Syntactic Simulations: computational modeling of the evolution of syntax

Bronwyn Woods

Swarthmore College Department of Linguistics

April, 2007

Abstract

Two closely related questions in linguistics are how language evolved and to what extent it evolved as a genetically specified, innately coded cognitive module. The nativist approach to language evolution, which postulates the existence of such an innate mechanism, stands in contrast to the cultural approach which attributes the evolution of language to a complex interaction of the process of language transmission and the evolutionary pressures on languages themselves. With the advent of new investigative techniques such as computer simulation, many researchers have begun to explore more thoroughly this second approach. I present evidence for cultural evolution derived from computational simulations, exploring how this can inform the investigation into the genesis of language. I also explore the possibility of an account of language evolution informed by both nativist and cultural theories. [1]

1 Introduction: The Evolution of Language

The question of the origin of language is one that has, until recently, been largely passed over by the field of linguistics. Historically it has been an area plagued by rampant unfounded speculation followed by general disregard. In 1866 the Societe de Linguistique de Paris went as far as to ban scientific discussion of the subject (*Christiansen and Kirby* 2003). Similarly, the Linguistic Society of America didn't publish a full-length article on language evolution for the first 76 years of its existence (*Newmeyer* 2002). In recent years, however, the question has reemerged with great vigor. A large part of this resurgence is due to new tools and approaches, such as computational modeling and simulation.

So how do humans come to possess a system of communication orders of magnitude more complex than that of any other organism? The strong nativist approach would say that language comes from the "language organ", a genetically innate module that specifies grammar (*Cowie* 1999). Another approach is to claim that the complexities of language are largely a product of the social process of language transmission over many generations (*Kirby* 2002; *Kirby and Hurford* 2002; *Christiansen et al.* 2002; *Kirby and Christiansen* 2003). In this paper I will argue that the clearest picture of both synchronic and evolutionary aspects of language comes from a combination of these two views. To conclude, I will present a possible mechanism by which cultural and biological aspects of language can form a single coherent account of the evolution of language.

1.1 Challenges in studying the evolution of language

The body of knowledge about the evolution of the human species is large and growing. Why is it that we know so little about the origins of language, one of humanity's most defining features? It turns out that the problem of language evolution is very difficult. While it is inherently interdisciplinary, many of the techniques employed by the relevant fields of biology, archeology and linguistics are not applicable to the specific problem.

Language does not leave physical evidence like fossils or stone tools. Language change does leave evidence in languages themselves, and historical linguistics makes use of this to deduce information about languages no longer spoken. However, the rate of linguistic change means that the effective time limit on this method is on the order of 2000 years. Biologists often deduce information about traits of a species by comparing them to similar traits in related organisms, but language is a uniquely human ability, making this approach difficult. These factors together present a hurdle to the field of language evolution, as it lacks evidence with which to guide investigation (*Newmeyer* 2003; *Lieberman* 2000).

Another consideration that has traditionally discouraged linguists from pursuing the field is the incompatibility of evolutionary accounts of language development with the uniformitarian hypothesis. The uniformitarian hypothesis makes the claims that all languages are, in some sense, equal and that the basic nature of human language has remained constant throughout history (*Newmeyer* 2002). A strong coherence to uniformitarianism is at odds with the idea of the incremental development of language, and such an attitude has frequently made its way into research on language evolution (*Newmeyer* 2002, 2003). The uniformitarian hypothesis is increasingly controversial, but is also widely accepted in the linguistic community. As Edward Sapir said in 1921, "When it comes to linguistic form, Plato walks with the Macedonian swineherd, and Confucius with the head-hunting savage of Assam" (1921, 29). This view is easily defensible when considering languages of modern, or even recorded, times. However, when this view is extrapolated historically to the genesis of language, it necessitates the assumption that language sprang fully formed from non-existence to full complexity without any intermediate steps – an unlikely event. The challenge to uniformitarianism presented by evolutionary accounts of language development is thus a hurdle for the field of language evolution.

In the face of this difficulty, researchers have drawn on a wide array of approaches. Perhaps we can, in fact, learn something about human language from studying the linguistic abilities of other species. Under the view that ontology recapitulates phylogeny, language

acquisition may inform theories of language development. Advances in neuroscience and genetics allow us to make inroads into determining the foundation of human cognitive uniqueness. As I will show in this paper, another promising approach is the use of computer simulations. These techniques are helping to overcome the problem for language evolution posed by sparsity of evidence. Nevertheless, theories of language evolution are often difficult to support or revoke, largely based on the lack of empirical data.

It is unlikely that we will be able to prove conclusively the correctness of any particular account of how human language came into existence. This being the case, the question becomes: What value does studying language evolution have aside from providing fodder for debate? The answer, I believe, is that theories of language evolution can inform theories of synchronic linguistics. If our goal is to understand the incredibly complex system that is human language, we cannot hope to achieve it without having some idea of the processes by which that system developed. A theory of language evolution is not only an attempt to understand the past, but also a step toward understanding the present.

1.2   Approaches to language evolution

A theory of the evolution of language goes hand in hand with a theory of the cognitive and social architecture of language. As such, the question of language origin is closely tied to another controversial question in linguistics: to what degree is language innately coded? On this issue there are two general approaches: nativist and empiricist.

The nativist position was popularized by Chomsky in the late 1950s. Originally presented as an alternative to the dominant behaviorist approach of the time, Chomskyan nativism dictates that humans possess innate, domain specific knowledge of language and that this knowledge is encoded in a module known as Universal Grammar (UG) (*Cowie* 1999). The empiricist view challenges the existence of UG, and may also deny the innateness or domain specificity of linguistic knowledge. I will present here what I will call

the cultural approach – that many of the qualities of language that the nativist assumes must be dictated innately can in fact be derived from the dynamics of language transmission and learning over many generations. In the following sections I will present a more comprehensive exploration of the nativist and cultural approaches. Though here I present their differences, I will argue later that these approaches need not be mutually exclusive and that the best approach may be a theory encompassing elements of both theories.

1.2.1   Nativism

According to Fiona Cowie (1999), three of the claims made about language acquisition and proficiency by Chomskyan (or strong) nativism are the following:

(DS)   **Domain Specificity:** Learning a language requires that the learner's thoughts about language be constrained by principles specific to the linguistic domain.

(I)   **Innateness:** The constraints on learners' thoughts during language-learning are innately encoded.

(UG)   **Universal Grammar:** The constraints and principles specified in (DS) as being required for language-learning are to be identified with the principles characterized in the Universal Grammar.

(*Cowie* 1999, 176)

The last of these, (UG), is perhaps the most controversial. Universal Grammar, in the nativist framework, is a postulated innate module that specifies principles of language without any environmental input. UG severely restricts hypotheses a learner can make about the structure of language. It is partly this aspect of UG that makes it a convenient feature of a theory of the cognitive basis for language. The argument for UG relies heavily on the poverty of stimulus argument. Human languages are infinite – there is no limit to the number of grammatical sentences. It is not possible for a learner to hear all, or even

a small fraction, of the possible utterances. The poverty of stimulus argument maintains that there is insufficient information present in the input that language learners receive to uniquely determine a grammar unless the search space is severely restricted. According to nativist theories, this restriction is provided by UG. In reference to the grammatical rules which they hypothesize must be present in the innate language module, Lightfoot and Chomsky make the following claims:

> None of this is the result of training or even experience. These facts are known
> without training, without correction of error, without relevant experience.
>
> (*Chomsky* 1990, 640)
>
> Children do come to know these things, and this knowledge is indeed part of
> the output of the language acquisition process, but it is not part of the input,
> not part of the "evidence" for the emerging system.
>
> (*Lightfoot* 1989, 322-23)

The claim made by the poverty of stimulus argument is certainly true at least to some degree, though there can be some argument over its ramifications. The question of how much information is actually present in children's linguistic input is controversial. Supporters of the poverty of stimulus argument for nativism claim to be able to identify grammatical features that are impossible to derive from the input available. Their critics point out the lack of empirical evidence to show that the necessary information is actually missing (*Cowie* 1999). Nevertheless, the poverty of stimulus argument is powerful. For instance, it is a mathematical fact that infinitely recursive systems cannot be learned with only positive evidence (Gold's theorem). Human language is arguably such a system, and many researchers claim that children are not exposed to negative evidence as regards grammaticality. Komarov and Nowak (2002) explore this idea and frame a mathematical argument that language would be unlearnable were there no restriction on possible grammars. This is consistent with the idea of UG, though a counter-argument is that such

a restriction need not be UG in the nativist sense, and could instead derive from more general cognitive constraints on human learning. The nativist view is that such general restraints are insufficient to account for the complexity seen in natural language.

In general, nativist theories of language de-emphasize the role that learning plays in language development. Assuming an innate UG suggests that it would be more accurate to think of children as acquiring language rather than learning it. Anderson and Lightfoot (2002) make this point explicitly.

> The functional properties of our language organ develop along a regular maturational path, such that it seems appropriate to see our linguistic knowledge as 'growing' rather than being 'learned'.
>
> (*Anderson and Lightfoot* 2002, 220)
>
> Eventually, the growth of language in a child will be viewed as similar to the growth of hair: just as hair emerges at a particular point in development with a certain level of light, air and protein, so, too, a biologically regulated language organ necessarily emerges under exposure to a random speech community.
>
> (*Anderson and Lightfoot* 2002, 40)

Many nativist theories of linguistics tend to sideline discussion of language evolution. If language is genetically determined, then it is for geneticists or evolutionary biologists to explain the genesis of the mechanism. In addition, before one can explore the evolution of UG, one must have a clear idea of what it is, and this is far from the case (*Newmeyer* 2003). Chomsky's version of UG is perhaps dominant in the field, but he gives little encouragement to the study of its evolution. He is opposed to the idea that UG may have developed through natural selection:

> We know very little about what happens when 10^{10} neurons are crammed into something the size of a basketball with further conditions imposed by

the specific manner in which this system developed over time. It would be a serious error to suppose that all properties, or the interesting properties of the structures that evolved, can be "explained" in terms of natural selection.

(*Chomsky* 1975, 59)

Nativist theories are not inherently at odds with evolutionary accounts of the development of language. Pinker and Bloom, for instance, make a case for the development of language through natural selection (*Pinker and Bloom*, 1990; *Pinker*, 1994).

Evolutionary theory offers clear criteria for when a trait should be attributed to natural selection: complex design for some function, and the absence of alternative processes capable of explaining such complexity. Human language meets these criteria.

(*Pinker and Bloom* 1990, 707)

Misgivings about acknowledging language as a product of natural selection are frequently founded in a misunderstanding of evolutionary processes. Many arguments against natural selection for language give no more explanation than that arguments for natural selection are intuitively unlikely. For instance, one common argument is to state that there could have been no selective benefit for language in a half formed state, and that human language must have arisen in one giant, sudden, leap.

Another argument is that complexities of language such as subjacency constraints and the like can have no selective benefit and so could not have evolved through natural selection. Such conclusions are unfounded. Outside of the domain of language, the existence of unique, highly complex characteristics is readily and rightly attributed to natural selection. Pinker (1994) gives the example of the elephant's trunk. The elephant's trunk is quite complex and entirely unique, and yet unremarkable from an evolutionary point of

view. In the evolution of a complex attribute, there need not be selective benefit for each behavior or ability that attribute allows.

Nevertheless, there is a persistent belief that language is somehow qualitatively different from elephants' trunks. And, in fact, there is a factor that comes into play with language that does not influence elephants' appendages. Unlike most animal communication systems, human language is learned. This crucial quality introduces a dynamic which makes the relationship between human language and innate UG opaque. That is, the relationship of UG to the fully mature language of a human adult is mediated greatly by the linguistic environment and input. As Brighton et al. (2005a) argue, even were we to obtain a complete understanding of the human language faculty (UG), we would not be able to understand human language without also understanding the history and dynamics of language as a learned cultural phenomenon.

1.2.2   Empiricism and the cultural evolution of language

While the nativist approach relies on the genetic evolution of the human brain, other approaches emphasize the importance of cultural evolution. The evolutionary power of natural selection on the genome is almost universally recognized. The application of similar ideas to the evolution of cultural phenomenon is less widely accepted. Nevertheless, the processes and effects of natural selection need not be limited to the genetic domain.

Natural selection, in Darwinian terms, is survival of the fittest. In very simplistic terms, an entity that is more fit will survive and reproduce while entities which are less fit will fail to reproduce. Eventually the population will consist entirely of entities of the first type. Richard Dawkins suggests that this process may more accurately be seen as "survival of the stable" (1989). Evolution works on replicators – entities that persist by making accurate copies of themselves. For instance, genes are replicators. Though we might think of evolution as working on the level of organisms, organisms themselves are not replicators as they do not duplicate themselves precisely. Organisms are the vehicle

for the persistence of genes. Evolutionarily successful genes are those which persist by being accurately copied and transmitted from one vehicle to the next. Genes that cannot reliably pass from an organism to its offspring, or prevent that organism from having offspring, cannot persist. This powerful process is the foundation of genetic evolution, but the logic behind natural selection does not require that replicators be genes or that they be transmitted through sexual reproduction.

A great deal of research and writing explores the ideas of natural selection in the cultural domain. Understanding the elusive giant that is human culture is a daunting task, and approaches to cultural evolution vary with respect to how close a parallel they draw to genetic evolution. Dawkins proposes the idea of cultural replicators, which he calls memes, which are parallel to genetic replicators. Examples of memes, he says, are "tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches" (1989, 192). The literature on memetics proposes a variety of similar definitions of a meme, for instance:

- "Instructions for carrying out behavior, stored in brains (or other objects) and passed on by imitation." (*Blackmore*, 1999, 43)

- "A unit of cultural inheritance, hypothesized as analogous to the particulate gene, and as naturally selected by virtue of its 'phenotypic' consequences on its own survival and replication in the cultural environment." (*Dawkins*, 1999, 297)

The idea that culture can be broken down into these small units of selection is by no means universally accepted. Richerson and Boyd make the explicit argument that cultural variants are not replicators. They dispute Dawkin's idea of memes: "[meme] connotes a discrete, faithfully transmitted gene-like entity, and we have good reasons to believe that a lot of culturally transmitted information is neither discrete nor faithfully transmitted" (2005, 63). In contrast, Susan Blackmore (1999) argues that the meme is a powerful

and productive idea, and one appropriately applied to culture. Blackmore argues that memes are transmitted through humans' remarkable capacity for imitation. Nevertheless, she emphasizes that not every aspect of culture can or should be attributed to memes and that much of human behavior is more Skinnerian, derived from reinforcement. One particular aspect of memetic theory is particularly pervasive in this debate: what is the unit of culture that qualifies as a replicator?

Even considering the definitions presented above, it is far from clear what precisely constitutes a meme, let alone how these units come together to create what we see as culture. Nikolaus Ritt (2004) points out that smaller units are more likely to qualify as replicators. He also notes that it is more difficult to build complex cultural phenomenon out of these smaller units than larger potential memes such as religious beliefs and stories. He hypothesizes that this is why most authors discussing memetic theories of culture suggest larger memes. This problem is, of course, not unique to the cultural application of natural selection. When Darwin initially developed his theory of natural selection, the gene has not yet been discovered. We have since discovered the gene's status as the unit of genetic selection, but the fact is that a gene is not an isolated entity. A gene is merely a stretch of DNA whose boundaries are defined by observing its behavior. As Dawkins notes,

> The 'gene' was defined, not in a rigid all-or-none way, but as a unit of convenience, a length of chromosome with just sufficient copying-fidelity to serve as a viable unit of natural selection. If a single phrase of Beethoven's ninth symphony is sufficiently distinctive and memorable to be abstracted from the context of the whole symphony and used as the call-sign of a maddeningly intrusive European broadcasting station, then to that extent it deserves to be called one meme.
>
> (*Dawkins*, 1989, 195)

Nevertheless, identifying the potential unit of selection for cultural evolution is a daunting task. I will not establish a position on the issue of whether culture as a whole is made up of replicators, or even to what extant it is influenced by natural selection. That would be an investigation tangential to this paper. However, inasmuch as language is a cultural phenomenon, I will explore a subset of the cultural debate.

Dawkin's notion of a meme is especially well suited to language. Richerson and Boyd's assertion that cultural information is not faithfully transmitted does not hold with respect to language. In fact, it is exactly this property of faithful transmission, even in the face of what appears to be impossible odds, that lends support to nativist theories of language. The very existence of the field of linguistics is evidence for our intuitive belief that language is a system composed of discrete, formalizeable rules. These rules of syntax, phonology and more are transmitted accurately and with astonishing force and clarity from one generation to the next.

What exactly do I mean when I speak of linguistic memes, and in what ways are they parallel to genes? Before we can answer this question, we must define what is required for something to be an evolutionary replicator. The most common definition of a replicator (derived from *Dawkins*, 1989) stipulates three qualifications: fidelity, fecundity and longevity.

- Fidelity: A replicator must have high copying fidelity. Otherwise it cannot persist. However, for evolution to occur, replicators cannot be copied flawlessly in all cases.

- Fecundity: A replicator must produce copies of itself. The more copies a replicator produces, the more successful it is.

- Longevity: A replicator must persist for a period of time with its characteristics intact. Successful replicators are those which persist through duplication, and a certain lifespan is essential for replication to occur.

A gene possesses all of these characteristics. A gene has high copying fidelity. The process of DNA replication is very accurate, though with a small chance of mutation. Genes which are successful replicators have high fecundity. A gene persists within an organism throughout its life, giving the gene sufficient longevity to pass from the organism to its children. The question is whether there are units of language that also fulfill the criteria.

Nikolaus Ritt presents a theory of language as a collection of memes. He examines aspects of language such as phonemes, morphemes, morpheme clusters, categories and rules as potential replicators. He concludes that all these components are likely to deserve replicator status. For instance, he points out that phonemes are long-lived constituents of an individual's linguistic competence. Once acquired, they resist even conscious efforts to alter them during, for instance, second language acquisition. Larger units such as syntactic categories also seem promising as linguistic replicators. Different speakers and speech communities are remarkably consistent in their classification of words into syntactic categories such as 'noun' (*Ritt*, 2004, 146), demonstrating the copying fidelity of these categories. On the complex end of the spectrum, linguistic rules may also be considered replicators. Syntactic rules, for instance, have high copying fidelity between speakers. Similarly, they have high fecundity, being passed on to each new learner of a language. Like phonemes, syntactic rules also have sufficient longevity for replicator status, remaining in the linguistic competence of a native speaker throughout his life. Ritt (2004) discusses the replicator status of each of these components in depth, and I will refer the interested reader to his work. For the purposes of the current discussion I will assume the presence of linguistic replicators such as those just discussed.

If we can identify linguistic replicators, how does this lead to a theory of language evolution? How close a parallel can we draw to genetic replicators and genetic evolution? Just as a complex of genes has as its phenotypic representation an organism, a complex of linguistic memes has as its phenotype a particular linguistic competence. The vehicle for the persistence of genes is an organism. The vehicle for the persistence of linguistic

memes is the human brain. Successful linguistic memes must copy themselves from one human's brain to others. The parallel between genes and linguistic memes is striking, but by no means perfect. A major difference is the means of transmission. Whereas genes pass from an organism to its biological descendants, linguistic memes pass from one brain to another through the process of learning. The reproductive success of a gene is dependent on the reproductive success of the organism of which it is a part. This is not the case for linguistic memes. A human with more children may have the opportunity to provide linguistic input to more language learners, but this is incidental. A successful linguistic meme need not positively affect the fitness of the human who possesses it. It need only be able to pass through the process of linguistic production followed by language learning. The closest analogy may be to consider each linguistic competence as an organism whose environment is the human brain. Natural selection will lead to languages well adapted for their environment – well able to survive and replicate in the human brain.

Approaching language evolution from the perspective of cultural evolution is a promising approach. I will explore research that uses computational simulation to demonstrate that processes such as those described above can lead to the development of complex linguistic universals. However, before I get to that point, I will explore in general terms the use of computational simulations in the field of language evolution.

2 Evolutionary Simulations

Complex adaptive systems lend themselves well to computational simulation. Such systems are difficult to understand intuitively because of the large number of factors involved. Many evolutionary systems exhibit emergent behavior which is difficult to predict. Modeling these systems mathematically presents difficulties, since it is difficult to account for the behavior of an entire population or to allow for randomness. Computers allow convenient simulation of large populations and permit the inclusion of elements of randomness.

Gilbert and Doran present a variety of reasons for why simulation is useful for studying social phenomena (1994c, 20-22):

- The dynamics of social systems may involve a set of complicated interdependencies among a large number of components or units.

- The terms and examples of [simulations] must be explicitly expressed. Definitions must be complete and underlying assumptions must be coherent.

- Social experimentation using simulation models can explore a wide variety of conditions, variables, policies and underlying assumptions. The conduct of these experiments is less costly and more efficient than similar experimentation in the laboratory or the real world, provided that the phenomena under study are sufficiently well understood to be accessible to specification, and the limits of the procedure are made as explicit as the model itself.

These points demonstrate some of the great advantages of simulation, but also point out some of the limitations. As stated in the third point above, simulations are appropriate provided that the phenomena are "sufficiently well understood to be accessible to specification". Can simulations be useful in areas such as language evolution where we have very little information or understanding of the phenomena? I will argue that they can be, though the limitations of simulation must not be forgotten.

It is important to realize that though simulations are a vital tool for developing theories and hypotheses, they cannot provide definitive answers. That is, a simulation of an event cannot show how that event actually happened. A simulation demonstrates what occurs given a certain set of assumptions and conditions, but does not show that those assumptions and conditions are correct or complete. There are, however, two particularly useful results that can be gained from simulations: a simulation can verify or refute

causal assumptions of a theory, and a simulation can help to isolate or identify necessary conditions.

Theories of phenomena such as language evolution often make causal claims that are difficult to verify empirically. These claims are often supported by thought experiments or reasoning. Simulations can make this reasoning more explicit, verifiable and replicable. Successful simulation can show that an assumption is reasonable or that a process is feasible. Perhaps more importantly, unsuccessful simulation shows that a theory is incomplete or that its assumptions are flawed. Since it is relatively easy to alter the conditions of simulations, they can also help to isolate sufficient or necessary conditions for a particular result. Failure of a simulation to show the desired result may show that the theory is missing a vital component. Systematically altering the parameters of a successful simulation can help to determine which parameters are essential to the simulated process. Empirical investigation of the system being simulated can confirm or disprove the predictions made in such a way. Essentially, a simulation is a well-formed hypothesis that makes predictions that may be empirically investigated.

Though there are may types of computer simulation, evolutionary systems are well suited to multiagent simulation. As the name implies, multiagent simulations model a group of distinct agents, each acting according to a set of rules. There is no central control guiding the progress of the simulation. Each agent functions as a discrete entity. This sort of simulations allows for a design that focuses on the individual rather than global perspective. Rather than setting parameters for the global or macroscopic performance of a system, the experimenter instead can focus on the behavior of the microscopic, individual components of the system. This is appropriate for simulations of social and cultural phenomenon. It may be difficult or impossible to describe the system globally, and the individual is often the unit of interest. Also, many social and cultural systems demonstrate emergence, or complex, unpredictable behavior resulting from simple behavior on the individual level. Agent-based models can simulate emergent behavior. Multiagent

simulation is clearly a good choice for simulating the cultural evolution of language.

2.1   Simulation in the field of language evolution

The usefulness of simulations is recognized be researchers in many sections of the field of language evolution. I will briefly outline here some of the research being done on the evolution of phonetics and semantics using computational simulation. The research presented here is not meant to be comprehensive, but merely to demonstrate how widely simulation can be applied. For more detailed exploration of this research, I direct the reader to the papers cited.

Bart de Boer (2000; 2002) uses computer simulation to explore self-organization in vowel systems. He uses multi-agent simulations. Each agent has a vowel system, with vowels specified according to features such as formant frequencies. The agents in de Boer's simulations play imitation games. One agent produces a vowel. A second agent perceives this vowel and produces the vowel in its own internal system that is closest to the one perceived. The first agent perceives this new vowel and checks to see if it recognizes it as the same vowel originally produced. If it does, the game succeeds, if not, the game fails. Each agent receives feedback regarding the success of the game and adjusts its vowel inventory accordingly. Agents can adjust their inventories by taking actions such as adding, removing and merging vowels. At first, the vowel inventory of the community of agents is very spread out and non-uniform. After some time, however, the vowel system of the community becomes clustered. This clustering occurs differently in different runs of the simulation. However, de Boer shows that different types of vowel systems occur in simulation with frequencies very similar to those observed in extant human languages. These simulations lend support to the theory that aspects of language such as phonetic systems can emerge through multi-agent interactions and self-organization, rather than as a result of innate tendencies.

Language is more than a structured system of sounds or symbols. In order for language to have communicative function, those sounds and symbols must be associated with meaning. They must be grounded in a semiotic system. Luc Steels (1998; 2002) presents research that uses computational methods to explore the development of such symbol grounding. He presents two hypotheses for the development of a semiotic system – either new symbols become grounded in a Prue-existing system of meanings, or the symbols and meanings develop concurrently and influence each other's development. Steels uses robotic agents playing what he calls language games to explore this question. The agents in his simulations are embodied in cameras, allowing them to observe their environment, which consists of shapes of different colors and sizes arranged in two dimensions. The agents attempt to communicate through language games – one agent identifies an object using terms in its lexicon. A second agent 'hears' these terms and attempts to identify the object being referenced. Both agents receive feedback regarding the success of the communication and attempt to adjust their system of symbol-meaning pairings accordingly. Steels finds that shared systems of meanings can emerge through this interaction between agents, supporting the theory of co-development between symbols and meanings.

The research mentioned here demonstrates the pervasiveness of simulation techniques in the field of language evolution. The breadth of this research is extreme, and research in one area may inform that in the others. However, due to the depth of literature available in each subdiscipline, I constrain the focus of this paper to the area of syntactic evolution.

3 Definition of Terms

Terms such as language and grammar have very obvious meanings to many, and yet giving a exact definition can be difficult. To allow for precise discussion, I will discuss here the implications of each term as I will use it throughout this paper.

Fundamentally, language is a means of communication. In order for it to serve that

purpose, it must provide a mapping between a set of meanings and a set of utterances. This mapping is what I will refer to as language. It should be clear that this definition covers a lot of mappings that are nothing like natural human languages. For instance, a possible language would be one in which each individual meaning was expressed by a unique, unrelated string. Human languages, however, have many characteristics that the language just mentioned did not have. For instance, human languages are compositional, recursive and can express an infinite set of possible meanings. The human languages currently spoken in the world are only a subset of the languages that would be learnable by a human. The languages actually spoken in the world change, but they all stay within the set of languages that I will refer to as possible human languages. Figure 1 shows this relationship between possible languages, possible human languages, and extant human languages.
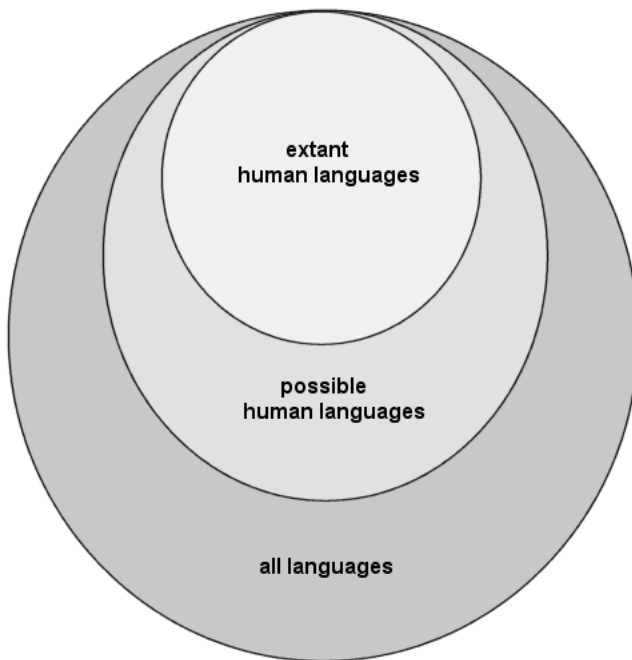

Fig. 1: The relationship between types of language. The set of extant human languages is a subset of the set of possible human languages. These sets are both subsets of the set of all possible languages, as defined here.

In order for agents to use language they must have some linguistic knowledge. Typi-

cally, this knowledge is considered to consist of two parts: the lexicon and the rules of the grammar. For the purposes of this paper, *grammar* refers to both of these components. I will refer to the component parts of a grammar as the lexical entries and rules.

4   The Iterated Learning Model

Language transmission and evolution is a complex system. It is difficult, on the basis of intuition, to understand the interaction between the learning process and the development of complex linguistic structures. I will propose that understanding this system is a problem well suited to computational modeling techniques, but in order to motivate the approaches I will explore, I will first present a high level description of how the process of language transmission might result in compositionality.

Compositionality is the property by which the meaning of an utterance is a function of the meaning of its parts. In other words, compositional languages can construct utterances out of smaller, meaning-bearing components. Compositionality is an absolute universal in human languages. That is, every human language exhibits this property (*Brighton and Kirby*, 2006). As such, compositionality is a prime candidate for genetic explanation. However, this does not disprove the possibility that compositionality could arise through cultural evolution.

In theory, language need not be compositional. Each meaning could have a unique, unanalyzable signal associated with it. Though this is not a type of language that we consider a possible human language, consider the situation where a group of agents communicates using such a system. Now consider the situation where this language is being transmitted from generation to generation within the population. In each generation, learners will be exposed to a subset of the meaning/signal pairs. Because learners are only exposed to a limited number of utterances, there will certainly be parts of the language that they do not hear. When these learners later attempt to produce strings for

meanings they haven't heard, they will by necessity invent new signals that are unlikely to correspond to what may have already been present in the language at some earlier time. Such a language cannot be stable. It will change quickly and unpredictably from generation to generation, and string/meaning correspondences will be frequently lost and reinvented. But now consider the situation where, by chance, some learner hears two signals that have both a similarity in form and in the meanings they convey. Assuming this learner has the capacity to generalize, he can form a rule for the production of a set of meanings based on the similarity he observed. Eventually, this learner will produce output for the next generation to learn from. Assuming he is equally likely to express any meaning, and remembering that the rule he deduced can apply to multiple meanings, he is likely to produce signals that were formed using the rule. A linguistic structure that applies to a broader range of meanings has a greater chance of being exhibited to language learners in every generation. The learners are therefore likely to make the same generalization and form the same rule. As this shows, compositional structures in language make good evolutionary replicators (*Hurford* 2000) and are likely to persist.

The sort of description given above is useful for gaining an intuition of how cultural evolution of language might proceed, but the system of language transmission is far too complex for thought experiments to be very helpful. The need for keeping track of behavior of agents and their grammars over many generations is a problem well suited to computer simulation. The prominent framework for exploring cultural evolution of language through simulation is the Iterated Learning Model (ILM) (*Kirby and Hurford* 2002). As a framework for exploring the dynamics of language transmission between generations, the ILM requires a set of simulated agents, organized into generations. The agents are prompted to communicate, with periodic replacements of one generation of agents with the next.

Though the implementation of the model varies widely over different experiments, the basic structure is the same. In each implementation, there is some space of meanings.

This can be finite or infinite and represented simply as a range of numbers (*Tonkes and Wiles* 2002) or symbolically with predicate logic (*Kirby and Hurford* 2002). Additionally, the agents in the simulation have some fundamental vocabulary for expressing utterances – the signal space. The basic procedure of the ILM can be summarized as the following sequence of actions:

1. Agent A, the speaker, generates a set of utterances corresponding to some set of meanings.

2. Agent B, the learner, hears the output from agent A and uses some learning algorithm to deduce a grammar.

3. Agent A is removed from the simulation, agent B becomes the speaker, and agent C is introduced as the new learner

4. Repeat steps 1-3 for a number of generations

The following sections will elaborate on each step of this process.

4.1   Step 1: Generation of utterances

When an agent, as the speaker, is asked to generate an utterance for some meaning m, the agent will, if possible, produce a signal s which expresses m according to its grammar. However, agents begin with no knowledge of language; initially, they have no grammar. It is therefore inevitable that at some point in the simulation (clearly in the first generation and possibly later), agents' grammars will be incapable of generating an utterance corresponding to m. In this case, agents invent a signal, drawing on the signal space of the experiment. This invention is often done randomly, though agents may be allowed to draw on information in their grammar about similar meanings (*Kirby* 2002).

4.2   Step 2: Learning

One of the most widely varied parameters among ILM implementations is the learning algorithm used by the agents to derive their grammar. In all cases, however, the learner is given the meaning m and the signal s as input to its algorithm. The fact that learners are provided with the intended meaning as well as the signal stands out as an unrealistic aspect of the model. Clearly a complete simulation of language transmission would require a more realistic model. As Kirby and Christiansen (2003) point out, simulations would ideally include contexts which would be public to all agents as well as meanings, which would be private. Much work has been done on this aspect of modeling language communities. I will briefly discuss one such experiment and why the results are encouraging for ILM simulations.

Steels and Kaplan (2002) present an experiment in which they demonstrate the ability of a group of visually grounded agents to develop a common lexicon. The problem they confront is the the Gavagai problem (presented by Quine, 1960): if a speaker says Gavagai while pointing to a rabbit in a field, how is the hearer to know whether Gavagai means *rabbit, animal, white* or any number of other aspects of the scene? Steels and Kaplan create a simulation in which agents are physically grounded – they exist in the form of cameras located in a simple environment. The agents play a language game by communicating about aspects of the scene and receiving feedback based on the success of communication. The agents are able to develop a common lexicon for describing the objects in their environment and are even able to adapt this lexicon as the environment changes.

Steels and Kaplan present their experiment as a first step in an area that is open for significant further study. However, their results suggest that it may be an acceptable simplification to provide internal meanings to learners in ILM simulations. If we assume that agents are able to develop a common lexicon, it makes sense to separate and ignore that step of learning and focus instead on the development of syntactic structure. Agents

are given a meaning m and a signal s. Their task is then to modify their grammar such that it can account for the mapping between m and s.

4.3   Step 3: Introduction of new agents

New agents are periodically introduced into the simulation as new learners. The start state of these agents is identical to that of the agents that started the simulation. There is no genetic evolution of the agents in that there is no way that existing agents can influence the starting state of new agents. Since all agents are identical, any change that is observed in the language of the population must derive from the dynamics of the transmission and learning process.

4.4   Step 4: Iteration

To observe the effects that language transmission might have on its form, steps 1-3 must be repeated many times. The number of repetitions depends heavily on the details of the simulation, but frequently the simulation is run until the language of the population stabilizes, or remains effectively the same between generations.

In the next section I will describe several experiments that have been done using the iterated learning model in various configurations. These explore issues such as compositionality, recursively and irregularity. In section 6 I will review the issue of what we can draw from these simulations and where we should proceed with caution. Following this I will explore how evidence from these experiments weighs into the debate about the evolution and innateness of language.

5   ILM Simulations and Results

As I have discussed above, a common argument against empiricist views of language evolution has been that no general theory of learning can account for the complexity seen

in human language. In this section I will discuss a number of simulations and their results. Each experiment seeks to demonstrate that some language universal may be explained through cultural evolution. As the experiments are similar in form, I will present the first in a greater level of detail to create a clear picture of how such experiments generally work. For the later experiments I will focus more closely on the relevant results and implications.

The progress of ILM simulations can be presented on several different levels. Examining the internal representations of the agents' grammars periodically as the simulation progresses allows for a low level exploration. This is useful for tracking the development of a feature in a way that is intuitively easy to follow. Examining the internal representations of the agents shows the progression of the language of the community in an explicit form. This is the type of representation that I will use to present the results of the following ILM simulations. Other possibilities are to examine the simulations at a more general level through the use of various visualization techniques, and graphing the progress of selected aspects of the simulation with respect to each other or to time. I will present this type of analysis in later sections as I explore how the simulations may be interpreted.

5.1   Compositionality

As mentioned before, compositionality is a common feature of every human language. This property is necessary for a finite language to have infinite expressibility, as human languages do. Being such a fundamental feature of human language, compositionality is among the targets of many simulations. I will discuss a simulation by Simon Kirby, the developer of the iterated learning model described above. The data and procedure I present here are a summarization of the information presented in Kirby (2002).

Kirby uses a simple form of predicate logic to define the meaning space for his simula-

tions. His meaning space consists of a set of predicates, such as *hates*, *loves* and *admires*, and a set of arguments such as *gavin*, *pete*, *heather* and *mary*. Meanings are presented to the agents as in (1).

(1)   likes(gavin, mary)

The agents use what is called a subsumption learning algorithm to deduce their grammar. When an agent hears a signal/meaning pair it first simply memorizes the mapping, checking for duplicates. So, an agent might incorporate (2) into its grammar. In this representations, the $S/$ indicates that what follows is a complete sentence. The string on the right hand side of the arrow is that which corresponds to the meaning on the left. In this example, English words are used for clarity, but in the actual simulation, the strings are generated randomly and do not correspond to any real human language.

(2)   S/likes(gavin, mary) → gavinlikesmary

Agents who can do nothing but memorize such mappings have no capacity to generalize. As explained in section 4, ability to generalize is necessary for the ILM to work. Kirby allows the agents to generalize by, when possible, subsuming two rules into one – the agents make the least general generalization that can account for both rules.

In one generation of the simulation, the speaker produces 50 utterances, with randomly chosen meanings. There are 100 possible meanings, which means that no learner can hear every meaning. As expected, at the beginning of the simulation, the languages that the agents learn are idiosyncratic and unstable. The grammar of the first learner in the simulation is shown in Figure 2. This grammar is almost entirely non-compositional. A chance similarity, however, has allowed for the creation of a single rule. The learner heard the string *flg* meaning *admires(mary, john)* and *fnv* meaning *admires(pete, john)*. From this input, the learner deduced a single rule: the meaning *admires(X, john)* can be expressed by *fY*, where Y is the string that represents person X in the meaning. The

learner has evidence that two objects, mary and pete, can take the place of X. Therefore, *mary* and *pete* become part of a class of words, called *A*, and that class is incorporated into a rule. In the notation presented here, the presence of a class name on the right hand side of the arrow indicates that any member of that class can be inserted at that point in the utterance. The variable paired with that class name, in this case x, also appears on the left hand side, indicating which portion of the meaning changes based on the selected word.

Generation 1	
S/detests(john, gavin) → nqb	S/detests(john, pete) → fu
S/hates(heather, mary) → b	S/detests(mary, gavin) → qqq
S/loves(mary, pete) → k	S/hates(gavin, john) → w
S/admires(john, mary) → u	S/admires(gavin, mary) → h
S/detests(pete, john) → ayj	S/hates(heather, gavin) → jrx
S/likes(heather, gavin) → g	S/hates(pete, mary) → r
S/loves(pete, john) → vcs	S/likes(gavin, pete) → qi
S/likes(john, pete) → os	S/admires(gavin, john) → j
S/loves(heather, gavin) ?→ e	S/detests(john, mary) → f
S/likes(mary, gavin) → ke	S/detests(heather, pete) → wkm
S/admires(john, gavin) → hy	S/detests(pete, mary) → sm
S/admires(pete, heather) → dx	S/loves(heather, john) → I
S/admires(gavin, pete) → x	S/hates(john, heather) → xf
S/likes(heather, mary) → d	S/loves(mary, gavin) → bni
S/detests(heather, john) → m	S/admires(gavin, heather) → yn
S/hates(heather, pete) ?→ yya	
	S/admires(x, john) → f A/x
	A/mary → lg
	A/pete → nv

Fig. 2: The grammar at generation 1. For the most part, the grammar consists of memorized strings for each meaning, but a chance similarity has resulted in the formation of one rule. The similarity of *flg* meaning *admires(mary, john)* and *fnv* meaning *admires(pete, john)* has resulted in the creation of the category *A* for *mary* and *pete*. Figure taken from (*Kirby* 2002)

In the first generation the grammar of the learner is almost exclusively non-compositional, but this quickly starts to change. As the simulation progresses, agents begin to make more generalizations and have grammars that are somewhat compositional and productive. Fig-

ure 3 shows the grammar of an agent at generation 14. It is clear that this agent has made

many generalizations from the data it heard. However, its grammar is still quite unsys-

tematic. We can see that the agent has created six word classes, A through F. Though

nouns and verbs appear in separate classes, many of the predicates and objects occur in

multiple classes and it is unclear how the members of each class are related. Also, the

agent still has many meanings which is has simple memorized which it cannot construct

compositionally.

Generation 14	
S/hates(pete, john) → a	A/gavin → b
S/p(john, x) → A/x B/p	A/mary → ni
S/likes(gavin, pete) → lw	A/john → y
S/hates(heather, john) → z	A/heather → x
S/p(x, mary) → l B/p A/x	A/pete → h
S/p(pete, gavin) → dx, E/p	B/loves → y
S/admires(heather, mary) → hhi	B/hates → n
S/likes(mary, pete) → h	B/likes → z
S/p(x, heather) → F/p A/x	B/detests → m
S/hates(gavin, mary) ?→ rw	C/pete → t
S/detests(gavin, john) → vow	C/gavin → yo
S/hates(heather, gavin) → s	C/heather → gpi
S/detests(x, y) → D/x A/y	C/john → d
S/hates(mary, x) → D/x rs	D/heather → kr
S/hates(heather, pete) → kw	D/gavin → q
S/likes(heather, gavin) → ufy	E/hates → c
S/loves(x, y) → A/y A/x	E/detests → rp
S/likes(x, y) → l C/y A/x	F/detests → r
S/p(x,y) → C/x B/p n A/y	F/admires → ud

Fig. 3: The grammar at generation 14. This grammar is somewhat compositional, but
still shows much of the arbitrariness of earlier generations. We can see here that
some rules have been created that abstract over several predicates. In these cases,
the predicate position is represented by p and the category from which to choose
a word is represented (as before) by p/X. Figure taken from (*Kirby* 2002)

Figure 4 shows the state of the language at generation 112. We can see that there

has been a dramatic reduction in the size of the agent's grammar. The agent's grammar

no longer contains any memorized complete meanings, but rather has two rules by which

to construct all utterances. The number of classes has also decreased since generation
14. There is now only one class of verbs, though the nouns are still divided between two
classes.

Generation 112	
S/p(x,y) → C/y B/p n A/x	B/loves → xfh
S/p(x,y) → A/y C/x B/p n	B/hates → n
	B/admires → srw
A/gavin → b	B/likes → z
A/mary → ni	B/detests → m
A/pete → re	C/heather → fkm
A/john → y	C/pete → t
	C/mary → ns
	C/gavin → yo
	C/john → d

Fig. 4: The grammar at generation 112. This grammar is very compositional and relatively
        stable. Figure taken from (*Kirby* 2002)

The language of generation 112 is, as Kirby emphasizes, highly regular. This grammar
is stable for thousands of generations, until eventually a series of changes results, at
generation 7944, in the grammar shown in Figure 5. We can see that in this grammar
has only one rule for producing sentences (in OSV order) and two classes of words (which
correspond to nouns and verbs). This grammar remains entirely stable.

Generation 7944	
S/p(x,y) → v A/y g A/x B/p n	
A/gavin → gw	B/hates → n
A/john → gbb	B/loves → c
A/pete → k	B/detests → m
A/heather → gyt	B/admires → srw
A/mary → pd	B/likes → z

Fig. 5: The grammar at generation 7944. This grammar is entirely compositional, with
        only one sentence creation rule and two classes of words. Figure taken from (*Kirby*
        2002)

Kirby's experiment shows the development of compositional structure in a population
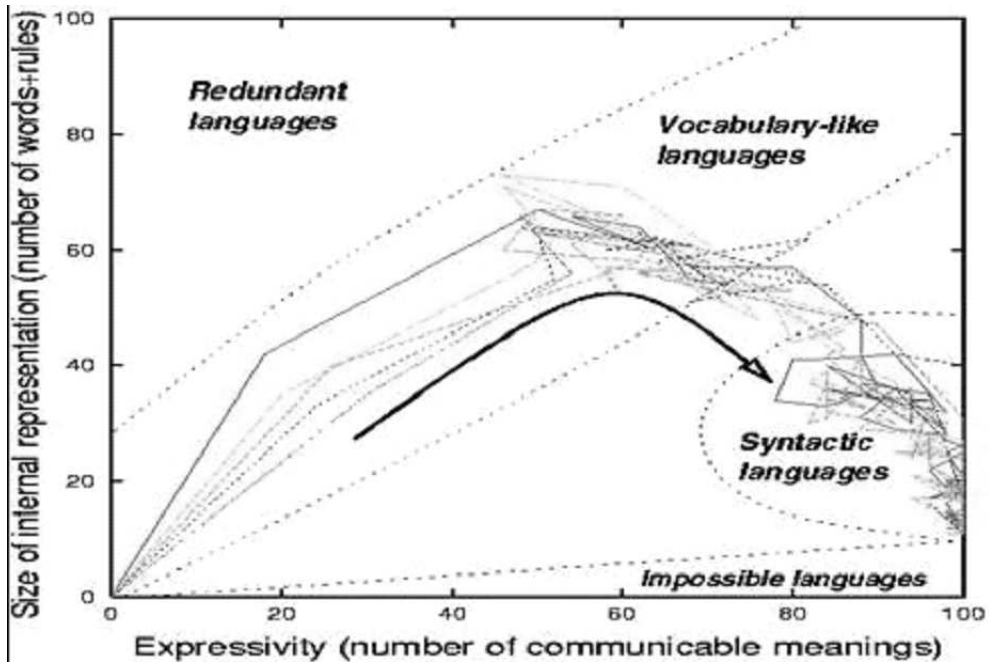
Fig. 6: A graph of the size versus expressivity of grammars in the simulation. Figure
        taken from (*Kirby* 2002)

with no genetic evolution. Clearly, the agents are capable of learning non-composition
grammars, since for generations they do. The agents introduced at the end of the simu-
lation are identical to those introduced at the beginning, but the language they learn is
qualitatively different. This difference results from the dynamics of language transmission
combined with properties of the agents' learning algorithm.

Examining the grammars of the agents helps to gain an understanding of the develop-
ment of the agents' language on a microscopic level. A more macroscopic understanding
is better obtained by examining a graph of the agents' grammars over several simulation
runs. This graph is shown in figure 6. One thing to note is that the regular, compositional
grammars are much smaller than the primarily lexical grammars early in the simulation.
We can see that the languages in figure 6 make a sudden shift from being vocabulary-like
to being syntactic as they increase to maximum expressivity. Kirby (*Kirby* 2002; *Kirby
and Hurford* 2002) hypothesizes that the driving force behind this shift is the *learning
bottleneck* (discussed further in section 7). As mentioned before, no learner in the simula-

tion can hear all possible meanings. This creates a bottleneck in the learning process since any language, if it is to remain stable between generations, must be able to be learned from only a fraction of the possible meaning/utterance pairs it specifies. Because compositional languages are able to pass through this bottleneck and still be learned, they are more evolutionarily stable than their non-compositional counterparts. It is therefore not surprising that compositional languages replace the earlier lexical languages.

Many experimenters have replicated the results of this experiment, often with the goal of isolating necessary conditions. I will discuss this research in sections 7 and 8. I will discuss one additional experiment here, to make a crucial point about cultural evolution: just as in genetic evolution, there is a distinction to be made between the genotype and phenotype of linguistic agents. The linguistic genotype of an agent is its internal representation of language. This is what we have been examining in the discussion above. The linguistic phenotype of an agent is the set of utterances that its grammar produces. Just as in genetic evolution, natural selection can only occur based on phenotypic consequences of the genotypic makeup. That is, linguistic memes can only be selected for or against based on the effect they have on the exterior, phenotypic, language. The representation internal to an individual agent is only important insofar as it can produce the evolutionarily successful phenotype.

An experiment by Hurford (2000) highlights this point. As in Kirby's simulation, Hurford's agents have the capacity to generalize. But Hurford limits the probability that the agents will employ this ability to 25%. That is, three quarters of the time, the agents make no attempt to generalize what they learn and simply memorize string-meaning pairs. Nevertheless, the language that develops by the end of the simulation is entirely compositional. Since learners always memorize a large proportion of the strings they hear, the internal grammars of the agents still have, in addition to compositional rules, many individual string-meaning mappings. But crucially, these memorized mappings comply with the same compositional grammar. That is, the external language of the

community is entirely compositional despite the individuality of each agent's grammar. This servers to emphasize the point that we are interested in the cultural evolution of languages themselves, not of any characteristics of the individual agents. A linguistic meme is successful if it persists. The grammars of linguistic agents are the vehicle for persistence of the linguistic meme, but these grammars need not be uniform, efficient, or identical to support the meme's survival.

These experiments are two of many that demonstrate the emergence of compositional structure in the ILM framework. Other experiment modify the details of the simulation. Some represent the meaning space using n-dimensional vectors rather than structured predicate logic (*Tonkes and Wiles* 2002). Many use different learning algorithms, including Minimum Description Length modeling (*Brighton and Kirby* 2001a,b), neural network training (*Tonkes and Wiles* 2002), and other connectionist techniques (*Smith* 2003). The results are robust to these changes in the modeling parameters. In section 8 I will explore how looking for similarities in the biases of these varied learning algorithms can help to inform our understanding of the forces driving cultural evolution.

5.2   Recursivity

Human languages are infinitely expressive. Recursivity, with compositionality, allows this without requiring language users to also possess an infinitely large grammar. Like compositionality, recursivity is universal among human languages. A simple extension to the experiment described above allows us to explore how this feature can also emerge from cultural evolution.

Kirby (2002) extends the experiment describes above to allow for recursivity by introducing predicates into the meaning space which can take other predicates as arguments. For instance, a possible meaning in this simulation is (3), meaning *john believes that heather praises mary.*

(3)   S/believes(john, praises(heather, mary))

Meanings of the type included in the last experiment, with only a single predicate, are termed degree-0 meaning. Meanings containing a single embedded predicate are degree-1, while meanings containing a doubly embedded predicate are degree-2 In this simulation, the speakers are now prompted to produce meanings of degree-0, as well as degree-1 and degree-2.

As before, the grammars of the initial generations are quite idiosyncratic and unstable. For instance, the grammar of the first generation contains over 100 rules, mostly simple mappings from complete meanings to strings. However, as before, after a number of generations the grammar dramatically decreases in size. Figure 7 shows the grammar of an agent in generation 115 of this simulation.

Generation 115	
S/p(x,q) → S/q C/p gp B/x d	B/pete → f
S/p(x,p) → stlw A/p B/y B/x	B/heather → v
	B/gavin → eks
A/loves → r	B/mary → k
A/admires → i	B/john → a
A/hates → wja	C/says → fdbtl
A/detests → w	C/decides → b
A/likes → btl	C/believes → o
	C/knows → z
	C/thinks → t

Fig. 7: The grammar at generation 115. The two sentence rules allow for recursiveness, and the first rule allows for complete sentences to be included within it. Figure taken from (*Kirby* 2002)

This grammar has two sentence rules. One is similar to the rule developed in the first simulation. The other, crucially, has a category for subordinating verbs such as *say* and *believe* and allows for an embedded *S*. This makes the grammar recursive, and infinitely expressive.

Though the agents in the experiment were only ever prompted to produce meanings of at most degree-2, the grammar that results is capable of expressing meanings of infinite

degree. This is also the case in natural languages. Human languages are infinitely recursive (theoretically, the sentence *"He says that she says that he says that she says... something"* is grammatical with an infinite number of repetitions of the he says/she says construction. Of course, an infinite sentence could never actually be uttered). However, an individual learner only ever hears sentences of some finite degree of recursion.

5.3   Irregularity

The experiments presented above attempt to isolate potential causes of regular patterns considered to be linguistic universals. However, the fact is that these patterns are not entirely universal throughout language. Languages are filled with irregularity. For instance, the generally compositional system of the English past tense marker is disrupted by irregular verbs such as *go* and *see*. The simulations in the previous sections do not allow for irregularity. Kirby and Hurford (2002) present a simulation that attempts to demonstrate the irregularity can be present in stable culturally evolved languages. The following description is based on their paper.

Kirby and Hurford make three major modifications to the structure of the experiments as described so far. Firstly, in situations where an agent's grammar allows for multiple strings to express the same meaning, the agent will always produce the shorter string. Secondly, there is a small chance that utterances will not be produced correctly and a character will be dropped from the utterance. Lastly, and perhaps most importantly, the probability with which meanings are expressed is not uniform. Some meanings are more likely to be expressed, and are therefore expressed with a greater frequency.

The grammar of the agents in this simulation follows a familiar pattern. For the first generations, the grammars of individual agents are idiosyncratic and not uniform. Eventually, though, a stable grammar develops. However, in this case, some of the meanings are consistently expressed irregularly. That is, though they do not follow the composi-

	a_0	a_1	a_2	a_3	a_4
b_0	**g**	**s**	kf	jf	uhlf
b_1	**y**	jgi	ki	ji	uhli
b_2	yq	jgq	kq	jq	uhlq
b_3	ybq	jgbq	kbq	jbq	uhlbq
b_4	yuqeg	jguqeg	kuqeg	juqeg	uhluqeg

Fig. 8: Results from Kirby and Hurford (2002) showing the emergence of irregularity in an ILM model. The irregular forms, shown in bold, express the most frequent meanings. Figure taken from (*Kirby and Hurford* 2002)

tional pattern, the irregular string-meaning pairs are stable between generations. The irregular strings correspond to the meanings that are expressed with greater frequency. This is unsurprising since these are the meanings that are likely to be expressed frequently enough to be heard reliably by learners in every generation. This pattern also corresponds to the pattern of irregulars in natural language, which are often the most frequent words (for instance, *go* and *to be* in English).

Figure 8 shows the results of the Kirby and Hurford simulation. In this case, the meanings consist of a combination of two components (Kirby and Hurford suggest that these two components could correspond, for instance, to shape and color or action and tense). One component is represented by the range a_0 to a_4 while the other is represented by b_0 to b_4. The meaning components with lower numbered subscripts are expressed with greater frequency. We can see in the figure that the irregular forms are indeed the most frequent. They are also the shortest strings, which is, anecdotally, also similar to natural language.

The results of this simulation lend support to the idea that cultural evolution can lead to many features of human language. This is especially true since the pattern of regularity in this experiment parallels that seen in natural language. It seems likely that the emergence of irregularity is linked to the uneven frequency with which meanings were expressed. However, this is an area that calls for further investigation, as the other changes made to the simulation could easily have also had an effect. I will discuss the care that

must be taken in interpreting simulations in section 6.

6   Interpreting Simulation Results

ILM simulations show very appealing results, but simulation has limits that should be considered along with the results. Also, given the specificity and simplification of the ILM simulations, it is necessary to consider the applicability of simulation results to natural language. I will discuss in this section several facets of the previously presented simulations and consider some of the possible limitations on data obtained through simulation. Simulations are greatly idealized and simplified, which means that we should carefully examine their results, but even with these caveats, data from simulations are valuable.

6.1   Relationship of Simulated Language to Natural Language

The languages in the simulations are very simplified. Can we draw parallels between these languages and natural human languages? Of course, the features of compositionality, recursivity and irregularity are the most relevant parallels. However, there are other features of the simulated languages which show interesting similarities to natural language.

Consider the grammar that resulted in the example experiment on compositionality (figure 6). The single rule in this grammar is: $S/p(x,y) \rightarrow v\ A/y\ g\ A/x\ B/p\ n$. This rule is compositional, but there are components that seem to be entirely functional without bearing meaning – the v, g, and n specified as necessary in the rule. In more complex systems, such as that developed in the recursivity experiment, these "meaningless" components can serve to distinguish and identify different rules. It would be relatively simple to determine how these components became fossilized during the simulation, but this is not the interesting question. It is more interesting to consider whether there are parallel components in natural language. It seems reasonable that there might be. Extant human languages contain grammatical constructions that carry meaning about the form of the

utterance rather than any referent of the words themselves.

In many ways, the languages of the simulations are parallel to natural language. On the most basic level, they are a pairing between meaning and form. Of course, much of the detail of how that meaning is established in glossed over in the assumption that agents have access to both during learning. Crucially, the simulated languages show patterns of compositionality, recursiveness and irregularity. Other smaller patterns of similarity can be noted: in addition to the parallel mentioned above, parallels can be seen in some experiments in the development of competing systems of regularity. These additional, smaller, similarities which are not the specific target of experiments are tempting indications of the validity of simulation results for natural language exploration. However, with the amount of simplification present in simulation, it is difficult to know whether these similarities are meaningful or merely figments of the construction of the experiments.

6.2   Problem of Overspecification

A frequent concern when using computational simulation is that the simulations may be overspecified. The fear is that the simulation may be unintentionally pre-programmed to show the desired behavior, or that the simulated agents may be able to "cheat" in ways not intended by the simulator. There are several ways of attempting to identify and eliminate overspecification. One is to work to isolate and identify the particular properties of the simulation that lead to the desired outcome. I discuss this technique shortly, both in general and for Kirby's ILM simulations specifically.

In situations involving emergent behavior, another technique for analyzing the results is to compare the results to the general expected behavior in emergent situations. For instance, if exploring the emergence of a particular characteristic within a population, we might expect to see some sort of s-curve. A conflicting pattern of data might indicate that an undesirable factor or process was influencing the simulation. Of course, such a

comparison is only useful when we can predict with some confidence what sort of pattern we expect out data to follow. In the case of ILM simulations, this does not seem to be the case.

Though the development of syntax was surely an emergent phenomenon, ILM simulations themselves do not show typical emergent behavior. ILM simulations are inherently vertical; that is, there are only two (or at most several) agents "alive" at any given time. We are not interested in the emergence of a characteristic within a population of agents. In some ways, it would be more accurate to view the set of meanings in the simulated language as the "population". The simulations presented in this paper explore the emergence of characteristics in languages themselves. Analyzing the data with respect to an assumed model of emergence would require considering the simulations and their results from a somewhat different perspective. This might be an interesting direction of inquiry, but has not yet been investigated.

Nevertheless, we must consider the possibility that the ILM simulations presented here were overspecified in some way. Tonkes and Wiles (2002), for instance, point out that the learning algorithm used by Kirby in his experiments was originally developed for use in Natural Language Processing, suggesting that perhaps it was unduly specialized for the task. However, Tonkes and Wiles also present results from their own experiment indicating that similar results can be obtained with a less language-specific learning algorithm.

In general, the best evidence that the results of the simulations described here should not be discounted due to such concerns is that they have been replicated by many experimenters using a variety of formalisms, learning algorithms and experimental setups (*Brighton et al.*, 2005b; *Tonkes and Wiles*, 2002; *Christiansen and Ellefson*, 2002). There is also an intuitive argument to be made for the validity of the ILM simulations. Agents, identical in structure, who are introduced late in the simulation learn a qualitatively different language than those introduced earlier. Some change occurs in the language of the community. Isolating whatever causes that change is essential to making full use of the

simulation results.

6.3   Isolating Sufficient Conditions

Simulations that show desirable results are interesting, but the most useful information to be extracted from successful simulations is the identification of the crucial conditions. As mentioned in section 2, one of the advantages of simulations is that they can be a good tool for identifying the assumptions and conditions that are essential to a theory. One way of doing this is to systematically vary the parameters of the simulation and observe the effect that this has on the results. When changing a parameter causes the simulation to fail, it is an indication of the importance of that feature.

Several people have explored Kirby's ILM simulation in this way. Kirby comes to the conclusion that the learning bottleneck is crucial to cultural evolution. Research done by Brighton, Smith and Kirby (2005b) supports this claim. On the other hand, Tonkes and Wiles (2002) challenge the preeminence of the learning bottleneck. This issue is discussed further in section 7.

Though there is no genetic evolution in ILM models, the agents do possess a learning algorithm. The structure of this algorithm clearly will have an effect on the simulation process and the resulting languages. What are the necessary attributes that a learning algorithm must possess for language-like communication systems to culturally evolve? Is stipulating necessary, possibly unique, components of the learning algorithm (an innate characteristic) different than stipulating the existence of UG? These issues are explored in section 8.

As mentioned above, simulations are greatly simplified from reality. This issue has been discussed with regards to lexicon formation in section 4. Though the simplification of simulations may be detrimental to their applicability, it is also beneficial. One advantage of studying simulations is that they allow us to isolate and examine aspects of a process

in a way that is not possible in the real world. In the following sections, I will discuss attempts to isolate the crucial components to the success of the ILM simulations presented previously.

7   The Learning Bottleneck

Kirby emphasizes the importance of the learning bottleneck in the process of cultural evolution of compositional languages. In natural language, which is both compositional and recursive, there is a bottleneck by necessity. A learner cannot be exposed to all meaning signal pairs of such as infinite language. In simulations of language evolution, the meaning space is often greatly restricted, with the result that a learning bottleneck must be artificially imposed by restricting the number of utterances to which a single learner is exposed. Nevertheless, the bottleneck seems to be crucial in stimulating the evolution of compositional languages. In the presence of a learning bottleneck, compositional languages are comparatively more evolutionarily successful because only they are able to pass through the learning bottleneck intact.

It is reasonable to assume that a learning bottleneck would have been in play during the evolution of human language. For there to be no bottleneck, all learners must be exposed to all meaning signal pairs possible in the language of the community. Even if this language were finite, practical considerations suggest that as soon as the community of language users became sufficiently large or diverse, the number of expressible meanings would increase to the point where it would be unlikely for each member of the community to be exposed to them all. At this point, this bottleneck would create an environment in which compositional languages could reap selective advantage.

Though it is easy to intuitively identify the learning bottleneck as a crucial factor for the evolution of compositionality, the simulations presented so far do not explore this assumption. Nevertheless, simulation is an ideal forum for exploring such assumptions

about necessary conditions. Brighton et al. (2005b) present a variation on Kirby's ILM simulations (described in section 5) which aims to isolate and identify the necessary conditions for the emergence of compositionality. I will describe their experimental framework here and discuss their results as relates to the importance of the learning bottleneck. In section 8 I will discuss their results as they relate to necessary learning biases of individual agents.

Brighton et al. use a more abstract representation for the language in their simulation than that of the simulations presented above. Meanings are represented as vectors in a meaning space defined by two parameters: F and V. F represents the number of features that each meaning has and V represents the number of values that each feature can take on. To demonstrate this, consider the meaning space M defined by F=2 and V=2:

M = { (1, 1), (1, 2), (2, 1), (2, 2) }

Signals are represented by strings of letters. The signal space is defined by the alphabet (Σ) and the maximum length of a string (l_{max}). As a concrete example, consider the signal space defined by $\Sigma = \{a, b\}$ and $l_{max} = 2$.

S = {a, b, aa, ab, ba, bb}

Unlike the language described in section 5, languages which use these meaning and signal spaces cannot be recursive. However, this model of language does allow for compositional languages. Brighton et al. give the following example, which they call $L_{compostional}$.

$L_{compostional}$ = { ⟨(1,1), ac⟩, ⟨(1,2), ad⟩, ⟨(2,1), bc⟩, ⟨(2,2), bd⟩ }

Using this model of language, Brighton et al. conduct an ILM simulation similar in structure to those already discussed. In each generation, the 'adult' agent agents it prompted to produce a set of meaning-signal pairs which is then given to the learner

agent as training data. After this, the adult agent is removed from the simulation, the learner becomes the adult and a new learner is introduced.

The learning algorithm of this simulation is structured to allow manipulation to explore various learning biases. The grammar of each agent is represented as a matrix, A. This matrix contains the strengths of associations between both partial and complete meanings and signals. The rows of the matrix represent all the possible components of the meanings in M. A component of a meaning is a vector that may be underspecified. That is, for each feature of the meaning, the component vector can contain either the value of that feature for that meaning, or an unspecified wild card value. The columns of the matrix represent all the possible components of the signals in S. A signal component, parallel to a meaning component, is a possibly underspecified string. The entries of this matrix represent the association between each component meaning and component signal. Initially, all entries in the matrix are set to zero. Learning is the process of adjusting these associations based on the input, according to a set of parameters.

When an agent is given a meaning-signal pair from which to learn, it adjusts each entry in its matrix according to specified parameters where C_m and C_s are the sets of component meanings and signals specified by the meaning-signal pair.

$$\Delta a_{ij} = \begin{cases} \alpha \text{ if } i \in C_m \text{ and } j \in C_s \\ \beta \text{ if } i \in C_m \text{ and } j \notin C_s \\ \gamma \text{ if } i \notin C_m \text{ and } j \in C_s \\ \delta \text{ if } i \notin C_m \text{ and } j \notin C_s \end{cases}$$

When the agents are prompted to produce a signal to correspond to a given meaning, they find the sets of meaning components which are analysis of the given meaning. They evaluate these sets of meaning components with respect to the possible signal strings on the basis of connection strength. The signal they produce is that with the highest score.

Varying the values for α, β, γ and δ in the learning algorithm alters the learning biases

of the agents. The results of such variation are described in the next section. For the purpose of demonstrating the importance of the learning bottleneck, Brighton et al. set these parameters to $\alpha = 1, \beta = -1, \gamma = -1, \delta = 0$.

To explore the impact of the learning bottleneck, Brighton et al. ran their ILM simulation under two conditions. In the first, learners were exposed to all meanings in the meaning space. In this condition, there was no bottleneck. In the second condition, learners were only exposed to a subset of meanings. The graph in figure 9 shows the results comparison of the results in the bottleneck and no bottleneck conditions.
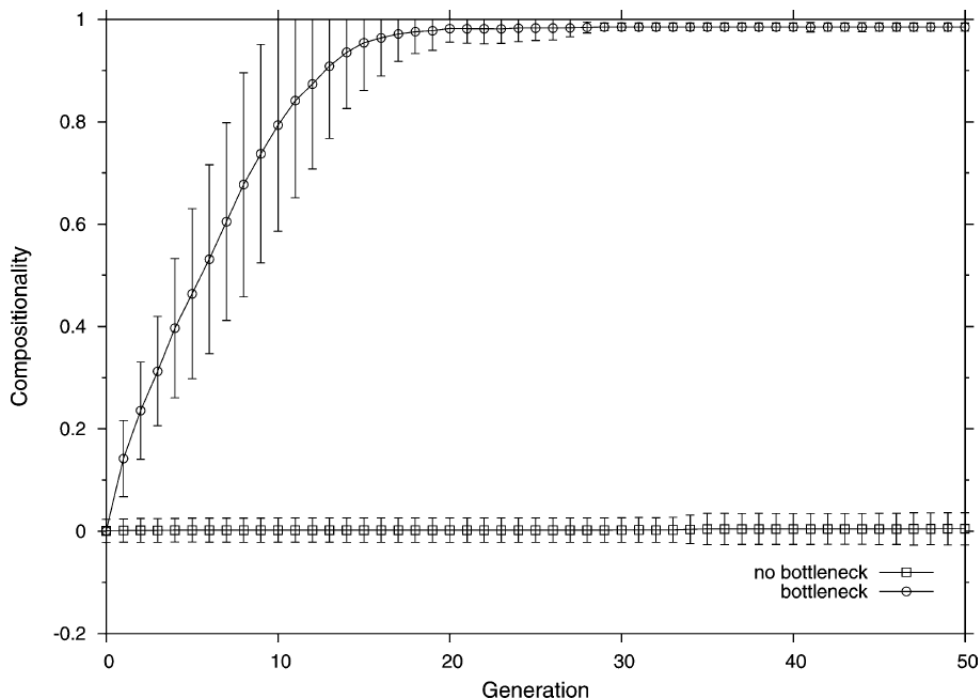

Fig. 9: A comparison of the results of a simulation with and without a bottleneck on transmission. Compositional language only emerges when a bottleneck is present. Figure taken from (*Brighton et al.* 2005b).

In this experiment, compositionality is evaluated as a value between 0 and 1. For details on the calculation of this value, reference Brighton et al. (2005b). Figure 10 shows fragments of noncompositional and compositional grammars from a run of Brighton's simulation.

Meaning	Signal in initial language	Signal in final language
(3,3,3)	db	def
(3,3,2)	cfc	ded
(3,2,3)	cfh	daf
(3,2,2)	deg	dad
(2,3,3)	fbg	fef
(2,3,2)	gae	fed
(2,2,3)	chg	faf
(2,2,2)	cbc	fad
(1,3,3)	cgg	gef

Fig. 10: Fragments of a noncompositional and a compositional grammar. The former is the initial grammar of the agents at the beginning of the simulation. The latter is the final grammar of the agents at the end of the simulation. The noncompositional grammar has a compositionality score of -0.025 while the compositional grammar has a compositionality score of 0.991. Figure taken from (*Brighton et al.* 2005b)

We see in figure 9 that without a learning bottleneck, compositional language does not arise. When a learning bottleneck is imposed, however, compositional language is able to evolve through the process of language transmission. The results of this experiment indicate that the learning bottleneck is indeed important for the cultural evolution of compositionality. Brighton et al. do not specify the severity of the bottleneck they impose. It is likely that there is a crucial range of bottleneck sizes in order for the bottleneck to simulate the evolution of compositionality. A bottleneck that is too restrictive will not allow any language to be reliably transmitted. A bottleneck that is too nonrestrictive will not provide sufficient advantage to compositional languages and will likely result in language similar to those which arise in the no bottleneck condition.

8   Learning Biases

Grammatical universals exist, but I want to suggest that their existence does not imply that they are prefigured in the brain like frozen evolutionary accidents [...] they have emerged spontaneously and independently in each evolv-

ing language, in response to universal biases in the selection processes affecting

language transmission.

(*Deacon* 1997, 115-116)

The type of learning algorithm used in ILM simulations is very influential. The learning biases of the agents in the simulations correspond to biases that we assume human cognition must have for human language to develop. In some ways we can draw parallels between these learning biases and the UG postulated by strong nativist theories. UG is an explanation for regularities and complexities of human language. Learning biases are an alternative explanation for the same phenomenon.

Implied in the nativist proposal of UG is the idea that were we to know the structure of the cognitive UG module we would necessarily know the structure of human languages. The problem of how this might be accomplished is what Kirby, Smith and Brighton (2004) refer to as the problem of linkage. If instead of UG, we hypothesize a learning bias for each agent – what Kirby and Christiansen (2003) call a Universal Bias – we can look to cultural evolution to solve the problem of linkage. We have seen that the learning biases in simulations lead to the development of language universals through cultural evolution. What aspects of the learning bias are important for this to occur?

Brighton et al. (2005b) explore this question by modifying the parameters of the learning algorithm described in the previous section. They find that two things are crucial: an ability of generalize, and a prejudice toward one-to-one mappings of meanings to signals.

The capacity to generalize has already been mentioned as a prerequisite for ILM simulations to be successful. For language to pass through the learning bottleneck, the complete grammar must be inducible from a limited amount of input which requires generalizations. Brighton et al. demonstrate this by comparing the performance in ILM simulations of agents with and without the ability to generalize. Recall that the behavior

of the agents' learning algorithm is determined by the parameters α, β, γ and δ. It turns out that the propensity to generalize is determined by the relationship of α and δ. For details of why this is, see Smith (2003) and Brighton et al. (2005b). In brief, the values of α and δ are related to the ability to generalize as follows:

$\alpha > \delta$ Agents have the ability to generalize.

$\alpha = \delta$ Agents do not generalize reliably but are able to generalize on occasion.

$\alpha < \delta$ Agents do not have the ability to generalize.

Brighton et al. tested the condition in which $\alpha > \delta$ (reliable generalization) against that in which $\alpha = \delta$ (unreliable generalization). Their results are shown in figure 11.
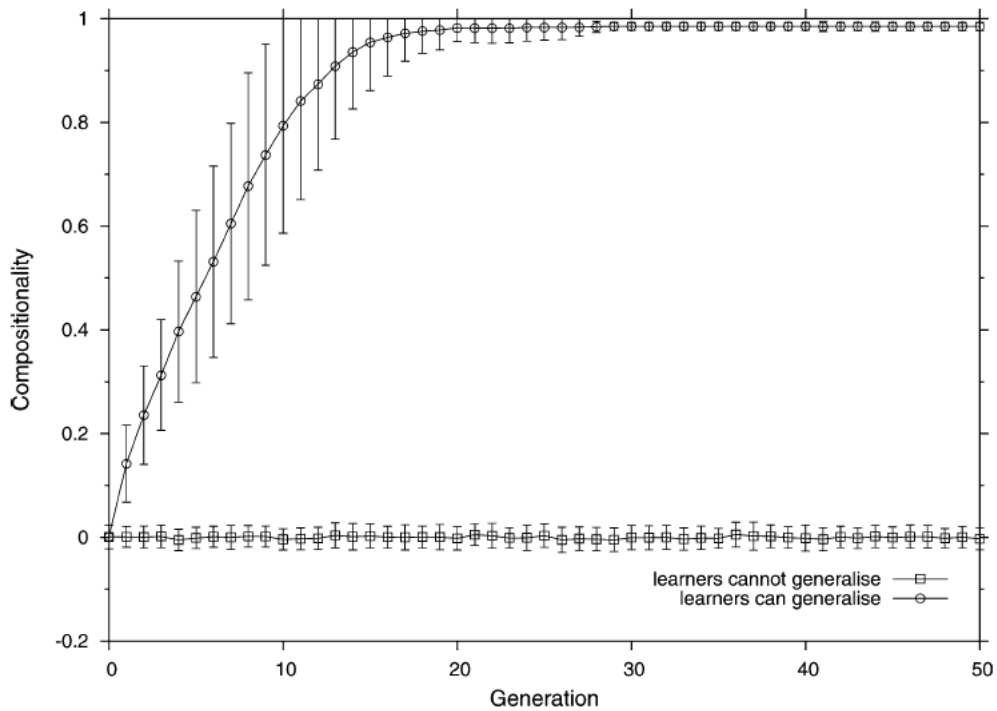

Fig. 11: A comparison of the results of a simulation with and without a reliable ability to generalize. Compositional language only emerges with generalization. Figure taken from (*Brighton et al.* 2005b).

We can see from these results that the ability to generalize reliably is indeed important for the cultural evolution of compositionality.

The necessity of a bias to generalize is intuitively easy to understand. The necessity of a prejudice toward one-to-one mappings is less intuitively obvious. But consider a situation in which learners have no prejudice against many-to-one meaning to signal mappings. These learners would be amenable to learning a language which, in the extreme, represented all meanings by a single string. Though this language is infinitely expressive in a technical sense, it is obviously useless for communication. In the simulation done by Brighton et al., learners without a bias against many-to-one mappings developed precisely such a language. Brighton et al. also argue for the necessity of a bias against one-to-many mappings.

Once we propose learning biases as essential to the development of language, it is important to consider whether humans demonstrate such biases. Extant human languages show both one-to-many and many-to-one mappings between signals and meanings. Though on the surface this seems to contradict the idea that humans have a bias against such mappings, Brighton et al. (2005b) argue that this is not the case. As they say, properties of the final language do not necessarily reflect properties of the learning biases of linguistic agents. They also cite psychological and historical linguistic evidence to support their claims that humans possess a bias for one-to-one mappings. For instance, studies with children demonstrate that they are more likely to associate novel words with novel objects rather than postulate many-to-one mappings of words to objects or objects to words.

Many other researchers have examined the question of what learning biases play critical roles in the success of ILM simulations. For instance, Smith (2003) reaches the same conclusions as Brighton et al. as to the importance of the ability to generalize and the preference toward one-to-one mappings. He goes on to classify neural net style learners into groups of *constructors*, *maintainers*, *learners* and *nonlearners*. Constructors can start from no language and develop a structured linguistic system as we have seen in the simulations presented here. Maintainers can, if presented with input generated by such

a structured system, transmit the system to the next generation even with some noise or disruption. Learners can learn the system and transmit it, but only in the absence of noise. Nonlearners lack even the capacity to learn the structured system presented to them. The answer to the question of what places any given learning algorithm into one of these four groups is prerequisite to a complete understanding of the mechanics of cultural transmission. Without this information, it is difficult to make claims about the cognitive machinery that humans must have for the progression cultural evolution. However, with the suggestions made above, we can make a start at identifying what constraints on cognition may be sufficient, or even necessary, for the development of language through cultural evolution.

Once we establish that cultural accounts of language evolution require specific learning biases, we must address the question of what differentiates such biases from the nativist UG. Perhaps one answer to this is that learning biases may have scope outside of the linguistic domain. The ability to generalize is likely to have wide applicability. It is possible that a bias for one-to-one mappings could be useful outside the linguistic domain as well. If we postulate that these biases are not specific to language, we should be able to find evidence of the same biases in other areas of human cognition.

9   The Genome

I argue that the evidence for cultural evolution is sufficient to make the strong nativist approach untenable. On the other hand, there is evidence for an innate component to language that I have not yet discussed. There is clearly something unique to humans that gives us a linguistic ability not possessed by any other species. The purely cultural approach would suggest that this is entirely the result of general learning biases and the cultural processes demonstrated in the simulations. However, cultural evolution of language takes many generations. Processes such as creolization or the formation of native

sign in Nicaraguan schools for the deaf (*Senghas* 1995) occur in at most a few generations. It would be difficult to make the claim that cultural evolution alone could account for these processes occurring in such a short time. In reality, it is equally untenable to claim that genetic changes did not play a role in language evolution as to claim they were the sole factor.

Considerations such as those mentioned above lend impetus to the search for a theory that is neither entirely nativist nor exclusively cultural. Many researchers have suggested that an evolutionary process known as the Baldwin effect might allow for such a compromise (*Yamauchi* 2004; *Turkel* 2002; *Kirby and Hurford* 1997; *Briscoe* 2003). In the following sections I will discuss the Baldwin effect and its possible applicability to language evolution.

9.1   The Baldwin effect

The Baldwin effect was proposed over 100 years ago by the psychologist James Baldwin as a means of incorporating "a new factor" into the theory of evolution (1896). Baldwin's original idea was expanded upon and developed by Simpson (1953), who coined the name, and Waddington, who had concurrently developed similar ideas (*Yamauchi* 2004). Very generally, the Baldwin effect proposes a means by which cultural or learned behavior can become genetically assimilated.

Natural selection gives selective benefit to the ability to learn. Consider the behavioral fitness landscape for an organism. Each point on this landscape represents the selective advantage of a certain set of behavioral characteristics. Evolution is essentially a hill-climbing search over fitness landscapes. If behavioral characteristics are entirely innate, each organism will stay in a single position on the fitness landscape throughout its life. Genetic mutations and recombinations will cause individuals in a population to be spread out on the fitness landscape, and those located at higher points will have a selective

benefit. If the fitness landscape is smooth - that is, if incremental changes to the behavioral makeup of an organism in a particular direction results in progressively increasing fitness - the evolutionary search will be able to find the local maxima. If, on the other hand, the fitness landscape is spiky, with large areas of equal fitness and local peaks of higher fitness, the evolutionary search will progress randomly through the flat spaces ad will be less likely to reach the maxima.

Consider now the introduction to this fitness landscape of a population of organisms that have the capacity to acquire learned behavior. That is, individual organisms may move around the fitness landscape during their lifetimes. On the spiky fitness landscape described above, the ability of organisms to explore a portion of the surrounding fitness landscape effectively smooths the spikes. Individuals placed innately near to a fitness spike will be likely to find the spike and reap the selective benefit.

While there is a selective pressure for the ability to learn, the Baldwin Effect describes the reverse pressure - that for learned behavior to become innate. Again, consider the population of agents on a fitness landscape. These agents are able to acquire learned behaviors, meaning that if there is a fitness peak in their vicinity they are likely to find it within their lifetimes. Once they find the peak, they gain a selective advantage. Now imagine that a genetic change allows a certain agent to find the fitness peak more quickly than the others. Though many of the agents will reach the peak eventually, this first agent will be there for longer and therefore have greater fitness. The genetic change that would allow this speedy approach to the peak is the incorporation of a previously learned behavior into the genome. That is, agents who, based on their genetically determined behavior, begin closer to the fitness peak will have a selective advantage over those who must search a larger area.

Following Turney (1996), we can look at this part of the process as a result of learning being costly. For example, the experimentation necessitated by the learning process is potentially dangerous. As Turney says, it can be advantageous to instinctively avoid

snakes rather than learning to do so through trial-and-error. To continue with the snake example, we can make the following comparisons. Individuals who are able to learn to avoid snakes have an advantage over those who aren't able to acquire this behavior (learning is advantageous). Individuals who instinctively avoid snakes have an advantage over those who must learn to do so (learning is costly). By this reasoning, innate mechanisms are selectively preferable over learned mechanisms. But of course, the ability to learn remains advantageous for the smoothing effect that it has on spiky fitness landscapes. The competing pressures to develop support for learned and instinctive behaviors creates an evolutionary balancing act. Turney describes how such a situation might play out:

> If it is possible for the behavior to be performed by an instinctive mechanism, it will usually be advantageous for such a mechanism to evolve, since instinctive mechanisms tend to be less expensive than learned mechanisms. However, when a new behavior is first evolving, an instinctive mechanism may require the population to make a large evolutionary leap, while a learned mechanism may be able to arise in smaller evolutionary increments. Learning may allow the behavior to eventually become common and robust in the populations, which then gives evolution the time required to find an instinctive mechanism to replace the learned mechanism. In summary, at first learning is advantageous, but later it is not.
>
> (*Turney*, 1996, 137)

Learned behavior can guide evolution toward peaks of fitness that it might not otherwise be likely to find. As Yamauchi says, "learning paves the path of the evolutionary search so that the burden of the evolutionary search is eased" (2004, 2). In 1987, Hinton and Nowlan performed a computer simulation of this process, showing that the ability to learn aided populations in evolving more quickly toward maximum fitness (*Hinton and Nowlan* 1987).

9.1.1   Hinton & Nowlan simulation of the Baldwin Effect

Hinton and Nowlan's simulation of the Baldwin Effect is quite simple, but is important as
an early demonstration of learning guiding evolution. The agents in Hinton & Nowlan's
simulation are neural nets with 20 potential connections. On the creation of an agent,
each connection (or, in genetic terms, allele) is specified at either 1 (on), 0 (off) or ?
(unspecified). The unspecified alleles are set by learning during the lifetime of the agent.
Initially, each allele for each agent is set at random with a probability of 0.5 for ? alleles
and 0.25 for alleles of 1 and 0.

The fitness landscape of the simulation is such that there is a single spike of increased
fitness corresponding to a single configuration of alleles. During the lifetime of an agent,
it is given a certain number of learning trials. On each trial, it randomly sets each of
its ? connections to either 0 or 1. If the agent ever discovers the high fitness setting,
it stops the learning process. Otherwise, it continues to randomly set its connections on
each trial. Agents are allowed 2^{10} trials, meaning that agents with 10 correct connections
and the rest unspecified are likely to find the correct setting within their lifetimes. Figure
12 shows the fitness landscape of the simulation. Without learning, there is only a single
peak. With learning, the fitness landscape is smoothed with areas of increased fitness on
either side of the peak.

After each generation, new agents are produced by selecting pairs of existing agents
and using a single crossover point to merge their initial settings. That is, the crossover
point is chosen and the offspring gets the connection setting of one parent up to the
crossover point and the connection settings of the other parent after the crossover point.
Agents are selected at random, with agents that achieved the high fitness setting gaining
significant advantage in the selection process. This advantage is proportional to $1 + \frac{19n}{1000}$
where n is the number of learning trials remaining once the agent discovers the correct
setting. This means that agents who can achieve the high fitness setting with fewer trials
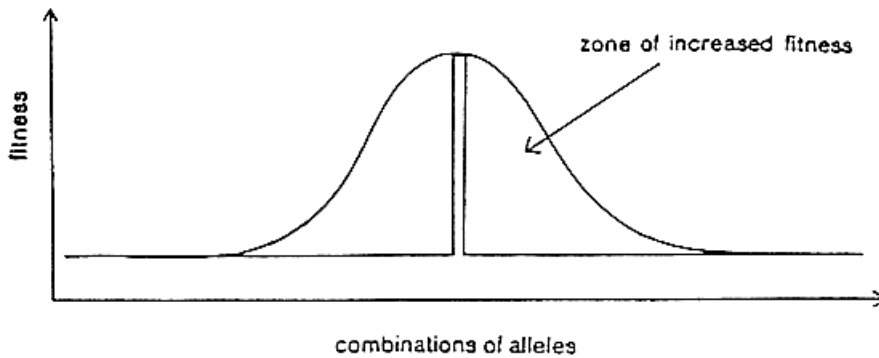
Fig. 12: The fitness landscape of the Hinton & Nowlan simulation. There is a single spike of increased fitness, with areas of increased fitness on either side as the result of learning. Figure taken from Hinton & Nowlan (1987).

have a selective advantage.

Figure 13 shows the distribution of alleles during the simulation. We can see that the incorrect alleles are quickly eliminated from the population. The frequency of correct alleles increases, but a few unspecified alleles remain. These unspecified alleles are able to remain since agents are able to learn the correct settings with only a few learning trials.

With learning, the population evolves a stable state in which agents are able to reach the maximum fitness with a limited number of learning trials. However, when agents were not allowed to learn, the fitness peak was never found. An evolutionary search relies on small shifts toward fitness maxima resulting in increased fitness. This is not the case in the spiked landscape of this simulation. With a mostly flat fitness landscape, the evolutionary search proceeds at random. It is unlikely for an agent to chance upon the correct allele settings. Even if an agent does happen to find the fitness peak, it is unlikely that the correct settings will be maintained into the next generation. When the fit agent mates with an agent that has not found the fitness peak, crossover will destroy the high fitness allele configuration. In the framework of the Hinton and Nowlan simulation, learning guides evolution toward a fitness peak that evolutionary search without learning could not find.
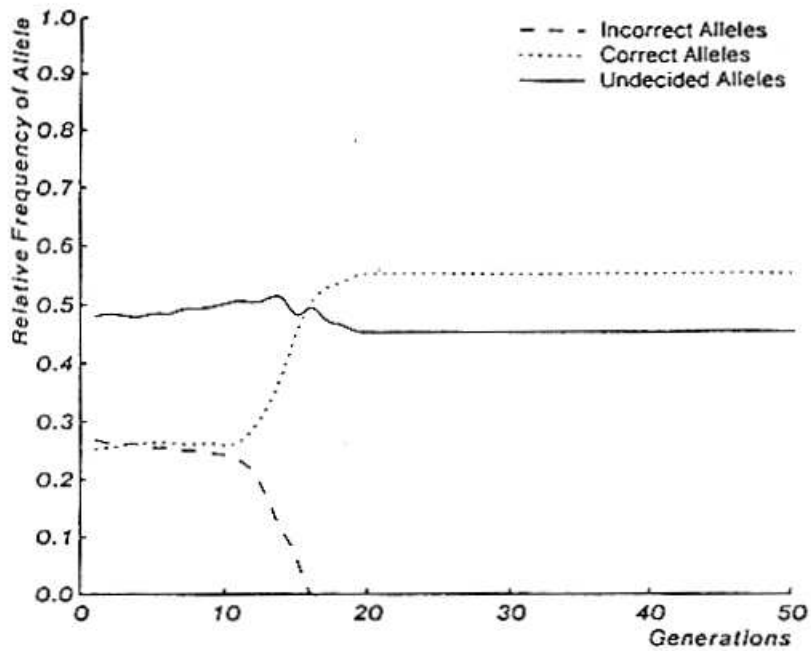
Fig. 13: The results of the Hinton & Nowlan simulation. The population stabilizes with mostly correct alleles. Some unspecified connections remain as agents are able to learn the correct settings for these alleles with only a few trials. Figure taken from Hinton & Nowlan (1987).

The simulation by Hinton and Nowlan demonstrates the Baldwin Effect by showing that learning can guide evolution toward an otherwise unreachable fitness peak. However, their simulation only shows the effect for a single, extreme, fitness landscape. As mentioned before, learning can be costly. We might expect to find fitness landscapes on which the Baldwin Effect is not beneficial to the evolutionary search. Ancel (2000) explores this question. She finds that the Baldwin Effect is only beneficial in certain fitness landscapes. She classifies these landscapes according to two conditions:

Condition1 Plasticity expedites the search from an initial population distribution to the first encounter with the optimum phenotype.

Condition2 Condition1 is observed for initial genotype distributions sufficiently distant from the target. (*Ancel*, 2000, 318)

In other words, phenotypic plasticity, of which the ability to learn is an example, expedites evolutionary search in situations where the initial placement of agents on the fitness landscape if sufficiently distant from the position of maximum fitness. The Baldwin Effect is most helpful on fitness landscapes which are highly peaked, such as those we have discussed. As the fitness landscapes become smoother, the benefit of the Baldwin Effect to evolutionary search becomes negligible.

Ancel (2000) demonstrates that learning can actually impede the fixation of a beneficial genotype in a population. The costliness of learning is important to the success of the Baldwin Effect. If learning not costly enough, it can keep suboptimal genotypes from being eliminated by natural selection. On the other hand, if learning is too costly, it will not aid the evolutionary search in the first place.

We have evidence that the Baldwin Effect can guide evolution toward fitness peaks. But we have also seen that the Baldwin Effect is not advantageous in all situations. Is the Baldwin Effect likely to have played a role in the evolution of language?

9.2   Baldwin Effect as Applied to language

Several researchers propose the Baldwin Effect as a promising possible explanation for language evolution. Pinker and Bloom (1990) make this argument:

> When some individuals are making important distinctions that can be decoded with cognitive effort, it could set up a pressure for the evolution of neural mechanisms that would make this decoding process become increasingly automatic, unconscious, and undistracted by irrelevant aspects of world knowledge. [...] The process whereby environmentally induced responses set up selection pressures for such responses to become innate, triggering conventional Darwinian evolution that superficially mimics a Lamarckian sequence, is sometimes known as "the Baldwin effect".
>
> (*Pinker and Bloom* 1990, 722)

The Baldwin effect provides a mechanism by which we can merge nativist and cultural ideas about language evolution. The two perspectives become complementary rather than antagonistic. For instance, we can look to cultural evolution to account for the genesis of language, as its strength lies in accounting for the generation of structure without specifically dedicated language mechanisms. But if we accept that such cultural behavior can eventually become innate, we can also look to more nativist ideas to explain quick processes such as creolization that do not lend themselves well to cultural explanations.

We need not hypothesize that the Baldwin effect played a role only during the genesis of language. If we accept the framework of the process, we can assume that this co-evolution of cultural and genetic aspects of language has been a constant force in language evolution from the first linguistic communities into modern times. As Richardson and Boyd put it,

> The most plausible explanation [for the evolution of language] is that simple culturally transmitted language arose first, and then selection favored a

> [...] special-purpose psychology for learning, decoding, and producing speech,
> which in turn gave rise to a richer, more-complex language.
>
> (*Richerson and Boyd* 2005, 193)

Such a sequence of events, referred to as a Baldwinian chain (*Yamauchi* 2004), allows us to develop a picture of language evolution that is simultaneously incremental, genetic and cultural. The ability of learning to guide evolution toward spiky areas of maximum fitness means that even learners with a very primitive or nonexistent genetic linguistic ability can reap the selective benefits of a linguistic system. Whatever innate language mechanism that may now exist need not have been fully formed for language to emerge.

10   Conclusions

Language is one of the most defining features of humanity, yet we know surprisingly little about how it came about. The question of language evolution was ignored for years, and is still frequently sidestepped by the dominant linguistic theories. Though it is by no means universal, a popular view of language evolution assumes the nativist approach, which maintains that there is some mechanism in the human brain that specifies language, which either evolved or emerged spontaneously. I have argued that such a strong nativist view is untenable in light of the evidence for cultural evolution.

Cultural evolution proves to be a very powerful force that almost certainly played a role in the evolution of language. Cultural evolution can explain many of the phenomenon treated as necessarily innate by nativist theories. Additionally, cultural evolution challenges the conclusiveness of the poverty of stimulus argument. Languages that are evolutionarily successful must be able to be transmitted through the learning bottleneck. This selects for languages containing complex features that allow for such a compression of information. Rather than being a hindrance to the learning of the complexities of language, we can now see sparsity of input as an evolutionary cause of these complexities.

It is not a tenable argument to maintain that there is nothing innate that allows humans to learn language. The simple fact that we are the sole species to be able to master the system proves this point. The more interesting question is whether the innate qualities that allow us to use language are specific to that task. If we accept the Universal Grammar proposed by Chomskyan nativism, the answer to this question must be yes. However, evidence from simulations of the cultural evolution of languages themselves indicates that the innate features may instead take the form of general learning biases that could have applications outside the domain of language learning and use. These learning biases place restrictions on the grammars that linguistic agents hypothesize, and over many generations this results in structured features of the community language.

As we understand it now, however, cultural evolution cannot account for the entirety of human language. For one thing, there are many features of language that have not been explored in this framework. Though we can hypothesize that these might follow similar patterns, we lack the evidence to be sure. Also, simulations of cultural evolution are extremely simplified versions of what occurs in the real world. The results of these simulations might be more conclusive if the simulations accounted for a more complex and accurate model of reality. Finally, cultural evolution, at least in simulation, is a process that takes generations to occur. It has been suggested that the changes in language through cultural evolution proceed more quickly when there is horizontal transmission within a generation as well as vertical transmission between generations. Nevertheless, it seems unlikely that processes such as creolization can be accounted for solely by cultural evolution.

The Baldwin effect provides a framework in which we can hypothesize a composite theory taking ideas from both nativist and cultural approaches. Perhaps there is a specific, genetically innate language ability, but if there is, it likely initially co-evolved with a culturally evolved linguistic system. This sort of co-evolution between languages themselves and the human cognitive capacity is, I believe, the idea that will ultimately prove

most fruitful for understanding the genesis of human language.

As covered at the beginning of this thesis, the evolution of language as an academic discipline is faced with a general lack of evidence and surplus of speculation. As such, it is an area with few definitive answers, a situation which is unlikely to be resolved in the foreseeable future. Nevertheless, the exploration of the evolution of language can inform theories of modern language and play a role in the debate over nativism. If we can establish the explanatory power of cultural evolution, we can explore the implications that such a process might have on both the development and current state of the cognitive basis for language. Computer simulation shows great potential in this area. Simulations are necessarily simplistic, but as the body of work in this area continues to grow, there will be room to explore simulations that are more complex and realistic both in their portrayal of the social dynamics of language transmission and of the scope of languages themselves. Such work will solidify understanding of the linguistic present while it sates curiosity over the linguistic past.

References

Ancel, L. W., Undermining the baldwin expediting effect: Does phenotypic plasticity accelerate evolution?, *Theoretical Population Biology*, *58*, 307–319, 2000.

Anderson, S., and D. Lightfoot, *The Language Organ*, Cambridge University Press, 2002.

Baldwin, J. M., A new factor in evolution, *American Naturalist*, *30*, 441–451, 536–553, 1896.

Blackmore, S., *The Meme Machine*, Oxford University Press, 1999.

Brighton, H., and S. Kirby, The survival of the smallest: Stability conditions for the cultural evolution of compositional language, in *ECAL01*, edited by J. Kelemen and P. Sosik, pp. 592–601, Springer-Verlag, 2001a.

Brighton, H., and S. Kirby, Meaning space determines the stability of culturally evolved compositional language, *Tech. rep.*, Language Evolution and Computation Research Unit, Department of Theoretical and Apllied Linguistics, University of Edinburgh, 2001b.

Brighton, H., and S. Kirby, Understanding linguistic evolution by visualizing the emergence of topographic mappings, *Artificial Life*, *12*, 229–242, 2006.

Brighton, H., S. Kirby, and K. Smith, Cultural selection for learnability: Three hypotheses concerning the characterisitc structure of language, in *Language Origins: Perspectives on Evolution*, edited by M. Tallerman, Oxford University Press, 2005a.

Brighton, H., K. Smith, and S. Kirby, Language as an evolutionary system, *Physics of Life Reviews*, *2*, 177–226, 2005b.

Briscoe, T., Grammatical assimilation, in *Language Evolution*, edited by M. Christiansen and S. Kirby, Oxford University Press, 2003.

Chomsky, N., *Reflections on Language*, Pantheon Books, 1975.

Chomsky, N., On the nature, acquisition and use of language, in *Mind and Cognition: A Reader*, edited by W. G. Lycan, pp. 627–45, Blackwell, Cambridge, MA, 1990.

Christiansen, M., and M. Ellefson, Linguistic adaptation without linguistic contraints: The role of sequential learning in language evolution, in *The Transition to Language*, edited by A. Wray, Oxford University Press, 2002.

Christiansen, M., and S. Kirby, Language evolution: consensus and controversies, *Trends in Cognitive Science*, *7*, 300–307, 2003.

Christiansen, M., R. Dale, M. Ellefson, and C. Conway, The role of sequential learning in language evolution: Computational and experimental studies, in *Simulating the Evolution of Language*, edited by A. Canglelosi and D. Parisi, Springer-Verlag, 2002.

Cowie, F., *What's Within? Nativism Reconsidered*, Oxford University Press, 1999.

Dawkins, R., *The Selfish Gene*, second ed., Oxford University Press, 1989.

Dawkins, R., *The Extended Phenotype*, Oxford University Press, 1999.

de Boer, B., Self-organization in vowel systems, *Journal of Phonetics*, *27*, 2000.

de Boer, B., Evolving sound systems, in *Simulating the Evolution of Language*, edited by A. Cangelosi and D. Parisi, chap. 4, pp. 79–97, Springer, 2002.

Deacon, T., The symbolic species: The co-evolution of language and the brain, in *W. W. Norton*, 1997.

Gilbert, N., and J. Doran (Eds.), *Simulating Societies: The computer simulation of social phenomenon*, UCL Press, 1994c.

Hinton, G., and S. Nowlan, How learning can guide evolution, *Complex Systems*, *1*, 495–502, 1987.

Hurford, J., Social transmission favours linguistic generalization, in *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, Cambridge University Press, 2000.

Kirby, S., Learning, bottlenecks and the evolution of recursive syntax, in *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, edited by T. Briscoe, Cambridge University Press, 2002.

Kirby, S., and M. Christiansen, From language learning to language evolution, in *Language Evolution*, edited by S. Kirby and M. Christiansen, Oxford University Press, 2003.

Kirby, S., and J. Hurford, Learning, culture and evolution in the origin of linguistic contraints, in *ECAL97*, edited by P. Husbands and I. Harvey, pp. 493–502, MIT Press, 1997.

Kirby, S., and J. Hurford, The emergence of linguistic structure: An overview of the iterated learning model, in *Simulating the Evolution of Language*, edited by A. Canglolosi and D. Parisi, Springer-Verlag, 2002.

Kirby, S., K. Smith, and H. Brighton, From ug to universals: linguistic adaptation through iterated learning, *Studies in Language*, *28*, 587–607, 2004.

Komarova, N., and M. Nowak, Population dynamics of grammar acquisition, in *Simulating the Evolution of Language*, edited by A. Cangelosi and D. Parisi, Springer-Verlag, 2002.

Lieberman, P., *Human Language and our Reptilian Brain: The subcortical bases of speech, syntax, and thought*, Harvard University Press, 2000.

Lightfoot, D., The child's trigger-experience: Degree-0 learnability, *Behavioral and Brain Sciences*, *12*, 321–34, 1989.

Newmeyer, F., Uniformitarian assumptions and language evolution research, in *The Transition to Language*, edited by A. Wray, Oxford University Press, 2002.

Newmeyer, F., What can the field of linguistics tell us about the origins of language, in *Language Evolution*, edited by M. Christiansen and S. Kirby, Oxford University Press, 2003.

Pinker, S., *The Language Instinct*, HarperCollins, 1994.

Pinker, S., and P. Bloom, Natural language and natural selection, *Behavioral and Brain Sciences*, *13*, 707–784, 1990.

Richerson, P., and R. Boyd, *Not by Genes Alone: How culture transformed human evolution*, University of Chicago Press, 2005.

Ritt, N., *Selfish Sounds and Linguistic Evolution: A Darwinian Approach to Language Change*, Cambridge University Press, 2004.

Sapir, E., *Language: An introduction to the study of speech*, Harcourt, Brace and company, 1921.

Senghas, A., Children's contribution to the birth of nicaraguan sign language, Ph.D. thesis, MIT, 1995.

Simpson, G., The baldwin effect, *Evolution*, *7*, 110–117, 1953.

Smith, K., Compositionality from culture: the role of environment structure and learning bias, *Tech. rep.*, University of Edinburgh, 2003.

Steels, L., Synthesising the origins of language and meaning using co-evolution, self-organization and level formation, in *Approaches to the Evolution of Language: Social and Cognitive Bases*, edited by J. Hurford, C. Knight, and M. Studdert-Kennedy, Edinburgh University Press, 1998.

Steels, L., Grounding symbols through evolutionary language games, in *Simulating the Evolution of Language*, edited by A. Cangelosi and D. Parisi, chap. 10, pp. 211–226, Springer, 2002.

Steels, L., and F. Kaplan, Bootstrapping grounded word semantics, in *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, Cambridge University Press, 2002.

Tonkes, B., and J. Wiles, Methodological issues in simulating the emergence of language, in *The Transition to Language*, edited by A. Wray, Oxford University Press, 2002.

Turkel, W., The learning guided evolution of natural language, in *Linguistic Evolution through Language Acquisition*, edited by T. Briscoe, Cambridge University Press, 2002.

Turney, P., Myths and legends of the baldwin effect, in *Proceedings of the Workshop on Evolutionary Computation and Machine Learning*, pp. 135–142, International Conference of Machine Learning, 1996.

Yamauchi, H., Baldwinian accounts of language evolution, Ph.D. thesis, University of Edinburgh, 2004.