Culture shapes the evolution of cognition

Bill Thompson[a,b,1], Simon Kirby[b], and Kenny Smith[b]

[a]Artificial Intelligence Laboratory, Vrije Universiteit Brussel, B-1050 Brussels, Belgium; and [b]School of Philosophy, Psychology & Language Sciences, University of Edinburgh, Edinburgh EH8 9YL, United Kingdom

A central debate in cognitive science concerns the nativist hypothesis, the proposal that universal features of behavior reflect a biologically determined cognitive substrate: For example, linguistic nativism proposes a domain-specific faculty of language that strongly constrains which languages can be learned. An evolutionary stance appears to provide support for linguistic nativism, because coordinated constraints on variation may facilitate communication and therefore be adaptive. However, language, like many other human behaviors, is underpinned by social learning and cultural transmission alongside biological evolution. We set out two models of these interactions, which show how culture can facilitate rapid biological adaptation yet rule out strong nativization. The amplifying effects of culture can allow weak cognitive biases to have significant population-level consequences, radically increasing the evolvability of weak, defeasible inductive biases; however, the emergence of a strong cultural universal does not imply, nor lead to, nor require, strong innate constraints. From this we must conclude, on evolutionary grounds, that the strong nativist hypothesis for language is false. More generally, because such reciprocal interactions between cultural and biological evolution are not limited to language, nativist explanations for many behaviors should be reconsidered: Evolutionary reasoning shows how we can have cognitively driven behavioral universals and yet extreme plasticity at the level of the individual—if, and only if, we account for the human capacity to transmit knowledge culturally. Wherever culture is involved, weak cognitive biases rather than strong innate constraints should be the default assumption.

nativism | evolution | culture | language

A central debate in cognitive science concerns the nativist hypothesis: the proposal that certain universal features of human behavior can be explained by a biologically determined cognitive substrate consisting of "reliably-developing conceptual primitives, content-specialized inferential procedures, representational formats that impose contentful features on different inputs, domain-specific skeletal principles" (ref. 1, p. 309). The nativist hypothesis has been advanced for numerous psychological phenomena, such as concepts (2), folk psychology (3), music perception (4), and religious belief (5).

Perhaps the most widely known example of nativist reasoning comes from Chomsky's work on language: Linguistic nativism proposes a domain-specific faculty of language that strongly constrains which languages can be learned (6). Linguistic nativism is sometimes taken as the most successful example of nativist reasoning (1), and proof that nativist explanations are necessarily true in at least some domains.

The presence of innate domain-specific constraints on cognition would clearly require an explanation. Such constraints have been persuasively argued to be likely products of natural selection (7). Specifically dealing with constraints on language learning, Pinker and Bloom argue that coordinated constraints on variation may facilitate communication and therefore be adaptive: "the requirement for standardization of communication protocols dictates that . . .many grammatical principles and constraints must accordingly be hardwired into the [language acquisition] device" (ref. 7, p. 720). This leads naturally to what we will call strong linguistic nativism: an account proposing that

there are features of our biologically inherited cognitive machinery that provide hard constraints on what behaviors can be acquired, and that these constraints are domain-specific in the sense that they have evolved via biological evolution under pressure for enhanced communication; in other words, they have evolved for language.

Recent work has questioned the nativist position, either by reassessing the challenges facing language learners (8) or by questioning the evidence for certain types of language universals (9). Here we focus instead on the evolutionary reasoning behind strong nativism. Language, like many other human behaviors, is underpinned by social learning and cultural transmission alongside biological evolution (10, 11): Learners acquire the language of their speech community via a process of learning from observations of the linguistic behavior of that community. Consequently, language is a product of at least two evolutionary processes— biological evolution of the language faculty, and cultural evolution of languages.

In this paper, we test the evolutionary plausibility of strong nativism, by setting out a general model of the interactions between biological and cultural evolution. Two instantiations of this model show that culture can facilitate rapid biological adaptation of the language faculty yet does not deliver hard constraints on learning. Even when strong universal tendencies in behavior emerge, there are few circumstances where strong domain-specific innate constraints on cognition evolve, but many in which culture bootstraps the rapid fixation of weak, defeasible inductive biases.

Evolutionary Perspectives on Linguistic Nativism

How should we expect biology and culture to interact to shape language and the language faculty? One possibility is that the

Significance

A central debate in cognitive science concerns the nativist hypothesis: the proposal that universal human behaviors are underpinned by strong, domain-specific, innate constraints on cognition. We use a general model of the processes that shape human behavior—learning, culture, and biological evolution— to test the evolutionary plausibility of this hypothesis. A series of analyses shows that culture radically alters the relationship between natural selection and cognition. Culture facilitates rapid biological adaptation yet rules out nativism: Behavioral universals arise that are underpinned by weak biases rather than strong innate constraints. We therefore expect culture to have dramatically shaped the evolution of the human mind, giving us innate predispositions that only weakly constrain our behavior.

rate of language change far exceeds that of biological evolution, preventing genetic assimilation of linguistic features into the language faculty (12): the "moving target" argument. Chater et al. (12) demonstrate this point convincingly in a series of simulation models showing that biological evolution cannot encode innate constraints for linguistic features that change rapidly as a result of external factors (e.g., language contact). However, some aspects of language exhibit stable statistical regularities, or "language universals": These are at the center of nativist reasoning and commonly thought to be held stable by constraints on learning (13). Chater et al. also simulate this scenario, finding that, when language is influenced more by the innate constraints of learners than by external factors (a possibility they consider implausible but which might be assumed under the nativist hypothesis), strong constraints on learning do evolve. This work highlights the necessity of understanding how genetic and cultural processes interact to shape language: The plausibility of strong linguistic nativism is contingent on the relationship between individual-level biases and population-level languages.

Another approach to the evolution of constraints on learning is provided by ref. 14, who consider the evolution of constraints on the size of the set of languages considered by learners. They provide two general results. Firstly, there is selection in favor of a language faculty that reduces the size of the search space to levels that allow a population to converge on a shared grammar, as predicted by ref. 7. However, selection will not lead to the most constrained possible learner, because there are costs to being overly constrained (i.e., inability to acquire one of several languages at use in the population). This balance of selective pressures yields learners whose language faculty is permissive, allowing them to learn the largest possible set of languages, but which is nonetheless constraining enough to permit convergence within a population.

Both of these models present important limits on strong nativism. However, both models only consider the evolution of strong constraints on learning: innate mechanisms that effectively dictate whether or not a particular grammar can be acquired. This restriction to "nativism or nothing" rules out a broad class of possible forms of innateness. Many aspects of human cognition (15), and language acquisition in particular (16), may be better characterized by soft constraints: probabilistic inductive biases that can impose a continuum of preferences ranging from weak to strong. Probabilistic inductive biases in acquisition have been proposed to account for universals concerning word order generalizations (13) and hierarchical phrase structure (17) in syntax, suffixing and prefixing asymmetries in morphosyntax (18), and patterns of vowel harmony (19) and velar palatalization (20) in phonology.

A well-known property of cultural evolution is that, under a wide range of circumstances, weak inductive biases acting on learning can have strong effects in the cultural system as the effects of those biases accumulate (10, 21, 22): Given enough time, a weak bias in favor of a particular trait can eventually drive that trait to fixation. Probabilistic constraints combined with culture therefore potentially provide an alternative to nativism or nothing: Behavioral universals can be underpinned by weak biases at the level of the individual.

Here we present two models that allow us to study the biological evolution of a capacity for learning a culturally transmitted behavior. These models combine learning, culture, and evolution in a general way and allow us to look for conditions under which strong nativism might be evolutionarily plausible. We focus on the case where language change is influenced solely by learning, and not by external factors, because it is here that linguistic nativism has received support from evolutionary reasoning, and construct our models around a well-understood, general model of cognition: Bayesian inference. Bayesian models of cognition allow us to explicitly model the influence of innate

inductive biases (of various strengths) and environmental input on inference (23, 24), while generalizing over the particular psychological mechanisms that could implement inductive bias. Contemporary approaches to linguistic nativism have cast the debate directly in terms of inductive biases (25, 26), and several core topics of linguistic nativism have been addressed with Bayesian cognitive models.

Our models are maximally abstract: Language learning is reduced to acquisition of a single linguistic feature. Acquisition of natural languages involves learning about many statistically intertwined linguistic features simultaneously; however, this simplification allows us to derive general evolutionary results that relate inductive biases to linguistic structures.

Model 1

In Model 1, we assume a language can be described by a set of discrete absolute principles and potentially variable parameters (27). This perspective has a long pedigree in linguistics, has been described as "the bona fide theory of innateness" (ref. 28, p. 451), and enjoys wide support as a model for nativism, in the broad sense that it posits universal principles and narrowly restricted options for cross-linguistic variation (28–30).

Model Details. Let F be a linguistic feature that can vary in a binary manner across languages (e.g., F could represent harmony in the ordering of heads and complements across phrasal types). Then, a (potentially infinite) space of possible languages is carved into two possible types, T_0 and T_1, based on whether they obey the feature or not (e.g., T_1 languages exhibit harmonic ordering across phrase types, whereas T_0 languages do not). There are two corresponding classes of utterance, U_0 and U_1, which are produced by speakers and portray whether the underlying language is of type T_0 or T_1, according to the following likelihood function, where ϵ gives the probability of noise on production,

$$\Pr(U_x|T_x) = 1 - \epsilon, \quad \Pr(U_{y \neq x}|T_x) = \epsilon, \quad [1]$$

for $x, y = 0, 1$. Bayesian learners assess a particular hypothesis h by combining the likelihood of observed dataset d in light of that hypothesis, $\Pr(d|h)$, with the prior probability of the hypothesis independent of the data observed, $\Pr(h)$, to obtain a posterior probability of that hypothesis: $\Pr(h|d) \propto \Pr(d|h)\Pr(h)$. Here, we interpret the set of hypotheses as the set of possible grammars or language types, so that $h \in \{T_0, T_1\}$. Data $d = \{d_1, \ldots, d_N\}$ are sets of utterances from which learners must learn a language: The likelihood of a dataset is simply the product of the likelihood of the individual utterances $\Pr(d|h) = \prod_{k=1}^{N} \Pr(d_k|h)$, with $d_k \in \{U_0, U_1\}$ for $k = 1, \ldots, N$. The prior probability distribution over grammars reflects the innate biases of learners, and simply assigns probabilities to the two language types: $\Pr(h = T_0) = 1 - \alpha$, $\Pr(h = T_1) = \alpha$. Extreme values of α ($\alpha \approx 0$ or $\alpha \approx 1$) would correspond to a principle in standard terminology: a strong or absolute constraint on the type of language that can be acquired by learners, ruling out one type of language; $\alpha \approx 0.5$ would correspond to an unconstrained learner who can learn either language type.

To model cultural transmission, we use "iterated learning" (21, 22): The data that a learner is exposed to are produced by another individual who learned in the same way. In other words, once the learner has settled on a particular hypothesis, they will produce utterances in line with that hypothesis, from which others can learn.

Finally, we model biological evolution by specifying a set of genes that collectively determine the prior bias of a learner, and which are inherited by new learners, subject to mutation. We treat the prior as a "polygenic" trait (31): The genome of a learner is a set of n genes whose alleles each encode small effects

in favor of one language type or the other; the prior probability of a particular hypothesis is simply the proportion of genes that have the allele promoting that hypothesis. If a learner's genetic endowment includes exactly i genes favoring languages of type T_1 (and $n-i$ favor type T_0), then $\alpha_i = i/n$. Genetic transmission is under selection: In line with other models (14, 32), we assume individuals reproduce with a probability directly proportional to their ability to communicate with the rest of the population; individuals who have the same language type are deemed to communicate successfully. Let g_i be the proportion of the population at generation t made up by learners with prior α_i, and let a_i be the probability that a learner with this prior will acquire a language of type T_1. The fitness of a learner with prior α_i is given by

$$f_i = a_i c + (1 - a_i)(1 - c), \qquad [2]$$

where c is the population frequency of language type T_1.

Without cultural transmission, the outcome of language acquisition is determined entirely by the learner's prior preferences, so $a_i = \alpha_i$. However, when cultural transmission is included in the model, the outcome of language acquisition also depends upon the linguistic data a learner encounters, which in turn depend upon the distribution of language types at the previous generation, $c^{(t-1)}$ (see *Methods* for details of how a_i is calculated in the cultural models). We assume initially that learners acquire their language type by observing utterances produced by a single, randomly chosen individual in the previous generation, and, in *Supporting Information*, describe a version of our model that relaxes this assumption. We analyze two versions of the cultural model, by contrasting two varieties of Bayesian learner that are known to lead to different cultural evolutionary dynamics (21): "MAP learners" weigh up the posterior probability of both language types and select the winner [i.e., the hypothesis with the highest posterior probability—the maximum a posteriori (MAP) estimate]; "sample learners" pick a language according to its posterior probability. Given the model

we have described, the dynamics of biological evolution in the population is given by

$$g_i^{(t+1)} = \frac{1}{\phi} \sum_{j=0}^{n} g_j f_j m_{i,j} \quad \text{for } j = 0, 1, \ldots, n \qquad [3]$$

where $m_{i,j}$ is the probability that the offspring of a learner with prior α_j will, through mutation (see *Methods*), inherit α_i, and $\phi = \sum_{i=0}^{n} g_i f_i$ is the average fitness of the population.

Results. The initial state of each simulation is an unbiased population of learners (i.e., whose genes encode $\alpha = 0.5$). A result favoring the strong nativist hypothesis would be one in which natural selection for linguistic coordination leads to the final population's prior bias consisting of extreme values, $\alpha \approx 0$ or $\alpha \approx 1$, effectively ruling out one language type.

Fig. 1 shows results of numerical and agent-based simulations of these processes. In the baseline model (Fig. 1A) without cultural transmission (comparable to innate signaling in noncultural organisms), after a few thousand generations of gradual change, biological evolution eventually leads to the emergence of strong innate biases for a single language type, which maximizes the communicative utility of that language.

We see a radically different pattern of results when we add cultural transmission to the model. In cultural populations of MAP learners, one language type rapidly comes to dominate the population: Convergence in cultural populations occurs an order of magnitude faster than in acultural populations. However, this strong cultural preference is not directly reflected in the population's genes. Instead, learners have a very weak bias favoring the majority language type, and could easily acquire the unattested language type given appropriate data. The emergence of a strong cultural universal thus does not imply, nor lead to, nor require, a strong innate constraint. In contrast, under cultural transmission in sampling populations, neither strong innate constraints nor strong language universals emerge. When we
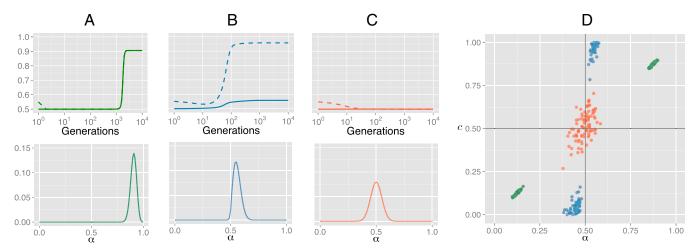

Fig. 1. Results for the evolutionary model. (*A*) Results of a numerical model of an infinitely large population, where there is no cultural transmission of language: Individuals select a language with probability given by their genes (i.e., α). (*Upper*) The evolution of the population's mean innate bias, α (solid line), which directly determines the proportion of individuals using languages of type T_1 (c, dashed line). (*Lower*) The distribution of α in the population after the model has converged, demonstrating a strong bias in favor of one language type. (*B*) Results from the model where language is culturally transmitted and in which learners pick the MAP hypothesis (again, solid line shows mean α, and dashed line shows c). (*C*) The same plots for learners who sample a hypothesis from the posterior distribution. Results of the numerical model give the equilibrium distribution of biases and languages, which reflect the stable outcome of evolution given these conditions. (*D*) Results from 300 agent-based simulations, giving the values of α and c (the same variables plotted in *A*–*C*, *Upper*) after 10,000 generations in acultural populations (green), and in 1,000 generations in cultural populations of MAP (blue) and sample (red) learners. Results supporting strong universals ($c \approx 0$ or 1) underpinned by strong nativism ($\alpha \approx 0$ or 1) would appear in the top left or bottom right corners of this graph but never occur when culture is included in the model; instead, we see strong universals underpinned by weak biases (in MAP populations) or no universals and no constraints on learning (in sampling populations); $n = 100$, $\epsilon = 0.01$, $\mu = 0.001$, $N = 2$.

Thompson et al.

include cultural transmission in the model, under no conditions do we see the evolution of strong universals underpinned by strong innate constraints.

Why is this happening? Weak biases in (MAP) individuals are amplified by cultural transmission, driving large effects at the population level (10, 21, 22). In our initial populations, every individual communicates equally well (or poorly): Reproduction occurs at random. Drift of the genes ensues, moving the population away from the perfect $\alpha = 0.5$ equilibrium (in the numerical analysis, we must specify a minor asymmetry in the initial linguistic distribution $c^0 = 0.55$ to mimic this stochastic drift process). At this point, cultural transmission unmasks the tiny biases of individuals, resulting in large effects on the population's culture: A linguistic universal begins to emerge. Natural selection then favors nonneutrality in the direction of the emerging universal. However, cultural evolution can also mask relative strength of bias (21, 22): Both weak and strong biases can drive strong universals and reliable acquisition of the dominant language; consequently, there is no selection in favor of stronger biases. The combination of unmasking and masking by cultural evolution leads to a balance of forces: Mutation pressure inherent in the genetic model causes drift toward neutrality in the prior, but natural selection keeps individuals away from perfect neutrality. As a result, MAP populations settle on the weakest possible biases that nevertheless ensure convergence on a single language type, leading to universals.

Sampling does not lead to the amplifying effect normally associated with cultural evolution; rather, the culture of sample learners tracks their biases (21). Weak biases provided by drift are never unmasked: There is no consistent selective pressure for a particular bias that would result from a cultural universal generated by unmasking. Neither strong innate constraints nor strong universals emerge: Genes and culture drift, in lockstep, toward a distribution determined by mutation.

Supporting Information presents a range of variations on this model, testing alternative basic assumptions about learning, culture, and biology. We test cases where learners learn from multiple teachers, where genomes vary continuously, where one language is functionally superior, and where both MAP and sample learners exist in the population. In the latter case, as predicted in ref. 33, MAP learners outcompete sample learners, suggesting the pattern of results for MAP learners is the more robust prediction. In none of these variants does evolution lead to nativization (Figs. S1–S5).
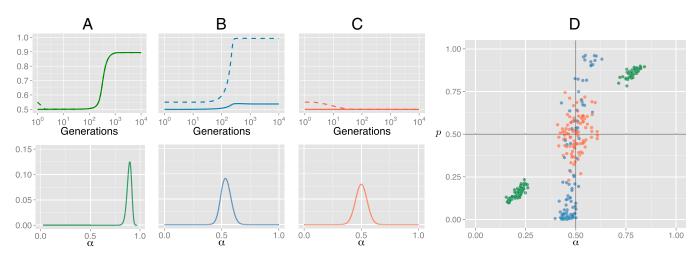
Model 2

Model 1 assumes learners acquire a discrete linguistic feature by choosing between just two hypotheses. This discreteness abstracts away from the continuous variation often thought to be characteristic of human learning. To address this limitation, we reformulate the model to reflect an alternative view of linguistic knowledge: Language acquisition reflects statistical inference over probabilistic relationships between grammatical categories.

Model Details. Let $\mathcal{L}$ be a simple stochastic grammar that specifies constraints on the possible orderings of grammatical categories X and Y. During acquisition, the learner infers a probabilistic generalization about the ordering of these categories: We could interpret this model of acquisition as applying to proposed universals concerning suffixing and prefixing, or the ordering of verbs and their objects, for example.

Let p be the probability of ordering $S \rightarrow X\ Y$, such that $\Pr(S \rightarrow X\ Y | \mathcal{L}) = p$, and $\Pr(S \rightarrow Y\ X | \mathcal{L}) = 1 - p$. The learner must make inferences about the underlying probabilities of the two ordering patterns by inducing an estimate $\hat{p}$: The set of hypotheses a learner considers is infinitely fine-grained, with the learner entertaining all values of p in the range [0,1]. We consider a noncultural model and the two types of Bayesian learner as before. *Methods* details the model of Bayesian inference we adopt, and describes a flexible scheme for specifying a prior distribution over this hypothesis space based on the learner's genes. For direct comparison with Model 1, we explore the case where the expected value of the prior over p reflects the proportion of the learner's genes promoting ordering XY. Formally, $\alpha_i = \mathrm{E}[\Pr(p|i)] \approx i/n$. Again, fitness reflects the ability to communicate by coordinating on an ordering of X and Y.

Results. Evolution in Model 2 (Fig. 2) reproduces the key results seen in Model 1. In acultural populations, evolution leads to fixation of strong inductive biases favoring one particular fixed ordering for X and Y. Including culture in the model radically changes the outcome of evolution. In populations of MAP learners, we observe rapid fixation of weak inductive biases that drive a strong cultural universal. In populations of sample learners, again, evolution leads nowhere: no fixation of domain-specific inductive biases and no cultural universals. Again, results hold under a broad parameter range, being particularly robust in MAP populations (see *Supporting Information* for testing).

Fig. 2. Evolution in model 2. (*A–C*) Equilibrium distribution of priors α (*Lower*) and timecourse dynamics (*Upper*) for the mean of the population's prior (α, solid lines) and expected use of ordering $S \rightarrow X\ Y$ (p, dashed lines), in acultural (*A*), and cultural (*B*, MAP learners; *C*, sample learners) populations. (*D*) Results of 300 agent-based simulations of the model (100 acultural, 100 MAP, 100 Sample) in populations of 100 learners after 1,000 generations: Points show the mean of the population's prior (α) and expected use of ordering $S \rightarrow X\ Y$ (p) in the final population; $n = 100$, $\mu = 0.001$, $N = 20$, $\lambda = 20$.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Discussion

These models show that cultural transmission radically changes the evolution of constraints on learning, rendering strong linguistic nativism untenable on evolutionary grounds. On the one hand, unmasking facilitates rapid evolution of domain-specific biases: Due to culture, the population-level consequences of those biases are amplified and visible to selection. However, masking makes evolving strong constraints unlikely: Given that weak constraints have equivalent effects to strong constraints, there is little or no selection for stronger constraints (34). Note that we do not rule out strong innate constraints on language learning that are domain-general, i.e., have not evolved for the purpose of constraining learning to facilitate communication. For example, we might expect strong evolved constraints that are truly domain-independent [e.g., constraints on statistical learning mechanisms (8) or principles of efficient computation (35) that apply to language and other systems], or which are a consequence of architectural constraints that are a byproduct of our evolutionary history [e.g., spandrels, or developmental constraints (35)]. We may also expect adaptations in the peripheral machinery for language (e.g., vocal anatomy and associated neural machinery), which may follow different evolutionary dynamics (36).

Weak biases are defeasible: The cultural environment can easily overrule these dispositions. This removes any apparent paradox in the idea that we can have biologically driven behavioral universals but nevertheless extreme plasticity at the level of the individual [see, e.g., work demonstrating that the visual cortex can be recruited for language processing in congenitally blind adults (37)]. Under an account assuming that behavioral universals can only be explained in terms of strong innate constraints, such individual-level plasticity is puzzling; however, under an account where cultural evolution mediates between the biases of individuals and behavioral universals, such plasticity is, in fact, predicted.

Weak constraints are also highly evolvable: Evidence for recent rapid adaptation in humans (38) may reflect rapid fixation of weak biases rather than the construction of strongly constraining domain-specific cognitive modules. Our models predict an increase in the rate and number of cognitive adaptations with the onset of culture in human evolution (11) and that the genetic underpinnings of these adaptations may be difficult to detect.

Conclusions

Culture mediates between the biases of individual learners and population-level tendencies or universals. This radically changes the predictions we should make about the language faculty, or any other system of constrained cultural learning: Specifically, the evolution of strong domain-specific constraints on learning is ruled out. Rather, the behavioral universals that these constraints are invoked to explain can instead be produced by weak biases, amplified by cultural transmission. Although we have framed our model in terms of language and linguistic nativism, the same account may be applicable to any behavior that is the product of interactions between culture and biology: Wherever cognition has been shaped to acquire culturally transmitted behaviors, our arguments should apply. We anticipate that cultural transmission may be amplifying the effects of learning biases in many domains of human behavior, mimicking the effects of strong innate constraints and inviting nativist overinterpretation; identifying these domains is a key priority. The default explanation of shared, universal aspects of language or other cultural behaviors should be in terms of weak innate constraints.

Methods

Model 1. To calculate acquisition probabilities in the model with cultural transmission, we must account for the range of possible datasets d each learner could encounter and the inferences she would make in each case, so that

$$a_i = \sum_d \Pr\left(d|c^{(t-1)}\right) \Pr_{induce}(T_1|d, \alpha_i). \quad [4]$$

Assuming a single teacher is randomly sampled from the previous generation, the probability of learning from a speaker of T_1 is $c^{(t-1)}$ [and of T_0 is $1 - c^{(t-1)}$], so the likelihood, $\Pr(d|c^{(t-1)})$, of observing dataset d is

$$\Pr\left(d|c^{(t-1)}\right) = c^{(t-1)}\Pr(d|T_1) + \left(1 - c^{(t-1)}\right)\Pr(d|T_0). \quad [5]$$

Induction probabilities for sample learners follow posterior probabilities directly, so $\Pr_{induce}(T_1|d, \alpha_i) = \Pr(T_1|d, \alpha_i)$. For MAP learners, these are

$$\Pr_{induce}(T_1|d, \alpha_i) = \begin{cases} 1, & \text{if } \Pr(T_1|d, \alpha_i) > \Pr(T_0|d, \alpha_i) \\ 0.5, & \text{if } \Pr(T_1|d, \alpha_i) = \Pr(T_0|d, \alpha_i) \\ 0, & \text{otherwise.} \end{cases} \quad [6]$$

During reproduction, each of a learner's n genes may mutate with independent probability μ into a gene of the opposite type. Because there are $\binom{n}{i}$ possible genomes with i genes favoring T_1, it is nontrivial to calculate the mutation probabilities between priors. However, it can be shown that

$$m_{i,j} = \sum_{k=0}^{n-i} \binom{n-i}{j-i+k}\binom{i}{k}\mu^{j-i+2k}(1-\mu)^{n-j-i+2k}. \quad [7]$$

We obtain the equilibrium distribution of biases $\mathbf{g}^*$ and language types c^* by specifying initial distributions $\mathbf{g}^0$ and c^0, and iterating the recursion Eq. **3** to find the resulting numerical solutions $\mathbf{g}^*, c^*$ that satisfy $\mathbf{g} = \mathbf{g}^*, c = c^* \Leftrightarrow \mathbf{g} = \mathbf{g}^{(t+1)}, c = c^{(t+1)}$, where $\mathbf{g} = (g_0, g_1, \ldots, g_n)$ is a probability vector for the g_i s, and superscripts index time points. In all numerical analyses of both models, $g_i^0 \propto \binom{n}{i}$ for $i = 0, \ldots, n$.

Model 2. The learner's prior over p follows a Beta distribution with parameters β^1 and β^0,

$$\Pr\left(p|\beta^1, \beta^0\right) = \text{Beta}(p; \beta^1, \beta^0) = \frac{p^{\beta^1-1}(1-p)^{\beta^0-1}}{B(\beta^1, \beta^0)}. \quad [8]$$

The data a learner observes consists of N independent utterances that exemplify an ordering pattern for X and Y, of which some number $y \in 0, \ldots, N$ exemplify the specific ordering $S \rightarrow XY$. The likelihood for $y|N, p$ is the binomial distribution

$$\Pr\left(y|\hat{p}_j\right) = \text{Binomial}(y; N, p) = \binom{N}{y}p^y(1-p)^{N-y}. \quad [9]$$

The posterior for p given y is also a Beta distribution,

$$\Pr\left(p|y, \beta^1, \beta^0\right) = \text{Beta}\left(p; \beta^1 + y, \beta^0 + N - y\right). \quad [10]$$

Let $\mathbf{p} = (\hat{p}_0, \hat{p}_1, \ldots, \hat{p}_n)$ give the expected outcome of language acquisition for each class of learner at generation t: $\hat{p}_i$ is the expected outcome for a learner with i genes promoting ordering XY. Without culture, $\hat{p}_i$ reflects the expected value when sampling randomly from the prior, given by the prior mean,

$$\hat{p}_i = \alpha_i = \text{E}\left[\Pr\left(p|\beta_i^1, \beta_i^0\right)\right] = \frac{\beta_i^1}{\beta_i^1 + \beta_i^0}. \quad [11]$$

With culture, $\hat{p}_i$ is conditional on the data a learner sees. Summarizing the possible datasets by the possible counts y (which are sufficient statistics), then, in the sample learner model,

$$\hat{p}_i^{(t+1)} = \sum_{j=0}^n g_j \sum_{y=0}^N \Pr\left(y|\hat{p}_j\right) \text{E}\left[\Pr\left(p|y, \beta_i^1, \beta_i^0\right)\right], \quad [12]$$

where the outcome of acquisition given y reflects the expected value when sampling from the posterior, the posterior mean,

$$\text{E}\left[\Pr\left(p|y, \beta_i^1, \beta_i^0\right)\right] = \frac{\beta_i^1 + y}{\beta_i^1 + \beta_i^0 + N}. \quad [13]$$

Likewise, in the MAP learner model,

Thompson et al.

$$\hat{p}_i^{(t+1)} = \sum_{j=0}^{n} g_j \sum_{y=0}^{N} \Pr\left(y|\hat{p}_j\right) \arg\max_p \Pr\left(p|y, \beta_i^1, \beta_i^0\right), \qquad \textbf{[14]}$$

where the outcome of acquisition reflects the maximum a posteriori estimate, or the posterior mode, which is known to be

$$\arg\max_p \Pr\left(p|y, \beta_i^1, \beta_i^0\right) = \frac{\beta_i^1 + y - 1}{\beta_i^1 + \beta_i^0 + N - 2}. \qquad \textbf{[15]}$$

Fitness can be computed, at the mean field level, with

$$f_i(\mathbf{p}) = \hat{p}_i \overline{p} + \left(1 - \hat{p}_i\right)(1 - \overline{p}), \qquad \textbf{[16]}$$

where $\overline{p} = \mathbf{p} \cdot \mathbf{g}$ is the overall expected value for p in the population. Thus, the recursion for biological dynamics becomes

$$g_i^{(t+1)} = \frac{1}{\phi} \sum_{j=0}^{n} g_j f_j(\mathbf{p}) m_{j,i}. \qquad \textbf{[17]}$$

Finally, we must specify the relationship between genes and prior parameters, β^1 and β^0. Let β^1 and β^0 reflect the ratio of gene types in the genome, smoothed by a constant λ,

$$\beta_i^1 = \frac{i + \lambda}{n - i + \lambda}, \quad \beta_i^0 = \frac{n - i + \lambda}{i + \lambda}. \qquad \textbf{[18]}$$

The uniform prior ($\beta_i^1 = \beta_i^0 = 1$) is given by a uniform distribution of gene types as in Model 1, because $(i + \lambda)/(n - i + \lambda) = (n - i + \lambda)/(i + \lambda) = 1 \Leftrightarrow i = n - i = n/2$. The parameter $\lambda \geq 0$ controls the relationship between asymmetry in the gene distribution and the shape of the prior distribution over p. See *Supporting Information* for details and analysis (Figs. S6–S8). Here we set $\lambda = 20$, which for our $n = 100$ example ensures the approximately linear relationship between the balance of genes in the genome and the prior.

1. Tooby J, Cosmides L, Barrett HC (2005) *The Innate Mind: Structure and Content*, eds Carruthers P, Laurence S, Stich S (Oxford Univ Press, Oxford), pp 305–337.
2. Carey S (2009) *The Origins of Concepts* (Oxford Univ Press, Oxford).
3. Leslie AM (1994) *Mapping the Domain Specificity in Cognition and Culture*, eds Hirschfeld LA, Gelman SA (Cambridge Univ Press, Cambridge, UK).
4. Justus T, Hutsler JJ (2005) Fundamental issues in the evolutionary psychology of music: Assessing innateness and domain specificity. *Music Percept* 23(1):1–27.
5. Boyer P (1994) *The Naturalness of Religious Ideas: A Cognitive Theory of Religion* (Univ Calif Press, Berkeley, CA).
6. Chomsky N (1965) *Aspects of the Theory of Syntax* (MIT Press, Cambridge, MA).
7. Pinker S, Bloom P (1990) Natural language and natural selection. *Behav Brain Sci* 13(4):707–784.
8. Monaghan P, Christiansen MH (2008) *Corpora in Language Acquisition Research: History, Methods, Perspectives*, ed Behrens H (John Benjamins, Amsterdam), pp 139–164.
9. Evans N, Levinson SC (2009) The myth of language universals: Language diversity and its importance for cognitive science. *Behav Brain Sci* 32(5):429–448, and discussion (2009) 32(5):448–494.
10. Boyd R, Richerson PJ (1985) *Culture and the Evolutionary Process* (Univ Chicago Press, Chicago).
11. Laland KN, Odling-Smee J, Myles S (2010) How culture shaped the human genome: Bringing genetics and the human sciences together. *Nat Rev Genet* 11(2):137–148.
12. Chater N, Reali F, Christiansen MH (2009) Restrictions on biological adaptation in language evolution. *Proc Natl Acad Sci USA* 106(4):1015–1020.
13. Culbertson J, Smolensky P, Legendre G (2012) Learning biases predict a word order universal. *Cognition* 122(3):306–329.
14. Nowak MA, Komarova NL, Niyogi P (2001) Evolution of universal grammar. *Science* 291(5501):114–118.
15. Griffiths TL, Chater N, Kemp C, Perfors A, Tenenbaum JB (2010) Probabilistic models of cognition: Exploring representations and inductive biases. *Trends Cogn Sci* 14(8):357–364.
16. Hsu AS, Chater N, Vitányi PMB (2011) The probabilistic analysis of language acquisition: Theoretical, computational, and experimental analysis. *Cognition* 120(3):380–390.
17. Culbertson J, Adger D (2014) Language learners privilege structured meaning over surface frequency. *Proc Natl Acad Sci USA* 111(16):5842–5847.
18. St Clair MC, Monaghan P, Ramscar M (2009) Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cogn Sci* 33(7):1317–1329.
19. Finley S, Badecker W (2012) Learning biases for vowel height harmony. *J Cogn Sci* 13(3):287–327.
20. Wilson C (2006) Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cogn Sci* 30(5):945–982.
21. Griffiths TL, Kalish ML (2007) Language evolution by iterated learning with Bayesian agents. *Cogn Sci* 31(3):441–480.

22. Kirby S, Dowman M, Griffiths TL (2007) Innateness and culture in the evolution of language. *Proc Natl Acad Sci USA* 104(12):5241–5245.
23. Fitch WT (2014) Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Phys Life Rev* 11(3):329–364.
24. Spelke ES, Kinzler KD (2009) Innateness, learning, and rationality. *Child Dev Perspect* 3(2):96–98.
25. Culbertson J (2012) Typological universals as reflections of biased learning: Evidence from artificial language learning. *Lang Linguist Compass* 6(5):310–329.
26. Pearl L, Sprouse J (2013) Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Lang Acquis* 20(1):23–68.
27. Chomsky N (1987) *Knowledge of Language: Its Nature, Origin and Use* (Foris, Dordrecht, The Netherlands).
28. Yang CD (2004) Universal Grammar, statistics or both? *Trends Cogn Sci* 8(10):451–456.
29. Baker M (2001) *The Atoms of Language: The Mind's Hidden Rules of Grammar* (Oxford Univ Press, Oxford), p 288.
30. Crain S, Khlentzos D, Thornton R (2010) Universal Grammar versus language diversity. *Lingua* 120(12):2668–2672.
31. Rockman MV (2012) The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution* 66(1):1–17.
32. Briscoe T (2000) Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language* 76(2):245–296.
33. Smith K, Kirby S (2008) Cultural evolution: Implications for understanding the human language faculty and its evolution. *Philos Trans R Soc Lond B Biol Sci* 363(1509):3591–3603.
34. Deacon TW (2010) Colloquium paper: A role for relaxed selection in the evolution of the language capacity. *Proc Natl Acad Sci USA* 107(Suppl 2):9000–9006.
35. Chomsky N (2005) Three factors in language design. *Linguist Inq* 36(1):1–22.
36. de Boer B (2002) *Simulating the Evolution of Language* (Springer, New York), pp 79–97.
37. Bedny M, Pascual-Leone A, Dodell-Feder D, Fedorenko E, Saxe R (2011) Language processing in the occipital cortex of congenitally blind adults. *Proc Natl Acad Sci USA* 108(11):4429–4434.
38. Hawks J, Wang ET, Cochran GM, Harpending HC, Moyzis RK (2007) Recent acceleration in human adaptive evolution. *Proc Natl Acad Sci USA* 104(52):20753–20758.
39. Bullock S (2001) Smooth operator? *Understanding and Visualising Mutation Bias*, eds Kelemen J, Sosik P (Springer, Prague), pp 602–612.
40. Perreault C, Moya C, Boyd R (2012) A Bayesian approach to the evolution of social learning. *Evol Hum Behav* 33(5):449–459.
41. Smith K (2009) *Iterated Learning in Populations of Bayesian Agents* (Cogn Sci Soc, Austin, TX), pp 697–702.
42. Burkett D, Griffiths TL (2010) Iterated Learning of Multiple Languages from Multiple Teachers, eds Smith ADM, Schouwstra M, de Boer B, Smith K (World Sci, Singapore), pp 58–65.

PSYCHOLOGICAL AND
COGNITIVE SCIENCES