The evolution of vocabulary

Kenny Smith

Language Evolution and Computation Research Unit,

School of Philosophy, Psychology and Language Sciences,

The University of Edinburgh,

Adam Ferguson Building, 40 George Square, Edinburgh EH8 9LL

kenny@ling.ed.ac.uk

Tel: 0131 650 3950

Fax: 0131 650 3962

Abstract

Human language is unique among the communication systems of the natural world. The vocabulary of human language is unique in being both culturally-transmitted and symbolic. In this paper I present an investigation into the factors involved in the evolution of such vocabulary systems. I investigate both the cultural evolution of vocabulary systems and the biological evolution of learning rules for vocabulary acquisition.

Firstly, vocabularies are shown to evolve on a cultural time-scale so as to fit the expectations of learners — a population's vocabulary adapts to the biases of the learners in that population. A learning bias in favour of one-to-one mappings between meanings and words leads to the cultural evolution of communicatively-optimal vocabulary systems, even in the absence of any explicit pressure for communication. Furthermore, the pressure to conform to the biases of learners is shown to outweigh natural selection acting on cultural transmission. Human language learn-

ers appear to bring a one-to-one bias to the acquisition of vocabulary systems. The functionality of human vocabulary may therefore be a consequence of the biases of human language learners.

Secondly, the evolutionary stability of genetically-transmitted vocabulary learning biases is investigated using both static and dynamic models. A one-to-one learning bias, which leads to the cultural evolution of optimal communication, is shown to be evolutionarily stable. However, the evolution *de novo* of this bias is complicated by the cumulative nature of the cultural evolution of vocabulary systems. This suggests that the biases of human language learners may not have evolved specifically and exclusively for the acquisition of communicatively-functional vocabulary.

Key words: language, communication, cultural evolution, learning bias

1 Introduction

Sophisticated communication systems abound in the natural world. Bees communicate the distance and direction of nectar sources to other bees using an elaborate dance (von Frisch, 1974). Various species of birds use alarm calls to warn conspecifics of approaching predators (for example, chickens, see Evans *et al.*, 1993), and songs to mark out territory and attract mates (Catchpole & Slater, 1995). Similarly, alarm calls are used by monkeys (for example, vervet and Diana monkeys, see Cheney & Seyfarth, 1990; Zuberbuhler *et al.*, 1997) and certain species of gibbon use song for territory maintenance, pair bonding and sexual advertisement (Raemaekers *et al.*, 1984). Chimpanzees use a system of facial expressions and vocalisations to perform social functions (de Waal, 1989), augmented with a system of gestural communication (Tomasello & Call, 1997).

Among this dizzying array of communicative behaviour, human language appears to be unique — more sophisticated, more expressive, more flexible, more diverse. Hockett (1960) began the trend of identifying *design features* of language, in an effort to characterise the aspects of language which distinguish it from non-human communication systems. Some important design features of language are:

Recursion: An expression of a particular type can be a subpart of a larger expression of that type. For example, sentences can be embedded within other sentences, as in [s I think [s he is rude]].

Compositionality: The meaning of an expression is a function of the meaning of its parts and the way in which they are combined. For example, the meaning of the expression "John kicked Mary" is dependent on the meanings of the subparts "John", "kick", the "-ed" suffix and "Mary".

Symbolicism: The form of a signal is arbitrarily related to its meaning. For example, there is no iconic relationship between the English word "apple" and the fruit it denotes.

Cultural transmission: Language is learned from the behaviour of other individuals, as opposed to being genetically transmitted.

We see echoes of these design features in non-human communication systems. The repeated subunits of bird song and the long calls of the gibbon hint at a weak form of conjunctive recursion. Bee dance is compositional — the meaning of the dance (the direction and distance to the food source) is dependent on the angle (for direction) and length (for distance) of the straight portion of the dance. Alarm call systems are generally taken to be symbolic, in that there is an arbitrary association between the form of an alarm call and the predator that it denotes. Learning plays some role in bird song formation.

However, it has been argued (in, e.g., Tomasello *et al.*, 1997; Tomasello, 1999) that true cultural transmission does not exist in non-human species, and, with particular relevance to the study of language, that there is no convincing evidence that acoustic experience plays any role in the development of call structure in primates (Hauser, 1996, p315 in particular), although it may play some role in determining the determining the response individuals make to calls, and also serve to narrow the contexts in which calls are used (Seyfarth & Cheney, 1997).

These design features of human language therefore make it unique among naturally-occurring communication systems — while we may draw parallels with the design features of non-human communication systems, the co-occurrence and ubiquity of these features singles human language out. The emergence of this unique communication system has been described as a major transition in evolution (Maynard Smith & Szathmáry, 1995) — the genesis of a new system of information transmission.

How did language come to be, and why is it so different from the communication systems of non-human species? In an attempt to demonstrate that well-understood processes of natural selection can be extended to an account of the evolution of language, Martin Nowak and colleagues have begun to expand the standard assumptions of evolutionary game theory and develop mathematical and computational models for understanding the evolution of language and the language faculty (see, e.g., Nowak *et al.*, 1999, 2001; Nowak & Komarova, 2001). A similar effort, primarily based around computational modelling techniques, has been underway in the field of evolutionary linguistics for some time (see, e.g., Hurford, 1989, 1991; Kirby, 2002b).

A common element of these models has been a primary focus on the evolutionary dynamics arising from the cultural transmission of language. It has been shown that the cultural transmission of linguistic form, under circumstances where learners are exposed to input data which underspecifies the target language, leads to the emergence of recursive and compositional language (Kirby, 2002a; Brighton, 2002). Using a rather different approach, it has been shown that natural selection acting on cultural transmission can lead to the cultural evolution of lexical or grammatical coherence within a population (Nowak *et al.*, 1999, 2001).

In this paper I will present research relating to the cultural evolution of vocabulary systems, and the biological evolution of vocabulary acquisition strategies. The research outlined here therefore represents an investigation into some of the pressures at play in the evolution of a communication system which is culturally transmitted and symbolic — two of the design features of human language. This paper centres around a computational model which can be considered as an extension of the model introduced in Nowak *et al.* (1999) and further analysed in, e.g., Komarova & Nowak (2001). Nowak *et al.* assume that the cultural evolution of vocabulary systems is driven by natural selection acting on cultural transmission. I consider an alternative pressure acting on vocabulary systems, arising from the biases of learners. I then model the biological evolution of such learning biases.

In section 2 I introduce the model of vocabulary and vocabulary acquisition. In section 3 I discuss a range of vocabulary acquisition biases. In section 4 I compare the forces arising from learner biases and natural selection of cultural variants. The behaviour of populations experiencing both pressures is primarily determined by the biases of language learners — natural selection

is a secondary force. Finally, in section 5 I examine the evolution of learning biases.

2   Vocabulary, communication, and acquisition

A vocabulary is considered to be a system mapping between a set of meanings $\mathcal{M} = \left\{ m_1, m_2, \ldots m_{|\mathcal{M}|} \right\}$ and a set of signals $\mathcal{S} = \left\{ s_1, s_2, \ldots s_{|\mathcal{S}|} \right\}$. Meanings and signals are unstructured, atomic units. A linguistic agent is defined by an *association matrix A*, an $|\mathcal{M}| \times |\mathcal{S}|$ matrix. The entry a_{ij} from this matrix gives the strength of the association between the ith meaning m_i and the jth signal s_j. The production and reception behaviour of an individual is determined by that individual's association matrix.

During production, an individual is prompted with a meaning and required to produce a signal. When prompted with the meaning m_i, the individual produces a signal according to the production function $p(m)$:

$$p(m_i) = s_{\left(\operatorname{argmax}_j (a_{ij}) \right)} \tag{1}$$

i.e. the agent produces the signal which has the strongest association with meaning i. In the event that multiple associations have equal weight, one is selected at random from among the equally-weighted alternatives. By a similar process, on receiving the signal s_j the agent interprets that signal according to:

$$r(s_j) = m_{(\operatorname{argmax}_i (a_{ij}))} \tag{2}$$

An individual's A matrix therefore defines that individual's production be-

6

haviour p and reception behaviour r. If p is interpreted as a probabilistic function $p(s_j|m_i)$, which gives the probability of producing signal s_j given meaning m_i, and r is similarly interpreted as a probabilistic function $r(m_i|s_j)$ then the communicative accuracy between a speaker P using production function $p(s|m)$ and a hearer R using reception function $r(m|s)$ is given by:

$$ca\,(P,R) = \frac{\sum_{i=1}^{|\mathcal{M}|} \sum_{j=1}^{|\mathcal{S}|} p\,(s_j|m_i) \cdot r\,(m_i|s_j)}{|\mathcal{M}|} \tag{3}$$

assuming all meanings are equally frequent and equally important. In other words, the communicative accuracy between speaker P and receiver R is the average probability of the speaker producing a signal for a given meaning m_s, and the hearer interpreting the received signal as meaning $m_h = m_s$. The two-way communicative accuracy between two individuals A and B acting in turn as speaker and hearer is then:

$$ca'\,(A,B) = \frac{ca\,(A,B) + ca\,(B,A)}{2} \tag{4}$$

The kth individual's communicative accuracy with respect to a population of N individuals is given by:

$$\overline{ca}\,(k) = \frac{1}{N-1} \left(\sum_{l=1}^{l=N} ca'\,(k,l) \right) \tag{5}$$

where $l \neq k$, i.e. the average communicative accuracy of acting as speaker and hearer with each of the other $N-1$ members of the population.

An individual arrives at its A matrix by sampling the observable linguistic behaviour of other members of the population. We consider observable behaviour to be meaning-signal pairs. Based on the observation of meaning m_t

7

paired with signal s_u, a learner updates its A matrix according to:

$$\Delta a_{ij} = \begin{cases} \alpha \text{ if } i = t \text{ and } j = u \\[1em] \beta \text{ if } i = t \text{ and } j \neq u \\[1em] \gamma \text{ if } i \neq t \text{ and } j = u \\[1em] \delta \text{ if } i \neq t \text{ and } j \neq u \end{cases} \tag{6}$$

α therefore specifies how to change the strength of the association between meanings and signals which co-occur, β specifies how to change the association strength between a meaning and a signal where the meaning occurs in the absence of that particular signal, and so on. The four-tuple $(\alpha \ \beta \ \gamma \ \delta)$ defines a *learning rule*, which specifies how the association matrix should be changed according to observed linguistic behaviour.

3   Learning biases, synonymy and homonymy

An optimal communication system is a system which will lead to perfect communication between two individuals using it ($ca'(A, B) = 1$ if individuals A and B use the optimal system). What properties must a learning rule have if an individual using that learning rule is to be capable of acquiring an optimal communication system? It has been shown (Smith, 2002a) that, assuming $|\mathcal{M}| = |\mathcal{S}|$, learning rules capable of acquiring optimal systems must obey the property $\alpha + \delta > \beta + \gamma$, i.e. in order to acquire an optimal system, associations between meanings and signals which appear together or are absent together must be strengthened more than those between meanings and signals which

do not co-occur. For the purposes of this paper, I will restrict myself to those rules which obey this property — optimal communication should at least be a possibility for all learners.

There is still a large range of variation within this set of learning rules. If we restrict ourselves to the case where $\alpha, \beta, \gamma, \delta \in \{-1, 0, 1\}$ there are 31 rules which meet the property given above (Smith, 2002a). Each of these rules has a *learning bias* — for each rule, vocabularies with certain properties are easier or harder to learn than other vocabularies. We are interested in learning bias with respect to two properties: synonymy and homonymy.

Synonymy occurs where one meaning can be expressed with two or more words (signals). It is debatable whether true synonymy occurs at all in language — different words tend to convey different meanings, even if that difference is rather subtle (for example, the difference between "buy" and "purchase" is largely one of formality). In terms of meaning-signal mappings, synonyms are one-to-many mappings from meanings to signals.

The biases of A matrix learning rules with respect to synonymy is determined by the relationship between α and β. When $\alpha > \beta$, the learning rule is biased against synonyms. Consider a 2×3 association matrix trained on the meaning-signal pair $\langle m_1, s_1 \rangle$. Assuming association weights of 0 prior to learning, the matrix after learning will be:

$$\begin{pmatrix} \alpha & \beta & \beta \\ \gamma & \delta & \delta \end{pmatrix} \tag{7}$$

$\alpha > \beta$ ensures that only s_1 will be produced for m_1 — one-to-many mappings

9

from meanings to signals are avoided. If $\alpha \leq \beta$ then synonyms are not avoided. If $\alpha = \beta$ then s_1, s_2 and s_3 will produced with equal probability for m_1. If $\alpha < \beta$ then s_2 and s_3 will be produced for m_1. In either case, synonymy (a one-to-many mapping from meanings to signals) is introduced.

Homonymy occurs where multiple distinct meanings are expressed using a single word, and is rife in language — most words in a dictionary have several possible meanings, some of which are apparently not related. In terms of meaning-signal mappings, homonyms are many-to-one mappings from meanings to signals.

The biases of learning rules with respect to homonymy is determined by the relationship between γ and δ. When $\delta > \gamma$, the learning rule is biased against homonymy. Consider the A matrix given in 7 above, formed on the basis of the observation of the pair $\langle m_1, s_1 \rangle$. When $\delta > \gamma$ s_1 will *not* to be produced for m_2 — many-to-one mappings from meanings to signals are avoided. $\delta = \gamma$ leads to neutrality with respect to homonymy — s_1 may be produced for both m_1 and m_2. Finally, $\delta < \gamma$ leads to a bias in favour of many-to-one, homonymous mappings — m_2 maps to s_1, which is associated with m_1 in the observed meaning-signal pair.

As stated above, for an optimal system to be learnable, the constraint $\alpha + \delta > \beta + \gamma$ must hold. However, when $|\mathcal{S}| > |\mathcal{M}|$ the bias against synonymy ($\alpha > \beta$) must also hold. The bias against synonymy is therefore critical: without it, optimal systems are unlearnable, assuming that there are more possible signals than meanings.[1] The bias against homonymy is not critical for optimal systems to be learnable, assuming noise-free transmission. However, it will be demonstrated later in the paper that the bias against homonymy becomes

important in the context of iterated cultural transmission — learners should, ideally, be biased against synonyms and against homonyms, and in favour of one-to-one mappings between meanings and signals.

4 The cultural evolution of vocabulary systems

In this section I will explore two possible pressures driving the cultural evolution of vocabulary systems. These pressures arise from 1) the biases of language learners, and 2) natural selection.

The biases of learners lead to a pressure which favours variants of the culturally-transmitted trait which conform to the learner biases — in other words, if learners prefer to acquire a particular cultural characteristic then, over time, that characteristic will come to dominate the population to the exclusion of other characteristics. This has been termed *cultural selection* (Cavalli-Sforza & Feldman, 1981) or *directly biased transmission* (Boyd & Richerson, 1985). With respect to the cultural evolution of vocabulary, we should expect that the biases of learners will result in a pressure for vocabulary systems which conform to those biases.

The second pressure arises from selection acting on mature, enculturated individuals. If we assume that individuals with different cultural variants differ in their probability of surviving and acting as models for learners, or differ in their probability of being chosen as models by learners, then those cultural variants which have the highest probability of being transmitted to the next generation will tend to increase in number. Boyd & Richerson (1985) term this *natural selection of cultural variants*. With respect to the cultural evolution

11

of vocabulary, if we accept the view espoused by Pinker & Bloom (1990) or Nowak *et al.* (1999) that natural selection can see the consequences of communication then we should expect natural selection of cultural variants to favour communicatively-functional vocabulary systems.

The goal of this section is to address two questions. Firstly, do learner biases result in significant cultural evolution, and what kind of communication systems evolve? Secondly, can natural selection acting on cultural transmission optimise the communicative functionality of a population's vocabulary?

4.1   The Iterated Learning Model

To investigate these question I will use an Iterated Learning Model (or ILM, a term introduced in Brighton & Kirby, 2001). The ILM is an idealised model of the cultural transmission of language. In the ILM individuals acquire their linguistic competence based on observations of the linguistic behaviour of other individuals. In this paper the linguistic behaviour of interest is a vocabulary system, and individuals are modelled using the association matrix model given in section 2. For all results presented in this paper, there are 10 meanings and 10 signals ($|\mathcal{M}| = |\mathcal{S}| = 10$). Assuming a generational model of population turnover, the ILM proceeds as follows:

Initialisation: Create an initial population at generation 0, $pop_{g=0}$, of N individuals[2]. Each individual has an A matrix of the appropriate size with initial association weights of 0. All initial individuals use a particular learning rule, as discussed below.

Iteration: The iteration process consists of five steps:

1. Evaluate the communicative accuracy of every member of pop_g according to equation 5.

2. Create a new population pop_{g+1} of N individuals, each with initial association weights of 0 and the same learning rule as the rest of the population.

3. Each member of pop_{g+1} picks e cultural parents[3] from pop_g. Cultural parents are either chosen at random, or are selected according to their evaluated communicative accuracy. In the second condition, the probability of individual i being selected to act as a cultural parent is $\overline{ca}\,(i)\,/\sum_j \overline{ca}\,(j)$

4. Each member of pop_{g+1} observes and learns from the linguistic behaviour of its cultural parents. The learner observes the complete set of meaning-signal pairs generated by each of its cultural parents and updates its A matrix according to its learning rule. Meaning-signal pairs are generated by the parent for every $m \in \mathcal{M}$, according to the production process outlined in section 2. Noise is added to each meaning-signal pair[4] with probability ρ.

5. Replace pop_g with pop_{g+1}. Return to 1.

There are therefore three parameters of variation:

(1) The learning rule that individuals in the population use.

(2) The manner in which cultural parents are chosen. This can be at random, or dependent on the communicative accuracy of potential cultural parents. In the latter case there is natural selection acting on cultural transmission, in favour of functional vocabulary systems.

(3) ρ, the probability of noise.

I will present results for three learning rules. All of these rules satisfy the constraint $\alpha + \delta > \beta + \gamma$, and individuals using these rules are therefore capable of acquiring optimal vocabularies. Furthermore, for all three rules

13

$\alpha > \beta$, and learners using these rules are therefore biased against acquiring synonyms. The rules vary in terms of their biases with respect to homonymy. The three learning rules are:

(1) The $h+$ learning rule, biased in favour of homonymy. For this rule, $\alpha = 1, \beta = -1, \gamma = 1, \delta = 0$.

(2) The $h?$ learning rule, neutral with respect to homonymy. For this rule, $\alpha = 1, \beta = 0, \gamma = 0, \delta = 0$.

(3) The $h-$ learning rule, biased against homonymy. For this rule, $\alpha = 1, \beta = -1, \gamma = -1, \delta = 0$.[5]

For the results presented in the following two sections, each population is initialised with one of these three rules, and all individuals at subsequent generations use the same learning rule as the initial population. Populations are therefore homogeneous in this respect — we are concerned for the moment in the cultural evolution of vocabulary systems in a population with a shared and fixed learning capacity.

4.2 Learning bias alone

The first step is to identify the consequences of the biases of various learning rules in the context of iterated cultural transmission. To this end, runs of the ILM were carried out for each of the three learning rules outlined above. Various levels of noise were used ($\rho \in \{0, 0.05, 0.1\}$). For all the results presented in this section, learners select their cultural parents at random from among the individuals in the previous generation — an individual's communicative success does not influence its chances of acting as a cultural parent, and therefore

there is no natural selection acting on cultural transmission.

Figure 1 presents the results of the computational simulations. Ten runs of the ILM were carried out for each set of conditions. The graphs plot the mean and standard deviation of average communicative accuracy $(\frac{1}{N}\sum_i \overline{ca}\,(i))$ in each population, averaged over the ten runs, against time in generations. Three results are apparent from this figure.

(1) For populations of $h+$ learners, communicative accuracy remains at chance levels. This reflects the fact that populations of such individuals converge on a fully homonymous communication system, where every meaning is expressed using the same signal.

(2) For populations of $h?$ learners, communicative accuracy reaches intermediate levels. Higher levels of noise slightly reduce the final levels of communicative accuracy reached. The performance of $h?$ learners reflects their convergence on a partially ambiguous vocabulary system, as discussed below.

(3) For populations of $h-$ learners, communicative accuracy reaches optimal levels. This reflects the fact that these populations converge on shared, one-to-one mappings between meanings and signals. The addition of noise does not prevent such systems emerging.

The biases of individual learners therefore influence the behaviour of the population as a whole — as predicted by Cavalli-Sforza & Feldman (1981) and Boyd & Richerson (1985), populations converge on cultural characteristics which match their learning biases. In populations of $h+$ learners, where individuals are biased in favour of acquiring many-to-one, maximally homonymous vocabularies, such systems emerge over iterated learning events, resulting in chance

15

levels of communicative accuracy. It has been shown that populations of $h+$ learners are in fact incapable of even preserving an optimal initial communication system when $\rho > 0$ — noise during transmission introduces many-to-one mappings, which preferentially spread through populations of $h+$ learners, destroying any optimal initial system (Smith, 2002a) .

In populations of $h-$ learners, individuals are biased in favour of acquiring one-to-one mappings between meanings and signals, and consequently a shared, one-to-one system of mappings evolves over time, even in the presence of noise. Such systems lead to optimal levels of communicative accuracy.

Individuals who learn using a h? rule are neutral with respect to homonyms — many-to-one mappings from meanings to signals are as learnable as one-to-one mappings. Consequently, the behaviour of populations of h? learners is dependent on the initial, random system of meaning-signal mappings embodied in the meaning-signal pairs produced by the initial population. This random assignment will become shared among the population through the process of iterated learning. While $h-$ learners remove the many-to-one elements of the initial random system, and $h+$ learners add more homonyms, h? learners simply preserve the initial level of homonymy – the population's eventual communication system will embody the same number of many-to-one mappings as the initial random behaviour. Consequently, the populations converge on systems yielding the communicative accuracy we would expect given a random assignment of signals from $|\mathcal{S}|$ to meanings from $|\mathcal{M}|$ with replacement (from Oliphant, 1999):

$$ca \approx \left(1 - \left(1 - \frac{1}{|\mathcal{S}|}\right)^{|\mathcal{M}|}\right) \tag{8}$$

For $|\mathcal{M}| = |\mathcal{S}| = 10$ this evaluates to a communicative accuracy of approximately 0.651. The average final communicative accuracy for the runs where $\rho = 0$ and $\rho = 0.05$ are 0.66 and 0.656 respectively, which corresponds closely to this expected level of communicative accuracy. The average final communicative accuracy for the case where $\rho = 0.1$ is 0.616, reflecting the fact that a high level of noise introduces further many-to-one mappings, which are not eliminated by h? learners.

To summarise, the biases of language learners are important. The population's vocabulary system will be shaped by those learner biases, with systems which conform to the biases of individual learners being more likely to be successfully transmitted, and therefore coming to dominate the population.

4.3   Learning bias plus natural selection

From the perspective of functionality of a population's vocabulary system, having the correct learning bias (a homonymy-avoiding bias) is important when individuals learn from randomly-selected cultural parents. But can the incorrect learning bias be remedied by natural selection of cultural variants? In other words, if learners preferentially acquire the communication systems of successful communicators, can the populations converge on optimal communication systems in spite of a suboptimal learning bias?

To investigate this question, runs of the ILM were carried out using the same three learning rules as described above. Various levels of noise were used ($\rho \in \{0, 0.05, 0.1\}$). For all ILMs presented in this section, learners select their cultural parents according to the communicative accuracy of those potential

parents — the probability of an individual being selected to act as a cultural parent is given by $\overline{ca}\,(i)\,/\sum_j \overline{ca}\,(j)$.

Figure 2 presents the results of the computational simulations. As before, ten runs of the ILM were carried out for each set of conditions.

Three results are apparent from this figure:

(1) Populations of $h+$ learners continue to converge on fully homonymous communication systems, yielding chance levels of communicative accuracy, despite natural selection.

(2) Populations of $h?$ learners continue to converge on systems which yield intermediate levels of communicative accuracy, although the addition of natural selection leads to slightly increased average communicative accuracy.

(3) Populations of $h-$ learners continue to reach optimal levels of communicative accuracy, reflecting the emergence of a shared, one-to-one mapping between meanings and signals.

Populations of $h+$ and $h-$ learners essentially behave in the same way regardless of whether increased communicative success impacts on an individual's chances of becoming a cultural parent or not. Natural selection does impact on the performance of populations of $h?$ learners, acting to eliminate individuals using vocabularies with higher levels of homonymy from the pool of potential cultural parents. The addition of noise also plays an important role in this process — as with natural selection acting on genetic transmission, natural selection of cultural variants requires variation within the population (Boyd & Richerson, 1985), and noise acts to reintroduce the variation that selection eliminates. It may be that stronger selection pressure or a certain noise level

can improve the performance of populations of $h?$ learners further still.

4.4 Summary

To summarise the results presented in this section, the biases of learners have consequences for a culturally-transmitted vocabulary system, and learning bias is the key determinant of cultural evolution even when natural selection acts on cultural transmission. Populations of learners with absolutely the wrong learning bias (the $h+$ bias) converge on communication systems which yield very poor communicative accuracy, regardless of whether they learn from randomly-chosen cultural parents, or selectively from the population's better communicators. Populations of learners with a homonymy-neutral learning bias ($h?$ learners) converge on intermediate levels of communicative accuracy, although natural selection of cultural variants can improve the functionality of communication systems in such populations, given appropriate levels of noise. Finally, populations of learners with a bias in favour of one-to-one mappings ($h-$ learners) always converge on optimal communication systems, regardless of noise and even without the help of natural selection. A similar result, for a different learning model, is presented in Smith (2002b).

There are in fact good reasons for expecting learning bias to be a more powerful force than natural selection of cultural variants. Natural selection requires variation in the population — natural selection becomes irrelevant in a completely homogeneous population. Cultural selection resulting from learning bias is also dependent on variation to a certain extent — for example, in a population of $h+$ learners converged on an optimal, one-to-one vocabulary system, variation must be introduced before the learner bias in favour of

19

homonymy can take effect. However, learning bias can also *introduce* variation. In a population of $h-$ learners converged on an all-to-one, maximally homonymous vocabulary, variation is immediately introduced — no $h-$ learner can acquire such a system, and will acquire some other system, immediately introducing variation. Furthermore, the variation introduced by learning bias tends to be the *right kind* of variation. $h-$ learners will tend to introduce new one-to-one mappings, which is the right kind of variation in terms of their learning bias, and $h+$ learners will introduce new many-to-one mappings, the type of mapping they prefer to acquire. Biased learning therefore introduces variation which it itself can feed off, and furthermore introduces variants which are likely to persist. Natural selection has no such advantage — variation must be provided by some other source (mutation, or errors during learning), and variants which are introduced are not guaranteed to be the kind which will persist (notwithstanding the debate on directed mutation, see, e.g., Sniegowski & Lenski, 1995).

5   The evolution of learning bias

The results outlined in the previous section indicate that having the right learning bias is crucial — in populations with the wrong learning bias, functional communication systems never emerge, whereas in populations with the appropriate learning bias, optimal systems always evolve through cultural processes.

Can the appropriate learning bias evolve for the communicative payoff it offers? To be more precise, will natural selection acting on genetic transmission lead to the evolution of learning biases which result, through cultural pro-

cesses, in the emergence of optimal communication systems?

5.1 A static analysis

The first step in answering this question is to identify which learning biases are evolutionarily stable (Maynard Smith & Price, 1973; Harley, 1981; Maynard Smith, 1982) — which learning biases are such that a population adopting that learning bias will not be invaded by some other learning bias under the influence of natural selection? This breaks down into two sub-questions: 1) what language will a population consisting entirely of individuals with a particular learning bias have?; 2) what level of communicative accuracy will some individual inserted into such a population have?

Question 1 has been addressed in Section 4. Populations of $h+$ learners converge on fully homonymous vocabularies, yielding chance levels of communicative accuracy. Populations of $h?$ individuals converge on vocabularies exhibiting some intermediate level of homonymy, yielding intermediate levels of communicative accuracy. Populations of $h-$ individuals converge on unambiguous, one-to-one vocabularies, yielding optimal communicative accuracy.

Assuming that there is noise on cultural transmission $(rho > 0)$,[6] a static analysis suggests that there are two evolutionarily stable learning biases — $h?$ and $h-$. The $h+$ bias cannot be evolutionarily stable. A population of $h+$ individuals will have a maximally homonymous communication system, and chance levels of communicative accuracy. Any $h?$ learners inserted into this population will acquire this system, and will therefore attain equal levels of communicative accuracy — $h?$ individuals can invade a $h+$ population by drift.

21

Any $h-$ learners inserted into this population will be unable to acquire the population's vocabulary, due to their bias against homonymy, and will acquire a random system of mappings between meanings and signals. However, this random system will yield the same (chance) level of communicative accuracy as the $h+$ population's maximally ambiguous system, so $h-$ individuals can also invade a $h+$ population by drift.

A population of $h?$ individuals will converge on a partially homonymous system. Any $h+$ individuals introduced into such a population will, due to their bias in favour of homonyms and noise on transmission, fail to learn the partially homonymous system and instead acquire a more homonymous vocabulary. Consequently, $h+$ individuals will have lower communicative accuracy and will be selected against. $h-$ individuals introduced into $h?$ populations will suffer the opposite problem — they will be unable to acquire certain highly homonymous parts of the population's vocabulary. They will therefore have lower communicative accuracy than $h?$ individuals, and will be selected against. It therefore seems, under a static analysis, that the $h?$ bias is evolutionarily stable.

Is the $h-$ bias evolutionarily stable? $h-$ populations will converge on unambiguous vocabularies. Any $h+$ individuals introduced into such a population will preferentially acquire any homonymous mappings introduced by noise, will suffer reduced communicative accuracy and will be selected against. $h?$ individuals will acquire (neutrally, rather than preferentially) homonymous mappings introduced by noise and will also be selected against. The $h-$ bias should therefore also constitute an ES learning bias.

A numerical analysis confirms these expectations. Table 1 give the expected

communicative accuracies of an infinitely small proportion of individuals with a particular learning bias after exposure to the vocabulary system of a population of individuals with another learning bias when communicating with individuals who possess the majority learning bias. The majority population is assumed to have converged on the type of vocabulary system which its bias favours (fully homonymous systems for populations of $h+$ learners, partially homonymous systems for populations of $h?$ learners, unambiguous vocabularies for populations of $h-$ learners). The results in the tables are based on 1000 evaluations for each possible combination of majority and minority genotype. As can be seen from the table, $h+$ does not constitute an ES learning bias — individuals with some other learning bias communicate as well as individuals with the majority learning bias, and therefore will not be selected against. $h?$ and $h-$ are evolutionarily stable — individuals with some other learning bias introduced into populations in which these rules are in the majority suffer from reduced communicative accuracy, and will be selected against.

5.2   A dynamic analysis

Is this static analysis sufficient? There are two reasons for thinking that it is not. Firstly, the payoff associated with the different learning biases is frequency-dependent — it may be that, given a certain proportion of invading mutants, biases which appear to be evolutionarily stable in the static analysis given above are in fact unstable. Frequency-dependence is common in social coordination problems (see, e.g., Cavalli-Sforza & Feldman, 1983).

Secondly, and more significantly, cultural evolution in populations is *cumulative*, and the payoff associated with different learning biases is therefore time

dependent. Vocabularies develop over time in populations, and the communicative payoff a particular individual receives will be dependent on the vocabulary system that it is attempting to acquire, which is in turn dependent on the learning biases of individuals in preceding generations of the population. To take a concrete example, the communicative payoff accruing to an individual with the $h-$ bias will be dependent on the learning biases which have been present in the population it is born into, and how long those biases have been present. A $h-$ individual born into a population which has previously contained only $h+$ learners will have low communicative accuracy when communicating with another $h-$ individual born into the same population — both will use random systems, due to their difficulty in learning completely ambiguous vocabularies, and will therefore communicate with chance levels of accuracy. Two $h-$ individuals in the initial generation of a fully $h-$ population will be in a similar position — as there will be no established vocabulary in a such a population, they will be unlikely to arrive at overlapping vocabularies. However, another two $h-$ individuals born into the same population a few generations later will be in a very different situation — the population's vocabulary will have been shaped, by the biases of the previous generations, so as to fit their own learning bias. Consequently, they will communicate more successfully with one another.

Does the cumulative effect of cultural evolution alter the picture when we consider which learning strategies are evolutionarily stable? To investigate this question, the ILM described in section 4 was modified to allow the genetic transmission of learning biases to be investigated. In the Evolutionary Iterated Learning Model (EILM) the learning rule an individual uses is considered to be genetically transmitted, and populations can be genetically heterogeneous.

The EILM proceeds as follows:

Initialisation: Create an initial population $pop_{g=0}$ of N individuals[7]. Each individual has an A matrix of appropriate size with initial association weights which are either a) set so as the individual produces some predefined vocabulary V or b) all zero, so as the agent produces every meaning-signal pair with equal probability. The proportion of the initial N individuals which have the $h+$, $h?$ and $h-$ rules is a parameter of variation.

Iteration: The iteration process consists of 5 steps:

1. Evaluate the communicative accuracy of every member of pop_g according to equation 5.

2. Create a new population pop_{g+1} of N individuals with initial association weights of 0. Each of the N individuals in pop_{g+1} inherits its learning rule from a member of pop_g. Biological parents are selected according to their evaluated communicative accuracy, with the probability of individual i being selected to act as a biological parent given by $\overline{ca}\,(i)\,/\sum_j \overline{ca}\,(j)$.

3. Each member of pop_{g+1} picks e cultural parents[8] from pop_g. Cultural parents are chosen at random.

4. Each member of pop_{g+1} observes and learns from the linguistic behaviour of its cultural parents. The learner observes the complete set of meaning-signal pairs generated by each of its cultural parents and updates its A matrix according to its inherited learning rule. Meaning-signal pairs are generated by the parent for every $m \in \mathcal{M}$, according to the production process outlined in section 2. Noise is added to each meaning-signal pair with probability ρ.[9]

5. Replace pop_g with pop_{g+1}. Return to 1.

There are two main parameters of variation in the model. Firstly, the initial agents either have a certain shared pre-configured vocabulary V, or they have no shared vocabulary. Secondly, the genetic makeup of the initial population can be varied. In the following two sections, the impact of these two parameters will be explored. In section 5.2.1, the case where initial populations are pre-converged on some predefined vocabulary are explored. In this situation, can a small proportion of individuals with a particular learning rule invade a population using some other learning rule? In Section 5.2.2, the case where initial populations have no established vocabulary is explored. In this situation, can a small proportion of individuals with a particular learning rule invade a population using some other learning rule, and how does this biological evolution interact with ongoing cultural evolution in the population?

5.2.1 A dynamic analysis with converged culture

Runs of the EILM were carried out to establish whether a small proportion of individuals with a particular learning rule could invade a population using some majority learning rule, where that majority population is pre-converged on some shared vocabulary V which is optimal in terms of the learning bias of the majority population. To this end, populations were initialised such that 90% of individuals in that population had genotypes encoding a particular learning rule, with the remaining 10% having another learning rule. All possible pairwise combinations of majority and minority populations were investigated. Majority populations of $h+$ individuals are pre-converged on a fully homonymous vocabulary, majority populations of $h?$ individuals are pre-converged on partially homonymous vocabularies, and majority $h-$ populations are pre-converged on unambiguous vocabularies.

Twenty runs of the EILM were carried out for each experimental condition. The results of these simulations are summarised in table 2. The tables display the average final communicative accuracy of the populations (the average value of $\frac{1}{N} \sum_i \overline{ca}(i)$ for the final generation of twenty simulations, where each simulation run was allowed to proceed to a stable state), the number of populations diverging from the majority genotype (i.e. the number of successful invasions), and the number of simulation runs converging on an optimal system.

As can be seen from table 2, the results for majority $h-$ populations are in line with the static model — $h-$ populations are evolutionarily stable. The results for majority $h+$ and $h?$ populations deviate somewhat from the predictions for the static model.

Majority $h+$ populations can be invaded through drift by $h?$ individuals (as predicted by the static analysis), and this does not change the population's vocabulary system. $h+$ populations are also invaded by $h-$ individuals. However, this invasion occurs more frequently, and always results in the invaded population converging on an optimal vocabulary system. In the cases where the $h-$ genotype comes to dominate the population, the minority subpopulation of $h-$ learners fail to acquire the majority vocabulary and instead converge, through cultural processes, on a partially shared, somewhat functional vocabulary system, due to their bias in favour of one-to-one mappings. This partially shared system gives them higher communicative accuracy than individuals with the $h+$ genotype, who prefer to acquire ambiguous, many-to-one systems. Consequently, the $h-$ genotype is selected for and comes to dominate the population. The population then converges on a shared optimal communication system resulting in maximal levels of communicative accuracy. Due to noise on transmission, $h+$ learners cannot acquire such a system and

continue to be selected against.

Why does the $h-$ genotype not always invade populations of $h+$ learners? The scenario outlined above requires that the numbers of $h-$ learners are maintained during the early stages of the process of cultural evolution — prior to arriving at a partially shared, partially functional communication system, $h-$ individuals have no communicative advantage over $h+$ individuals. Genetic transmission in the population during the early stages is therefore random, and the $h-$ genotype is prone to elimination through drift. In cases where this genotype is not eliminated by drift, individuals with the $h-$ genotype eventually receive a communicative payoff and come to dominate the population. However, when drift drives down the numbers of $h-$ individuals the process of cultural evolution is impeded, and the $h-$ genotype drifts out of the population.

This process is illustrated in Figure 3, which shows two sample runs of the EILM, with a majority of $h+$ individuals in the initial population. Time in generations is plotted against the population's average communicative accuracy, the proportion of individuals with the $h-$ genotype, and the relative communicative accuracy of $h-$ individuals.[10]

Figure 3(a) shows a run where drift eliminates the $h-$ genotype before cumulative cultural evolution resulting from the $h-$ bias gets underway — the relative communicative accuracy of $h-$ individuals remains around 1 until they are eliminated.

Figure 3(b) illustrates a run where the population converges on the $h-$ genotype and an optimal communication system. During the first 50 generations of the simulation, genetic transmission is random — the relative communica-

tive accuracy of $h-$ individuals hovers around 1, and consequently individuals with the $h-$ genotype are no more likely to breed than $h+$ individuals. However, drift happens to maintain the $h-$ genotype in the population. After 50 generations, the subpopulation of $h-$ individuals begin to converge on a partially shared vocabulary, and $h-$ individuals show slightly improved relative communicative accuracy. $h-$ individuals are selected to breed with greater frequency, and the proportion of individuals with the $h-$ genotype increases. This in turn leads to a greater communicative payoff for individuals with $h-$ genotypes, as a functional vocabulary begins to take shape. The $h-$ genotype comes to dominate the population, and an optimal vocabulary is constructed as a consequence of the $h-$ learning bias.

As indicated by the results in table 2, a similar process allows $h-$ individuals to invade $h?$ populations. However, this invasion occurs less frequently (in two runs as opposed to four). This is due to the fact that it takes longer for a $h-$ population to converge on a system which is more communicatively functional than the vocabulary already in place in the majority population. In majority $h+$ populations this process is fairly rapid — the baseline level of communicative accuracy in such populations is low. However, in majority $h?$ populations the level of communicative accuracy which must be reached is higher, and consequently the $h-$ minority is more likely to be eliminated by drift or weak selection prior to this baseline being reached.

5.2.2 A dynamic analysis with random initial culture

The results outlined in the previous section illustrate that the static analysis breaks down due to the cumulative nature of cultural evolution. The

static analysis is based on the expected communicative accuracy of the first generation of invaders relative to the majority population. As the dynamic model demonstrates, the communicative accuracy that a particular learning bias leads to changes over time, dependent on the biases in the population at previous generations. Learning biases which appear to be evolutionarily stable in the static analysis can therefore turn out to be unstable, as is the case with the $h?$ strategy. Furthermore, in certain situations the static analysis suggests that selection will be neutral, where in fact the dynamic model shows that certain learning biases can change the population's culture so as to generate a selective advantage for themselves, as happens in the case where $h-$ genotypes invade majority $h+$ populations.

These factors are likely to be particularly pronounced when a population is not pre-converged on a particular vocabulary system. The set of EILM simulations outlined in the previous section were repeated, with the initial population having a random vocabulary — all association strengths for each individual in the initial populations were set to 0, so that each initial individual produces each meaning-signal pair with equal probability. Under this set of circumstances, with a clean slate for cultural evolution, which learning biases are evolutionarily stable?

As before, twenty runs of the EILM were carried out for each experimental condition. The results of these simulations are summarised in table 3.

Dealing first with majority $h+$ populations, the results from the full dynamic model are broadly similar to the case where the initial population is pre-converged on a fully homonymous language. $h+$ populations are prone to invasion by both $h?$ and $h-$ individuals, and $h-$ individuals are more likely

30

to successfully invade $h+$ populations due to their ability to shape the population's culture so as to give themselves a selective advantage.

The results for majority $h?$ populations are somewhat different. $h?$ populations are prone to invasion by both $h+$ and $h-$ individuals. The average final communicative accuracy of $h?$ populations is largely determined by the bias of the invading genotype. When the invaders have the $h+$ bias, communicative accuracy is low. This is due to the impact of $h+$ individuals in the early stages of the simulations — $h+$ individuals, even in small numbers, tend to introduce homonyms which $h?$ individuals cannot eliminate. In extreme cases, the mixed population will converge on a maximally homonymous vocabulary. The eventual genetic makeup of the population depends on the level of homonymy that the population converges on in the early stages of cultural evolution. If the population converges on a partially homonymous vocabulary then $h+$ individuals cannot invade — $h+$ individuals cannot acquire such systems and therefore suffer decreased communicative accuracy. However, if the presence of small numbers of $h+$ individuals causes the population to converge on a fully homonymous vocabulary, $h+$ learners can invade by drift — $h+$ individuals can acquire the maximally homonymous system, do not suffer any communicative disadvantage, and will not be selected against. This happens in 2 of the 20 simulations carried out.

When a small number of $h-$ individuals are introduced into majority $h?$ populations the final communicative accuracy is significantly higher, due to the elimination of homonyms in the early stages of the simulation, resulting from the $h-$ learning bias. $h-$ individuals also frequently invade majority $h?$ populations when there is pre-established culture — $h-$ subpopulations converge on less ambiguous communication systems which lead to more successful com-

munication and result in selection for $h-$ genes. As with the case for the $h-$ genotype invading $h+$ populations, this selective advantage is dependent on drift in the early stages of the runs preserving $h-$ genes in sufficient numbers for cumulative cultural evolution resulting from the $h-$ learning bias to take effect.

Finally, majority $h-$ populations resisted invasion by $h+$ and $h?$ individuals in all simulations carried out. During the early stages of these runs, the majority $h-$ individuals begin to converge on a functional communication system. The presence of cultural variability in the population, resulting in part from noise and in part from the fact that the population has not yet converged on a vocabulary, impedes the acquisition performance of $h+$ and $h?$ individuals, who tend to acquire homonyms, and therefore $h-$ individuals are selected for.

5.3   Summary

The main result presented in this section is that only the $h-$ learning bias is evolutionarily stable under a wide range of circumstances. The static analysis suggests that the $h?$ bias should also be evolutionarily stable. However, the static analysis ignores the cumulative nature of the cultural evolution of vocabularies. The two dynamic analyses reveal that this cumulative cultural evolution can lead to shifts away from learning biases which appear stable under the static analysis — $h+$ individuals can invade $h?$ populations under certain circumstances, and $h-$ individuals can invade $h+$ and $h?$ populations under a wide range of circumstances, due to their ability to change the population's culture to their own advantage.[11] This suggests that a learning bias in favour of one-to-one mappings between meanings and signals can evolve

32

biologically.

However, the evolution of the $h-$ learning bias in populations where it is not initially dominant is always dependent on an initial phase of genetic drift. Subpopulations of $h-$ individuals converge on shared, functional vocabulary systems, but the cultural evolution of such systems is cumulative and takes time. During the early stages of the cultural construction process individuals with the $h-$ genotype receive no communicative advantage relative to individuals with other genotypes, and consequently do not enjoy increased reproductive success. The $h-$ subpopulation is therefore prone to elimination by drift during the early stages, and optimal communication systems will only evolve if drift preserves $h-$ genes in sufficient numbers for the construction process to get underway — the $h-$ bias may be evolutionarily stable once established, but its evolution is not straightforward.

6 Discussion

This paper presents research on the evolution of culturally-transmitted, symbolic vocabulary systems. This investigation is based on an abstract model of vocabulary systems and vocabulary acquisition, introduced in section 2. In section 3 I identify the properties of rules which are capable of acquiring optimal communication systems. Such rules must be biased against acquiring synonyms, one-to-many mappings from meanings to signals. There are a range of biases regarding homonymous, many-to-one mappings from meanings to signals — learning rules can be biased in favour of acquiring homonyms, or neutral with respect to homonymy, or biased against acquiring homonyms. The combination of a bias against synonyms and a bias against homonyms

33

amounts to a bias in favour of one-to-one mappings between meanings and signals.

In section 4 I investigate the cultural consequences of these learning biases. A population's communication system changes over time to fit the biases of individual learners. In populations of learners biased in favour of acquiring homonymous vocabularies ($h+$ learners), this pressure arising from learning bias leads to the evolution of maximally ambiguous communication systems, which provide only chance levels of communicative accuracy. In contrast, in populations of learners who are biased in favour of acquiring one-to-one mappings between meanings and signals ($h-$ learners), communicatively optimal, unambiguous communication systems evolve through cultural processes. In populations of learners who are neutral with respect to homonymy ($h?$ learners), partially ambiguous communication systems evolve, which offer intermediate levels of communicative accuracy. These results hold regardless of whether learners learn from randomly-chosen cultural parents, or selectively learn only from more successful communicators — natural selection of cultural variants is comparatively weak in comparison to the pressures arising from the biases of individual learners.

Finally, in section 5 I describe an investigation into the evolution of learning biases themselves. Learning rules, and therefore learning biases, were assumed to be genetically transmitted, with natural selection acting on genetic transmission. Under this set of circumstances, a learning bias in favour of unambiguous, one-to-one vocabulary was found to be the most stable — no other learning biases were evolutionarily stable, being prone to invasion by individuals with the one-to-one bias, while the one-to-one bias was evolutionarily stable under a wide range of circumstances. However, the evolution of one-to-

one learning biases, and consequently optimal communication, is dependent to some extent on genetic drift — in a population pre-converge on a suboptimal communication system, or with no established communication system, there is no immediate advantage to the one-to-one bias.

What do these models tell us about the evolution of vocabulary systems, and in particular the evolution of culturally-transmitted vocabulary in humans? There are two main issues. Firstly, the Iterated Learning Model outlined in section 4 highlights the importance of the biases of individual language learners in shaping a culturally-transmitted communication system. Secondly, the Evolutionary Iterated Learning Model outlined in section 5 highlights some factors likely to be at play in the evolution of learning biases such as those found in humans.

What biases do human language learners bring to the vocabulary acquisition task? It has been suggested that language learners have a general expectation that language should embody a one-to-one mapping between underlying semantic structures and surface forms. It has been argued that this learning bias derives from a *maxim of clarity* (Slobin, 1977), a preference for *transparency* (Langacker, 1977) or a preference for *isomorphism* (Haiman, 1980). This general bias manifests itself at all levels of linguistic structure (in morphology, in the lexicon, and in syntax).

The application of such biases to the acquisition of lexical items has received empirical verification. Biases against synonymy in human infants have been demonstrated in a series of experiments (e.g., Kagan, 1981; Markman & Wachtel, 1988). This bias against synonyms has been called the *Mutual Exclusivity* bias (Markman & Wachtel, 1988; Markman, 1989) or the *Principle of Contrast*

(Clark, 1988, 1990). Biases against homonymy have also received attention, and it has been experimentally demonstrated that children are biased against acquiring homonymous lexical items (e.g., Beveridge & Marsh, 1991; Mazzocco, 1997).

There is therefore good empirical evidence that children are biased against acquiring synonyms and homonyms — in sum, it appears that human infants are biased in favour of one-to-one mappings between meanings and words. If this is the case, then why is it that language apparently exhibits no synonyms, but homonymy is common?

Firstly, these biases are not envisaged as being absolute — they are an initial expectation on the part of the child, which can be overridden given sufficient evidence in the linguistic stimulus. Secondly, we should expect other pressures acting on the vocabulary of language to affect synonyms and homonyms in different ways. Langacker (1977) suggests two other pressures acting on vocabulary — a speaker preference for *signal simplicity* (that is, minimising articulatory effort during speech) and a learner preference for *code simplicity* (minimising the distinct number of surface forms which must be memorised). The preference for signal simplicity will act to introduce homonyms — minimisation of articulatory effort will reduce the number of available distinct words, therefore increasing the likelihood that distinct meanings will come to be expressed by a homonymous word. A pressure for code simplicity will also tend to introduce homonymy — if a single surface form can be used to refer to multiple meanings, then this reduces the number of distinct words which must be memorised. Interestingly, code simplicity *disfavours* synonymy — memorising two distinct words to express a single meaning increases the learning burden, perhaps unnecessarily. Therefore, even if we expect language learners

to be equally biased against synonyms and homonyms, other pressures can act to reintroduce homonyms and further weed out synonyms.

It seems likely that human language learners are biased in favour of acquiring vocabulary systems which associate meanings with signals in a one-to-one fashion, although other factors counteract this bias, leading to the observed vocabulary systems of language, which are not perfectly one-to-one. We should therefore expect the human vocabulary acquisition bias to lead, through cultural processes, to communicatively optimal vocabulary systems (in the absence of competing pressures), or systems which are at least communicatively functional (when there are competing pressures, such as pressures for signal and code simplicity). This brings us to a second question — is this vocabulary acquisition bias in humans a language-specific adaptation which evolved for the communicative payoff it potentially yields?

The Evolutionary Iterated Learning Models outlined in section 5 sheds some light on this issue. There are two points to be made. Firstly, the $h-$ bias in favour of one-to-one mappings between meanings and signals is the most stable evolutionarily — small numbers of individuals with this learning bias can invade majority populations of homonym-neutral ($h?$) or homonymy-preferring ($h+$) learners, and majority populations with this bias are resistant to invasion by other learning biases. In this sense, the $h-$, one-to-one learning bias is the most likely bias to evolve.

However, the EILM simulation results also illustrate a negative point — if populations are converged on a suboptimal communication system, or there is no established communication system present in the population, natural selection will not favour individuals who are predisposed to learn an optimal

37

system. In the simulation results shown here, a benevolent period of genetic drift is required to break the cycle of suboptimal communication, at which point a functional system begins to emerge and being biased to acquire an optimal communication system becomes advantageous. The division of large populations into multiple competing groups might increase the likelihood that one subpopulation would arrive, by drift, at the optimal bias and subsequently out-compete other populations (as suggested by the shifting balance theory, see, e.g., Wright, 1931; Coyne *et al.*, 1997; Whitlock & Phillips, 2000).[12] However, the point remains that, in a suboptimally communicating population, being able to learn an optimal communication system confers no immediate advantage. Biases tailored to the acquisition of an optimal system must be preserved in sufficient numbers for cumulative cultural evolution to begin if they are to come to dominate such populations. If the preservation of appropriate biases in a population is unlikely, either due to their low initial numbers or some cost associated with that bias, appropriate biases are unlikely to evolve.

This conclusion casts doubt on the view that the human vocabulary-learning bias is a domain-specific adaptation. This however does not force us to conclude that human vocabulary-learning biases arose through extra-adaptive mechanisms. An interesting possibility is that the bias is a consequence of a domain-general cognitive capacity, which arose as a consequence of multiple benefits it provided, including a communicative payoff. What would such a general cognitive capacity be?

One possibility is that the one-to-one bias is a consequence of the evolution of a sophisticated *theory of mind*, an understanding of other individuals as intentional agents. Tomasello (e.g., Tomasello & Call, 1997; Tomasello, 1999) argues that the human capacity for this type of reasoning is uniquely sophis-

38

ticated among the primates. Tomasello argues that this cognitive capacity could have evolved under an array of selection pressures, including selection for communication, cultural learning in general, cooperation, and tool use. This view ties in well with the literature on human language learning biases, which typically view such biases as a consequence of a child's theory of mind and their understanding of the nature of the communicative act. Armed with such an understanding, children should expect language to exhibit a one-to-one mapping between meanings and signals, and this indeed appears to be their initial expectation. In other words, the human vocabulary acquisition bias may not have evolved specifically and exclusively for the acquisition of communicatively-functional vocabulary, but rather be a consequence of a more general cognitive capacity which evolved due to a raft of benefits it provided, including perhaps communication.

How well do the results presented here extend from a simple model of vocabulary to a fuller treatment of the evolution of sophisticated, syntactically-structured language? This remains to be seen. However, this paper highlights two factors. Firstly, the biases of learners have a significant impact on the structure of culturally-transmitted communication systems, and pressures deriving from learner biases potentially outstrip other pressures, such as natural selection of cultural variants. Secondly, formal computational or mathematical models constitute a framework which can profitably be applied to the study of the evolution of language, one of the defining characteristics of our species.

Acknowledgements

References

BEVERIDGE, M. & MARSH, L. (1991). The influence of linguistic context on young children's understanding of homophonic words. *Journal of Child Language* **18**, 459–467.

BOYD, R. & RICHERSON, P. J. (1985). *Culture and the Evolutionary Process.* Chicago, IL: University of Chicago Press.

BRIGHTON, H. (2002). Compositional syntax from cultural transmission. *Artificial Life* **8**(1), 25–54.

BRIGHTON, H. & KIRBY, S. (2001). The survival of the smallest: Stability conditions for the cultural evolution of compositional language. In: *Advances in Artificial Life: Proceedings of the 6th European Conference on Artificial Life* (KELEMEN, J. & SOSÍK, P., eds.). Berlin: Springer-Verlag, pp. 592–601.

BRISCOE, E. (ed.) (2002). *Linguistic Evolution through Language Acquisition: Formal and Computational Models.* Cambridge: Cambridge University Press.

CATCHPOLE, C. K. & SLATER, P. J. B. (1995). *Bird Song: Biological Themes and Variations.* Cambridge: Cambridge University Press.

CAVALLI-SFORZA, L. L. & FELDMAN, M. W. (1981). *Cultural Transmis-*

sion and Evolution: A Quantitative Approach. Princeton, NJ: Princeton University Press.

CAVALLI-SFORZA, L. L. & FELDMAN, M. W. (1983). Paradox of the evolution of communication and of social interactivity. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 2017–2021.

CHENEY, D. & SEYFARTH, R. (1990). *How Monkeys See the World: Inside the Mind of Another Species.* Chicago, IL: University of Chicago Press.

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

CLARK, E. (1990). On the pragmatics of contrast. *Journal of Child Language* **17**, 417–431.

COYNE, J. A., BARTON, N. H. & TURELLI, M. (1997). Perspective: a critique of Sewall Wright's shifting balance theory of evolution. *Evolution* **51**(3), 643–671.

DE WAAL, F. B. M. (1989). *Peacemaking among Primates.* Cambridge, MA: Harvard University Press.

EVANS, C. S., EVANS, L. & MARLER, P. (1993). On the meaning of alarm calls: Functional reference in an avian vocal system. *Animal Behaviour* **46**(1), 23–38.

HAIMAN, J. (1980). The iconicity of grammar: Isomorphism and motivation. *Language* **56**(3), 515–540.

HARLEY, C. B. (1981). Learning the evolutionarily stable strategy. *Journal of Theoretical Biology* **89**, 611–633.

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

HOCKETT, C. F. (1960). Logical considerations in the study of animal communication. In: *Animal sounds and communication* (LANYON, W. E. &

TAVOLGA, W. N., eds.). Washington, DC: American Institute of Biological Sciences, pp. 392–430.

HURFORD, J. R. (1989). Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua* **77**(2), 187–222.

HURFORD, J. R. (1991). The evolution of critical period for language acquisition. *Cognition* **40**(3), 159–201.

KAGAN, J. (1981). *The second year.* Cambridge, MA: Harvard University Press.

KIRBY, S. (2002a). Learning, bottlenecks and the evolution of recursive syntax. In: Briscoe (2002), pp. 173–203.

KIRBY, S. (2002b). Natural Language from Artificial Life. *Artificial Life* **8**(2), 185–215.

KOMAROVA, N. & NOWAK, M. A. (2001). The evolutionary dynamics of the lexical matrix. *Bulletin of Mathematical Biology* **63**(3), 451–484.

LANGACKER, R. W. (1977). Syntactic reanalysis. In: *Mechanisms of Syntactic Change* (LI, C. N., ed.). Austin, TX: University of Texas Press, pp. 57–139.

MARKMAN, E. M. (1989). *Categorization and naming in children: problems of induction.* Cambridge, MA: MIT Press.

MARKMAN, E. M. & WACHTEL, G. F. (1988). Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology* **20**, 121–157.

MAYNARD SMITH, J. (1982). *Evolution and the Theory of Games.* Cambridge: Cambridge University Press.

MAYNARD SMITH, J. & PRICE, G. R. (1973). The logic of animal conflict. *Nature* **146**, 15–18.

MAYNARD SMITH, J. & SZATHMÁRY, E. (1995). *The major transitions in*

evolution. Cambridge: Oxford University Press.

MAZZOCCO, M. M. (1997). Children's interpretations of homonyms: a developmental study. *Journal of Child Language* **24**(2), 441–467.

NOWAK, M. A. & KOMAROVA, N. L. (2001). Towards an evolutionary theory of language. *Trends in Cognitive Sciences* **5**(7), 288–295.

NOWAK, M. A., KOMAROVA, N. L. & NIYOGI, P. (2001). Evolution of universal grammar. *Science* **291**, 114–117.

NOWAK, M. A., PLOTKIN, J. B. & KRAKAUER, D. C. (1999). The evolutionary language game. *Journal of Theoretical Biology* **200**(2), 147–162.

ODLING-SMEE, F. J., LALAND, K. N. & FELDMAN, M. W. (1996). Niche construction. *American Naturalist* **147**(4), 641–648.

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

PINKER, S. & BLOOM, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences* **13**(4), 707–784.

RAEMAEKERS, J. J., RAEMAEKERS, P. M. & HAIMOFF, E. H. (1984). Loud calls of the gibbon (*Hylobates lar*): repertoire, organization and context. *Behaviour* **91**(1–3), 146–189.

SEYFARTH, R. M. & CHENEY, D. L. (1997). Some general features of vocal development in nonhuman primates. In: *Social influences on vocal development* (SNOWDON, C. T. & HAUSBERGER, M., eds.). Cambridge: Cambridge University Press, pp. 249–273.

SLOBIN, D. I. (1977). Language change in childhood and history. In: *Language Learning and Thought* (MACNAMARA, J., ed.). London: Academic Press, pp. 185–221.

SMITH, K. (2002a). The cultural evolution of communication in a population of neural networks. *Connection Science* **14**(1), 65–84.

SMITH, K. (2002b). Natural selection and cultural selection in the evolution of communication. *Adaptive Behavior* **10**(1), 25–44.

SMITH, K. (2003). The Transmission of Language: models of biological and cultural evolution. PhD Thesis, The University of Edinburgh.

SNIEGOWSKI, P. D. & LENSKI, R. E. (1995). Mutation and adaptation: the directed mutation controversy in evolutionary perspective. *Annual Review of Ecology and Systematics* **26**, 553–578.

STEELS, L. & KAPLAN, F. (2002). Bootstrapping grounded word semantics. In: Briscoe (2002), pp. 53–73.

TOMASELLO, M. (1999). *The cultural origins of human cognition.* Harvard: Harvard University Press.

TOMASELLO, M. & CALL, J. (1997). *Primate Cognition.* Oxford: Oxford University Press.

TOMASELLO, M., CALL, J., WARREN, J., FROST, G., CARPENTER, M. & NAGELL, K. (1997). The ontogeny of chimpanzee gestural signals: A comparison across groups and generations. *Evolution of Communication* **1**(2), 223–259.

VOGT, P. & COUMANS, H. (2003). Investigating social interaction strategies for bootstrapping lexicon development. *Journal of Artificial Societies and Social Simulation* **6**(1). http://jasss.soc.surrey.ac.uk/6/1/4.html.

VON FRISCH, K. (1974). Decoding the language of the bee. *Science* **185**, 663–668.

WHITLOCK, M. C. & PHILLIPS, P. C. (2000). The exquisite corpse: a shifting view of the shifting balance. *Trends in Ecology and Evolution* **15**(9), 347–348.

WRIGHT, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.

ZUBERBUHLER, K., NOË, R. & SEYFARTH, R. (1997). Diana monkey long-distance calls: messages for conspecifics and predators. *Animal Behaviour* **53**(3), 589–604.

Footnotes

[1]The cardinality of both sets in the real world is constrained by factors such as the number of semantic distinctions which can be made (which restricts $|\mathcal{M}|$) and the maximum practical word length (which restricts $|\mathcal{S}|$). It is far from clear how the cardinality of these two sets should then be set in the model. For the purposes of this paper I will assume that $|\mathcal{S}| \geq |\mathcal{M}|$. The case where $|\mathcal{S}| < |\mathcal{M}|$ would require the ability to prioritise over meanings, which is not a factor in this model. This case will therefore be ignored.

[2]$N = 200$ for all ILMs outlined in this paper.

[3]$e = 3$ for all ILMs outlined in this paper. Different values of e yield qualitatively similar results.

[4]In order to add noise to a meaning-signal pair $\langle m_i, s_j \rangle$, s_j is replaced with a randomly-selected $s_k \in \mathcal{S}$, where $k \neq j$.

[5]This learning rule implements 'lateral inhibition', mentioned in the context of the evolution of vocabulary systems in, for example, Oliphant (1999), Steels & Kaplan (2002) and Vogt & Coumans (2003).

[6]In the interest of brevity I will omit the analysis for the case when cultural transmission is noise-free. Firstly, noise-free transmission is unlikely. Secondly, the analysis for noise-free transmission produces results which are broadly similar to that for noisy transmission.

[7]$N = 200$ for all EILMs outlined in this paper.

[8]As in the ILMs outlined in the previous section, $e = 3$ for all EILMs

outlined here.

[9]As for the static analysis, I assume that cultural transmission is not noise-free — $\rho = 0.1$.

[10]Relative communicative accuracy is the average communicative accuracy of individuals with the $h-$ genotype divided by the average communicative accuracy of individuals with the $h+$ genotype.

[11]The ability of biased learning rules to change a population's culture to their own advantage constitutes a cultural analogue of niche construction (Odling-Smee *et al.*, 1996).

[12]Cumulative cultural evolution in small populations is also more rapid, meaning that the period of benevolent drift required in a small population is shorter than that required in a large population. An evaluation of the impact of genetic and cultural population size is given in Smith (2003), Chapter 4.

Table 1

Static numerical analysis, $\rho = 0.1$. Entries give average values (over 1000 tests) of $ca'(A, B)$, where individual A has the minority genotype and B has the majority genotype, both being exposed to the same vocabulary. The case where the minority and majority genotypes are the same shows the expected communicative accuracy accruing to an individual with the same genotype as the majority genotype — this is the baseline value which determines whether mutant genotypes will be selected for or against. $h?$ and $h-$ are both evolutionarily stable.

Minority genotype

		$h+$	$h?$	$h-$
	$h+$	0.1	0.1	0.1
Majority genotype $h?$		0.45	0.63	0.61
	$h-$	0.88	0.98	1.0

Table 2

Dynamic analysis with converged culture, $\rho = 0.1$. Each cell gives the average final communicative accuracy in the twenty populations, the number of populations converging on the *minority* genotype, and the number of populations converging on a communicatively-optimal vocabulary. Majority $h-$ populations are completely stable, while majority $h?$ populations are occasionally invaded by $h-$ individuals.

<div align="center">Minority genotype</div>

		$h+$			$h?$			$h-$		
Majority genotype	$h+$		NA		0.1	2	0	0.28	4	4
	$h?$	0.63	0	0		NA		0.68	2	2
	$h-$	0.99	0	20	0.99	0	20		NA	

49

Table 3

Dynamic analysis with random initial culture, $\rho = 0.1$. Only the $h-$ bias is evolutionarily stable.

	Minority genotype								
	$h+$			$h?$			$h-$		
$h+$	NA			0.1	1	0	0.55	10	10
$h?$	0.23	2	0	NA			0.87	8	8
$h-$	0.99	0	20	0.99	0	20	NA		

Majority genotype
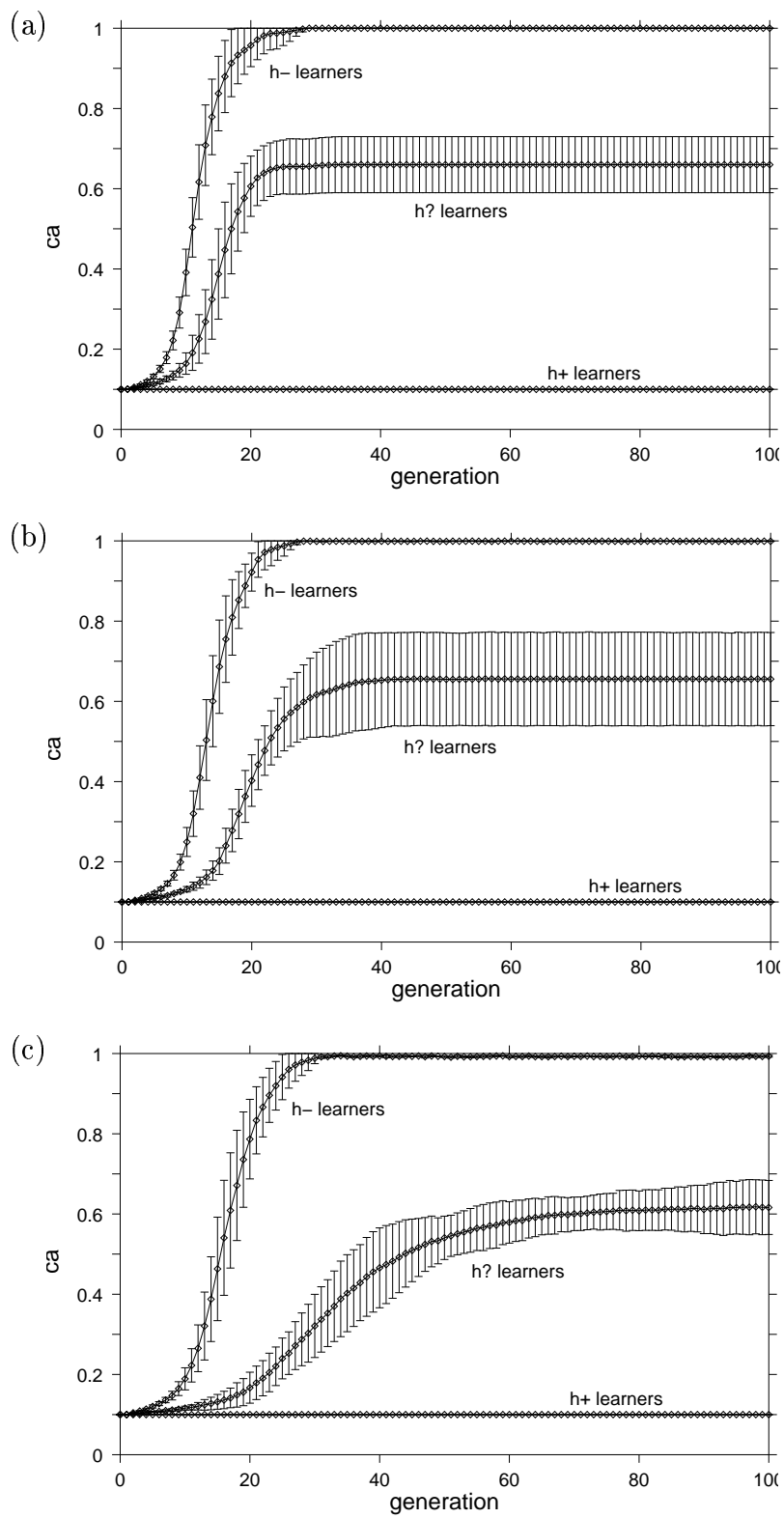
Fig. 1. Cultural evolution as a consequence of learner biases, with various levels of noise. (a) gives results for $\rho = 0$, (b) gives results for $\rho = 0.05$, (c) gives results for $\rho = 0.1$.
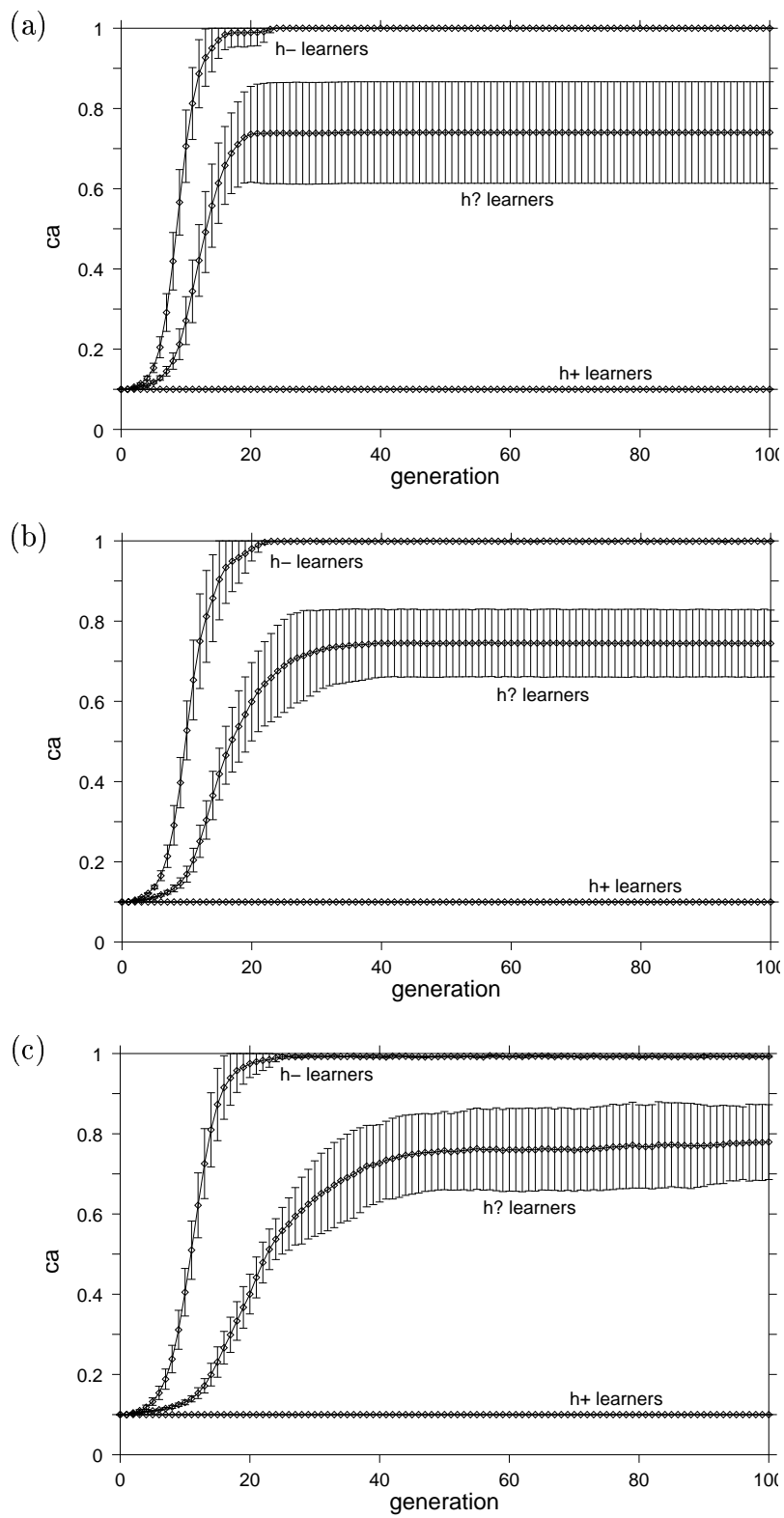
51

Fig. 2. Cultural evolution as a consequence of learner biases and natural selection of cultural variants, with various levels of noise. (a) gives results for $\rho = 0$, (b) gives results for $\rho = 0.05$, (c) gives results for $\rho = 0.1$.
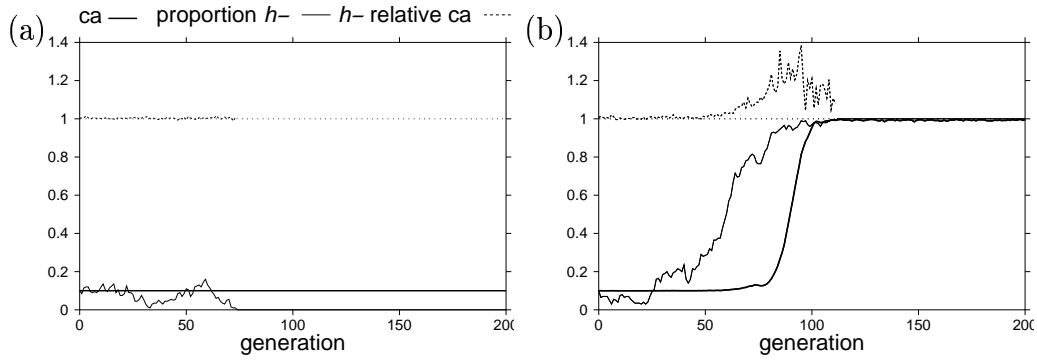
Fig. 3. The evolution of learning biases in a majority $h+$ population converged on a fully homonymous system. (a) shows a run where drift eliminates $h-$ learners before the cultural construction process gets underway. (b) shows a run where drift preserves $h-$ learners in sufficient numbers for their learning bias to take effect. They are then selected for, and the population converges on an optimal vocabulary system.

53