Natural selection and cultural selection in the evolution of communication

Kenny Smith
Language Evolution and Computation Research Unit
Department of Theoretical and Applied Linguistics
University of Edinburgh
Adam Ferguson Building
40 George Square
Edinburgh, UK

kenny@ling.ed.ac.uk
Phone: +44 131 650 6658
Fax: +44 131 650 3962

1

Abstract

It has been postulated that aspects of human language are both genetically and culturally transmitted. How might these processes interact to determine the structure of language? An agent-based model designed to study gene-culture interactions in the evolution of communication is introduced. This model shows that cultural selection resulting from learner biases can be crucial in determining the structure of communication systems transmitted through both genetic and cultural processes. Furthermore, the learning bias which leads to the emergence of optimal communication in the model resembles the learning bias brought to the task of communication by human infants. This suggests that the iterated application of such human learning biases may explain much of the structure of human language.

Key words: communication, language, evolution, culture, learning

1   Introduction

Language is transmitted from generation to generation within a speech community. The precise nature of the intergenerational transmission remains a contentious issue. The transmission of language down the generations involves at least some cultural transmission – under normal circumstances children learn the language of their speech community through exposure to the linguistic behavior of that community. The most influential linguistic theories of modern times assume genetic transmission of the language faculty between generations in addition to this cultural transmission – language learners come to the language acquisition task equipped with some genetically-encoded Language Acquisition Device (Chomsky, 1987). The research outlined in this paper represents an attempt to understand the types of interactions which may occur between cultural transmission and genetic transmission of communication systems within a communicating population. Specifically, in this paper I argue that cultural selection resulting from learning biases is key in determining the structure of communication systems, such as language, which are both genetically and culturally transmitted.

In Section 2 the literature on the evolution of communication and language is reviewed. This review reveals a paucity of models on gene-culture interactions which are simple enough to be easily understood yet detailed enough to test hypotheses regarding the evolution of communication or language in the real world. In Section 3 such a model is proposed. Sections 4 and 5 present results generated by this model. These results suggest that cultural selection resulting from the biases inherent in the model of the learner is crucial in determining the structure of the emergent communication systems, and that natural selection is unable to override these biases. In Section 6 the learning biases of the model are examined in detail and related both to other agent-based models of the evolution of communication and to the communication-specific learning biases observed in humans. Finally, in the concluding section it is suggested that much of the structure of language may be best explained in terms of cultural evolution resulting from a pre-adapted learning mechanism.

2   The literature

The literature on the evolution of communication and language can be roughly divided into two main areas – that which addresses the question "when should we expect to see communication or language?" and that which

addresses the question "what structure should we expect communication or language to exhibit?".

The first question has typically been addressed by theoretical biologists but has recently been tackled by researchers using agent-based modeling techniques. Researchers in this area are concerned with the interlocking issues of signal costs and honesty (for example, from biology Zahavi (1975), Zahavi (1977), Krebs and Dawkins (1984) and Grafen (1990), from agent-based modeling Wheeler and de Bourcier (1995), Bullock (1997), Noble (1998)) and altruism (from agent-based modeling Ackley and Littman (1994), Oliphant (1996) and Reggia, Schulz, Wilkinson, and Uriagereka (2001)). These important issues will not concern us in this paper, other than to say that the presence of honest communication has been identified as a possibility in at least some circumstances by some researchers.

The second question can be viewed as consisting of three subquestions: "what structure would we expect communication or language to exhibit if it were shaped by purely biological processes?"; "what structure would we expect communication or language to exhibit if it were shaped by purely cultural processes?"; "what structure would we expect communication or language to exhibit if it were shaped by both biological and cultural processes?".

There has been some research on the first question, both by biologists (see Hauser (1996) for a summary) and agent-based modelers. For example, Werner and Dyer (1992), Levin (1995) and Di Paolo (1997) show that agents can coordinate their actions or internal states optimally or near-optimally using innate communication systems given selection pressure for that coordination. Werner and Todd (1997) show that the reverse can also be true – agents can violate the innate expectations of receivers given innate signalling behavior and selection pressure for such violation. Cangelosi and Parisi (1998) show that an efficient biologically-transmitted communication system can emerge even without direct selection pressure, effectively due to evolution of internal representations and genetic drift of a communication system on top of this evolved substrate. However, human language, our ultimate object of study, consists of at least some learned component and these models are therefore of limited utility in understanding it.

The second question, concerning the structure of communication given purely cultural transmission processes, has received a considerable amount of attention in recent years. It appears that such processes may only be relevant to the study of communication in humans, given that Hauser states that "although call structure [in non-human primates] changes ontogenetically, no study has provided convincing evidence that acoustic experience is

4

causally related to such changes" (Hauser, 1996, p315). Consequently, much of the research into this area has been carried out by linguists or cognitive scientists using agent-based modeling techniques to explain the cultural evolution of features of human language such as syntax (e.g. Batali (1998), Batali (in press), Hurford (2000), Kirby (2000), Kirby (in press), Brighton and Kirby (2001), see Hurford (2002) and Kirby and Hurford (2002) for an overview), regularity and irregularity (e.g. Kirby (2001), Worden (in press)) and other language universals (e.g. Christiansen and Devlin (1997), Kirby (1998)). Less work has been carried out on the more basic issue of the cultural evolution of non-syntactic communication systems. However, agent-based models directed at this issue can be found in Hutchins and Hazelhurst (1995), Oliphant and Batali (1997), Oliphant (1999) and Steels (1999).

This substantial body of literature presents a persuasive argument that the features of communication and language can be explained in terms of cultural processes. However, this work does have its weaknesses. Typically each paper considers a single model of learning. This lack of comparison between learning mechanisms makes it difficult to identify the biases of the chosen model of learning. Secondly, these models assume a degree of pre-existing mental apparatus, including a learning mechanism. This mental apparatus presumably evolved, although not necessarily for its later role in language-processing. But how might the evolution of learning mechanisms interact with the resulting cultural processes? Such models are not designed to address this question.

Finally, there is a small body of literature investigating the question of the structure of communication systems emerging through a mixture of genetic and cultural processes. Pinker and Bloom (1990) and Dor and Jablonka (2000) introduce hypothetical scenarios under which positive interactions between natural selection and cultural transmission lead to language. However, human intuitions regarding the behavior of such complex adaptive processes are notoriously poor and a formal model is desirable.

In an early paper MacLennan and Burghardt (1994) consider how reinforcement learning might interact with natural selection in the evolution of vocabulary-like systems. However, given that "a series of studies beginning with Brown and Hanlon (1970) have demonstrated that there is little reliable correlation between the grammaticality of children's utterances and the sorts of responses to these that their parents give" (Bloom and Gleitman, 2001), the relevance of models involving reinforcement learning to our understanding of language, our ultimate object of study, is doubtful. Batali (1994) considers interactions between selection and learning in populations

5

of neural networks. The languages Batali's networks attempt to learn are externally determined, rather than emerging from the populations of agents themselves, and are therefore not truly culturally transmitted. This makes the relevance of this model to the field of human language less clear. Cangelosi (1999) uses neural networks to investigate gene-culture interactions in the evolution of symbolic communication systems. However, as in an earlier model (Cangelosi and Parisi, 1998), the structure of the communication system is determined by genetic factors with learning playing little role. Kirby and Hurford (1997), Turkel (in press) and Yamauchi (2001) consider possible interactions between natural selection and learning in the evolution of an innate language acquisition device and a language. However, their representations of language and learning are so abstract as to make any claims about the structure of human language difficult. Finally, in a recent paper Kvasnička and Pospíchal (1999) model interactions between natural selection and learning of culturally-emergent communication systems in a population of neural networks. This model is a step in the right direction, detailed enough to allow hypothesis to be formed about the structure of communication systems in the real world, abstract enough to be analyzable. However, the model suffers from two defects. Firstly, only one learning mechanism is considered. Secondly, only one level of selection pressure is considered. This means that the relationship between the learning bias of the learning mechanism and the forces of natural selection remain unclear.

This paper presents a model of the interactions between processes of biological transmission and cultural transmission in the evolution of simple communication. The model avoids defects of earlier models in investigating, in detail, the relationship between different selection pressures, different learning biases and different strengths of learning bias. This allows us to address the hypothesis, suggested by previous models, that natural selection and learning will interact positively to create optimal communication systems. The ability to manipulate the various pressures proves to be essential in understanding the key determinants of the behavior of the system. The model is simple enough to be analyzed but detailed enough to provide a starting point in understanding how these issues might apply to the evolution of communication and, ultimately, language, in the real world.

3   The model

The model consists of a simple model of communication (Section 3.1), a model of a communicative agent (Section 3.2) and a model of genetic and

cultural transmission (Section 3.3).

3.1 The communication system

For the purposes of this model communication systems are mappings between a set of unstructured meanings m and a set of unstructured signals s. A communication system consists of a production function, $p(m)$, mapping from m to s, and a reception function, $r(s)$, mapping from s to m.

3.1.1 Communicative Accuracy

A measure of communicative accuracy can be defined for such communication systems. Given a signaler, P, producing signals using the function $p(m)$ and a receiver, R, interpreting signals using the function $r(s)$, the accuracy of communicating the meaning $m_i \in m$ between the two individuals, $ca(P, R, m_i)$, is:

$$ca(P, R, m_i) = \left\{ \begin{array}{ll} 1 & \text{if } r(p(m_i)) = m_i \\ 0 & \text{otherwise} \end{array} \right\}$$

When $ca(P, R, m_i) = 1$ the communication is successful[1] . A population's communicative accuracy can be estimated by taking the average $ca(P, R, m_i)$ for a random sample of P, R and m_i. In a population possessing an *optimal communication system* $ca(P, R, m_i) = 1$ for any choice of P, R and m_i.

3.1.2 Ambiguity

Such communication systems can be classified in terms of the degree of ambiguity they exhibit in the mapping from meanings to signals. Ambiguity arises when signals which are perceptually indistinguishable are associated with distinct meanings. A communication system of the type outlined above will be termed:

- *Unambiguous* if every meaning is associated with a distinct signal or signals.

---

[1] Note that this assumes that meanings are functionally distinct. For example, if two meanings m_i and m_j result in the same behavior on the part of the receiver and $r(p(m_i)) = m_j$ then the communication would be measured as a failure but could, at the behavioral level, be considered a success.

- *Partially ambiguous* if some, but not all, meanings are associated with identical signals.

- *Fully ambiguous* if all meanings are associated with identical signals.

These terms are more formally defined below.

3.2 The communicative agent

Feedforward neural networks are used to model communicative agents. The structure of the network used is shown in Figure 1. Networks with this configuration will be referred to as *imitator* networks. The input to the imitator network is considered to be the meaning to be communicated and the imitator's output is considered to be the signal used by that agent to communicate the input meaning, with the precise nature of the meaning-signal mapping determined by the connection weights in the network.
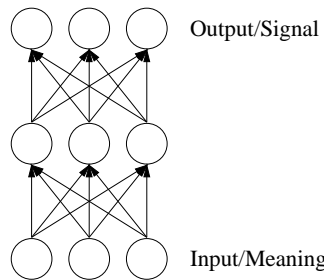

Figure 1: The structure of the neural network.

Communication systems therefore map from 3D meaning vectors to 3D signal vectors. Binary vectors are used, giving 2^3 possible meanings and 2^3 possible signals. A subset of the set of possible meaning vectors are considered to be communicatively relevant situations, where "communicatively relevant" means that agents receive a fitness payoff for communicating about those situations. For all simulations outlined in this paper, the set of communicatively relevant situations, m, consists of the unit vectors (1 0 0), (0 1 0) and (0 0 1). The set of available signals, s, consists of all 2^3 possible binary signal vectors.

Neural networks were chosen to model communicative agents for several reasons. Firstly, there is some tradition of using neural networks in research on the evolution of communication – neural networks of some form are used by Batali (1994), Hutchins and Hazelhurst (1995), Batali (1998), Cangelosi

8

and Parisi (1998), Cangelosi (1999), Livingstone and Fyfe (1999) and Kirby and Hurford (2002). Continuing this tradition provides several benefits. In particular, using a similar model allows the results of this research to be more easily related to previous research and the generality of the results of earlier simulations to be tested.

Secondly, well-established mechanisms exist for training neural networks to learn input-output mappings (i.e. backpropagation). Using an established learning mechanism reduces the amount of novel elements contained in the model, as well as allowing our understanding of that mechanism to be expanded.

Finally, using neural networks allows both genetically-transmitted and culturally-transmitted information to influence, in principle, the eventual behavior of agents in the model. Some of the assumptions arising from the choice of neural networks are somewhat dubious – for example, the assumptions that there is a one-to-one correspondence between genotype and phenotype and that genetic information merely provides a starting point for unconstrained learning in the phenotype. Two points can be raised in defense of these assumptions. Firstly, these assumptions are not unprecedented in the computational modelling literature – see for example the models described in Belew, McInerney, and Schraudolph (1992), Nolfi, Elman, and Parisi (1994), Batali (1994) and Rolls and Stringer (2000). Secondly, using this approach avoids some even more arbitrary assumptions that would be required to model combined influences of genes and culture in a more abstract model. These assumptions will, however, be returned to in the concluding section.

The disadvantage of using feedforward networks is that the slightly contorted reversal process outlined below is required in order to allow bidirectionality. Why is bidirectionality desirable when modeling communication? It is a fundamental assumption of modern linguistics (originating with Chomsky (1965)) that production and reception depend upon a common underlying knowledge of language - an individual's linguistic *competence*. Competence can be distinguished from *performance*, which determines how the structures underlying competence are accessed during reception and production. This competence-performance distinction is maintained here, with an agent's competence being encoded in the set of connection weights in their neural network and their performance being determined by the production and reception processes used to access this competence.

3.2.1 Production and reception

Producing the signal associated with a given meaning $m_i \in m$ in such imitator agents is straightforward — the given meaning is used as the input to the network and activations are propagated forward through the network to give a real-valued output pattern of activation, which is thresholded at 0.5 to give the binary signal associated with the given meaning.

The deterministic nature of the feedforward network during production means that the definition of ambiguity for communication systems can be formally stated. Communication systems used by neural networks will be termed:

- *Unambiguous* if $p(m)$ is a one-to-one, or an injective, function.

- *Partially ambiguous* if $p(m)$ is a many-to-one function, but the range of $p(m)$ is not a singleton set.

- *Fully ambiguous* if the range of $p(m)$ is a singleton set.

Reception is slightly more complex, given that the networks are not bidirectional. All $m_i \in m$ are propagated through a given agent's network to produce a real-numbered output pattern of activation for each meaning. Each output pattern is given a confidence rating, corresponding to how closely that pattern matches the received signal, $s_r \in s$. The meaning which produces the signal closest to s_r, according to the confidence measure, is chosen as the interpretation of s_r. This method is based on the method used by Batali (1998) and Kirby and Hurford (2002) for producing outputs for similar networks.

The confidence measure that a given real-numbered output vector, o, of length n matches a target binary vector t of length n is given by $C(t|o)$. $C(t|o)$ is simply the product of the confidence scores for each individual node $1...n$ in the output vector i.e.

$$C(t[1 \ldots n] | o[1 \ldots n]) = \prod_{i=1}^{n} C(t[i] | o[i])$$

where the confidence measure for node i is

$$C(t[i] | o[i]) = \left\{ \begin{array}{ll} o[i] & \text{if } t[i] = 1, \\ (1 - o[i]) & \text{if } t[i] = 0. \end{array} \right.$$

(Equations adapted from Kirby and Hurford (2002))

3.3 The transmission of communication systems

In this section a model of the genetic transmission of network connection weights and cultural transmission of communication systems is outlined. A genetic algorithm (Holland, 1975) is used to model the process of genetic transmission and is combined with an iterated learning model (Brighton and Kirby, 2001) which is used to model cultural transmission.

3.3.1 The genetic algorithm

The genetic algorithm has four key components:

1. A model of population turnover.

2. A model of genotypes, phenotypes and the mapping from genotype to phenotype.

3. Breeding based on an evaluation of communicative ability.

4. A method of recombination and mutation of genes during breeding.

These four components are described below.

Population turnover   A generational population model is used — at every time step of the simulation the entire population of size p is replaced by a new population of size p generated by breeding interactions between the members of the old population. $p = 100$ for all simulations outlined in this paper.

Genotypes and phenotypes   The phenotype communicative agent used is as outlined in Section 3.2 — a three-layer, feedforward neural network mapping from input meanings to output communicative signals. Each individual's initial connection weights are specified by their genotype — each agent's genotype consists of a string of real numbers, with each locus in the genotype mapping to a particular connection in the phenotype network. The real-numbered allele at a particular locus in the genotype determines the initial weight of the associated connection in the phenotype network. This mapping from genotype to phenotype is illustrated in Figure 2. The agents in the initial population have random alleles in the range $[-1, 1]$. There is no restriction on the range of real-numbered alleles beyond the initial population.
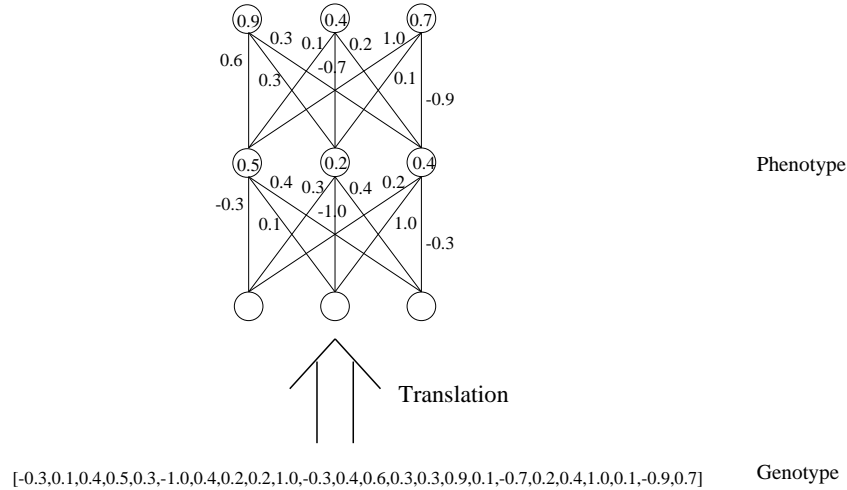
Figure 2: The mapping from genotype (a string of real numbers) to phenotype (a neural network). Bias node connection weights are shown in the associated node.

Selective breeding The probability of an agent breeding is determined by its success at communicating with other members of its generation of the population. The method of evaluating communicative success is given below:

- For each agent A in the population:

  1 Remove A from the population.

  2 Pick an agent B at random from the population.

  3.1 Pick a meaning, m_s, at random from the set of communicatively relevant situations.

  3.2 Call A the signaler and B the receiver.

  3.3.1 Generate the signal s_s, that the signaler associates with m_s (via the production mechanism outlined in Section 3.2.1).

  3.3.2 Identify the meaning, m_r, that the receiver associates with s_s (via the reception process outlined in Section 3.2.1).

  3.3.3 Compare m_r with m_s and score the success of the communication. If m_r is identical to m_s score the communication as a success and increment A's fitness. Otherwise, the communication is a failure.

12

3.4 Call A the receiver and B the signaler and repeat step 3.3.

3.5 Return B to the population.

4 Repeat steps 2 and 3 f times[2].

5 Return A to the population

This fitness assessment algorithm corresponds to the measure of communicative accuracy outlined in Section 3.1.1. Agents receive a reward both for understanding and being understood and the rewards for both are equally weighted. The fittest b individuals in the population breed with equal probability to produce the next generation of agents, where $0 < b \leq p$.

Recombination of genes Breeding involves recombination of the genes of two parents, via crossover, and mutation. Single-point crossover occurs with probability P_{cross} [3]. Point mutations occur on the newly formed genotype with probability $P_{mutation}$[4]. Mutation results in the value at the mutated locus being increased by a random real number in the range [-1,1][5].

3.3.2 Cultural transmission

Iterated learning models have been used to examine the cultural evolution of communication (Oliphant, 1999) and compositional language (Kirby and Hurford, 1997; Batali, 1998; Batali, in press; Kirby, 1999; Kirby, 2000; Kirby, 2001; Brighton and Kirby, 2001; Kirby and Hurford, 2002). In the iterated learning model "each generation of language user acquires its linguistic competence by observing the behavior of the previous generation" (Brighton and Kirby, 2001, p592). This acquired linguistic competence then governs the behavior which is observed by the subsequent generation. The iterated learning model resembles the cultural equivalent of a genetic algorithm, although typically there is no notion of fitness.

In terms of the current model, individuals at generation $N + 1$ observe and learn from the communicative behavior of generation N individuals. Each individual at generation $N + 1$ receives e exposures to the communication systems of the population at generation N. These exposures are

---

[2]$f = 6$ for all simulations outlined in this paper.

[3]For all simulations outlined in this paper $P_{cross} = 0.95$.

[4]For all simulations outlined in this paper $P_{mutation} = \frac{0.1}{L_g} = 0.0042$, where L_g is the length of the genome.

[5]This mutation operator, in conjunction with the unrestricted range of alleles, allows the possibility of the emergence of extremely large-valued alleles. However, in practice such alleles do not occur. In the simulations outlined in this paper all alleles remain within the range [-5.41,5.29].

randomly distributed among the fittest t members of generation N. During each exposure, the set of meaning-signal pairs of the N generation agent is used to train the generation $N + 1$ agent. The backpropagation algorithm was used to implement this learning process[6], with the starting point for learning being the connection weights specified in the learning agent's genotype. The learning agent's communication system will therefore be determined, at least to some extent, by the interactions between the processes of genetic transmission via breeding and cultural transmission via learning.

4   Results for imitators

In this section results generated by the model under two main parameter settings are presented:

1. Imitation learning and natural selection ($p = 100$, $20 \leq b \leq 100$, $t = 100$, $0 \leq e \leq 200$) (Section 4.1).

2. Selective imitation learning and natural selection ($p = 100$, $20 \leq b \leq 100$, $20 \leq t \leq 100$, $0 \leq e \leq 200$) (Section 4.2).

Under both general configurations the questions is asked: do the populations converge on communication systems resulting in high levels of communicative accuracy within the population? This was assessed by running 10 simulations under each parameter setting for a fixed number of generations (1000) and measuring the average communicative accuracy of the population, as defined in Section 3.1.1, for the last 10 generations of each run.

4.1   Imitation learning and natural selection

In this section we investigate the accuracy of the emergent communication systems in populations of size $p = 100$ for various numbers of learning episodes e and various amounts of selective pressure b. e ranges from 0 (no learning) to 200, and b ranges from 100 (no selection pressure on breeding) to 20 (very strong selection pressure on breeding). For all simulations outlined in this section, $t = 100$ – immature agents potentially sample and learn from the communicative behavior of the entire preceding generation, regardless of fitness.

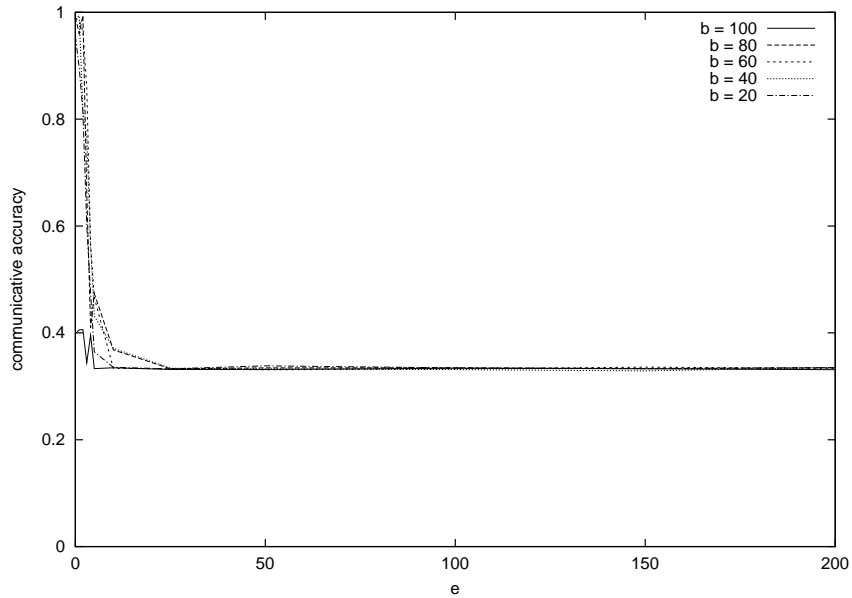[6]A learning rate of 0.5 and momentum of 0 were used.

Figure 3: Average communicative accuracy in populations of imitator agents after 1000 generations, for various values of e and b. Each point represents the average of 10 simulations.

The average communicative accuracy of the populations for the full range of values of e are shown in Figure 3. Figure 4 shows the communicative accuracy of populations where $0 \leq e \leq 25$. For extremely low values of e optimal communication systems can emerge given selection pressure on breeding ($b < 100$). Average communicative accuracy rapidly tails off as e increases. For $e > 5$ there is little difference between simulations where breeding is random ($b = 100$) and breeding is non-random ($b < 100$) and for $e > 25$ there is no difference, with all populations converging on communication systems resulting in average communicative accuracy of 0.33. This corresponds to the chance level of performance of a population attempting to communicate three meanings with a fully ambiguous communication system. How can the emergence of these suboptimal communication systems be explained?

Table 1 shows the success of imitator networks at acquiring systems of differing levels of ambiguity for a given number of exposures, e. As can be seen from the Table, systems exhibiting a higher degree of ambiguity are easier to acquire than systems exhibiting a lower degree of ambiguity for all
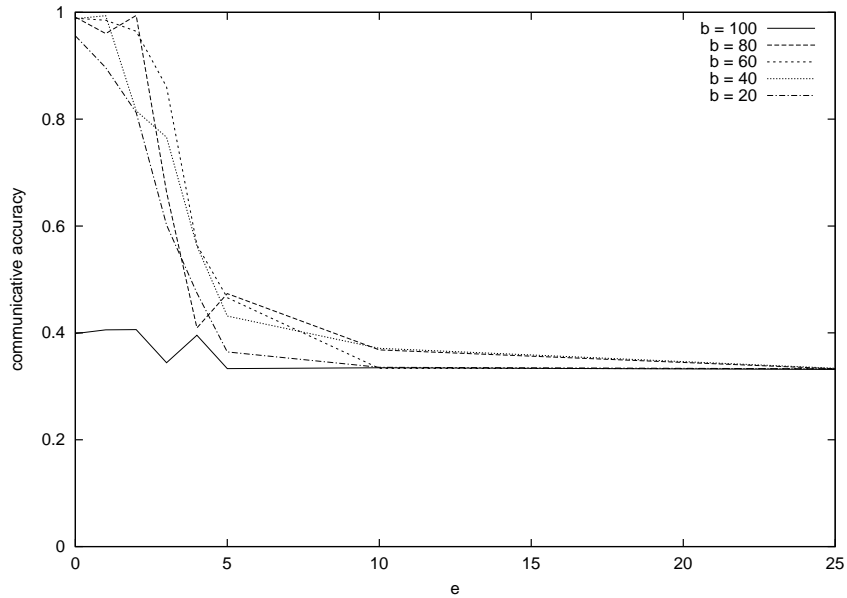
15

Figure 4: Average communicative accuracy in imitator populations, for small values of e.

values of e. The learning bias of imitator agents results from the imitator network architecture, as is discussed in Section 6.1. It should be noted that systems exhibiting a higher degree of ambiguity have an additional advantage, in that every exposure to an ambiguous system contains multiple exposures to the ambiguous signal. However, this is not the key factor, as can be seen by comparing success rates for fully ambiguous systems with low values of e to success rates for unambiguous systems with higher values of e – for example, five exposures to a fully ambiguous system (15 exposures to the ambiguous signal) gives a rate of success which can only be matched by 150 exposures to an unambiguous system, which gives 150 exposures to each of the three unambiguous signals.

As can be seen from Table 1, for extremely low e ($e < 5$) even fully ambiguous systems can not be reliably acquired. Figure 4 shows that levels of communicative accuracy significantly above the random level are only observed given $e < 5$ and selection pressure on breeding. In these circumstances learning is effectively disabled and natural selection is free to evolve the population's communication systems, resulting in levels of communicative accuracy above the chance level, with optimal communication for $e \leq 2$.

16

e	System Type		
	Fully Ambiguous	Partially Ambiguous	Unambiguous
1	24.9	0.4	0
2	50.1	0.6	0
3	76.0	1.6	0
4	90.9	1.8	0
5	97.9	1.4	0
10	100	0.4	0
25	100	1.5	0
50	100	32.9	13.3
100	100	93.1	82.8
150	100	99.5	97.9
200	100	100	99.8

Table 1: Percentage success at acquiring various types of system with various values of e. These results were empirically derived by generating 100 random communication systems of each type and training 100 networks with small random initial weights in the range $[-1, 1]$ on each system.

These optimal communication systems disappear given $e > 5$. Why is natural selection not developing optimal communication systems under these circumstances? It appears that the process of cultural transmission is overriding the process of natural selection. Fully ambiguous systems are always the easiest class of system to learn, and are therefore more likely to pass intact through the learning process. The repeated cultural transmission of communication results in the elimination of communication systems which do not conform to the learning biases of the agents – there is cultural selection in favor of systems which conform to the learner biases. In the case of imitator agents the learning bias happens to be in favor of communication systems which are extremely poor in terms of communicative accuracy. As we will see in Section 5, a different agent model leads to a different learning bias.

The learning bias of the agents in favor of ambiguous systems is a property of the imitator architecture (see Section 6.1), rather than the learning rate, which merely determines the strength of the bias for a particular value of e. Therefore the precise value of e at which cultural transmission overrides natural selection is dependent on the particular learning rate used – for example, if a lower learning rate was used then we would observe cultural transmission disabling natural selection only at larger values of e.

Importantly, however, natural selection would still be overridden by cultural selection at some point.

Why does natural selection not counteract the cultural adaptation of the communication systems to the learner biases and weed out poor communicators? Learning in the phenotype masks an individual's genetic makeup — with $e > 5$, no matter how good an agent's genes are, their effects are likely to be overtaken by learning, which almost fully determines an agent's communicative behavior. Shielding (Ackley and Littman, 1992) prevents natural selection from identifying good gene combinations and weeding out bad gene combinations. Only when $e < 5$ is natural selection not disabled by shielding of the genotype.

There are certain combinations of genes which make learning a particular communication system impossible — an agent's genes constitute the starting point for learning, and the backpropagation algorithm is sensitive to initial weights to a certain degree. Genetic drift does occasionally result in small numbers of agents being born whose genes are so good they cannot learn fully ambiguous communication systems. However, these agents must still communicate with their neighbors, and if those neighbors use a fully ambiguous system then using a better system to communicate with them yields no benefit[7]. The good gene combinations do not survive for long due to inter-breeding with agents whose genes allow them to acquire fully ambiguous systems. Cultural transmission leads to cultural stagnation in the simulated populations — the biases of the learners favor fully ambiguous communication systems and natural selection is powerless to counteract this.

4.1.1   Imitation learning and collapse

Cultural transmission not only prevents the development of an optimal communication system in the simulated populations — it prevents the maintenance of such a system. Figure 5 shows the average communicative accuracy of populations of imitator agents who start out with a shared, optimal, innate communication system — all the agents in the initial population have a hand-selected set of genes which encode an unambiguous communication system. $e = 200$ for all simulations in Figure 5. Various amounts of selection pressure (b) are used. $t = 100$, as in the simulations in the previous section. As can be seen from Figure 5, all populations collapse from using an unambiguous communication system to using a fully ambiguous communication

---

[7]Preferred interaction with genetically-related individuals might alleviate this problem somewhat, but was not investigated here.
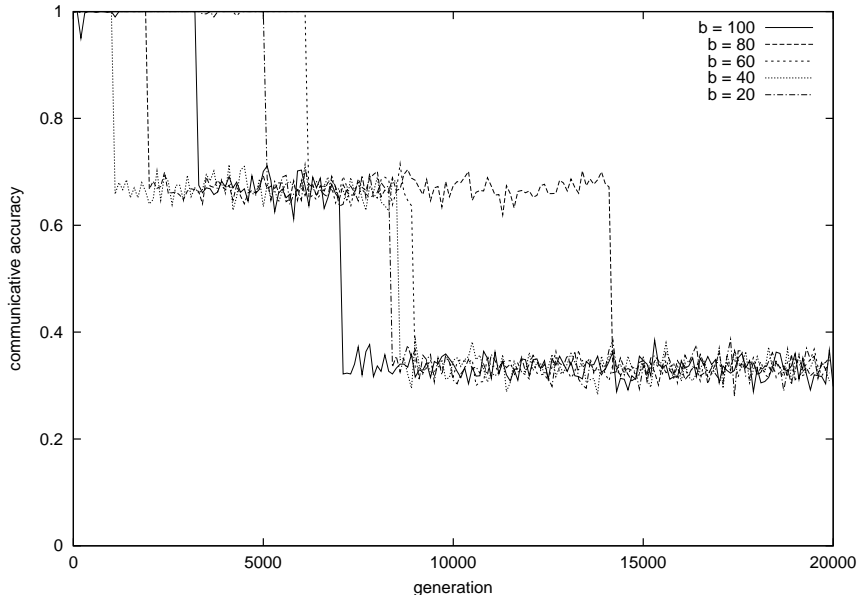
system within 15000 generations.

Figure 5: Average communicative accuracy of imitator populations over time, where $t = p = 100$.

As discussed above, learning in the phenotype almost completely masks an agent's genes but there are certain combinations of genes which make learning a particular communication system impossible. In each simulation shown in Figure 5 an agent will eventually be born whose genes are so bad that they cannot learn the unambiguous communication system in use by the rest of the population. This individual will learn a partially ambiguous or fully ambiguous communication system instead. Such agents will be unlikely to breed, given that their fitness will usually be lower than other agents in the population. Suboptimal communicators do have a negative effect on the fitness of optimally-communicating agents, given that those optimally-communicating agents suffer a penalty for not understanding or being understood by suboptimal communicators, although this will not usually depress the population's fitness enough to allow a suboptimal communicator to breed. However, while such individuals are unlikely to breed, their communication systems *will* be observed and learned from by agents in the next generation.

Table 2 shows the percentages of agents with random connection weights

19

System Type	% population
Unambiguous	2
Partially Ambiguous	25
Fully Ambiguous	73

Table 2: Percentage of initial population using systems of each degree of ambiguity.

in the range $[-1, 1]$ using communication systems of the three levels of ambiguity. Agents with random connection weights clearly tend to have a fully ambiguous communication system. This approximates the response of imitator agents to training on conflicting communication systems – training on conflicting data effectively randomizes the connection weights in the agents' network.

As discussed above, agents with bad genes will occasionally occur in the population due to genetic drift. The communication system of such agents will be observed by other agents in the subsequent generation. These individuals run the risk of acquiring a suboptimal communication system due to the randomizing effect of conflicting training data. If they do acquire a suboptimal system they will be unlikely to breed. Regardless of whether the suboptimal communicators breed or not, their communication systems will be observed by agents in the next generation. As increasing levels of ambiguity result in more successful cultural transmission, suboptimal communication systems spread through the population like a virus due to the processes of cultural transmission, until the whole population converges on a fully ambiguous communication system. Once again, natural selection is powerless to stop this process.

Note that $e = 200$ represents the best-case scenario for learning agents. $e = 200$ results in the highest level of learnability for unambiguous communication systems and also ensures that, at the early stages of collapse, suboptimal communication systems will constitute only a small part of an agent's observations. In populations where $e \leq 2$ the collapse phenomenon does not occur, but as discussed above the behavior of these populations is entirely determined by natural selection – they cannot truly be called learning populations.

4.2 Selective imitation learning and natural selection

The phenomenon observed in the simulations outlined in the previous section is purely a result of the bias of the learners towards acquiring fully ambigu-

20

ous communication systems. In this section an additional learning bias, a preference of learners to learn from successful communicators, is added to the model.

Natural selection is implemented in this model by only allowing the top b members of the population to transmit their genetic information to the next generation via breeding. Similarly, only the fittest t members of the population transmit their communication systems culturally to the next generation, through the process of being observed and learned from. In the simulations in the previous section $t = p$ – all members of the population participate in cultural transmission, regardless of fitness. However, in this section simulations are described where $t < p$ – an agent's participation in cultural transmission depends to some extent on their fitness.

In these populations of discriminating learners there are therefore 3 potential selection pressures operating on the evolving populations and communication systems:

1. *Natural selection* (when $b < p$), operating on genetic transmission, favoring genes whose phenotype realizations are successful communicators.

2. *Cultural selection for learnability*, operating on cultural transmission, favoring communication systems which conform to the learning bias for fully ambiguous systems.

3. *Cultural selection for communicative success* (when $t < p$), operating on cultural transmission, favoring communication systems which result in successful communication.

Selection pressures 1 and 3 are clearly related, although operating on different modalities of transmission. Selection pressures 2 and 3 operate in the same modality of transmission, but are in direct competition.

Figures 6, 7, 8, 9 and 10 show the communicative accuracy of the emergent communication systems in populations of size $p = 100$, for various values of b, t and e.

The addition of the cultural selection pressure for communicative success has clearly failed to have a significant impact on the emergent communication systems – for $e > 10$ the populations' communication systems tend to be fully ambiguous. For very low values of t and high values of e partially ambiguous communication systems do occasionally emerge. However, the behavior of the population is still dominated by the intrinsic learning bias of the agents, which favors fully ambiguous systems.
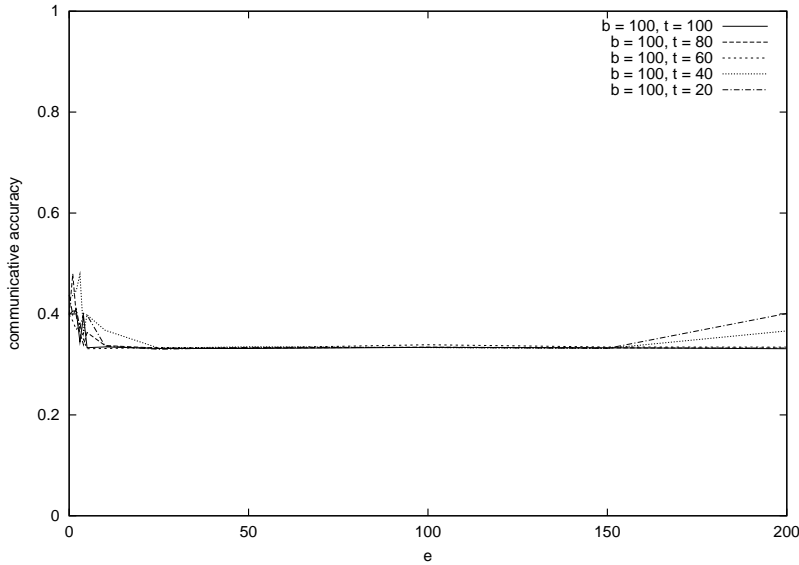
21

Figure 6: Average communicative accuracy in populations of imitator agents after 1000 generations, for various values of e and t ($b = 100$). Each point represents the average of 10 simulations.

4.2.1 Selective imitation and collapse

Figure 11 shows the communicative accuracy over time of imitator populations, for various values of b and t. $p = 100$ and $e = 200$ in all these simulations. As in the simulations outlined in Section 4.1.1, initially the populations are genetically converged on an optimal communication system.

As can be seen from Figure 11, the addition of selective imitation in the runs where $t < 100$ has failed to prevent the populations from moving away from the initial optimal communication system – in 3 runs the population has converged on a fully ambiguous system yielding chance levels of communicative accuracy, whereas in 2 runs ($b = t = 60$ and $b = t = 40$) the population has converged on a partially ambiguous communication system.

The populations have failed to maintain the original optimal system for the same reason as the populations discussed in Section 4.1.1 – shielding allows mutations to accumulate in the population, those mutations eventually prevent some agents acquiring the optimal communication system and the observation of that suboptimal behavior disturbs more individuals in subsequent generations. This results in the rapid spread of the communication
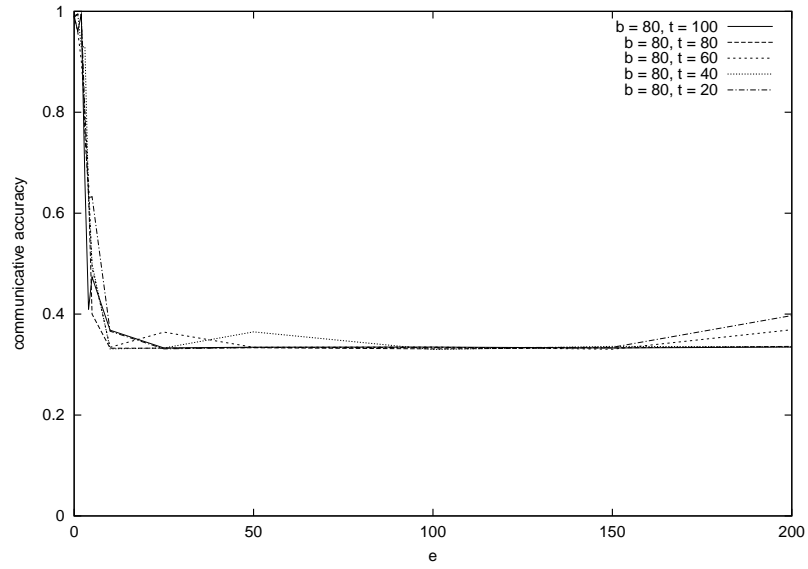
Figure 7: Average communicative accuracy in populations of imitator agents for various values of e and t ($b = 80$).
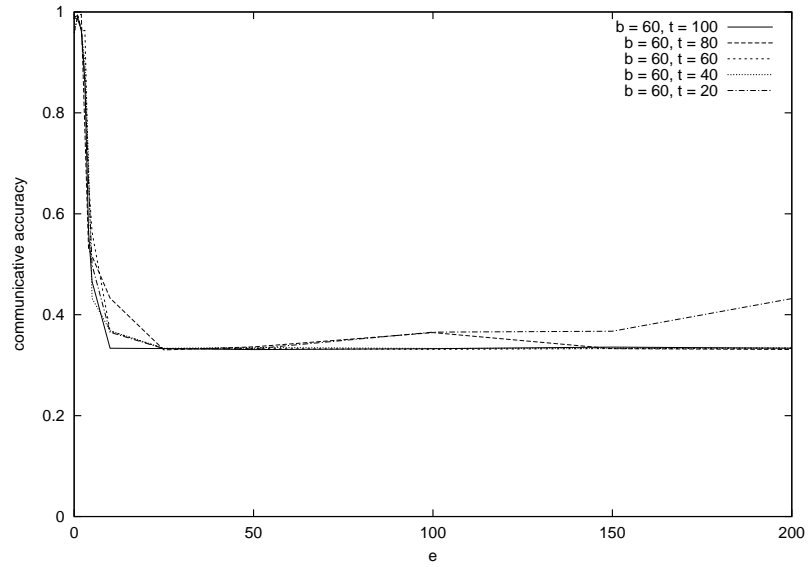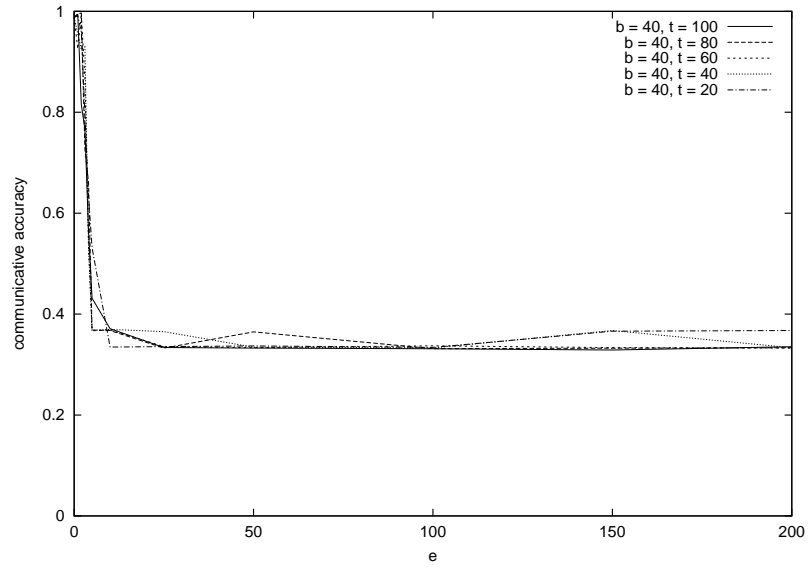

Figure 8: Average communicative accuracy in populations of imitator agents for various values of e and t ($b = 60$).
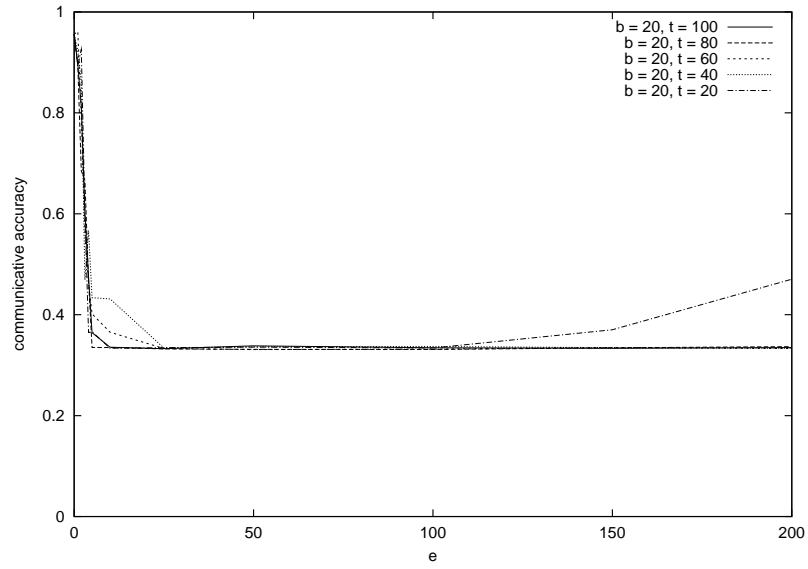
23

Figure 9: Average communicative accuracy in populations of imitator agents for various values of e and t ($b = 40$).

Figure 10: Average communicative accuracy in populations of imitator agents for various values of e and t ($b = 20$).
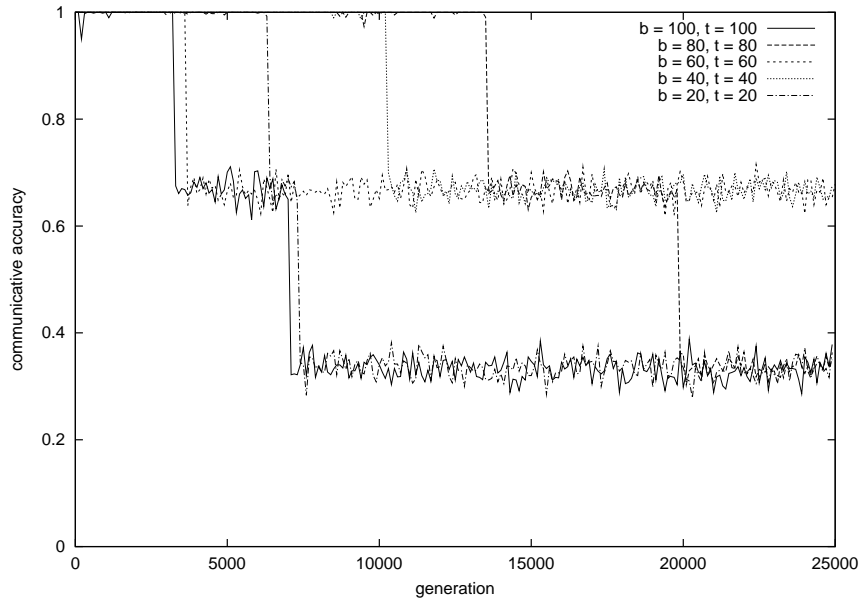
24

Figure 11: Average communicative accuracy of imitator populations over time.

systems which are most easy to acquire, which happen to be suboptimal in terms of fitness. Lower values of t makes it less likely that individuals with suboptimal communication systems will transmit those systems, therefore making it less likely that the population will collapse. However, as Figure 11 shows, collapses can occur given a sufficient number of mutation events at a single generation.

5 Tailoring the learning bias

In the simulations outlined in the previous section there were two learning biases — the intrinsic bias of the learners, which favors increased ambiguity, and the bias in favor of learning from successful communicators, which depended on t. In this section the model of a communicative agent is revised to build in a learning bias towards optimal, unambiguous communication systems. This bias results in the rapid and reliable emergence of such systems.

25

5.1 The new communicative agent

As outlined in Section 3.2, the communicative agents in all previous simulations were feedforward neural networks mapping from input meanings to output signals. Signal production for these imitator agents was merely a matter of propagating an input meaning pattern of activation through the network to produce an output signal. Reception was achieved by presenting all communicatively relevant meanings and selecting the meaning which maximizes confidence in the received signal. These networks are strongly biased in favor of acquiring fully ambiguous communication systems.

The new model of a communicative agent has exactly the same basic form as the imitator model, being a three-layer feedforward neural network. However, the crucial difference is that the new networks, which will be referred to as *obverter* (Oliphant and Batali, 1997)[8] networks, map from input signals to output meanings — the direction of the mapping has been reversed. Production and reception in these obverter networks operate as follows:

Production: Each of the set of possible signals is propagated through the network, producing a real-numbered output pattern of activation for each signal. The signal which produces the meaning closest to the meaning to be communicated, as determined by the confidence measure outlined in Section 3.2.1, is used to communicate the given meaning (as for imitator reception).

Reception: The received signal pattern is propagated forward through the network and the output pattern of activation is thresholded to produce a binary pattern of activation corresponding to that agent's interpretation of the received signal (as for imitator production).

The learning biases of these agents are shown in Tables 3 and 4. As can be seen from Table 3 these agents are strongly biased against learning fully ambiguous and partially ambiguous communication systems. Somewhat surprisingly, learnability never reaches 100%, even for unambiguous communication systems. It appears that certain unambiguous systems are unlearnable by obverter agents, while certain unambiguous systems are 100% learnable. The pattern to this learnable-unlearnable distinction is not important in this paper – the key point is that certain unambiguous systems are highly learnable whereas partially ambiguous and fully ambiguous systems are less learnable.

---

[8]Obverter networks are the equivalent of what Hurford (1989) termed Saussurean learners.

Table 4 shows that obverter networks with random weight settings in the range [-1,1] are strongly biased towards unambiguous communication systems — as discussed in Section 4.1.1, these random weight biases approximate the response of networks to exposure to conflicting communication systems.

e	System Type		
	Fully Ambiguous	Partially Ambiguous	Unambiguous
1	0.1	0.2	0.3
2	0.2	0.2	0.4
3	0.1	0.4	0.4
4	0	0.3	0.6
5	0	0.5	0.6
10	0	1.0	1.6
25	0	4.6	7.8
50	0	8.9	31.3
100	0	11.1	53.9
150	0	11.0	51.7
200	0	18.0	55.0

Table 3: Percentage success of obverter agents at acquiring various types of system with various values of e.

System Type	% population
Unambiguous	65
Partially Ambiguous	33
Fully Ambiguous	2

Table 4: The percentage of a population of obverter agents with small, random weights using communication systems of the given type

5.2   Obverter learning results in optimal communication

In the simulation runs plotted in Figure 12, the new obverter learner is substituted for the imitator learner used in previous sections. Excluding the change in the agent model, all other simulation details are identical to the simulation runs described in Section 4.1 — specifically, $p = 100$ and $t = 100$.
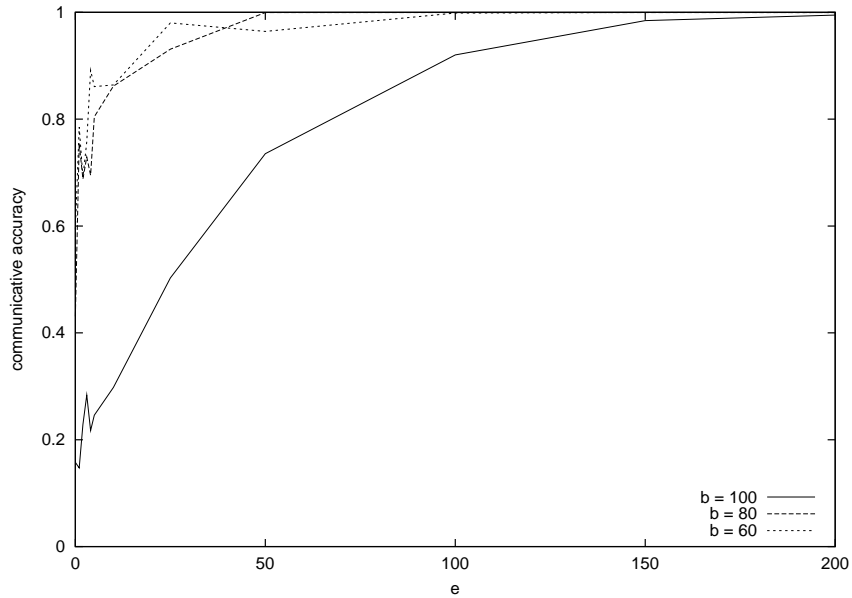
Figure 12: Average communicative accuracy in populations of obverter agents after 1000 generations, for various values of e and b ($t = 100$). Each point represents the average of 10 simulations.

Figure 12 shows a clear difference between simulation runs with no selection pressure on breeding ($b = 100$) and those with selection pressure on breeding ($b < 100$). For the runs with no natural selection, communicative accuracy is low with low e and increases as e increases. For $e \geq 100$ the populations reliably converge on optimal, unambiguous communication systems. As shown in Table 3, (a subset of the set of) unambiguous systems are highly learnable for these values of e and have a significant learnability advantage over ambiguous systems. As a result, ambiguous systems are selected against during cultural transmission until the populations converge on unambiguous systems.

In the runs with selection pressure on breeding the populations have higher communicative accuracy with $e < 100$ due to the development of innate communication systems through natural selection. As e increases the communicative accuracy of the population increases, indicating a positive interaction between natural selection and learning – natural selection favors genotypes which improve the learnability of unambiguous communication systems in use in the populations (i.e. the Baldwin Effect (Baldwin, 1896)).

As e increases this interaction decreases, and when $e \geq 100$ the interaction all but disappears.

Furthermore, populations of obverter agents are capable of maintaining such optimal communication systems indefinitely – unlike the populations shown in Sections 4.1.1 and 4.2.1 they do not suffer from the collapsing problem, even in the absence of selection pressure on breeding.

6 The key learning bias

Imitator agents cannot create or maintain optimal communication systems, even given a helping hand from natural selection, and obverter agents can construct and maintain optimal communication systems, given sufficient exposure, without help from natural selection. This is due to the inherent learning biases of the two types of agents (summarized in Tables 1, 2, 3 and 4). The relationship between these biases and the structure of the networks is explored in detail in Section 6.1. The key biases identified in Section 6.1 are discussed in terms of other models in Section 6.2.

6.1 The learning bias explored

In the terms of this paper, optimal communication systems are unambiguous mappings from meanings to signals – one-to-one (or injective) functions. Suboptimal systems are many-to-one or all-to-one functions. In terms of production and reception functions $p(m)$ and $r(s)$, in an optimal communication system:

1. $p(m)$ should be an injective function.

2. $r(s)$ should be a superset of the inverse of $p(m)$.

These two restrictions guarantee that every meaning is expressed using a distinct signal and that the reception process maps signals back on to the meanings they were originally intended to convey.

Feedforward neural networks learn many-to-one functions. Due to the deterministic nature of the feedforward propagation of activation values they cannot learn one-to-many mappings. The easiest function for a network to acquire is therefore an all-to-one mapping from inputs to outputs, the hardest learnable function is an injective (one-to-one) function and one-to-many mappings are unlearnable. The reversal process used to model reception behavior for imitators and production behavior for obverters is

similarly biased – it generates a function, which may be injective or many-to-one, based on the function the feedforward network has acquired. In general, if the network has acquired a function $f(x)$ which has a range y, then the reversal process ensures that element $y_i \in y$ will map onto a single element $x_i \in x$ such that $f(x_i) = y_i$ – in simple terms, the reversal process deterministically reverses the function acquired by the network.

In imitator agents the feedforward network learns functions from meanings to signals – it learns $p(m)$. Since it is a feedforward network it will be biased towards acquiring a many-to-one or all-to-one $p(m)$. As illustrated in Figure 13 and discussed in the caption, the maximally stable $p(m)$ for imitator agents is therefore an all-to-one fully ambiguous function. Imitators therefore do not have a bias in favor of the 1st feature (above) of an optimal system. Reception in imitators will be based on their acquired $p(m)$ – as shown in Figure 13, in the case of an all-to-one $p(m)$, in $r(s)$ the signal s_i that constitutes the range of $p(m)$ will map onto a single element from m. Therefore a population of imitators agents will tend to produce the same signal for every meaning and interpret the ambiguous signal as communicating one arbitrary meaning. This situation results in performance equivalent to random guessing.

In obverter agents the feedforward network learns functions from signals to meanings – it learns $r(s)$. As illustrated in Figures 14 and 15 the only culturally stable system has an injective $p(m)$ (point 1 above) and an $r(s)$ which includes at least the inverse of $p(m)$ (point 2 above). Obverter agents are therefore strongly biased in favor of acquiring systems with the properties of optimal communication systems.

6.2 Other models

As mentioned in Section 2, there are several other models where cultural processes result in the emergence of optimal communication. Do the learning mechanisms used in these models include biases in favor of the two properties outlined above? This kind of analysis often requires a great deal of familiarity with the model involved. However, this bias can be identified in certain other models.

Beginning with the models involving cultural transmission and no natural selection, the two key biases can be observed in the auto-associator networks of Hutchins and Hazelhurst (1995), the "obverter" learner of Oliphant and Batali (1997), which is capable of constructing an optimal system of communication from random behavior, (but not the "imitator", which is not) and in the "constructor" agents of Smith (2002), capable of construct-
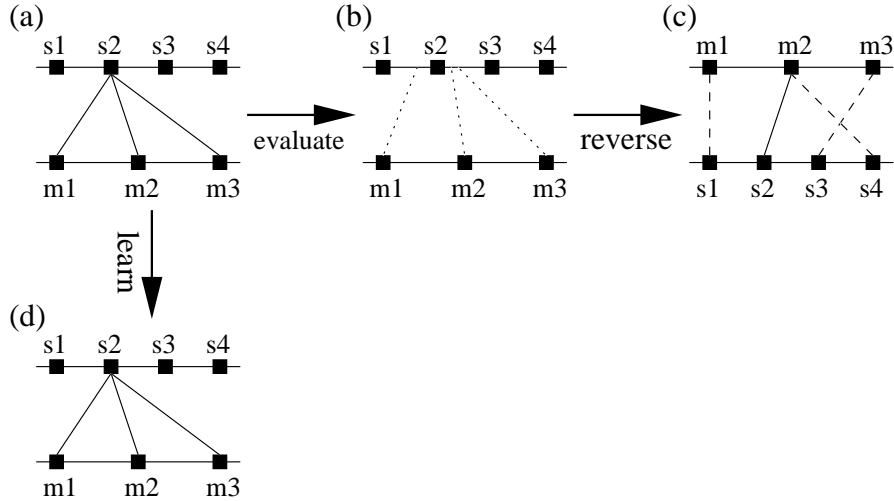
30

Figure 13: (a) is a representation of an imitator agent's feedforward network encoding an all-to-one $p(m)$ mapping three meanings onto a single signal, $s2$. The function from a domain of real numbers (input unit activations) to a codomain of real numbers (output unit activations) is represented by two lines, the lower line representing the domain, the upper representing the codomain. Squares represent particular points on the line corresponding to binary meanings or signals. Associations are shown with solid lines between elements in the domain and elements in the codomain. (b) represents the confidence-measuring step of the reversal process for the network underlying (a). In order to decide $r(s2)$, the real-number values of $p(m1)$, $p(m2)$ and $p(m3)$ are calculated. These real-numbered mappings are represented by dotted lines in (b). (c) represents the $r(s)$ derived from applying the reversal process to (a). $r(s2) = m2$ because $m2$ mapped closer to $s2$ than any other m in (b). The other associations are effectively random. The random nature of these mappings is represented by dashed lines. (d) represents the function acquired by an imitator network exposed to behavior generated by (a) – as it is an all-to-one function between meanings and signals it is easily learned by imitator agents. This is in fact the only stable function for imitators.
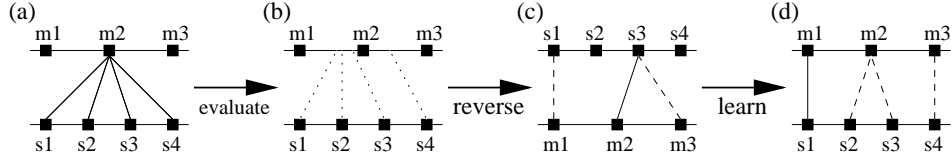
31

Figure 14: (a) represents an all-to-one $r(s)$ encoded in an obverter agent's feedforward network. As obverters map from signals to meanings this is the most learnable $r(s)$. (b) represents the confidence-measuring step of reversing this $r(s)$ to generate a $p(m)$ – as before, real-number mappings are shown as dotted lines. (c) shows the $p(m)$ derived from (a). $p(m2) = s3$ as $s3$ mapped closest to $m2$ in (b). The other associations are essentially random. The $p(m)$ in (c) produces the meaning-signal pairs $\{(m1, s1), (m2, s3), (m3, s3), \}$. The order of the meanings and signals in these pairs are reversed to train the next generation of obverter networks. (d) shows the $r(s)$ resulting from training an obverter network on the signal-meaning pairs $\{(s1, m1), (s3, m2), (s3, m3), \}$. $r(s1) = m1$, as expected. However, feedforward networks cannot learn one-to-many mappings so $r(s3)$ is effectively randomly assigned to a signal, in this case $m2$. As $s2$ and $s4$ are not represented in the training set they are effectively randomly assigned mappings. Notice that the mapping in (a) has been destroyed in (d) – (a) is not a culturally stable mapping.
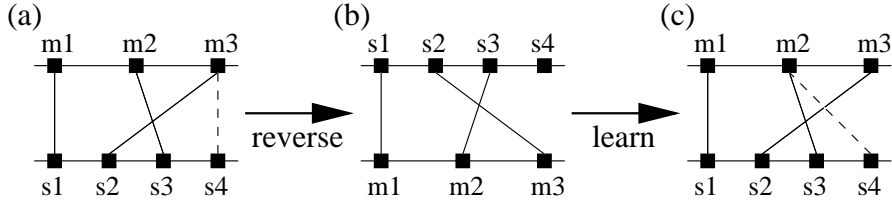

Figure 15: Only an unambiguous $p(m)$ is stable for obverter agents. (a) represents an obverter agent's $r(s)$. (b) is the $p(m)$ derived from reversal of (a) – it is an injective function. (c) illustrates that the $r(s)$ resulting from training the next generation of agents on data produced by (b) is effectively similar to (a) and will therefore lead to (b) once again – (a) and (b) are culturally stable. The only unstable aspect is the floating synonym $s4$. This synonym is highly unlikely to interfere with the mapping in (b) and the floating synonym phenomenon can be observed in the other obverter models outlined in the main text.

ing an optimal system, but not in any non-constructor agents. The class of constructor agents in these papers includes the Hebbian learner of Oliphant (1999), also capable of constructing an optimal system. The neural networks in Batali (1998) are essentially obverter agents, although the importance of their inherent bias is not identified.

The key bias can also observed in the model outlined in Kvasnička and Pospíchal (1999), which involves cultural transmission and natural selection. Given that the neural networks used by Kvasnicka and Pospichal are practically identical to the obverter network outlined in this paper, the behavior of their populations can probably be explained purely in terms of cultural processes. The absence of a contrastive learning bias or variance in natural selection pressure in their model obscures this fact. The models of MacLennan and Burghardt (1994) and Kirby and Hurford (1997) do have learning biases which are specifically directed towards optimizing acquired systems, but these models either model communication at a different level (Kirby and Hurford, 1997) or can be criticized for their use of reinforcement learning (MacLennan and Burghardt, 1994). Finally, it is worth pointing out that the theoretical models proposed in Pinker and Bloom (1990) and Dor and Jablonka (2000) do not take into account the role of learning biases in cultural evolution. The computational model outlined in this paper suggests that the consequences of such biases, which may be non-obvious, need to be taken into consideration.

The two key biases appear to be common, or at least recurring, in learning mechanisms capable of constructing optimal communication systems in the modeling literature. Oliphant (1999) claims that this kind of bias is in fact widespread in the natural world. But is there evidence that any species is actually biased in favor of learning one-to-one mappings in the domain of communication? As mentioned in Section 2, it is doubtful whether experience plays a role in determining the structure of communication in any non-human primate. However, the sole species in which experience definitely plays some role (humans) does appear to exhibit this bias. It has been suggested that word acquisition in humans is guided by the Contrast Principle (Clark, 1988), a bias in favor of one-to-one mappings between meanings and words. While Bloom (2000) suggests this principle is part of the human theory of mind, it can be conceived of as a communication-specific learning bias. The model suggests that some of the nature of the human communication system may be explicable in terms of cultural processes resulting from the iterated application of human learning biases. The role of such a learning bias in the evolution of syntax is a possible subject for future research - is such a bias sufficient for the cultural evolution of syntax, as well as simple

communication? If not, what other components are required?

7 Conclusions

This paper outlines a computational model of the emergence of communication in a population of communicative agents. As previous work suggests, natural selection alone is capable of evolving optimal, innate communication systems in such populations. However, the addition of cultural transmission of communication systems does not necessarily assist the emergence of optimal communication systems. The biases of the learners involved in the cultural transmission process result in cultural selection — communication systems which conform to the biases of the learners are more likely to be successfully transmitted than communication systems which do not.

The results of simulations in which cultural selection is in direct conflict with natural selection are outlined in Section 4.1. In these circumstances, cultural selection resulting from the intrinsic biases of the agents proves to be the determining factor in the emergent behavior of the simulated populations. This is a clear case of what Durham (1991) would term gene-culture *opposition*. In Section 4.2, a second cultural selection pressure was introduced which was in direct conflict with the intrinsic learning biases of the simulated agents. This secondary pressure failed to override the intrinsic biases of the learners. In Section 5 the model of the learning agent was modified to build in a bias towards optimal communication systems. In populations of such agents optimal communication systems rapidly and reliably emerged, due to the cultural selection pressures arising from the learners' biases. As discussed in Section 6, these learning biases can be explained in terms of the networks' structure.

This model has several limitations. The model of communication used, with three unstructured meanings and eight unstructured signals, is very simple, although there is no reason to expect these results not to hold for more complex models of communication. The absence of any environment outwith the agents means that communicative accuracy must be measured at the agent-internal representations. While this is not unreasonable for referential communication systems, a behavior-based measure of communicative success would be more appropriate for modeling non-referential communication. Finally, natural selection has a fairly limited role to play in this model, a matter which is discussed further below.

These results have implications for both computational modeling of gene-culture interactions and research into the origins and evolution of language.

34

For computational modelers, the clear implication is that a particular choice of agent model or learning model can have a fundamental impact on the behavior of the systems as a whole. This suggests that modelers should be aware of how specific their results are to their model of learning and be prepared to justify their model of learning and its associated biases in terms of the real-world system that is being modeled. The bias in favor of one-to-one mappings associated with the the obverter agent corresponds to a learning bias observed in humans (the Contrast Principle, as discussed in Section 6.2), suggesting that, in terms of learning bias, the obverter model is preferable to the imitator model as a model of human language learning.

More generally, the simulations outlined in this paper suggest that research into the origins and evolution of language should not underestimate the role of cultural selection in this process. These simulations give an illustration of the fact that the learning biases of individual learners can have profound and far-reaching effects when placed in the context of iterated cultural transmission, and that in certain circumstances these cultural processes can effectively nullify the influence of natural selection during genetic transmission.

This is not to say that natural selection can have no role in the explanation of the evolution of language. In the simulations outlined in this paper, natural selection is restricted to tinkering with the starting point for the learning process. In a more realistic model, all aspects of the learning apparatus would be genetically transmitted. It would therefore be possible for natural selection to develop learning algorithms, and therefore modify learner biases and determine the precise nature of cultural selection occurring during cultural transmission. Under these circumstances, can natural selection identify learning algorithms which result in cultural selection for optimal communication? Preliminary modeling work in this area (Smith, 2001; Smith, in prep) suggests that natural selection may be unable to reliably identify such biases due to the significant delay between the appearance of the beneficial genes and the establishment of widespread beneficial culture. The story of the evolution of language may therefore be best told in two parts, with the development of the necessary pre-adaptation of an appropriate learning mechanism occurring in a geological time scale and the development of language parasitic on this learning mechanism occurring in a historical time scale.

Acknowledgements

References

Ackley, D. and M. Littman (1992). Interactions between learning and evolution. In C. Langton, C. Taylor, J. Farmer, and S. Rasmussen (Eds.), *Artificial Life 2*, pp. 487–509. Redwood City, CA: Addison-Wesley.

Ackley, D. and M. Littman (1994). Altruism in the evolution of communication. In R. Brooks and P. Maes (Eds.), *Artificial Life 4: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, pp. 40–48. Redwood City, CA: Addison-Wesley.

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

Batali, J. (1994). Innate biases and critical periods: Combining evolution and learning in the acquisition of syntax. In R. Brooks and P. Maes (Eds.), *Artificial Life 4: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, pp. 160–171. Redwood City, CA: Addison-Wesley.

Batali, J. (1998). Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy, and C. Knight (Eds.), *Approaches to the Evolution of Language: social and cognitive bases*, pp. 405–426. Cambridge: Cambridge University Press.

Batali, J. (in press). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In E. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge: Cambridge University Press.

Belew, R., J. McInerney, and N. N. Schraudolph (1992). Evolving networks: Using the genetic algorithm with connectionist learning. In C. Langton, C. Taylor, J. Farmer, and S. Rasmussen (Eds.), *Artificial Life 2*, pp. 511–547. Redwood City, CA: Addison-Wesley.

Bloom, P. (2000). *How Children Learn the Meanings of Words*. MIT Press.

Bloom, P. and L. Gleitman (2001). Language acquisition. In R. A. Wilson and F. Keil (Eds.), *The MIT Encyclopaedia of the Cognitive Sciences*. Cambridge, MA: MIT Press.

Brighton, H. and S. Kirby (2001). The survival of the smallest: Stability conditions for the cultural evolution of compositional language. In J. Kelemen and P. Sosik (Eds.), *Advances in Artificial Life (Proceedings of the 6th European Conference on Artificial Life)*. Heidelberg: Springer-Verlag.

Brown, R. and C. Hanlon (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (Ed.), *Cognition and the Development of Language*. New York: Wiley.

Bullock, S. (1997). An exploration of signalling behaviour by both analytic and simulation means for both discrete and continuous models. In P. Husbands and I. Harvey (Eds.), *Fourth European Conference on Artificial Life*, pp. 454–463. Cambridge, MA: MIT Press.

Cangelosi, A. (1999). Modelling the evolution of communication: From stimulus associations to grounded symbolic associations. In D. Floreano, J. D. Nicoud, and F. Mondada (Eds.), *Advances in Artificial Life*, Number 1674 in Lecture notes in computer science. Springer.

Cangelosi, A. and D. Parisi (1998). The emergence of a 'language' in an evolving population of neural networks. *Connection Science 10*(2), 83–97.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1987). *Knowledge of Language: Its Nature, Origin and Use*. Dordrecht: Foris.

Christiansen, M. and J. Devlin (1997). Recursive inconsistencies are hard to learn: A connectionist perspective on universal word order correlations. In M. Shafto and P. Langley (Eds.), *Proceedings of the 19th Annual Cognitive Science Society Conference*, pp. 113–118. Lawrence Erlbaum Associates.

Clark, E. (1988). On the logic of contrast. *Journal of Child Language 15*, 317–335.

Di Paolo, E. (1997). An investigation into the evolution of communication. *Adaptive Behaviour 6*, 285–324.

Dor, D. and E. Jablonka (2000). From cultural selection to genetic selection: a framework for the evolution of language. *Selection 1(1-3)*, 33–55.

Durham, W. H. (1991). *Coevolution: Genes, Culture and Human Diversity*. Stanford, CA: Stanford University Press.

Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology 144*, 517–546.

Hauser, M. D. (1996). *The Evolution of Communication*. Cambridge, MA: MIT Press.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Cambridge, MA: MIT Press.

Hurford, J. R. (1989). Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua 77*, 187–222.

Hurford, J. R. (2000). Social transmission favours linguistic generalization. In C. Knight, M. Studdert-Kennedy, and J. Hurford (Eds.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pp. 324–352. Cambridge: Cambridge University Press.

Hurford, J. R. (2002). Expression/induction models of language evolution: dimensions and issues. In E. Briscoe (Ed.), *Linguistic Evolution Through Language Acquisition: Formal and Computational Models*. Cambridge: Cambridge University Press.

Hutchins, E. and B. Hazelhurst (1995). How to invent a lexicon: the development of shared symbols in interaction. In N. Gilbert and R. Conte (Eds.), *Artificial societies: the computer simulation of social life*. london: UCL Press.

Kirby, S. (1998). Fitness and the selective adaptation of language. In J. R. Hurford, M. Studdert-Kennedy, and C. Knight (Eds.), *Approaches to the Evolution of Language: Social and Cognitive Bases*, pp. 359–383. Cambridge: Cambridge University Press.

Kirby, S. (1999). Learning, bottlenecks and infinity: a working model of the evolution of syntactic communication. In K. Dautenhahn and C. Nehaniv (Eds.), *Proceedings of the aisb'99 symposium on imitation in animals and artifacts*. Society for the Study of Artificial Intelligence and the Simulation of Behaviour.

Kirby, S. (2000). Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In C. Knight, M. Studdert-Kennedy, and J. R. Hurford (Eds.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pp. 303–323. Cambridge: Cambridge University Press.

Kirby, S. (2001). Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation 5(2)*, 102–110.

Kirby, S. (in press). Learning, bottlenecks and the evolution of recursive syntax. In E. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge: Cambridge University Press.

Kirby, S. and J. R. Hurford (1997). Learning, culture and evolution in the origin of linguistic constraints. In P. Husbands and I. Harvey (Eds.),

Fourth European Conference on Artificial Life, pp. 493–502. Cambridge, MA: MIT Press.

Kirby, S. and J. R. Hurford (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi and D. Parisi (Eds.), *Simulating the Evolution of Language*. Springer Verlag.

Krebs, J. R. and R. Dawkins (1984). Animal signals: mind-reading and manipulation. In J. R. Krebs and N. B. Davies (Eds.), *Behavioural ecology: an evolutionary approach*. Oxford: Blackwell Scientific Publications.

Kvasnička, V. and J. Pospíchal (1999). An emergence of coordinated communication in populations of agents. *Artificial Life 5*(4), 319–342.

Levin, M. (1995). The evolution of understanding: a genetic algorithm model of the evolution of communication. *Biosystems 36*, 167–178.

Livingstone, D. and C. Fyfe (1999). Modelling the evolution of linguistic diversity. In D. Floreano, J. Nicoud, and F. Mondada (Eds.), *Advances in artificial life: fifth european conference on artificial life*, pp. 704–708. Berlin: Springer.

MacLennan, B. and G. Burghardt (1994). Synthetic ethology and the evolution of cooperative communication. *Adaptive Behaviour 2*, 161–187.

Noble, J. (1998). Evolved signals: Expensive hype vs. conspiratorial whispers. In C. Adami, R. Belew, H. Kitano, and C. Taylor (Eds.), *Artificial Life 6: Proceedings of the Sixth International Conference on Artificial Life*. Cambridge, MA: MIT Press.

Nolfi, S., J. L. Elman, and D. Parisi (1994). Learning and evolution in neural networks. *Adaptive Behavior 3*(1), 5–28.

Oliphant, M. (1996). The dilemma of saussurean communication. *BioSystems 37*, 31–38.

Oliphant, M. (1999). The learning barrier: Moving from innate to learned systems of communication. *Adaptive Behavior 7*(3/4), 371–384.

Oliphant, M. and J. Batali (1997). Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter 11*(1).

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

Reggia, J., R. Schulz, G. Wilkinson, and J. Uriagereka (2001). Conditions enabling the evolution of inter-agent signaling in an artificial world. *Artificial Life 7*(1), 3–32.

Rolls, E. T. and S. M. Stringer (2000). On the design of neural networks in the brain by genetic evolution. *Progress in Neurobiology 61*, 557–579.

Smith, K. (2001). The importance of rapid cultural convergence in the evolution of learned symbolic communication. In J. Kelemen and P. Sosik (Eds.), *Advances in Artificial Life (Proceedings of the 6th European Conference on Artificial Life)*, pp. 381–390. Heidelberg: Springer-Verlag.

Smith, K. (2002). The cultural evolution of communication in a population of neural networks. To appear in Connection Science.

Steels, L. (1999). *The Talking Heads Experiment*, Volume I. Words and Meanings. Antwerpen: Laboratorium. Special pre-edition.

Turkel, W. J. (in press). The learning guided evolution of natural language. In E. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge: Cambridge University Press.

Werner, G. and M. Dyer (1992). Evolution of communication in artificial organisms. In C. Langton, C. Taylor, J. Farmer, and S. Rasmussen (Eds.), *Artificial Life 2*, pp. 659–687. Redwood City, CA: Addison-Wesley.

Werner, G. and P. Todd (1997). Too many love songs: Sexual selection and the evolution of communication. In P. Husbands and I. Harvey (Eds.), *Fourth European Conference on Artificial Life*, pp. 434–443. Cambridge, MA: MIT Press.

Wheeler, M. and P. de Bourcier (1995). How not to murder your neighbour: Using synthetic behavioral ecology to study aggressive signaling. *Adaptive Behavior 3*(3), 273–309.

Worden, R. (in press). Words, memes and language evolution. In E. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge: Cambridge University Press.

Yamauchi, H. (2001). The difficulty of the baldwinian account of linguistic innateness. In J. Kelemen and P. Sosik (Eds.), *Advances in Artificial Life (Proceedings of the 6th European Conference on Artificial Life)*. Heidelberg: Springer-Verlag.

Zahavi, A. (1975). Mate selection - a selection for a handicap. *Journal of Theoretical Biology 53*, 205–214.

Zahavi, A. (1977). The cost of honesty (further remarks on the handicap principle. *Journal of Theoretical Biology 67*, 603–605.