Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language

Simon Kirby*[†], Hannah Cornish*, and Kenny Smith[‡]

*School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Edinburgh EH8 9LL, United Kingdom; and [‡]Division of Psychology, Northumbria University, Newcastle-upon-Tyne NE1 8ST, United Kingdom

Edited by Dale Purves, Duke University Medical Center, Durham, NC, and approved June 6, 2008 (received for review August 20, 2007)

We introduce an experimental paradigm for studying the cumulative cultural evolution of language. In doing so we provide the first experimental validation for the idea that cultural transmission can lead to the appearance of design without a designer. Our experiments involve the iterated learning of artificial languages by human participants. We show that languages transmitted culturally evolve in such a way as to maximize their own transmissibility: over time, the languages in our experiments become easier to learn and increasingly structured. Furthermore, this structure emerges purely as a consequence of the transmission of language over generations, without any intentional design on the part of individual language learners. Previous computational and mathematical models suggest that iterated learning provides an explanation for the structure of human language and link particular aspects of linguistic structure with particular constraints acting on language during its transmission. The experimental work presented here shows that the predictions of these models, and models of cultural evolution more generally, can be tested in the laboratory.

cultural transmission | iterated learning | language evolution

The emergence of human language has been cited by Maynard Smith and Szathmary (1) as the most recent of a small number of highly significant evolutionary transitions in the history of life on earth. The reason they give for including language in this list is that language enables an entirely new system for information transmission: human culture. Language is unique in being a system that supports unlimited heredity of cultural information, allowing our species to develop a unique kind of open-ended adaptability.

Although this feature of language as a carrier of cultural information obviously is important, we have argued that there is a second sense in which language is an evolutionary milestone: each utterance has a dual purpose, carrying semantic content but also conveying information about its own construction (2–5). Upon hearing a sentence, a language learner uses the structure of that sentence to make new inferences about the language that produced it. This process allows learners to reverse-engineer the language of their speech community from the utterances they hear. Language thus is both a conveyer of cultural information (in Maynard Smith and Szathmary's sense) and is itself culturally transmitted. This cultural transmission makes language an evolutionary system in its own right (2–3), suggesting another approach to the explanation of linguistic structure. Crucially, language also represents an excellent test domain for theories of cultural evolution in general, because the acquisition and processing of language are relatively well understood, and because language has an interesting, nontrivial, but well documented structure.[§]

During the past 10 years a wide range of computational and mathematical models have looked at a particular kind of cultural evolution termed "iterated learning" (4–13).

Iterated Learning. Iterated learning is a process in which an individual acquires a behavior by observing a similar behavior in another individual who acquired it in the same way.

Spoken (or signed) language is an outcome of iterated learning. Although in some circumstances aspects of language may be explicitly taught, acquired from a written form, or arise from deliberate invention, almost all the features of the languages we speak are the result of iterated learning. Models of this process (4–13) demonstrate that, over repeated episodes of transmission, behaviors transmitted by iterated learning tend to become 1) easier to learn, and 2) increasingly structured. Note that this process is cumulative and is not considered to arise from the explicit intentions of the individuals involved. Rather, this type of cultural evolution is an "invisible hand" process leading to phenomena that are the result of human action but are not intentional artifacts (14).

Although these models are indicative of the power of cultural evolution in explaining language structure, skepticism remains as to how well computational models of learning match the abilities and biases of real human learners. For example, responding to a growing body of computational models of the emergence of multiword utterances from unstructured randomness (5, 8, 10, 11, 15), Bickerton notes, "Powerful and potentially interesting although this approach is, its failure to incorporate more realistic conditions (perhaps because these would be more difficult to simulate) sharply reduces any contribution it might make toward unraveling language evolution. So far, it is a classic case of looking for your car-keys where the street-lamps are" (16, p. 522).

What is needed, therefore, is an experimental paradigm for studying the evolution of complex cultural adaptations using real human participants. Ideally, this paradigm should mirror previous computational and mathematical models and provide a test for the claim that iterated learning leads to adaptively structured languages. It should demonstrate whether cumulative adaptive evolution without intention is possible purely by virtue of cultural transmission.

In this paper, we implement such a paradigm and demonstrate cumulative, adaptive, nonintentional cultural evolution of an artificial language in a laboratory population of human participants.

Diffusion Chains. Diffusion-chain studies provide the best example of experimental treatments of iterated learning. In these experiments a participant observes some target behavior (provided by the experimenter) and then is required to replicate that behavior in some way that can be observed by a second participant. This second participant in turn attempts to replicate the first participant's

Author contributions: S.K., H.C., and K.S. designed research; H.C. performed research; S.K., H.C., and K.S. analyzed data; and S.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[†]To whom correspondence should be addressed. E-mail: simon@ling.ed.ac.uk.

[§]From a practical perspective it is also an ideal subject for study in that it is relatively straightforward to record and analyze precisely.

This article contains supporting information online at www.pnas.org/cgi/content/full/0707835105/DCSupplemental.

behavior for a third participant, and so on. (We refer to each iteration of this cycle as "1 generation.") Using this procedure, we can observe the diffusion of behavior through a chain of cultural transmission. The first reported use of this methodology was by Bartlett in 1932 (17), but only recently did researchers begin to apply this approach systematically (18–24).

The most recent, and arguably the most significant, instance of a diffusion-chain experiment is the work of Horner *et al.*, which explores the cultural transmission of tool-use strategies in populations of chimpanzees and children (24). Diffusion chains are set up in which an experimenter demonstrates 1 of 2 possible techniques for opening a puzzle box ("artificial fruit") to a participant. Subsequent participants observe their predecessor's box-opening behavior and then in turn become the model for the next generation. These experiments demonstrate clearly that both chimpanzees and children are capable of high-fidelity cultural transmission: the box-opening technique used by the last participant in the chains (of up to 10 individuals) is the same as that demonstrated to the first participant, with a chain of faithful transmission between the first and last participants.

Although these experiments show that cultural transmission can be studied empirically even in nonhumans, they do not support our claim that culture leads to cumulative nonintentional adaptation because the behavioral information that is being transmitted is drawn from a limited set of possibilities. For example, in the puzzle-box study, there are essentially 2 different strategies for opening the box. The task is not complex enough to demonstrate adaptation, let alone cumulative adaptation. In any case, both the strategies seem to be equivalently "adaptive" in cultural and environmental terms, in that both open the box and both are transmittable.

To get around these problems and to allow us to make a direct comparison with human language, we replicate the basic diffusion-chain design with a more complex artificial-language learning task of labeling visual stimuli with strings of written syllables (25, 26). To make this task tractable, we use adult human participants and observe the cultural evolution of the artificial language for 10 cultural generations.

This work bears some resemblance to a recent body of experimental work on the shared construction of communication systems (27–30). Of particular relevance is a recent paper by Selten and Warglien (30) that demonstrates that pairs of participants sometimes can create structured and efficient communication systems over the course of repeated interactions. The major difference between the experiments described here and the work of Selten and Warglien is the role of intentional design. In Selten and Warglien's experiments, as in those of Galantucci (27) and Garrod *et al.* (28, 29), participants interact repeatedly with the explicit goal of arriving at a shared system for communication. Therefore the systems they construct are the outcome of conscious design. Our diffusion-chain experiment allows us to explore whether structured languages can emerge without intentional design, as has been argued to be the case for language (14).

Design of Experiment 1. Participants are asked to learn an "alien" language made up of written labels for visual stimuli. The stimuli are pictures of colored objects in motion, and the labels are sequences of lowercase letters (see Fig. 1 for an example and the *Methods* section for more details).

For training purposes, the language to be learned (a set of string–picture pairs) is divided randomly into 2 sets of approximately equal size: the SEEN set and the UNSEEN set. A participant is trained on the SEEN set, being presented repeatedly with each string–picture pair in random order (see *Methods* for details). During subsequent testing, participants are presented with a picture and asked to produce the string they think the alien would give for that picture. Participants are tested on both the SEEN and UNSEEN sets in their entirety.

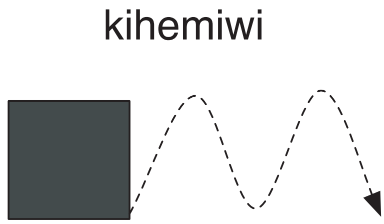

Fig. 1. An example string–picture pair.

The initial set of labels in the language is generated and assigned randomly, and the first participant in the experiment is trained on this random language. Subsequent participants are trained on the output of the final testing of the previous participant, which is re-divided into new SEEN and UNSEEN sets. Note that the experimental procedure is equivalent for all participants, despite the different sources of training data: at no stage are participants told that they are being trained on the output of another person, nor did any participants guess that the transmission of an acquired language was part of the experiment. Crucially, participants believe they are copying the input language as best they can; a posttest questionnaire revealed that many participants did not even realize that they were being tested on stimuli they had not seen in training, so that intentional design on the part of the participants is unlikely. To put it another way, the participants' goal is to reproduce the language, not improve to it in some way. (We return to this point in the *Discussion* section).

Our hypothesis is that we will observe cumulative adaptive evolution of the language being transmitted in this experiment; that is, we should see the emergence of adaptive structure in response to the pressure on the language to be transmitted faithfully from generation to generation. If this hypothesis is correct, we should see 2 things: 1) an increase in the learnability of the language over generations (i.e., a decrease in transmission error), and 2) the evolution of linguistic structure (i.e., an increase in predictability in the mapping between meanings and signals).

We devised 2 measures to test this hypothesis. First, we used a measure of string similarity to compare words in the languages of participants at adjacent generations (see *Methods*). The Levenshtein edit distance (31) between pairs of words (i.e., the smallest number of character insertions, replacements, and deletions required to transform 1 word into the other) provides a reasonable theory-neutral measure of distance. We normalized the edit distance for length of words so that identical strings have a distance of 0 and maximally distinct ones have a distance of 1. The mean distance between all of the words in a participant's output and the corresponding words in the previous generation's output gives a straightforward measure of the error in transmission of the language.

Second, we constructed a measure of linguistic structure based on measures of compositionality used in some computational models (12). Our aim was to quantify the degree to which the mapping between meanings (visual scenes) and signals (character strings) is systematic, an obvious hallmark of structure in human language. A language is systematic if patterns of similarity and dissimilarity in signals provide information about the relationship between the meanings those signals map on to. Accordingly, we calculated the correlation between all pairs of edit-distances in the set of signals and the corresponding distances between meanings (i.e., whether they differed in shape, color, and/or movement). By using Monte-Carlo techniques, we can calculate the extent to which this alignment between meaning and signal differs from the alignment we would expect to see by a random, unstructured assignment of signals to meanings (see *Methods* for details).

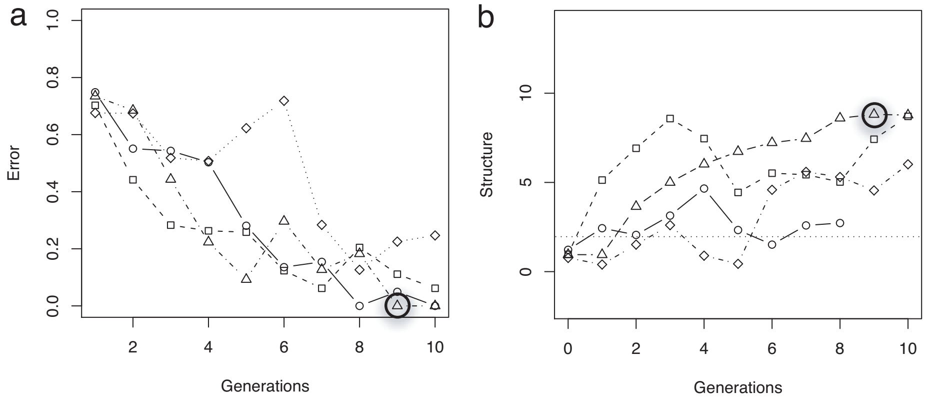

Fig. 2. Transmission error and a measure of structure by generation in 4 chains. *a* shows the increase in learnability (decrease in error) of languages over time. *b* shows structure in the languages increasing. The dotted line in *b* gives the 95% confidence interval so that any result above this line demonstrates that there is a nonrandom alignment of signals and meanings. In other words, structure in the set of signals reflects structure in the set of meanings. In 2 cases, this measure is not defined and therefore is not plotted (see *Methods*). The language discussed in the paper is circled.

Results of Experiment 1. The results of our first experiment, involving 4 separate diffusion chains of 10 participants each, are shown in Fig. 2. Each of these chains was initialized with a different random language. There is a clear and statistically significant decrease in transmission error between the initial and final generations (mean decrease 0.748, SD = 0.147; $t(3) = 8.656$; $P < 0.002$). This decrease confirms the first of our predictions: the language is adapting to become increasingly transmissible from generation to generation. Indeed, toward the end of some chains the language is transmitted perfectly: these participants produced exactly the same strings for every meaning as their predecessor, although they had not been exposed to the strings associated with half of those meanings.

How is this adaptation possible? Is any structural evolution of the language taking place as in the second of our 2 predictions? As Table 1 shows, the number of distinct strings in each language decreases rapidly. The initial random languages are completely unambiguous: every meaning is expressed by a distinct signal. The transmission process cumulatively introduces ambiguity as single strings are re-used to express more and more meanings. In other words, the languages gradually introduce underspecification of meanings. Clearly, the reduction in the number of strings must make a language easier for participants to learn, but the reduction alone cannot account for the results we see. For example, the reduction does not explain how, in some chains, participants are able to produce the correct signal for every meaning, including meanings drawn from the UNSEEN set.

The answer to this puzzle lies in the structure of the languages. The initial random language is, by definition, unstructured: nothing in the set of signals gives any systematic clue to the meanings being conveyed. The only way to learn this language is by rote. Equally, if a language is randomly underspecified, then rote learning is the only way it can be acquired. For example, if the same signal is used for a black spiraling triangle and a red bouncing square, then a learner must see this signal used for both of these meanings to learn it. Because we deliberately hold items back from the SEEN set, rote learning for all meanings is impossible. For learners to be able to generalize to unseen meanings successfully, there must be systematic underspecification.

We can observe exactly this kind of structure evolving by examining a language as it develops in the experiment. For example, by generation 4 in 1 of the diffusion chains, the string *tuge* is used exclusively for all pictures with an object moving horizontally. The distribution of the other strings in the language is more idiosyncratic and unpredictable at this stage. By generation 6, *poi* is used to refer to most spiraling pictures, but there are exceptions for triangles and squares. Blue spiraling triangles or squares are referred to as *tupin*, and red spiraling triangles or squares are *tupim*. In the following generation, these exceptional cases are reduced to the blue spiraling triangle and the red spiraling square. By generation 8 (shown in Fig. 3), and also for generations 9 and 10, the language has settled on a simple system of regularities whereby everything that moves horizontally is *tuge*, all spiraling objects are *poi*, and bouncing objects are divided according to shape.

It is precisely because the language can be described by using this simple set of generalizations that participants are able to label correctly pictures that they have never previously seen. This generalization directly ensures the stable cultural transmission of the language from generation to generation, even though each learner of the language is exposed to incomplete training data.

Table 1. Number of distinct words by generation in the first experiment

Generation	0	1	2	3	4	5	6	7	8	9	10
○ Chain 1	27	17	9	6	5	4	4	2	2	2	2
□ Chain 2	27	17	15	8	7	6	6	6	5	5	4
△ Chain 3	27	24	8	6	6	5	6	5	5	5	5
◇ Chain 4	27	23	9	10	9	11	7	5	5	4	4

Symbols correspond to those in Fig. 2.

Fig. 3. An example evolved language in the first experiment. This language exhibits systematic underspecification, enabling learners to reproduce the whole language from a fragment.

Table 2. Number of distinct words by generation in the second experiment

Generation	0	1	2	3	4	5	6	7	8	9	10
○ Chain 1	27	23	22	17	21	21	17	21	25	13	16
□ Chain 2	27	26	13	10	10	16	16	12	12	13	12
△ Chain 3	27	11	16	14	12	17	14	16	20	19	12
◇ Chain 4	27	19	19	17	19	17	22	23	21	27	23

Symbols correspond to those in Fig. 4.

Our structure measure confirms that the languages evolve to become more structured. As can be seen in Fig. 2*b*, significantly nonrandom structure in the mapping from meanings to signals emerges rapidly. Furthermore, the languages produced by the final generation are significantly more structured than the initial languages (mean increase 5.578, SD = 2.968, $t(3) = 3.7575$, $P < 0.02$).

Languages in this experiment are evolving to be learnable, and they are doing so by becoming structured. This development of structure confirms our hypothesis regarding the cultural evolution of language. However, we are interested in whether it would be possible for a language to evolve that is learnable and structured but also expressive, i.e., a language that would be able to label meanings unambiguously. Such a language cannot rely on systematic underspecification of meanings but instead must find some other means of gaining structure.

Design of Experiment 2. Accordingly, in the second experiment we made a single minor modification: we "filtered" the SEEN set before each participant's training. If any strings were assigned to more than 1 meaning, all but 1 of those meanings (chosen at random) was removed from the training data. This filtering effectively removes the possibility of the language adapting to be learnable by introducing underspecification: filtering ensures that underspecification is an evolutionary dead-end. This process, although artificial, is an analogue of a pressure to be expressive that would come from communicative need in the case of real language transmission.

Results of Experiment 2. As expected, under the modified regimen, the overall number of words in participants' output remains comparatively high throughout the experiment, as shown in Table 2. Fig. 4*a* shows how transmission error changes as the language evolves. Once again, it is clear that the languages are becoming more learnable over time (mean decrease 0.427, SD = 0.106, $t(3) = 8.0557$, $P < 0.002$) although it is not possible to introduce the kind of underspecification seen in Experiment 1. Furthermore, it is clear from Fig. 4*b* that, as in Experiment 1, the languages are becoming increasingly structured over time (mean increase, 6.805, SD = 5.390, $t(3) = 2.525$, $P < 0.05$). Because filtering rules out the generalizations that emerged in the previous experiment, a different kind of structure that does not rely on underspecification must be emerging.

If we examine the languages at particular stages in their cultural evolution, we can see exactly what this structure is. For example, Fig. 5 shows the language output by a participant at generation 9 in 1 of the diffusion chains. When one looks at this language, it immediately becomes clear that there is structure within the signals. We can analyze each signal as 3 morphemes expressing color, shape, and movement, respectively, with 1 exceptional irregularity (*renana* for a bouncing red circle). It turns out that this general structure emerges by at least generation 6 and persists to the end of the experiment, although the details change as some morphemes are lost or are reanalyzed from generation to generation [see supporting information (SI) Tables S1–S8 for the complete set of languages].

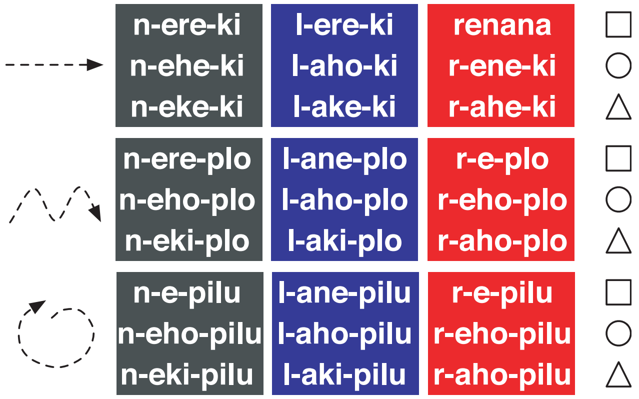

Fig. 5. An example evolved language in the second experiment. The language is structured: the string associated with a picture consists of substrings expressing color, shape, and motion, respectively. The hyphens represent 1 way of analyzing the substructure of these strings and are added purely for clarity; participants in the experiment always produced strings of characters without spaces or any other means of indicating substructure.

Discussion

What we have observed here under laboratory conditions is cumulative cultural adaptation without intentional design. Just as

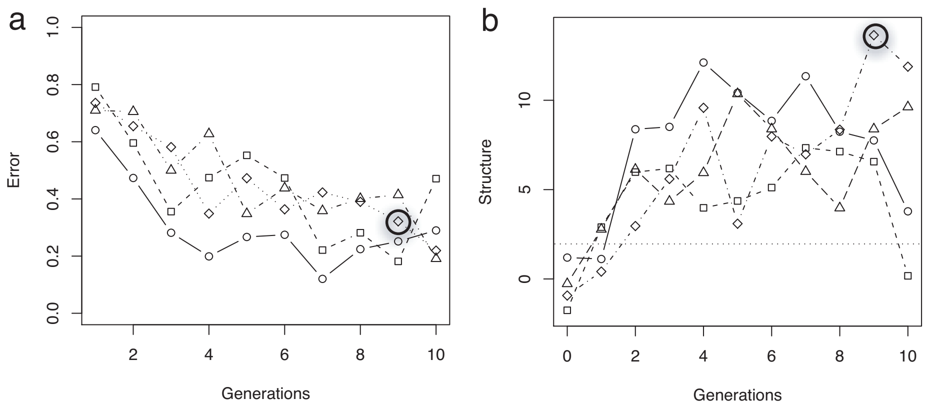

Fig. 4. Transmission error and structure by generation in the experiment in which ambiguous data were removed from the training set at each generation. *a* gives error for the whole language; *b* gives structure. These results show that, despite the blocking of underspecification, structure still evolves that enables the languages to become increasingly learnable. The language discussed in the paper is circled.

previous computational models have predicted (4–13), the culturally evolving language has adapted in a way that ensures its successful transmission from generation to generation, despite the existence of a bottleneck on transmission imposed by the incomplete exposure of each participant to the language. Cultural adaptation results in languages that circumvent this transmission problem by exploiting structure in the set of meanings to be conveyed. Note that this adaptation is cumulative with respect to learnability and structure but not with respect to expressivity: cumulative adaptation does not suggest that the languages necessarily become more functional with respect to communication.

In all our experiments we have shown that languages, by virtue of being culturally transmitted, become increasingly learnable and increasingly structured. An obvious question is: to what extent does the structure we see emerging resemble structures found in real human languages?

In the first experiment, we saw underspecification introduced into the language. This underspecification was not random but was systematic, in that similar meanings were given the same label. The form of the language reflected regularities in the visual scenes, namely that they consisted of shape, color, and motion. Of course, in the experiment this process ran unchecked and in some cases led to languages in which almost every meaning was expressed by a single signal.

The languages in our first experiment therefore could be seen as being counter-functionally ambiguous. However, there is another way of thinking about our results. Rather than seeing the emerging language as ambiguous, some participants thought it revealed something about the way the aliens saw the world. For example, in posttest discussions, 1 participant noted that "color is not important to these aliens." This observation suggests that the participants did not consider the language to be ambiguous, but instead thought that it reflected the distinctions in meaning that the aliens were interested in communicating. The collapse of distinctions based on color (which eventually occurred in all 4 replications of the first experiment) in favor of distinctions based on shape and movement is compatible with the literature on a shape bias, an expectation that words will refer to shapes of objects rather than to properties such as color or texture (32). It may be that, while adapting to become more learnable by eliminating semantic distinctions, the languages in the experiment retain the distinctions that seem most salient and/or likely to be labeled linguistically.

Systematic underspecification similar to that found in the experiments is an important feature of natural language. For example, in the class of nouns only proper names refer to specific entities. Other nouns are underspecified and typically correspond to natural classes. However, systematic underspecification is not the only way in which the structure of the set of meanings makes itself felt in linguistic expressions. Most obviously, natural languages exhibit the species-unique property of compositionality in syntax and morphology.[¶] The meaning of an expression normally is a function of the meanings of subparts of that expression and of the way the subparts are put together. It is precisely this property that we hypothesize allows language to be both learnable and expressive.

Expressivity in human language is assumed to be a consequence of the use of language for communication and also may be attributable to predispositions of child language learners (33, 34). In 1 computational model of iterated learning (8), an expressivity requirement is enforced simply by filtering out ambiguous meaning-strings from the data given to the learner, leaving a training set with a unique 1-to-1 mapping between meanings and strings. Although learners still are free to infer ambiguous strings, such ambiguity would not be transmitted to the following generation.

[¶]Arguably, the dance of honey bees (35) and the calls of Campbell's monkeys (36) are both minimally compositional. However, there is no evidence (as yet) for culturally transmitted or open-ended compositional communication outside our species.

We implemented exactly this filtering process in the second experiment, to dramatic effect, even though for the participants the conditions in this experiment were essentially identical to those in the previous experiment. As in Experiment 1, after being presented with string–picture pairs, the participants had to recall these pairs and generalize to unseen pictures. Nevertheless, unlike in the previous experiment, systematic compositional structure emerged. Rules evolved for constructing signals out of a combination of meaningful substrings, and these rules tended to be transmitted from generation to generation once they had emerged (see Tables S1–S8 for the full set of languages). The difference between these 2 experimental settings is simply that the second introduces a new adaptive challenge for the evolving language. To be transmitted faithfully from generation to generation, a language in this experiment must be both learnable and unambiguous. The learnability constraint is imposed by the participants in the experiment, and the ambiguity constraint is imposed by our additional filter.

The result is the evolution of exactly the type of structure that optimizes both these competing constraints: compositionality. The evolution of this structure reveals a key feature of cultural transmission: it gives rise to adaptive systems that respond to the pressures imposed by the transmission bottleneck that exists between the producer and learner of behavior. Crucially, this adaptation by the language maximizes its own transmissibility, and the adaptation can take place without intentional design on the part of the individuals involved. Participants in the second experiment could not be aware that ambiguous signals were being filtered, and yet a completely different sort of structure emerged. This finding demonstrates that adaptation can be independent of the intentions of individuals.

Finally, the difference between the 2 experiments also shows that the languages that emerge are not simply a reflection of the native language of the participants. A participant's first language may influence the learnability of a particular artificial language and therefore play a role in shaping the cultural evolution of those languages in our experiments. However, this explanation cannot be the whole story: if participants were merely stamping their own linguistic knowledge onto the data that they were seeing, there would be no reason we would find rampant structured underspecification in the first experiment and a system of morphological concatenation in the second.

Conclusions

We have shown that it is possible to study cumulative cultural adaptation in the laboratory. Using a diffusion-chain paradigm with an artificial-language learning task, we provide empirical support for computational and mathematical models of iterated learning that show language to be an adaptive system in its own right. We demonstrate the cumulative evolution of an adaptive structure without intentional design on the part of the participants in the experiment.

We can understand the linguistic structure emerging in these experiments as an adaptive response by language to the problem of being transmitted from generation to generation. In particular, language faces the problem of being reproducible from a subsample. In the first experiment, the language solves this problem by introducing systematic underspecification in the meaning-signal mapping. In the second experiment, the language faces the additional challenge of being transmitted despite filtering for ambiguity. Compositional structure is a potential solution to this particular transmission problem, and this structure emerges. It is important to reiterate that participants in the experiment did not intentionally design this solution; indeed, they were not even aware of the problem. Participants believed they were reproducing as best they could the language to which they were exposed. Just as biological evolution can deliver the appearance of design without the existence of a designer, so too can cultural evolution.

Methods

Eighty participants were recruited to participate in an ''alien language'' learning study. Each had to learn a language made up of written labels for visual stimuli. Participants were university students with no background in linguistics. The female:male ratio was 46:34, the mean age was 22.5 years, the minimum age was 18 years, and the maximum age was 40 years. The experiment was conducted in accordance with the ethics procedures of the Department of Linguistics and English Language at the University of Edinburgh. Participants carried out the experiment at a computer terminal and received written and verbal instructions (see *SI Text*). During training, participants were presented with string–picture pairs on the computer monitor. During testing, participants were presented with pictures on the monitor and were prompted to enter strings using the keyboard, with any sequence of alphanumeric characters being permissible.

Visual Stimuli. There were 27 possible stimuli to be labeled. Each was a colored object with an arrow indicating motion. Each object feature (shape, color, motion) varied over 3 possible values: square, circle, or triangle; black, blue, or red; horizontal motion, bouncing, or spiraling motion.

Labels. The set of labels in the initial language was generated and assigned randomly and was constructed by concatenating between 2 and 4 syllables (without spaces between) taken from a set of 9 simple consonant–vowel pairs. Because participants were free to enter any sequence of characters they chose during testing, subsequent labels were unconstrained.

Training and Testing Regimen. Each language (a set of 27 string–picture pairs, 1 string for each of 27 possible pictures) was divided randomly into 2 sets: the SEEN set (14 string–picture pairs) and the UNSEEN set (13 string–picture pairs). Each participant acquired the language in a single session comprising of 3 rounds of training with an optional 2-minute break between rounds. A single round of training consisted of 2 randomized exposures to the SEEN set, followed by a test. In the first 2 rounds this test phase contained only half the SEEN and half the UNSEEN items; the final test at the end of the third round (which was the only source for the next generation's language) consisted of all 27 pictures.

During each training pass through the SEEN set, participants were presented with each pair in a random order, with the string being displayed for 1 second followed by both string and picture being displayed for a further 5 seconds. During testing, participants were presented with a picture and prompted to type in the string they thought the alien would produce for that picture.

In the second experiment, the SEEN set was filtered before presentation to participants. Specifically, if any string labeled more than 1 picture, all but 1 of those string–picture pairs (chosen at random) was moved into the UNSEEN set. As a result, the training data seen by participants in the second experiment consisted of a purely 1-to-1 mapping from strings to pictures, even if the language of the previous generation included 1-to-many mappings.

Diffusion-Chain Design. The first participant in the experiment was trained on a language with randomly constructed labels. Subsequent participants were trained on the output of the final testing of the previous participant: the previous participant's final testing output was randomly redivided into a new SEEN and UNSEEN set.

Measure of Transmission Error. The mean distance between all the signals in a participant's output and the corresponding signals in the previous generation's output gives a measure of intergeneration transmission error, and is given by

$$E(i) = \frac{1}{|M|}\sum_{m \epsilon M} \mathrm{LD}(s_i^m,\ s_{i-1}^m)$$

where s_i^m is the string associated with meaning m by the participant at generation i, LD s_i^m, s_j^m is the normalized Levenshtein distance (31) between strings s_i^m and s_j^m, and the sum is over a set of meanings M of magnitude $|M|$.

Measure of Structure. For a particular language, a measure of structure is computed as follows. The distances between all pairs of strings in the language are calculated using normalized Levenshtein distance. In addition, the distances between all pairs of meanings also are calculated using a simple hamming distance (so that meanings differing in 1 feature have a distance of 1, meanings differing in 2 features have a distance of 2, and so forth). The Pearson's product-moment correlation between these 2 sets of distances then is calculated, giving an indication of the extent to which similar meanings are expressed using similar strings. To compare across different languages and to measure significance, it is necessary to compute a Monte Carlo sample of this measure under permutations of the strings over meanings. The graphs shown in the paper give the *z* score for the veridical correlation based on 1,000 randomizations. The dotted line on the graph therefore shows the 95% confidence interval that the observed mapping could be obtained by random assignment of signals to meanings. This measure is undefined when there is no variation in the Monte Carlo sample, for example when the language has only the same string for all meanings or for all but 1 of the meanings. In these cases, all possible reorderings are equally structured.

ACKNOWLEDGMENTS. H.C. was supported by an Economic and Social Research Council 1 + 3 Studentship. K.S. was supported by a British Academy postdoctoral fellowship.

1. Maynard Smith J, Szathmary E (1997) *The Major Transitions in Evolution* (Oxford Univ Press, Oxford, UK).
2. Deacon TW (1997) *The Symbolic Species: The Co-evolution of Language and the Brain* (WW Norton, New York).
3. Christiansen MH, Chater N (2008) Language as shaped by the brain. *Behav Brain Sci*, in press.
4. Kirby S (1999) *Function, Selection and Innateness: The Emergence of Language Universals* (Oxford Univ Press, Oxford UK).
5. Hurford JR (2000) Social transmission favours linguistic generalization, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, eds Knight C, Studdert-Kennedy M, Hurford JR (Cambridge Univ Press, Cambridge, UK), pp 324–352.
6. Kirby S, Smith K, Brighton H (2004) From UG to universals: Linguistic adaptation through iterated learning. *Stud Lang* 28:587–607.
7. Kirby S, Dowman M, Griffiths T (2007) Innateness and culture in the evolution of language. *Proc Natl Acad Sci USA* 104:5241–5245.
8. Kirby S (2000) Syntax without natural selection, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, eds Knight C, Studdert-Kennedy M, Hurford JR (Cambridge Univ Press, Cambridge, UK), pp 303–323.
9. Brighton H (2002) Compositional syntax from cultural transmission. *Artif Life* 8:25–54.
10. Zuidema W (2003) How the poverty of the stimulus solves the poverty of the stimulus, *Advances in Neural Information Processing Systems 15 (Proceedings of NIPS'02)*, eds Becker S, Thrun S, Obermayer K (MIT Press, Cambridge, MA), pp 51–58.
11. Smith K, Brighton H, Kirby S (2003) Complex systems in language evolution: The cultural emergence of compositional structure. *Adv Complex Syst* 6:537–558.
12. Brighton H, Smith K, Kirby S (2005) Language as an evolutionary system. *Physics of Life Reviews* 2:177–226.
13. Griffiths TL, Kalish ML (2007) Language evolution by iterated learning with Bayesian agents. *Cogn Sci* 31:441–480.
14. Keller R (1994) *On Language Change: The Invisible Hand in Language* (Routledge, London).
15. Batali J (2002) The negotiation and acquisition of recursive grammars as a result of competition among exemplars, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, ed Briscoe EJ (Cambridge Univ Press, Cambridge, UK), pp 111–172.
16. Bickerton D (2007) Language evolution: a brief guide for linguists. *Lingua* 117:510–526.
17. Bartlett FC (1932) *On Remembering* (Macmillan, Oxford).
18. Curio E, Ulrich E, Vieth W (1978) Cultural transmission of enemy recognition: One function of avian mobbing. *Science* 202:899–901.
19. Laland KN, Plotkin HC (1990) Social learning and social transmission of foraging information in Norway rats (*Rattus norvegicus*). *Anim Learn Behav* 18:246–251.
20. Galef BG, Jr., Allen C (1995) A new model system for studying animal traditions. *Anim Behav* 50:705–717.
21. Laland KN, Williams K (1997) Shoaling generates social learning of foraging information in guppies. *Anim Behav* 53:1161–1169.
22. Mesoudi A, Whiten A (2004) The hierarchical transformation of event knowledge in human cultural transmission. *J Cognit Cult* 4:1–24.
23. Mesoudi A, Whiten A, Dunbar RIM (2006) A bias for social information in human cultural transmission. *Br J Psychol* 97:405–423.
24. Horner V, Whiten A, Flynn E, de Waal FBM (2006) Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proc Natl Acad Sci USA* 103:13878–13883.
25. Gomez RL, Gerken L (2000) Infant artificial language learning and language acquisition. *Trends Cogn Sci* 4:178–186.
26. Hudson-Kam C, Newport E (2005) Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Lang Learn Dev* 1:151–195.
27. Galantucci B (2005) An experimental study of the emergence of human communication systems. *Cogn Sci* 29:737–767.
28. Garrod S, Doherty G (1994) Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition* 53:181–215.
29. Garrod S, Fay N, Lee J, Oberlander J, MacLeod T (2007) Foundations of representation: Where might graphical symbol systems come from? *Cogn Sci* 31:961–987.
30. Selten R, Warglien M (2007) The emergence of simple languages in an experimental coordination game. *Proc Natl Acad Sci USA* 104:7361–7366.
31. Levenshtein VI (1966) Binary codes capable of correcting deletions, insertions and reversals. *Sov Phys Dokl* 6:707–710.
32. Landau B, Smith LB, Jones SS (1988) The importance of shape in early lexical learning. *Cogn Dev* 3:299–321.
33. Slobin DI (1977) in *Language Learning and Thought*, Macnamara J ed (Academic, London), pp 185–221.
34. Smith K (2004) The evolution of vocabulary. *J Theor Biol* 228:127–142.
35. Von Frisch K (1974) Decoding the language of the bee. *Science* 185:663–668.
36. Zuberbühler K (2002) A syntactic rule in forest monkey communication. *Anim Behav* 63:293–299.