise and errors in the input and output. So far, all agents have enjoyed "perfect" communicative conditions: signals

are never corrupted in the process of writing from blackboard to database or database to blackboard, and all "unknowns" in one agent's database correspond to "knowns" in the other agent's database. Yet communication in the real world is notoriously fraught with difficulties in reception, transmission, and interpretation. A great deal could be learned about the robustness of coordination, meaning representation, communicative effectiveness by varying the amount of noise under which systems develop.

For instance, if a population is evolved under conditions in which a certain percentage of writes to the database or blackboard will be altered or damaged, does communication – and meaning – still evolve? What level of accuracy is necessary to develop optimum communication? Does communicative effectiveness degrade gradually as noise increases, or is there a threshold beyond which it doesn't work? What strategies can agents adopt (for instance, increased redundancy) that might counteract the effects of noise?

These questions make sense when analyzing evolution in noisy environments, but the addition of noise can be also used to measure the robustness of the scenarios evolved here. How would the optimally fit agents analyzed and developed in this work respond if forced into a noisy situation? Would they still be highly effective communicators, or would the presence of even a little bit of inaccuracy hurt them substantially? Answers to these questions can tell us a great deal about what "comes for free" in the process of language evolution. In other words, it could shed light on the question of what circumstances are necessary for adaptation to noise.

In general, many of the findings of this research are strongly suggestive of and consistent with certain theories of language evolution. In no case do they conclusively demonstrate that other theories are incorrect, but that wasn't the point of the research. Our goal was to explore this paradigm of computational simulation of the emergence of meaning, and to determine what factors made that emergence most likely. This knowledge was to form the basis of future explorations of the emergence of meaning, as well as to create a "map" of the intellectual landscape and suggest which paths of exploration would be most profitable. To that end, let us finally consider directions for further research: where do we go from here?

5.4 Directions for Future Research

The first goal of any future research in this specific paradigm should focus on determining what is necessary to create optimal performance on large databases (size 12 or larger). As long as our agents are limited in what size of databases they are effective on, it is very difficult to generalize findings beyond this specific paradigm. If one wants to, ultimately, make claims about the emergence of syntax due to making use of underlying structure in the database, then one needs to have databases complicated and long enough to have structure.

How can we accomplish this? We may find that it is beyond our capacity to evolve agents that perform optimally even on very large databases. But there are still multiple avenues that haven't been tried. Since many of the limitations seem to be in terms of lack of memory and/or processing capability, adding in rudimentary memory or more *processing* functions to the function set might generate improved performance. Similarly, since bottlenecks on output seem quite important, one might increase the output capabilities of individuals – allow them to print multiple digits at a time, or use characters that are not in the database, or use a structured, multi-layered blackboard. And since the problem was often due to coordination difficulties, agents could gain the ability to know what the other agent was looking at in the database or the blackboard. This is actually somewhat plausible in terms of human behavior – we rely a great deal on body language and inferences from the situation to understand what our conversational partners are thinking and where they are in the dialogue. By adding these capabilities in, we might diminish the coordination problems our agents currently fall prey to.

If optimal performance could be consistently generated for larger databases, then the field is wide open for directions to take this research. It would be fascinating to explore in what ways the internal structure of the database (analogous to a human's "knowledge base") affects the syntax that develops. What sort of structure is necessary to evolve human language-like syntax? What aspects of syntax are necessary and fundamental?

We have already discussed other directions of potentially fruitful and fascinating

future research. Modifying the structure of the agents so that identical twins do not speak to one another may reveal a great deal about population dynamics and the emergence of shared "languages." One could compare runs in which identical twins spoke to each other and runs in which individuals spoke to one or many distinct individuals in an attempt to piece apart the influence of genetic vs. cultural transmission. Explorations like this would nicely complement work already done (e.g. Oliphant, 1996) and shed further light on how shared languages are formed.

Another possibility, also touched on already, is to modify the input and output alphabets of agents in order to explore the possible emergence of reference. Additionally, we could add noise in order to examine the robustness of emergent communicative systems. Or we could give agents more working memory and pattern recognition ability in order to determine to what extent agent-internal capabilities are necessary for complex language. These are just some of the multiple avenues of future research that have been suggested by this work.

Finally, other genetic programming / computational simulation research on the nature of coordination would be valuable. Most of the conclusions we came to in this work rested on the difficulty of evolving coordination between two agents. Some work has been done in this field, but relatively little; and many of the findings suffer from generalizeability problems because they make a large number of assumptions in order to get their results (e.g. Batali, 1995, 1998; Burghardt & MacLennan, 1995). Thus, simulations exploring (for instance) the importance of information richness vs. quantity of data, or expressivity vs. interpretability, would be valuable in interpreting our results.

Many of the issues raised and tentatively explored here are still unanswered. It is my hope – and I think I have shown – that this paradigm may be effectively used to define the questions and discover answers that will shed more light on the path of human language evolution.