
Learning Bias, Cultural Evolution of Language, and the Biological Evolution of the Language Faculty

KENNY SMITH¹

The biases of individual language learners act to determine the learnability and cultural stability of languages: learners come to the language learning task with biases which make certain linguistic systems easier to acquire than others. These biases are repeatedly applied during the process of language transmission, and consequently should effect the types of languages we see in human populations. Understanding the cultural evolutionary consequences of particular learning biases is therefore central to understanding the link between language learning in individuals and language universals, common structural properties shared by all the world's languages. This paper reviews a range of models and experimental studies which show that weak biases in individual learners can have strong effects on the structure of socially-learned systems such as language, suggesting that strong universal tendencies in language structure do not require us to postulate strong underlying biases or constraints on language learning. Furthermore, understanding the relationship between learner biases and language design has implications for theories of the evolution of those learning biases: models of gene-culture coevolution suggest that, in situations where a cultural dynamic mediates between properties of individual learners and properties of language in this way, biological evolution is unlikely to lead to the emergence of strong constraints on learning.

Language is a socially learned system: children learn the language of their speech community, and this social learning underpins the diversity of the world's languages (Evans and Levinson 2009). However, at a first approximation, learning a language looks like an unusually challenging social learning problem, for two reasons. First, human languages are enormously complex systems, far outstripping the complexity of any documented animal communication system or indeed any other socially learned human behavior. Second, languages must be learned from evidence that is potentially deficient in various ways: for example, language learners learn from noisy linguistic data containing false starts, slips of

¹Language Evolution and Computational Research Unit, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh EH8 9AD, UK. E-mail: kenny@ling.ed.ac.uk.

the tongue, and other sorts of error; learners tend not to receive (and seem not to require) explicit instruction in the structure of the target language; learners do not receive explicit negative evidence that would be useful in ruling out plausible but incorrect hypotheses about the structure of the target language; learners receive incomplete data, in the sense that they are exposed to only a finite set of linguistic data, yet they are required to generalize to an infinitely expressive linguistic system (for a [critical] review of “poverty of the stimulus” claims, see Pullum and Scholz 2002). Despite these apparent difficulties, language acquisition seems to be relatively straightforward: all normally developing individuals acquire the language of their speech community, and do so during childhood, before they are able to master a range of other socially learned skills.

This mismatch between the apparent difficulties of language learning and the apparent ease with which language is acquired motivates the hypothesis that significant components of our knowledge of language must be innate: language learners come to the language acquisition task with some expectations about the nature of the system they are attempting to learn. In its strongest form (as presented by, e.g., Chomsky 1965; Piatelli-Palmarini 1989), language “learning” is considered to be illusory, at least for core structural components of language. Given the insurmountable difficulties of learning a language from data via relatively unconstrained learning processes, all interesting structural properties of language must be prespecified in the learner: “It is, for the present, impossible to formulate an assumption about initial, innate structure rich enough to account for the fact that grammatical knowledge is attained on the basis of the evidence available to the learner” (Chomsky 1965: p.58). Under this account, the cross-linguistic variation sustained by social learning is of a relatively superficial nature, limited to those aspects of language that *can* plausibly be learnt from data by relatively unrestricted processes of inference. The core of language, in particular grammar (the set of constraints on how words can be combined into sequences), is largely prespecified in the learner: innate constraints provide a small number of possible grammars that learners select among on the basis of observed behavior.

This theory also offers rather straightforward explanations of both language universals and the evolution of language in humans. Language universals are (putative; see Evans and Levinson 2009) common features observed across the world’s languages, and range from basic structural properties, often called design features (after Hockett 1960; e.g., languages utilize duality of patterning: a small number of meaningless sounds are recombined to form a large number of meaningful words, which are further recombined to form infinitely many meaningful sentences) to more subtle restrictions (e.g., languages that place the verb before the object in main clauses tend to have prepositions, rather than postpositions; Greenberg 1966). Under the account outlined above, these universals are simply a manifestation of species-wide constraints on language learning, i.e., universals are prespecified in the learner (“there is no doubt that a theory of language, regarded as a hypothesis about the innate ‘language-forming

capacity' of humans, should concern itself with . . . universals"; Chomsky 1965: p.30). The explanation of the evolution of language in the human species is then a biological account of the evolution of this species-wide learning device: because many of the structural properties of language are a reflection of hard-wired biological constraints, their evolution is necessarily and satisfactorily explained in terms of biological evolution under natural selection (Pinker and Bloom 1990).

However, a recent trend has been to revise these strong conclusions about the nature of language learning. This has in part been driven by a more careful consideration of the challenges inherent in language learning. For example, the linguistic data that learners are likely to encounter has been shown to have a surprisingly rich statistical structure that learners are capable of identifying and exploiting (for review, see Monaghan and Christiansen 2008). Furthermore, models of language learning as a process of rational inference show that some key aspects of language can in fact be acquired from realistic corpora of linguistic data by using domain-general learning techniques (e.g., Foraker et al. 2009; Frank et al. 2009).

Language transmission offers a parallel motivation for revising the assumption that language learning necessarily involves strongly constrained processes of learning. Because they are culturally transmitted, languages are potentially under selection for their learnability: even if the set of possible languages permitted by the human language learning device contains languages that are extremely hard to learn from data, cultural transmission over several episodes will lead to such languages being avoided in favor of more learnable languages (Christiansen and Chater 2008; Deacon 1997; Kirby 1999): as Zuidema (2003) appositely puts it, "the poverty of the stimulus solves the poverty of the stimulus." Thus, strong constraints on the space of possible languages do not have to be explicitly and directly hard-wired into learners, because cultural transmission will ensure that the languages which learners tend to encounter will be drawn from the subset of highly learnable languages.

There are a range of factors that might impact on the learnability and transmissibility of languages. This article reviews modeling and experimental work that shows how the biases of individual learners act to determine the learnability and cultural stability of languages.² Learners come to the language learning task with some biases that make certain linguistic systems easier to acquire than others, and as a consequence of cultural transmission these biases have effects on the types of languages we should expect to see in human populations. Understanding the cultural evolutionary consequences of particular learning biases is therefore central to understanding the link between language learning in individuals and language universals: in particular, a range of models

²Possibilities not considered here include functional considerations (either natural selection of language variants during cultural transmission [e.g., Nowak and Komarova 2001] or functional modification of the linguistic system during use [e.g., Croft 2000; De Beule and Bergen 2005; Pugliset al. 2008; Vogt 2005]) or linguistic prestige (e.g., Croft 2000; Labov 1963).

Table 1. A Biased Learning Rule

Parent 1 Language	Parent 2 Language	Probability of Offspring Acquiring	
		L1	L2
L1	L1	1	0
L1	L2	(1/2)(1 + B)	(1/2)(1 - B)
L2	L1	(1/2)(1 + B)	(1/2)(1 - B)
L2	L2	0	1

and experimental studies show that weak biases in individual learners can have strong effects on the structure of socially learned systems, suggesting that strong universals in language do not require us to postulate strong biases or constraints in language learning. Furthermore, understanding the relationship between learner biases and language design has implications for theories of the evolution of those learning biases.

Relationship Between Learning Bias and Distribution of Languages: Modeling Approaches

Weak Biases, Strong Effects. The classic model of biased learning, termed *directly biased transmission*, is provided by Boyd and Richerson (1985).³ This very general model provides a useful framework for discussing transmission models developed more specifically to explore the cultural evolution of language.

Assume there are two languages, L1 and L2. We are interested in the proportion of an infinitely large population that uses L1 at time t , denoted by p_t , and we assume that learners are exposed to the linguistic variants of two cultural parents. Boyd and Richerson provide the biased learning rule given in Table 1, where B gives the strength of the bias in favor of L1, $0 \leq B \leq 1$.

The proportion of the population using L1 after an episode of transmission is then given by the following equation:

$$\begin{aligned} p_{t+1} &= p_t^2(1) + p_t(1 - p_t)(1/2)(1 + B) + (1 - p_t)p_t(1/2)(1 + B) \\ &= p_t + Bp_t(1 - p_t) \end{aligned}$$

The key property of this type of transmission is that the frequency of L1 increases wherever $B > 0$ and $0 < p < 1$: $p = 0$ is an unstable fixed point, and whenever $p > 0$ the bias of learners drives L1 to fixation in the population, with the rate of change depending on the frequency of L1 in the population. Weak biases in this model have strong effects: any $B > 0$ will drive L1 to fixation, given sufficient time.

³Boyd and Richerson (1985) also provide a model of biased modeling, where cultural parents possessing one variant preferentially model another variant for learners. This model, and in particular the notion that different languages may differ in the extent to which they can be identified from the data they produce, has parallels in several models of language transmission, most notably the work of Partha Niyogi (e.g., Niyogi 2006). Such models are not discussed here.

Weak biases in learners can therefore have strong, possibly categorical, effects on the frequencies of competing linguistic variants in a language. Thus, language universals are not necessarily a direct reflection of strong biases in the language acquisition device, but may derive from weak bias in individual learners that have strong or even categorical effects as a result of cultural transmission.

Qualitatively similar results have been obtained in a range of more complex models developed specifically to explore the cultural evolution of various sorts of linguistic system. A recent trend (e.g., Frank et al. 2009; Foraker et al. 2009; Kemp et al. 2007; Xu and Tenenbaum 2007) has been to model language acquisition as a process of rational Bayesian inference. These Bayesian models of learning have corresponding models of cultural transmission via Bayesian social learning (Griffiths and Kalish 2007).

Given some linguistic data d , a Bayesian learner calculates the posterior probability of each language (hypothesis) h according to Bayes' law:

$$p(h | d) = \frac{p(d | h)p(h)}{\sum_h p(d | h)p(h)}$$

where $p(d | h)$ is the likelihood of the data d given h (i.e., the probability that a speaker of language h will produce the set of utterances d) and $p(h)$ is the prior probability of language h (i.e., the learner's belief, prior to encountering any linguistic data, that they will encounter language h). Although it is not necessary to do so, the prior probability distribution over languages can be naturally interpreted as innate constraints on the language learning process. Bayesian models of cultural evolution are therefore useful for exploring the relationship between prior bias and the outcomes of cultural evolution in populations, and indeed one of their main attractions is the explicit identification of prior bias, given by the term $p(h)$.

A learner's task is to select a language based on the posterior probability distribution after observation of some linguistic data, $p(h | d)$. One way to select a hypothesis is to select randomly among the hypotheses with maximum posterior probability (the set of MAP hypotheses, H):

$$p_{MAP}(h | d) = \begin{cases} 1/|H| & \text{if } h \in H \\ 0 & \text{otherwise} \end{cases}$$

Griffiths and Kalish (2007) provide a characterization of cultural transmission in populations of MAP learners that is broadly consistent with the predictions of the direct bias model outlined above. These results depend on the assumption that learners learn from data produced by a single model (a single cultural parent), and that there is some noise on transmission (i.e., speakers of L1 occasionally produce some data that can be interpreted as belonging to L2). Under these conditions,

cultural transmission in such populations will produce a distribution over languages⁴ centered on the language with the highest prior probability; thus, the distribution of languages will reflect the biases of learners, as in Boyd and Richerson's direct bias model. The distribution over languages is relatively insensitive to the strength of the prior preference in favor of particular languages: in common with the direct bias model, weak and strong prior biases in favor of a particular language can yield the same outcomes as a consequence of cultural transmission. Furthermore, factors other than the biases of learners, such as the amount of data to which learners are exposed, can influence the distribution of languages. Kirby et al. (2007) show that, given MAP hypothesis selection, the amount of data learners receive determines the extent to which the prior preference for certain languages is exaggerated. When learners receive small amounts of data, their prior preferences are exaggerated in the eventual distribution of languages: weak prior biases can have particularly strong effects. As the amount of data increases, this exaggeration in favor of the *a priori* more likely languages is attenuated. This result is consistent with earlier computational modeling work (e.g., Kirby 2000, 2001; Smith et al. 2003) that shows that a *bottleneck* on transmission (attempting to learn a large language from a relatively small amount of data) has consequences for language structure, in particular favoring languages that have generalizable structure: in these models, learners have a weak bias in favor of such languages, and the combination of the weak bias and the transmission bottleneck yields a strong effect in the resultant distribution of languages.

Weak Biases, Weak Effects? An alternative to MAP Bayesian hypothesis selection is *sampling*: learners select a language proportionately to its posterior probability:

$$p_{\text{samp}}(h \mid d) = p(h \mid d)$$

Sampling hypothesis selection yields a rather different set of cultural dynamics from both MAP Bayes and the direct bias model. Griffiths and Kalish (2007) show analytically that, in populations where learners learn from a single cultural parent, cultural evolution leads to a distribution over languages identical to the prior probability distribution over those languages; thus, regardless of starting frequencies of the languages or any other factors acting on cultural transmission (e.g., the amount of data learners receive), the stable end point of cultural evolution is solely determined by the prior probability of the various languages. There is an obvious contrast with the results outlined above for the direct bias and MAP Bayes models: under sampling hypothesis selection, weak biases do not

⁴The interpretation of this distribution over languages depends on whether we are considering transmission in a single diffusion chain, in which each generation consists of a single individual who learns from data produced by the preceding individual, or infinitely large populations in which each individual at generation $n + 1$ learns from data produced by a single individual in generation n . In the former case, the stable distribution described above is temporal: although the individual at generation n has a single language, averaging over a large number of generations yields a distribution of languages of the sort characterized above. In the large population case, the distribution described above gives the proportion of individuals at each generation using each language.

have strong effects; rather, the strength of prior bias directly determines the strength of its effects. Although Griffiths and Kalish emphasize that it is not necessary to equate prior bias with innateness (i.e., the prior bias in favor of a particularly linguistic feature could itself be influenced by learning in other domains), this result is consistent with a picture whereby language universals reflect, in a transparent fashion, the biases of language learners.

Given the mismatch between this result and the weak bias-strong effect models discussed above, it would be convenient to be able to discount the results for sampling Bayes learners, for example, as being a poor model of human learning. However, it is important to note that good fit has been obtained between a sampling model of transmission and human data in several studies of learning and cultural transmission (for review, see Griffiths et al. 2008b; also see discussion of Reali and Griffiths 2009, below). A more appropriate approach to understanding the generality of this result is to explore to what extent it depends on unduly restrictive assumptions about transmission in populations.

The proof presented by Griffiths and Kalish (2007) rests on the assumption that learners learn from a single model, and all learners in a population share the same prior bias. An area of ongoing work (Burkett and Griffiths 2009; Dediū 2009; Ferdinand and Zuidema 2009; Smith 2009) explores the consequences of relaxing these assumptions, bringing the model more in line with classic models of cultural transmission in populations (e.g., of the sort provided by Boyd and Richerson 1985). To focus on the case of learning from multiple parents (for an exploration of the consequences of transmission in populations of learners with heterogeneous biases, see Dediū 2009), Smith (2009) shows that when learners learn from multiple models rather than a single model, the direct link between prior bias and the outcomes of cultural transmission disappears. Instead, there is convergence to a distribution where one language dominates, with the dominant language being largely determined by initial frequency and relatively insensitive to strength of prior bias in favor of that language. Burkett and Griffiths (2010) note that this result depends on the assumption that learners attempt (as in Boyd and Richerson's direct bias model) to find a *single* grammar that accounts for all of their data: given that this data may come from multiple models, this assumption may be unwarranted. They provide a model in which the strength of this assumption can be varied in a continuous manner, yielding a smooth transition from situations in which strength of prior bias is relatively unimportant (learners attempt to find a single target grammar to account for their data, as in Smith 2009) to situations where the direct mapping from prior bias to stationary distribution re-emerges (learners learn a distribution over multiple languages, rather than a single language).

Summary. Table 2 summarizes the relationship between strength of bias in individual learners and the outcomes of cumulative cultural evolution in populations of such learners for these three models. The key point is that for several models of cultural transmission show that weak biases can have

Table 2. Summary of the Three Models

<i>Model</i>	<i>Bias Strength</i>	<i>Strength of Cultural Effect</i>
Direct bias (Boyd & Richerson)	Weak	Strong
	Strong	Strong
MAP Bayes	Weak	Strong
	Strong	Strong
Sampling Bayes	Weak	Weak
	Strong	Strong

strong effects as a consequence of cultural transmission. For language, this implies that strong universals tendencies in observed languages do not necessarily reflect strong constraints on language learning at the individual. However, the result for sampling Bayesian learning suggests that under at least some conditions that direct mapping from bias strength to the universals may in fact hold (i.e., strong universal tendencies do in fact reflect strong constraints on learning), and an ongoing and important area of work is to establish the generality of Griffiths and Kalish's result for transmission in sampling populations.

Relationship Between Learning Bias and the Distribution of Languages: Experimental Approaches

As well as studying the impact of learner biases or other factors on cultural evolution in abstract models, it is possible to set up simple cultural systems in the laboratory, by using a range of cultural diffusion methodologies (for reviews, see Mesoudi and Whiten 2008; Whiten and Mesoudi 2008). In cultural diffusion experiments, an artificial population is created and seeded with some behavior, and the change in that behavior over time is observed: this may be to establish the presence of cultural transmission (e.g., in nonhumans, as in Horner et al. 2006), to explore the social-learning strategies involved (e.g., Mesoudi 2008; Mesoudi and O'Brien 2008), or to explore the biases involved in social learning (Griffiths et al. 2008a; Kalish et al. 2007; Mesoudi et al. 2006).

These techniques also have been used to study the cultural evolution of artificial languages, by using a methodology that has been dubbed *iterated artificial language learning* (Kirby et al. 2008; Reali and Griffiths 2009; Smith and Wonnacott 2010)—this involves combining a simple diffusion chain design with an artificial language learning task. In a standard artificial language learning experiment (for review, see Gómez and Gerken 2000), participants are trained on some artificial language (e.g., a set of sequences of words or picture-label pairs following some abstract pattern) and then tested to see how well they have learnt this language (e.g., to what extent they can produce or identify sequences or labelings drawn from the target language). In an iterated artificial language learning experiment, as in the standard diffusion chain method, the language a participant produces during testing is simply used as the target language for the

next individual in a chain of transmission. The language therefore potentially changes as a result of being passed from individual to individual.

Learning Bias and Cultural Evolution of Artificial Languages. Reali and Griffiths (2009) and Smith and Wonnacott (2010) describe iterated artificial language learning experiments that show that biases that are hard to spot in individual learners, by using standard artificial language learning techniques, become apparent over repeated episodes of transmission: biases that seem weak when studied at the individual level can have strong effects on language structure. Both articles explore variability in language as a test case.

Languages tend not to exhibit unpredictable variation: all other things being equal, any variation in language tends to be conditioned on semantic, pragmatic, phonological, or sociolinguistic criteria (Givón 1985). To take an example from English, plurality on nouns in English, notwithstanding a relatively small set of exceptions, tends to be marked by the addition of an -s suffix to the noun: singular *dog*, plural *dogs*. However, the actual realization of the plural marker is somewhat variable: it can be realized as a voiced [z] (as in *dogs*), a voiceless [s] (as in *cats*), or syllabically ([Iz], as in *horses*). Importantly, this variability in the realization of -s is not random: there is a predictable rule relating the form of the plural marker to the final segment in the noun being plural marked: [Iz] is used after sibilants (as in the final [s] of *horse*), [s] is used after voiceless nonsibilants (as in the final [t] of *cat*), and [z] is used after voiced nonsibilants (e.g., [g] of *dog*). Thus, the variability in the realization of plural -s is *conditioned*, in this case on the linguistic (phonological) context, although conditioning on nonlinguistic (e.g., social) context is also possible (Labov 1963).

Why do languages tend to exhibit conditioned, rather than unpredictable (or *free*) variation? One possibility is that there is a simple mapping between properties of individuals and properties of language: the absence of free variation in language reflects the fact that language learners have strong or absolute biases against free variation, which can be observed using standard artificial language learning techniques. Child learners may indeed have such strong, readily observed biases: Hudson et al. (2005) show that, when trained on a system exhibiting unpredictable variation (namely, a language in which two variants of a meaningless marker word alternate randomly), 6-year-old children typically eliminate this unpredictable variation by eliminating one of the two variants entirely, producing only one version of the marker word on test (also see the wealth of evidence on child biases in favor of morphological regularity, e.g., Brown 1973; Vennemann 1978).

However, the models discussed in “Relationship Between Learning Bias and Distribution of Languages: Modeling Approaches” suggest at least two other possible explanations for the rarity of free variation in language. The first possibility is essentially a rather subtle variant on the position outlined by Hudson Kam & Newport, and hinges on the point that even strong biases may not be apparent in experimental designs which look only at individual isolate

learners, but may nonetheless have strong, observable consequences as a result of cultural transmission. Strong learner biases (in this case, in favor of predictable variation) may be hard to spot in the lab because such biases are only weakly apparent after exposure to large amounts of linguistic data: in Bayesian terms, given sufficiently large amounts of data, $p(d | h)$ will drown out differences in $p(h)$. However, as discussed under “Relationship between Learning Bias and the Distribution of Languages: Modeling Approaches”, these strong biases will become manifest over repeated episodes of cultural transmission (again in Bayesian terms, through iterated learning under either MAP or sampling hypothesis selection), leading to strong effects on language structure, as manifested, in this case, by the scarcity of free variation. For convenience, in the remainder of this section I refer to this as a case of *weak posterior bias* (indicating weak a posteriori effects of a strong bias in individual learners). To return to Hudson et al. (2005), in addition to demonstrating that child learners have a strong bias against variation, they show that no such bias is apparent in adult learners, and they conclude that the absence of free variation in language must therefore be due to the biases of child learners. However, this is not necessarily the case: it may be that their adult data reflects a case of weak posterior bias and that an equivalent experiment involving cultural transmission would in fact yield a strong signal of this strong underlying bias.

The second possible explanation for the absence of free variation in language (that I will refer to as the *weak prior bias* account, to contrast with the weak posterior bias account) is that learners may have a weak prior bias in favor of predictability, which only has strong effects when placed in a cultural context (assuming MAP hypothesis selection). The adult data provided by Hudson et al. would be compatible with the weak prior bias account: adult biases against variability are weak and therefore hard to spot in a standard artificial language learning experiment, but they might nonetheless have categorical consequences for language design as a consequence of their repeated application during transmission. The key point here is that the data provided by Hudson et al. from individual learners cannot arbitrate between these two possibilities and their preferred child-centered account and therefore does not allow us to conclude that the absence of free variation in language must necessarily be due to strong biases in child learners. To resolve these questions, we must investigate the consequences of cultural transmission explicitly. In this case, iterated artificial language learning provides the required methodology.

Reali and Griffiths (2009) describe an iterated artificial language learning experiment in which learners are required to learn a set of object-label pairs. In the initial language, exhibiting unpredictable variation, each object is paired with two labels (i.e., the choice of label is unpredictable). Adult learners exposed to this unpredictable variation seem to match the unpredictability of their input language reasonably closely: there is no statistical signature of regularization in the language they produce on test. However, as these languages are transmitted along diffusion chains they become increasingly predictable: one of the two labels for each object is lost, eventually yielding a final system that exhibits no variability. Although this

result is compatible with both weak posterior and weak prior biasing, Reali and Griffiths show that a somewhat better fit to their data is obtained by assuming a weak posterior bias (i.e., a strong prior bias against unpredictability,⁵ masked by large amounts of data during training, but becoming apparent through cultural transmission under sampling hypothesis selection).

Smith and Wonnacott (2010) present an iterated artificial language learning experiment based around a slightly different learning task: learners are required to produce descriptions of scenes involving one or two animals. An initial group of adult participants attempted to learn a target language in which nouns are marked for plurality using one of two plural markers, with the choice of plural marker varying unpredictably. Again, these isolate learners seem to capture this unpredictability fairly reliably—there is no strong statistical signature of a bias against unpredictability. None the less, and similarly to the Reali and Griffiths finding, the artificial languages increase in predictability over repeated episodes of transmission—after five generations, nine of 10 diffusion chains exhibit entirely predictable plural marker usage. However, unlike in Reali and Griffiths' experiment, this increase in predictability is not achieved by eliminating variants: both plural markers are retained in the majority of the diffusion chains. The difference from the Reali and Griffiths result is due to the availability of (rather minimal) linguistic context: over episodes of transmission, the usage of the two plural markers becomes conditioned on the noun being marked, such that plurality on some nouns is marked with one marker, and plurality on other nouns is marked with the alternative marker. The availability of conditioning context, absent from the Reali and Griffiths study, allows variability to be stable over repeated episodes of transmission in spite of a learner bias in favor of predictability: conditioned variability, of the sort witnessed in real languages, is the result. Again, this experimental result is consistent with either the weak posterior or weak prior bias accounts, and the work of identifying which is more plausible remains to be done.

Summary. As with the models discussed under “Relationship between Learning Bias and the Distribution of Languages: Modeling Approaches”, these experiments show that linking properties of language to properties of individual language learners is nontrivial. Although we might be tempted to assume that strong effects in language design must map straightforwardly to strong biases in language learners, the experimental data again shows that this conclusion is not safe: strong effects can emerge as a consequence of cultural transmission, even if individual-based experimentation might suggest that the biases of individual experimental participants are weak.

Evolutionary Implications

The modeling and experimental work described above addresses the extent to which we should expect observed distributions of language in human populations to

⁵Note that this could either a language-specific bias or simply a manifestation of a general expectation for events in the world to be somewhat predictable.

straightforwardly reflect the biases of the human language learning device. The general picture is that there is not necessarily a straightforward correspondence between typology and learner bias: under at least some models, weak biases in individual learners can have strong effects in populations. The same set of results has profound implications for theories of the evolution of those language learning capacities themselves. The consistent evolutionary prediction is that biological evolution should not deliver learning biases that are both strongly constraining and language-specific, for two reasons: the equivalence of the cultural outcomes of weak and strong learning biases mask bias strength from selection, and the frequency-dependent nature of communication payoffs weakens the selection pressure in favor of desirable learning biases when those biases are rare.

Shielding of Bias Strength Due to Cultural Evolution. The fact that weak biases can have strong effects (e.g., in Boyd and Richerson's model of directly biased transmission, or transmission under MAP Bayesian learning) poses a fundamental problem for the evolution of strong biases in favor of particular culturally transmitted systems. These models show that weak biases can have strong cumulative effects and that the stable outcomes of cultural transmission in populations can be identical under a range of bias strengths. Consequently, it is possible that biological evolution is neutral with respect to strength of bias.

Again, Boyd and Richerson's directly biased transmission model can be used to illustrate this point. Consider a population in which there are two strengths of bias in favor of L1, $B_1 > B_2 > 0$, with bias strength being genetically transmitted from parent to offspring. Further assume that individuals who acquire L1 receive some payoff relative to L2 learners: L1 has some *a priori* functional advantages over L2 (e.g., it might be more expressive than L2, or more concise). In such a population, a combination of biased learning and natural selection of cultural variants (L1 speakers are more likely to act as models than L2 speakers) will drive L2 out of the population, at which point any selective advantage associated with B_1 will disappear: in a culturally homogenous population, all learners learn the sole cultural variant in the population, regardless of bias strength (see Table 1). Thus, at cultural equilibrium there is selective neutrality with respect to bias strength, assuming biases of differing strengths are equally costly: if stronger biases are more costly (as is commonly assumed), then there will be selection in favor of weaker biases. More generally, Boyd and Richerson (1985) show that any selective advantage for biased learning relative to unbiased learning ($B = 0$) depends on cultural variation in the population, and cultural evolution under directly biased transmission eliminates that variation, thereby eliminating evolutionary pressure in favor of such a bias.

Smith and Kirby (2008) present a related result for the evolution of prior biases in populations of Bayesian learners. They consider an evolutionary extension of the MAP Bayes model of the type outlined above and show that if bias strength is genetically transmitted and selection operates on the stationary outcomes of cultural transmission (i.e., cultural change is fast relative to biological evolution), then selection is neutral over strength of bias: strongly

biased learners receive no selective advantage over weakly biased learners, because both are likely to occupy populations that are converged to the type of system matching their bias and are equally capable of acquiring such a system. This result does not hold for sampling Bayesian learners: populations of such learners converge to a mixed system, and strength of bias has a marked impact on ability to acquire the right system for sampling learners. Consequently, there is a payoff to having stronger biases in sampling populations: the shielding of bias strength by cultural evolution identified by Boyd and Richerson is dependent on models of cultural evolution (direct bias, MAP Bayes) that lead to indirect mappings from properties of individuals to properties of evolved cultural systems.

Coordination and the Evolution of Learning Biases. Boyd and Richerson's analysis of the evolution of direct bias assumes that there is some cultural variant that is *a priori* and always more functional than alternatives: selection will favor this bias up until the point where the cultural variance on which this selection depends is used up. However, the assumption of an *a priori* desirable cultural variant seems to apply less to coordination problems such as language. Although it may be the case that some linguistic variants are more functional than others (for an example, see below), this is potentially modulated by the requirement for language users to coordinate: using a more functional linguistic system is not necessarily advantageous if no one else in your population uses (produces or understands) that system. Thus, using the optimal linguistic system is potentially less important than using the linguistic system that other members of your population use.⁶ This has two consequences for the evolution of learning biases: it renders sampling Bayes hypothesis selection evolutionarily unstable (and consequently predicts that selection over bias strength should be neutral), and it makes evolving biases in favor of *a priori* desirable communication systems difficult.

Evolutionary Stability of Sampling and MAP Bayes. As discussed above, sampling and MAP hypotheses selection make different predictions regarding the strength of biases we should expect to see: strong biases in sampling populations, neutrality with respect to bias strength (or weak biases if strong biases are inherently costly) in MAP populations. Smith and Kirby (2008) show that, due to the fact that communicative payoffs require coordination, only MAP hypothesis selection is evolutionarily stable. MAP learners maximize their chances of acquiring the same system as other Bayesian learners exposed to the same data (regardless of whether they use MAP or sampling hypothesis selection). Sampling learners do not—they attribute some probability to hypotheses with lower posterior probability, and therefore reduce the probability of coordinating with other learners exposed to

⁶It is important to note that this is not necessarily always true. For example, Zuidema and de Boer (2009) show that functional, combinatorial sound systems can invade populations using less functional sound systems even if initially rare. Whether or not particular linguistic feature suffers from frequency dependence therefore needs to be established on a case-by-case basis, although it seems to be a reasonable default assumption that frequency dependence problems will apply.

the same data generating source. Consequently, only MAP hypothesis selection is evolutionarily stable for coordination problems like language: sampling populations can be invaded (rapidly) by initially rare MAP learners. This analysis therefore suggests that biological evolution should select for MAP learning, and consequently selection over bias strength should be neutral (as discussed above).

Scarcity Problems. Cumulative cultural evolution takes time; consequently, social learning may not be selected for when rare (Boyd and Richerson 1996). Cumulative cultural evolution eventually produces behaviors that are more complex and functional than those that can be discovered via asocial learning processes (also known as individual learning: learning via independent exploration of the environment without social influence). Consequently there is strong selective pressure in favor of ability to socially learn in populations in which such behaviors have been established via cultural processes. However, at the early stages of the process of cumulative cultural evolution, the advantages of learning socially will be limited (the behaviors available for social learning may be no more adaptive than those which could be achieved via asocial learning) and may be outweighed by the costs of social learning, resulting in selection in favor of purely asocial learning. Smith (2004) shows that this problem applies to the evolution of biased learning as well, for coordination problems such as language.

Smith (2004) considers the cultural transmission of vocabulary systems in populations in which learners have some genetically coded bias in favor of vocabularies with varying functionality. All other things being equal, a one-to-one vocabulary system is optimal in terms of communication: if two individuals share a one-to-one vocabulary system, the lack of ambiguity in the system enables a hearer to correctly identify the object communicated about by a speaker. In contrast, many-to-one vocabulary systems are less functional because they associate multiple objects with the same label; consequently, ambiguous words leave the hearer with some uncertainty as to the intended referent. As expected given the models discussed above, in populations that are homogeneous with respect to learning bias, the population's vocabulary system comes to reflect the biases of the learners, with consequences for communication in those populations: weak bias in favor of one-to-one mappings result in the cultural evolution of optimal communication systems; biases in favor of many-to-one mappings result in maximally ambiguous vocabularies, with consequently low levels of communicative accuracy. There is evidence from a range of sources that language learners do in fact have a general expectation that language should embody a one-to-one mapping between underlying semantic structures and surface forms (e.g., Langacker 1977; Slobin 1977), and this general bias manifests itself at all levels of linguistic structure (in morphology, in the lexicon, and in syntax; for reviews, see Smith 2003, 2004). The prediction of the models and experimental data outlined so far suggest that the consequence of such biases will be a linguistic system which is well adapted for communication.

Smith (2004) then considers populations that are heterogeneous with respect to learning bias, parents transmit their bias to their offspring, and reproduction is

proportionate to communicative success, with individuals whose vocabulary systems enable them to successfully communicate with other population members being likely to reproduce more. A reasonable expectation would be that selection acting on the transmission of learning biases can identify those learning biases which lead, via cultural processes, to functional vocabularies, resulting in the eventual biological evolution of one-to-one learning biases (of the type that humans seem to possess) and the cultural evolution of unambiguous vocabulary systems (as shown by Boyd and Richerson's general model of the evolution of direct bias for non-coordination problems). Although the results of several simulation runs of the coevolutionary model show that this is a possible outcome, it is however contingent on a period of genetic drift maintaining the advantageous bias in the population for sufficient time to allow its cumulative cultural effects to be felt: individuals bearing the one-to-one bias receive no fitness advantage in populations where they are rare, because other individuals are unlikely to share their vocabulary even if they share the same learning bias. Although individuals with the "desirable" bias do accrue some advantage if they remain in the population in sufficient numbers to begin the construction of a shared and unambiguous vocabulary, they are prone to elimination by drift at the early stages of this process.

Summary. All of these evolutionary analyses suggest that the evolution of strong domain-specific learning biases for language should not be the default assumption, but instead require special explanation: under a fairly wide range of circumstances, we should in fact expect biological evolution to deliver weak learning biases for language (also see Chater et al.; Christiansen, this issue).

Conclusions

Language is a socially transmitted system; consequently, the biases of language learners impact on its structure. There are two obvious implications of this. First, weak biases can have strong effects: near-categorical effects in language structure (language universals) do not necessarily reflect strong constraints on the learning system. Second, understanding the relationship between learner bias and eventual language design has implications for our understanding of the evolution of the capacity for language in humans: models that take account of cultural transmission effects make rather different predictions about the nature of the human language faculty from theories that focus on language learning as an individual-level phenomenon. In particular, dual-transmission models of language predict that we should not expect biological evolution of the language faculty to result in strongly constraining, domain-specific learning biases. Indeed, the problem such biases were originally invoked to explain—the challenge facing language learners in learning languages from data—may be largely attenuated by the very fact of language transmission: languages evolve to become learnable, removing some of the onus from biological evolution to explain how learners can solve the language learning problem.

Received 14 April 2010; revision accepted for publication 24 July 2010.

Literature Cited

- Boyd, R., and P. J. Richerson. 1985. *Culture and the Evolutionary Process*. Chicago, IL: University of Chicago Press.
- Boyd, R., and P. J. Richerson. 1996. Why culture is common but cultural evolution is rare. *Proc. Br. Acad.* 88:73–93.
- Brown, R. 1973. *A First Language*. Cambridge, MA: Harvard University Press.
- Burkett, D., and T. L. Griffiths. 2010. Iterated learning of multiple languages from multiple teachers. In *The Evolution of Language (EVOLANG 8): Proceedings of the 8th International Conference on the Evolution of Language*, A. D. M. Smith, M. Schouwstra, B. de Boer, and K. Smith, eds. Singapore: World Scientific.
- Chater, N., F. Reali, and M. H. Christiansen. 2009. Restrictions on biological adaptation in language evolution. *Proc. Natl. Acad. Sci. USA* 106:1015–1020.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Christiansen, M. H., and N. Chater. 2008. Language as shaped by the brain. *Behav. Brain Sci.* 31:489–509.
- Christiansen, M. H., F. Reali, and N. Chater. 2011. Biological Adaptations for Functional Features of Language in the Face of Cultural Evolution. *Human Biology* 83(2):247–259.
- Croft, W. 2000. *Explaining Language Change: An Evolutionary Approach*. London, U.K.: Longman.
- de Beule, J., and B. K. Bergen. 2006. “On the emergence of compositionality.” In *The Evolution of Language: Proceedings of the 6th International Conference*, A. Cangelosi, A. D. M. Smith, and K. Smith, eds. Singapore: World Scientific, 35–42.
- Deacon, T. 1997. *The Symbolic Species*. London, U.K.: Penguin.
- Dediu, D. 2009. Genetic biasing through cultural transmission: Do simple Bayesian models of language evolution generalize? *J. Theor. Biol.* 259:552–561.
- Evans, N., and S. C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32:429–492.
- Ferdinand, V., and W. Zuidema. 2009. Thomas’ theorem meets Bayes’ rule: A model of the iterated learning of language. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, N. A. Taatgen and H. van Rijn, eds. Austin, TX: Cognitive Science Society, 1786–1791.
- Foraker, S., T. Regier, N. Khetarpal, A. Perfors, and J. Tenenbaum. 2009. Indirect evidence and the poverty of the stimulus: The case of anaphoric “one.” *Cogn. Sci.* 33:287–300.
- Frank, M. C., N. D. Goodman, and J. Tenenbaum. 2009. Using speakers’ referential intentions to model early cross-situational word learning. *Psychol. Sci.* 20:578–585.
- Givón, T. 1985. Function, structure, and language acquisition. In *The Crosslinguistic Study of Language Acquisition*, vol. 2, D. Slobin, ed. Hillsdale, NJ: Lawrence Erlbaum, 1005–1028.
- Gómez, R. L., and L. A. Gerken. 2000. Artificial language learning and language acquisition. *Trends Cogn. Sci.* 4:178–186.
- Greenberg, J. H. 1966. *Universals of Language*. Cambridge, MA: MIT Press.
- Griffiths, T. L., and M. L. Kalish. 2007. Language evolution by iterated learning with Bayesian agents. *Cogn. Sci.* 31:441–480.
- Griffiths, T. L., B. R. Christian, and M. L. Kalish. 2008a. Using category structures to test iterated learning as a method for identifying inductive biases. *Cogn. Sci.* 32:68–107.
- Griffiths, T. L., M. L. Kalish, and S. Lewandowsky. 2008b. Theoretical and experimental evidence for the impact of inductive biases on cultural evolution. *Philos. Trans. R. Soc.* 363: 3503–3514.
- Hockett, C. F. 1960. The origin of speech. *Sci. Am.* 203:88–96.

- Horner, V., A. Whiten, E. Flynn, and F. B. M. de Waal. 2006. Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proc. Natl. Acad. Sci. USA* 103:13878–13883.
- Hudson Kam, C., and E. L. Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Lang. Learn. Dev.* 1:151–195.
- Kalish, M. L., T. L. Griffiths, and S. Lewandowsky. 2007. Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychon. Bull. Rev.* 14:288–294.
- Kemp, C., A. Perfors, and J. B. Tenenbaum. 2007. Learning overhypotheses with hierarchical Bayesian models. *Dev. Sci.* 10:307–321.
- Kirby, S. 1999. *Function, Selection and Innateness: the Emergence of Language Universals*. Oxford, U.K.: Oxford University Press.
- Kirby, S. 2000. Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, C. Knight, ed. Cambridge, U.K.: Cambridge University Press, 303–323.
- Kirby, S. 2001. Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity. *IEEE Trans. Evol. Comput.* 5:102–110.
- Kirby, S., H. Cornish, and K. Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proc. Natl. Acad. Sci. USA* 105:10681–10686.
- Kirby, S., M. Dowman, and T. L. Griffiths. 2007. Innateness and culture in the evolution of language. *Proc. Natl. Acad. Sci. USA* 104:5241–5245.
- Labov, W. 1963. The social motivation of a sound change. *Word* 19:273–309.
- Langacker, R. W. 1977. Syntactic reanalysis. In *Mechanisms of Syntactic Change*, C. N. Li, ed. Austin, TX: University of Texas Press, 57–139.
- Mesoudi, A. 2008. An experimental simulation of the “copy-successful-individuals” cultural learning strategy: Adaptive landscapes, producer-scrounger dynamics and informational access costs. *Evol. Hum. Behav.* 29:350–363.
- Mesoudi, A., and M. J. O’Brien. 2008. The cultural transmission of Great Basin projectile point technology I: An experimental simulation. *Am. Antiq.* 73:3–28.
- Mesoudi, A., and A. Whiten. 2008. The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philos. Trans. R. Soc. Lond. B* 363:3489–3501.
- Mesoudi, A., A. Whiten, and R. Dunbar. 2006. A bias for social information in human cultural transmission. *Br. J. Psychol.* 97:405–423.
- Monaghan, P., and M. H. Christiansen. 2008. Integration of multiple probabilistic cues in syntax acquisition. In *Trends in Corpus Research: Finding Structure in Data*, H. Behrens, ed. Amsterdam, Netherlands: John Benjamins, 139–163.
- Niyogi, P. 2006. *The Computational Nature of Language Learning and Evolution*. Cambridge, MA: MIT Press.
- Nowak, M. A., and N. L. Komarova. 2001. Towards an evolutionary theory of language. *Trends Cogn. Sci.* 5:288–295.
- Piattelli-Palmarini, M. 1989. Evolution, selection and cognition: From “learning” to parameter setting in biology and in the study of language. *Cognition* 31:1–44.
- Pinker, S., and P. Bloom. 1990. Natural language and natural selection. *Behav. Brain Sci.* 13:707–784.
- Puglisi, A., A. Baronchelli, and V. Loreto. 2008. Cultural route to the emergence of linguistic categories. *Proc. Natl. Acad. Sci. USA* 105:7936–7940.
- Pullum, G. K., and B. C. Scholz. 2002. Empirical assessment of stimulus poverty arguments. *Linguist. Rev.* 19:9–50.
- Reali, F., and T. L. Griffiths. 2009. The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition* 111:317–328.
- Slobin, D. I. 1977. Language change in childhood and history. In *Language Learning and Thought*, J. Macnamara, ed. London, U.K.: Academic Press, 185–221.

- Smith, K. 2003. *The Transmission of Language: Models of Biological and Cultural Evolution*. Ph.D. dissertation, The University of Edinburgh, Edinburgh, U.K.
- Smith, K. 2004. The evolution of vocabulary. *J. Theor. Biol.* 228:127–142.
- Smith, K. 2009. Iterated learning in populations of Bayesian agents. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, N. A. Taatgen and H. van Rijn, eds. Austin, TX: Cognitive Science Society, 697–702.
- Smith, K., H. Brighton, and S. Kirby. 2003. Complex systems in language evolution: the cultural emergence of compositional structure. *Adv. Complex Syst.* 6:537–558.
- Smith, K., and S. Kirby. 2008. Cultural evolution: Implications for understanding the human language faculty and its evolution. *Philos. Trans. R. Soc. B* 363:3591–3603.
- Smith, K., and E. Wonnacott. 2010. Regularization of unpredictable variation through iterated learning. *Cognition* 116:444–449.
- Vennemann, T. 1978. Phonetic analogy and conceptual analogy. In *Readings in Historical Phonology*, P. Baldi and R. N. Werth, eds. University Park, PA: The Pennsylvania State University Press, 258–274.
- Vogt, P. 2005. The emergence of compositional structures in perceptually grounded language games. *Artif. Intell.* 167:206–242.
- Whiten, A., and A. Mesoudi. 2008. Establishing an experimental science of culture: Animal social diffusion experiments. *Philos. Trans. R. Soc. Lond. B* 363:3477–3488.
- Xu, F., and J. B. Tenenbaum. 2007. Word learning as Bayesian inference. *Psychol. Rev.* 114:245–272.
- Zuidema, W. 2003. How the poverty of the stimulus solves the poverty of the stimulus. In *Advances in Neural Information Processing Systems*, vol. 15, S. Becker, S. Thrun, and K. Obermayer, eds. Cambridge, MA: MIT Press, 51–58.
- Zuidema, W., and B. de Boer. 2009. The evolution of combinatorial phonology. *J. Phon.* 37:125–144.