Princeton University Press

Chapter Title: COMMUNICATIVE TOOLS WITH RULES

Book Title: The Secret of Our Success
Book Subtitle: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter
Book Author(s): JOSEPH HENRICH
Published by: Princeton University Press. (2016)
Stable URL: https://www.jstor.org/stable/j.ctvc77f0d.16

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Princeton University Press is collaborating with JSTOR to digitize, preserve and extend access to *The Secret of Our Success*

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

CHAPTER 13

COMMUNICATIVE TOOLS WITH RULES

The origins and evolution of language have fascinated people for centuries.[1] Linguistic skills emerge nearly miraculously in young children as they effortlessly acquire a sophisticated communicative repertoire without instruction. More striking, even skilled orators cannot readily explain the underlying, often complex, grammatical rules or hierarchical structures that lie beneath the serial flow of sounds that issue forth from their mouths. They just do it. How can we explain both the ease with which children acquire languages as well as the intricate complexities of languages, which somehow emerge without the conscious understanding of language users themselves?

Converging lines of research from several fields now point to an answer: languages arise from long-term cumulative cultural evolution. Like other aspects of culture, including sophisticated technologies, rituals, and institutions, our repertoires of communicative tools—including spoken languages—have evolved via cultural transmission over generations to improve the efficiency and quality of communication, and to adapt to the details of local communication contexts, including physical environments and social norms (like taboos). Languages, then, are cultural adaptations for communication.

These communication systems had to adapt (culturally) to our brains, exploiting features of our ape cognition, and at the same time, created new selection pressures on our genes to make us better communicators.

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

These genetic evolutionary pressures were powerful, shaping both our anatomy and psychology: they pushed down our larynx to widen our vocal range, freed up our tongues and improved their dexterity, whitened the area around our irises (the sclera) to reveal our gaze direction, and endowed us with innate capacities for vocal mimicry and with motivations for using communicative cues, like pointing and eye contact.

The problem with many efforts to understand the evolution of language is that the lenses used are often focused too narrowly. By placing language within the context of our species' overall repertoire of communicative abilities and then seating this within culture-gene coevolution, we can begin to see the synergistic relationships between tools, practices, norms, communication, and language. Languages are a subset of culture that are composed of communicative tools (words) with rules (grammar) for using those tools. Many of the challenges to developing an account of the origins and evolution of language diminish once we (1) understand how cultural evolution can construct complex but highly learnable repertoires without inventors or intentional designs, and (2) realize that language was merely one element in a broad and synergistic process of culture-gene coevolution.

The idea is that, among our ancestors, cultural evolution accumulated, integrated, and honed many useful communicative elements over long stretches of time into increasingly complex repertoires—as it did with our technologies and institutions. Since none of this leaves a material record for paleoanthropologists to dig up, we can't directly observe this accumulation. However, if I'm right, the communication systems of living and historically known societies will reveal various signatures of cultural evolution, which we can spot if we look closely. We'll first see that communication systems are adapted to the challenges of local environments (they are cultural adaptations), which include both the social and physical environments. Then, we'll see that the conditions of cultural evolution can influence the size and complexity of communicative repertoires. Just as with technologies, the size and interconnectedness of speech communities influence their vocabulary sizes, phoneme inventories, and grammatical tools.

If you are not a linguist, this all probably sounds reasonable, or at least not obviously silly. Why shouldn't languages operate like every other domain of culture? Almost no one would deny that the tools used by Tasmanians were less complex and fewer in number than the tools used by Victorian Aborigines or by modern Australians in Melbourne. Few would deny that societies possess quite different kinds of institu-

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

tions and that these institutions could influence economic success. Most would agree that, at least in principle, it's possible that the institutions possessed by Pama-Nyungan or Inuit speakers could have facilitated their expansions. Linguists and linguistic anthropologists, by contrast, have often assumed that all languages are more or less equal, along all the dimensions that we might care about—equally learnable, efficient, and expressive.[2] Accompanying this assumption is the view that languages are not strongly shaped by nonlinguistic factors—that they are somehow hermetically sealed off from the rest of the world. Together these two assumptions have inhibited cultural evolutionary approaches to language. Recently, however, the cracks in these intellectual barricades have begun to multiply, as researchers use newly available language datasets to study languages as products of cultural evolution. Let's take a closer look.

Communicative Repertoires Culturally Adapt

There seems to be little argument about whether languages are learned from other people—that is, culturally transmitted. Even before birth, fetuses begin acquiring elements of the sounds and rhythms of the languages spoken by their mothers. Infants soon begin carefully watching the tongue and mouth movements of those around them and engage in vocal mimicry, providing themselves with immediate feedback for adaptive learning. Infant brains assemble specialized phoneme (sound) filters that allow them to effectively identify the sounds relevant in the local languages and to ignore the irrelevant ones. Around the end of their first year, children interpret pointing as a request to direct their attention at something and begin communicating with a combination of pointing and emotional reactions long before they talk. Infants and young children rapidly and unconsciously acquire elements of these communicative systems using many of the same cultural learning biases—based on cues of competence, reliability, and ethnicity (dialect)—that are used for learning about tools, practices, and social norms (see chapter 4). Learners watch their models using words or gestures *to do something*, like help a friend, request information, or obtain a particular object. They infer what their models are trying to do and imitate those goals, desires, and behaviors; meanings can be inferred from the rich context found in the routines of social life. Children's play often involves acting out scenes from adult social life, which uses language and other communicative elements in context and alongside noncommuni-

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

cative practices.[3] From the perspective of the learner, communicative systems are just one important aspect of culture—a relevant local skill governed by social norms—that has to be acquired, like fire starting, ritual practices, manioc detoxification, and taboo adherence. Of course, learning the local communicative repertoire gradually opens the floodgates to acquiring many more of the other locally relevant practices, beliefs, motivations, and ways of thinking.

Operating over eons, this cultural evolutionary process has equipped human populations with powerful toolkits for communication. These toolkits involve complex repertoires of elements, including gestures, body positions, facial expressions, and vocalizations, which include clicks, whistles, grunts, growls, hisses, and squeaks. Like our toolkits for food processing, these systems have adapted to the local environments of their populations and are influenced by the same kinds of factors that influence other aspects of culture. Let me start with some of the nonspeech elements of our repertoires, since these remind us that (1) spoken language represents only one component in our communication system, albeit an immensely important one in modern repertoires and (2) we humans can link meaning, by inferring the intentions of communicators, to an immense range of behaviors, not just to speech. Such systems may also provide insights into some of the earliest communicative systems used by our ancestors.

Face-to-face spoken conversations are interspersed with lots of communicative gestures, as well as facial expressions. People in different societies variously point with their index fingers, middle fingers, chins, or pursed lips. Be careful, though; at least two of these are considered rude in some place or another. We shake hands to greet others, or sometimes refuse to shake. If you are in Fiji, make sure to do a long shake (basically, hold hands for a bit). Elsewhere, people bow deeply, bow shallowly, or refuse to bow. To say "OK" or "good," I give the thumbs up or I circle my index finger and thumb in the OK sign. But, do NOT use the OK sign in Iran or Brazil, as President Richard Nixon mistakenly did (it's like giving the finger in the United States). We nod our heads to say "yes," but in some places that means "no" (e.g., Bulgaria). Some peoples subtly raise their eyebrows for "yes." Even ultra-verbal academics frequently use air quotes, an iconic gesture derived from the use of scare quotes in writing, to imply some disagreement with their terminology or its implications. I've seen many a humanities scholar, with a latte in one hand and a book in the other, struggle to communicate, unable to deploy air quotes to shield themselves from any undesirable implications of their words.

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

Figure 13.1. Plains Indian Standard Sign Language. The signs for "talk" and "crazy."

Sign and Whistle Languages

Gestures, however, are not merely a communicative sideshow, interwoven with spoken language. Fully fledged sign languages were common among hunter-gatherers, both in Australia and North America, and simpler gesture systems can be found across the globe, from African foraging bands to Christian monasteries. In North America, many groups of Plains Indians maintained both their own tribal sign languages, for hunting, warfare, and oratory at large ceremonies, and were conversant in Plains Indian Standard Sign Language. Plains Standard provided an important means of intertribal communication in a region of great (spoken) linguistic diversity. Using broad and open gestures, this language appears to be well adapted to communication over open plains, which favors a visual system over an auditory one, and for communication between individuals who might not want to approach each other too closely, as during intertribal communication. Probably because it was acquired after middle childhood, Plains Standard remained highly iconic, meaning that the signs contained elements of what they represented, and relied on much standardized pantomime (see figure 13.1). The sign for "bird," for example, involves flapping one's hands in a stylized way. For "eagle," you first make the sign for "bird," then the sign for "tail," and then for "black." So, "bird-tail-black" would be translated as "eagle."

The grammatical system of Plains Standard employed space instead of any temporal or sequential ordering. The equivalent of noun phrases (e.g., "this black dog") for subjects and objects can be constructed in space, at different locations, and then connected by the relevant verb sign. The hierarchical structure of spoken languages, which so impresses linguists, is relatively easy to construct if one can sketch "air pictures," as

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

they were called, and essentially draw the links. By most accounts, Plains Standard possessed levels of expressiveness comparable to those of spoken languages, although signed conversations were quite a bit slower than in spoken language. As European-descent peoples entered the Great Plains, they copied and used many Indians signs (see the *Boy Scout Handbook*).[4]

Among Aboriginal Australian hunter-gatherers, gestural signing systems of varying degrees of expressiveness and complexity have been documented across the continent, but particularly in the central desert. Some of these societies maintained fully fledged sign languages with the full expressiveness of their spoken counterparts. Elsewhere, the local sign systems possessed only a few dozen or a few hundred signs, which could be effective in particular contexts, like hunting, raiding, or during ritually imposed taboos on speaking, but didn't have the full expressiveness of spoken languages.

Interestingly, the most developed sign languages tend to occur in populations with ritually imposed taboos on speaking that endure for many months or years. Often, as among the Warlpiri, these taboos are associated with a period of mourning after the death of a close male relative. The length of the taboo period varies substantially with the closeness of the relationship, status of the deceased, and the unexpectedness of the death. For the closest female relatives, the taboo is maintained until the deceased spirit is avenged, which often takes about a year. Unlike Plains Standard, these sign systems are not particularly iconic and tend to follow grammatical rules derived from the local spoken tongue, although the sign vocabularies are heavily shared among groups. In the northern central desert, the closer groups live to one another, the higher the percentage of shared hand signs. This holds independent of the similarity of their spoken languages and suggests that these signs were diffusing independently from the spoken languages.[5]

For our purposes, hunter-gatherer signing systems illustrate three important points. First, they appear adapted to their local context—to the functions they serve. Plains Standard, with its large iconic gestures, has evolved to foster communication among adults over distances or at large ceremonies. By contrast, the subtle gestures of Australian sign languages appear to be, in part, solutions to communication when ritually imposed taboos prohibit speech. In places where, for example, women cannot speak for months after a relative dies, sign languages achieve the same level of expressiveness as the spoken language. Ergo, nonlinguistic social norms probably drove the development of these communicative

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

systems. Second, the complexity or expressiveness of Australian signing systems vary from group to group across the continent. Some groups have a limited number of signs (a limited vocabulary) and rather undeveloped grammatical systems, whereas others have the full deal. There's no argument that these various signing systems are equally complex or expressive. They aren't. Finally, the nature of the Australian systems suggests that they are learned by children, and thus the systems are not particularly iconic (unlike Plains Standard)—most hand signs lack any direct similarity to what they represent in the world. The reason is that iconicity aids adult learning, but when children learn languages they strip out the iconicity.

Remember, cultural evolution is smarter than we are, and so it has figured out many ways to facilitate communication. Now, let's consider whistle languages. Among remote populations from Turkey to Mexico, people can have a full conversation by whistling to each other, sometimes over long distances. Many of these whistled languages appear acoustically adapted for communication in steep mountainous terrain, where households might be relatively close, perhaps across a gorge, but an hour or more away in travel time. Some of these languages use finger whistles and can communicate easily over 2 to 3 kilometers and up to 10 kilometers under the right conditions. Other whistle languages use lip whistles for casual, close conversations. So, cultural evolution can and has built various languages out of whistles, because whistles were appropriate for the local context. Beyond whistles, anthropologists have found an incredible diversity of systems for communicating over distances using drums, horns, flutes, and gongs, among other techniques.[6]

Sonority

A successful communicator is one who can most effectively be understood, given the local social, environmental, or ecological conditions. As young or naïve learners focus on and learn from more successful communicators—who are using more effective communication tools—cumulative cultural evolution will gradually assemble sign or whistled repertoires, over time, in the same way that it hones kayaks, spears, and boomerangs. Given this, there's no reason to suspect that such cultural evolutionary processes somehow apply only to whistled or gestural sign languages, and not to typical spoken languages. Thus, spoken languages should—under the right circumstances—show some response to the local acoustic environments and to nonlinguistic social norms, just as

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

whistled and sign languages do. While researchers have done little work on such topics, there's some preliminary evidence.

Spoken languages vary in their *sonority*. The sonority of our voices decreases as the airflow used for speech is obstructed and is highest for open vowels, like the /a/ and lowest for so-called voiceless stops like the /t/ in *tin*. Pronounce each of these sounds and note the difference in the constriction of your airflow. Both vowels and consonants vary in sonority, but vowels generally have much higher sonority than consonants. This means that more sonorous languages tend to have more vowels (e.g., Hawaiian), while less sonorous ones pack the consonants together (e.g., Russian). For the same energy and effort, more sonorous speech sounds can be heard at greater distances and over more ambient noise than less sonorous ones.

If languages adapt culturally, then we can predict that in situations in which people do relatively more talking over greater interpersonal distances with more ambient noise and sound dispersion, languages will be more sonorous. Many environmental variables might influence this, but Robert Monroe, John Fought, and their colleagues reasoned that climate, and specifically temperature, might have a big effect. The idea is simple: in warmer climates, people work, play, cook, and relax outdoors. Compared to life indoors, living outside means that communicators more frequently face the challenges of distance, noise and poor acoustics. Their team generated measures of sonority from word lists for dozens of languages and then looked at the relationship between sonority and measures of climatic temperature, like the number of months per year when it's below 10°C (50°F).

It turns out that if all you know is climatic temperature, then you can account for about one-third of the variation in the sonority of languages. Languages in warmer climates tend to use more vowels than those in colder climates and rely more heavily on the most sonorous vowel, /a/. For consonants, languages in warmer climates rely more heavily on the most sonorant consonants, like /n/, /l/, and /r/. By contrast, languages in colder climates lean more heavily on the least sonorous vowels, as the /i/ in *deep*.[7]

This simple idea can have much nuance added to it. For example, not all warm climates are equally conducive to sonorous speech. In regions with dense forest cover, the advantages of high sonority might be less pronounced, or as the anthropologists Mel and Carol Ember have argued, very cold and windy climates may select against linguistic practices that involve opening one's mouth widely, due to the increased heat

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

loss. To this they added the idea that social norms about sexual restrictiveness might also influence sonority. Adding both of these nuances to the basic climatic temperature analysis, they managed to account for four-fifths of the variation in the sonority of languages (or at least of the few dozen languages they had data for). This is impressive, though it's important not to overstate the strength of these findings. This research program is just getting started. Nevertheless, something is clearly afoot here: the basic relationship between climatic temperature and sonority has been tested in several ways and stood up to retests by different researchers.[8]

Given the way cultural learning shapes our tools, diet, and toolkits to local environments, it shouldn't be surprising if cultural evolution similarly honed our phonemic toolkits.

Biologists have long been interested in how animal communication systems have adapted to the acoustics of a species' local habitat. Unlike language, much of nonhuman communication involves the deployment of a relatively fixed set of signals. However, research with monkeys, birds, and prairie dogs suggests that natural selection has calibrated these communication systems to diverse acoustic environments to make them more effective. Here, we have a parallel case in which selective cultural learning has—well outside of conscious awareness—adapted our communication systems to fit their local acoustic environments. Cultural evolution can solve many of the same adaptive problems as genetic evolution, only faster and without speciation.[9]

Culturally Evolving Complexity, Efficiency, and Learnability

Shifting around the sounds used by languages is one thing, but can cultural evolution actually aggregate, assemble, and integrate linguistic elements—tools and rules—in the same way it does for food-getting technologies or archery packages? Can cumulative cultural evolution endow languages with greater expressiveness or efficiency? The evidence below suggests that the answer is yes; but just as tool sophistication and toolkit sizes are limited by the systematic mistakes and misunderstandings inherent in the transmission of large and complex repertoires, so it is with languages. To survive and spread, languages too need to be learnable, especially by young children, and sometimes by adults in rapidly expanding groups. Thus, like other aspects of culture, we should expect vocabularies, phonemic inventories (repertoires of distinctive sounds),

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

and grammatical toolkits to be influenced by factors such as population size and interconnectedness, as well as by the details of social networks (class stratification and ethnicity), technologies (e.g., writing), and institutions (e.g., formal schools).

Vocabulary Size

Let's begin with vocabularies. Words are tools for communicating, and having the right words can increase the ease, speed, and quality of communication. Although precise measures of the vocabulary sizes of different languages are difficult to establish precisely, rough estimates are sufficient because the differences are so big. Many of the languages carefully documented by linguists and anthropologists working in small-scale societies contain between 3,000 to 5,000 words—that's the entire dictionary. Meanwhile, slim dictionaries for European languages often contain about 50,000 words. A solid college-level English dictionary contains about 100,000 words, and the full-size *Oxford English Dictionary* contains more than 300,000 entries. Of course, this is totally unfair since the dictionaries of nonliterate societies reside entirely in the heads of speakers (until an anthropologist arrives), but European dictionaries reside in books and online databases, and few or no actual speakers know all the words. However, the average American 17-year-old knows 40–60,000 words while their professors know about 73,000 words. These estimated in-the-head vocabularies contain 8 to 14 times more words than are available to the speakers of many small-scale languages, assuming they know all the words in their languages.[10]

These estimates no doubt exaggerate the effective differences in vocabularies, since many of the words missing from small-scale languages are words for pieces of technology, actions, or concepts— like satellites, sit-ups, and emoticons—that don't exist in small-scale societies. To get a perspective on this, let's consider two kinds of words that don't fall into this category: color terms and integer numbers.

Languages vary in their number of basic color terms. At one end of the spectrum, English has 11 basic terms: "black," "white," "red," "yellow," "green," "blue," "purple," "orange," "pink," "brown," and "gray." By contrast, many languages have no basic color terms, or only two, which are always "black" and "white" (or, more accurately, "light" and "dark"). In languages with only 2 or 3 color terms, a brilliant "blue" sky can be described as "black" and compared to the color of dark dirty water.[11] If a language has 3 basic color terms they are "black," "white," and "red,"

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

which usually partitions the entire color spectrum. Numerous languages have 5 color terms, which can be glossed as "white," "black," "red," "yellow," and "green-blue" (grue). Of course, there are always ways around a lack of particular color terms, for example, by referring to some object with the desired color. But, this involves some linguistic work and is thus often less efficient.[12]

Ancient literary works tell the same story as the cross-cultural research. The Old Testament, the works of Homer, and ancient Vedic poems are remarkably vague about color and at times devoid of colorful descriptions. Neither the sky nor the sea is "blue." The worlds of these works are largely cast in black, white, and red. Terms for green, blue, and yellow—as basic descriptive terms—don't emerge until later in the literary histories of these societies. It's amazing that these rich, poetic cultural treasures are nearly colorless.[13]

Thus, color terms provide a narrow but clear window into language evolution. The color spectrum is actually continuous, so the borders between colors are arbitrary—the physics of light does not actually provide discrete categories like "red" and "yellow." This could create a prickly problem for a learner to solve: what region of the visual spectrum does the label "red" actually apply to? How can I as a learner converge on the particular region of color space that you mean by "red"? The answer may lie in the fact that while the actual color spectrum is uniform, our visual system perceives it as containing contours. Think of these contours as perceptual clumps or bumps in our cognitive color landscape that stand out. These contours might be genetically adaptive, reflecting things we have particularly needed to distinguish over our evolutionary history; or they might reflect simply quirks in the system created by the physiological mechanisms needed for perceiving color (or, some of both effects). Nevertheless, these contours help learners triangulate in on the local meaning of "red" or "blue." Over time, color terms evolve culturally so as to maximize the perceptual similarity within a term;[14] that is, examples or shades of "red" are *perceptually* more similar to each other than surrounding shades, even if these "reds" aren't actually any closer based on objective measures of hue and saturation. To be clear, these innate physiological-perceptual constraints are nonlinguistic in origin but get exploited by cultural evolution to help us discretely partition the otherwise continuous color spectrum using words. As with so many cultural adaptations, the system looks engineered for communication about color and fitted to the constraints of our visual perception; but, of course, the only engineer was cultural evolution.[15]

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

As an English speaker, you have 11 basic color terms at your disposal because cultural evolution has figured out how to make them learnable. Meanwhile, Korean speakers have 14 color terms. However, most societies over human evolutionary history probably only had two or three color terms. Learning additional color terms influences our cognitive abilities and increases our gray matter, improving our discrimination and memory for different color shades *across* terminological boundaries at the expense of shades within the boundaries of one term. Of course, if you have more nonoverlapping terms, you have more terminological boundaries.[16]

Integer numbers provide a potentially even more potent example. Since you are reading this book, you have very likely received via cultural inheritance a linguistically encoded counting system that permits you to count without bound using only ten numerals, along with some rules to structure them. This system permits you to easily distinguish a pile of 27 cherries from a pile of 28 cherries. However, this ability has not been available for most of human history. To our knowledge, all societies have number words for 1, 2, and many, and numerous societies count "1," "2," "3," "many"—but that's it; they can't count any higher. They can't, for example, use their counting system to distinguish those two piles of cherries. This fits with research in cognitive science suggesting that 1, 2, and 3 have corresponding innate mental representations.

Beyond this, some small-scale societies augment their innate system using culturally learned body-part counting systems that count objects or people by assigning them to specific body parts in sequence, like counting on one's fingers (but also using elbows, knees, noses, eyes, etc.). In New Guinea, these systems top out at values as low as 12 but variously go as high as 22, 28, 47, or 74 in different societies. In some of these societies, there are ways to go through the cycle two or three times to obtain even higher values. Of course, populations have worked out ways around not having many integers, but groups with only limited numbers of integers usually rapidly adopt the more extensive counting systems of other languages, when they get the chance.[17]

In children, full comprehension of both integers and color terms develops relatively late, at least compared to other aspects of language. Interestingly, Western children now master their basic colors at a younger age than they did in generations past, suggesting that the cultural system is evolving to be better at transmitting this knowledge.[18] Acquiring these products of cumulative cultural evolution—integers and color terms—shapes our brains and influences our cognitive abilities. So, they

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

provide specific cases of how expanding vocabularies influences our "smarts."

More broadly, words are useful for thinking, so possessing a large vocabulary likely improves some kinds of problem solving. The active vocabularies of American adults, which is one component of IQ, have been increasing for at least a half century, and probably longer. While broader education at the university level can account for some of this growth, most of this expansion has occurred because adults have unintentionally continued to grow their vocabularies after their college years. This may have occurred because of shifts in the demands of many jobs in the United States during the twentieth century. The upshot is that American adults score higher on IQ tests than their parents or grandparents did at the same age, in part, because they have learned to use and understand more words.[19]

Finally, there's no reason why vocabularies couldn't have started expanding as soon as our ancestors were sufficiently good cultural learners. Nonhuman primates can already create mental categories or concepts for objects (banana), events (storm), and relational categories (rival), so cultural learning only needs to provide a means to connect these categories to particular gestures or sounds (that is, make words). Once connections are made, cultural evolution can cause these category-word relationships to become widely shared in a social group over generations.[20]

My point: words are like other aspects of culture, and consequently we should expect the vocabulary sizes of languages to expand with a population's collective brain. As both color terms and integers illustrate, such additions are not merely incidental, but may give us new cognitive abilities and raise our IQs.

Sound Inventories

Along with words, the inventories of phonemes used by societies are also influenced by the sociality of populations. Of course, the sounds of spoken languages are constrained, at least in the short run, to the anatomy of our mouths, tongues, and air passages, but within the range of possible sounds, the variation is impressive. Rotokas is spoken by about 4,000 people on an island off the coast of New Guinea. The sound inventory of Rotokas consists of 11 sounds composed of 5 vowels and 6 consonants. Hawaiian too has 5 vowels but combines them with 8 consonants. Among the Pirahã in Amazonia, the men's phonemic inventory parallels the Rotokas, with 11 sounds. Pirahã women, however, don't

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

distinguish two of the consonants that men do, so it appears that they may have the smallest known inventory, at 10 distinct sounds. By contrast, English has roughly 24 consonants and at least 12 vowel sounds, depending on the dialect. In southern Africa, !Xóõ speakers use roughly 47 non-click consonants and 78 click sounds and have a sound inventory of over 140 sounds. That's at least 10 times larger than Hawaiian, Pirahã, and Rotokas.[21]

So, languages vary massively in their number of phonemes. But, why do they vary?

Well, mostly we don't know, and detailed studies motivated by cultural evolutionary thinking are only just beginning. Nevertheless, one interesting factor may be that some people learn their languages in large communities with lots of strangers, where many interactions occur among people who share little in common. By contrast, in small-scale societies, children learn in relatively small and quite homogenous communities. When learners do not share much in common and have to learn from very different voices with substantial variation in pronunciations across individuals, learners are better able to triangulate in on sound contrasts, and getting those contrasts is more important to understanding meanings. In small communities, minor sound differences can be missed in the learning process, in part because the people who are communicating tend to share so much background knowledge and context. This effect may create a tendency for larger populations in highly integrated market societies to maintain more phonemes than smaller, more isolated speech communities.

Consistent with this, there is now evidence from several studies using different linguistic databases that show that *languages with more speakers tend to have more sounds—that is, larger inventories of phonemes*.[22]

Of course, the size of the phoneme inventory should interact with other aspects of language. If a language has more phonemes, it should be able to sustain efficient communication or expressiveness using shorter words. Since the selective processes of cultural evolution will favor greater communicative efficiency, languages that can support more total phonemes—due to larger and more diverse speech communities—can shorten their words. Indeed, several analyses also support this. Languages with more phonemes in their inventories tend to have shorter words.[23]

Overall, this growing body of evidence shows that larger and more interconnected populations tend to have more phonemes and shorter words. And while the analysis is not yet complete, there is also some rea-

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

son to suggest that the words of larger populations will be better optimized for communicative efficiency than those of smaller populations.

Routes to Grammatical Complexity

Here's a simple story; see if you can understand it.

> Girl fruit pick turn mammoth see
> Girl run tree reach climb mammoth tree shake
> Girl yell yell father run spear throw
> Mammoth yell fall
> Father stone take meat cut girl give
> Girl eat finish sleep

This story, modified from Guy Deutscher's *The Unfolding of Language* (p. 210), contains 23 different words arranged to *violate* English grammatical rules. Yet you probably more or less understood the story. Deutscher used some basic principles that are probably rooted in our ape cognition to construct the story in an understandable way without relying on English grammar. First, he used proximity: words for things that are closer in space are closer in speech. They "go together." This doesn't have to be the case in many languages, but it's our initial inference. Second, he used temporal sequencing: the events in the story follow the events in reality. Third, he relied on nonlinguistic causal structures: research with speakers from diverse languages suggests that actors (subjects) are first in thought. After actors, humans are next inclined to think of objects (patients) and, finally, actions (verbs). You can understand "girl fruit pick" more easily than "fruit girl pick" or "pick fruit girl," even though none are the subject-verb-object that English normally requires.[24]

So, you can achieve some degree of communication—even telling a story—with just some shared vocabulary. In fact, not only do you not need grammar, but you can do it even when the organization of the words violates the rules of your own language's grammar. This means that cultural evolution could gradually increase the complexity of a simple protolanguage (involving, say, 23 words) by incrementally adding grammatical rules and tools to its stockpile of words. We now know something of how this happens. Detailed historical and comparative studies show where specific grammatical elements come from and how they usually evolve. They typically come from first hijacking concrete words, then gradually bleaching them of their original meaning, and

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

often then honing them down to size, presumably to increase communicative efficiency. This process is called *grammaticalization*.[25] Here is a taste of grammaticalization:

1. Adjectives: Nouns and verbs, which are built on an ancient cognitive distinction between *things* and *actions* found in nonhuman primates and presumably in our ancient precultural ancestors, get deployed to build adjectives and adverbs. For example, a word for a rock-solid thing like "concrete" (cement, water, and gravel) can evolve to also operate as an adjective, meaning "not abstract," as in "Give me a concrete example." Someday, when new building materials replace concrete, this word may come to only mean "not abstract."
2. Tense: Verbs are sometimes gradually turned into markers of tense, for the past and future. The English word "gonna" provides a nice example of a word in transition to a grammatical element. Consider the sentence, "I'm gonna stay here and not move." Clearly, "gonna" is from the verb phrase "going to," but I just used it to express my intention to "stay." "Gonna" has not fully transitioned yet, but it appears to be doing what the word "will" did centuries ago. "Will" is derived from a verb meaning "want," which is a very common source for the future tense (e.g., in Swahili). Similarly, past tenses often evolve by modifying and then gradually trimming down the verb for "finish" or "done." In the unwritten Yasawan dialect of Fijian I've been studying, this can be seen in a not-yet-trimmed-down version, where the literal translation of "I have eaten" is "I am eating done."
3. Prepositions: Adverbs often get re-tasked to create prepositions, like "after" or "behind." For example, the body-part word for "back" or "buttocks" often gets first modified into an adverb like "behind" ("The sad events lay behind them"). Then, the adverb evolves into a preposition, as in "I put my laptop behind the door." In English, curiously, the euphuistic use of "behind" (as in "Get off your behind") to mean "buttocks" is turning this preposition into a body-part noun (which is rare).
4. Number and plural pronouns: Pluralization is often achieved by adding "all" to nouns and pronouns and then trimming it down. For example, in the dialect of English spoken in the southern United States, a second-person plural pronoun has emerged in the form of "y'all." This new pronoun, which many languages have but

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

no other dialects of English, arose as "you" crashed into an "all" that was added to remove any ambiguity about whether "you" was plural. Incidentally, "y'all" is not ungrammatical; it's an example of how languages evolve. If it feels wrong to you, that's because it violates one of your internalized linguistic social norms.

5. Subordinating conjunctions: Tools for marking subordinate clauses are often built from those for marking relative clauses, which are often built from verbal pointers (demonstratives like "this" and "that"). Verbal pointers in phrases like "that picture" or "this hat" evolve to mark relative clauses, as in "Zoey likes the picture *that* Jessica painted." These relative markers evolve into subordinating conjunctions, as in "I hope *that* Zoey likes the picture that Jessica painted."

Practically speaking, learners have to acquire the relevant social norms for how they can and cannot arrange words to communicate, but these rules are gradually altered as successful communicators, who have made adjustments to increase efficiency or expressiveness, are preferentially attended to and learned from. This can even be seen in relatively brief experiments in which people first had to learn an unfamiliar grammar and then had to use it to communicate. In one experiment, after only three days of interaction, the patterns of a just-learned grammar shifted in ways that increased communicative efficiency. Amazingly, the observed grammatical modifications parallel the actual distributions of these grammatical elements in real languages.[26]

All of this suggests that grammatical tools and rules should evolve like other aspects of culture. At some point, there must have been communicative repertoires that lacked some of the shared norms that languages have for referring to the past or future, marking number, or for embedding clauses. If true, we might find existing or historically known languages with smaller toolkits of grammatical tools or less efficient grammatical rules, just as we do for technologies. However, it may be that such tools, once they emerge in one language, rapidly spread to other languages, as good counting systems do. This would prioritize looking at more-isolated groups or languages, that may not have had time to adopt the new tools, or by looking back in time at past languages using ancient written sources.

Supporting this expectation, subordinating conjunctions, like "after," "before," and "because of," may have evolved only recently, in historical times, and are probably no more a feature of *human* languages than

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

composite bows are a feature of human technological repertoires. The tools of subordination seem less well developed in the earliest versions of Sumerian, Akkadian, Hittite, and Greek. This makes these languages slow, ponderous, and repetitious to read.[27] Subordination tools also seem to be missing in some living languages, including languages found in Australia (e.g., Warlpiri), the Arctic (Inuit), New Guinea (Iatmul) and the Amazon (Pirahã).[28] People use phrases that are subordinate to some degree, but this subordination is provided by a combination of concatenation and context, not by specialized grammatical tools like coordinating conjunctions. To see what I mean, here's subordination using only concatenation and context in action, in English: "I'm taking Sharon's son to the circus next week . . . she's the young woman we met last week . . . we were outside the Starbucks."

Philosophers have long been impressed by "language's" ability to embed phrases within larger hierarchical structures, often using grammatical tools for subordination. However, these findings suggest that much of the hierarchical structuring and fancy subordination that we see in most modern languages is the product of long-term cumulative cultural evolution. This is not to say that we humans don't have some souped-up innate abilities for dealing with hierarchical structures, which may also be useful for making tools or understanding social relationships, but merely that the elegant bits of grammar that permit us to fully harness these abilities were built by cultural evolution.

Finally, grammatical complexity, like other aspects of culture, is influenced by social institutions and adapts to the demands of the communicative contexts of a society. For example, larger populations with more heterogeneous speech communities and more adult language learners tend to have morphologically simpler grammars. In morphologically complex grammars, words have many different, often subtle, grammatical elements that must be inserted, at the front, back, or in the middle. These elements augment the meaning of the word, conveying tense, mood, number, gender, etc. Morphologically simpler languages do much less of this, but can (often) communicate the same information with additional words and word ordering. One interesting explanation for this is that larger populations have more adult language learners, and adult language learners are bad at learning morphologically complex languages.[29]

In summary, larger speech communities seem to possess bigger inventories of words, phonemes, and grammatical tools. That is, the languages possessed by these communities are more complex. Historical

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

evidence suggests that inventories of words and grammatical tools have expanded over centuries and millennia. Psychological evidence indicates that the contents of vocabularies and the rules of grammar, such as those related to color terms and integers, may alter our cognitive abilities, including both memory and perception. A combination of evidence indicates that modern world languages may in fact be rather unusual compared to most of the languages spoken over our species' evolutionary history.

Culturally Evolved to Be Learnable

As our communicative repertoires gradually expanded, there would have been both selective pressure on our genes to make our ancestors better at learning culture, including the elements of these emerging communicative systems, and pressure on the repertoire of communicative elements to be *more learnable*, especially by children. Here I'll briefly emphasize the role of the second partner in this duet—culture. Like other kinds of tools, which have evolved culturally to fit our hands, shoulders, and physical abilities; languages have culturally evolved to fit our brains, becoming more learnable by adapting themselves to the contours and constraints of our psychology (remember the color terms). Children are good at learning languages—in large part—because difficult-to-learn elements don't get learned, and ergo, don't get transmitted to the next generation. There are possible languages that are really hard for children to learn, but those lost out in competition to more learnable languages.[30]

Languages that more effectively exploit innate aspects of our psychology would be more learnable and less likely to be replaced by alternatives in the future. Computer simulations of the process of cultural transmission across generations among imperfect learners with limited memories favors the emergence of many of the key traits that distinguish languages from the communication systems of other species. As the size of vocabularies increased, the cultural evolutionary pressures on languages to become rule bound and regular would rise.[31] A learner might be able to acquire, for example, an idiosyncratic past-tense form for each verb in a language with a small lexicon, of say 50 words. But, they won't be able to do it for a language with 5,000 words. Thus, languages without such rules (syntax) didn't make it. They got beaten out by those languages with more-regular rules, like adding "-ed" for the past tense, that children could learn to apply. This will be the case both

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

because it's tough to remember all those specific cases, and because as vocabulary size increases many more specialized verbs will only be heard rarely. Irregular forms can persist (like "ate" instead of "eated"), but only on frequently used words, where the constant repetition eases such memory constraints.

Direct evidence for these processes in human evolutionary history is nonexistent. However, complementing the computer simulations mentioned above is indirect evidence from laboratory experiments like those described for the accumulation of technical skills in chapter 12. These experiments show that when people observe the performance of others using an artificial language and then make inferences from those observations, both structure (syntax-like forms) and compositionality (discrete words) emerge over many transmission events as people learn from each other over many laboratory generations. No single individual does much at all, and no one is trying to achieve this as a goal. It's an unconscious emergent product of cultural transmission over generations.

Research with children speaking real languages further confirms that regularity makes languages more learnable. Children speaking *more*-regular languages understand sentences, like "The horse kicked the cow," *better* than children who speak *less* regular languages. Specifically, Turkish- and English-speaking children outperformed children speaking the quite irregular Serbo-Croatian language. Of course, these performance differences vanish in later years. Nevertheless, such research suggests that not all languages are equally easy for kids (or adults) to learn.[32]

My point is that cultural evolution is a key reason why existing languages are so easily learned by children, and some of the recurrent features of languages, like syntax, are likely the result of cultural evolution working to keep languages learnable, especially as vocabularies expand.

The Synergies of Manual Skills, Norms, Gestures, and Speech

During our species' evolutionary history, coevolution was relevant in two important ways. First, increasingly complex communicative repertoires were coevolving with increasingly complex tools, practices, and institutions. This coevolutionary interaction was synergistic, since two or more of these domains of culture will create selection pressures for genes that improve our psychological abilities to acquire, store, organize, and retransmit valuable cultural information. This interaction would

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

have sharpened our abilities to culturally learn from others—improving our skills at inferring others' goals and intentions (to better learn from them), inferring underlying rules or norms, and learning complex sequences with hierarchical structure. Second, as noted earlier, these culturally accumulating communicative repertoires put selective pressures on genes for communicating: they pushed down our larynx to widen our vocal range, drove axonal connections from our neocortex down deep into our spinal cords to improve the dexterity of our hands and tongues, whitened the sclera of our eyes to reveal our gaze direction to cue others, and endowed us with reliably developing cognitive skills for vocal mimicry and communicative cueing, like pointing and eye gaze.

It's impossible to know in what behavioral domain cultural evolution first began to accumulate: in toolmaking, food processing, or gestural communication. It's also plausible that cumulative cultural evolution in some ancestral groups began with gestural repertoires while other groups, due to different ecological constraints, started to accumulate techniques for tool manufacture. The key point is that whatever came first would have likely catalyzed developments in the other domains. For example, human brains have axons that project from our neocortex directly to the ventral horn, where the motor neurons are, and then down deep into our spinal cords.[33] This anatomical change helps to explain our species' impressive manual dexterity and specifically our ability to learn tasks requiring manual dexterity. This change would have been favored by the cultural emergence of techniques requiring greater dexterity, like making nets, clothing, blades, or fire, or culturally transmitted skills and tactics, like using projectiles such as rocks, clubs, or spears for hunting or defense. Plausibly, either an emerging gestural communication system or increasingly complex toolmaking skills could have created genetic selection pressures favoring more direct axonal connections into the ventral horn and spinal cord. If gestural communication emerged first, it's plausible that toolmaking skills would have benefited from enhanced dexterity. If it were tools driving the boat, then enhanced gestural communication would have gotten the free ride. Better communicative skills also might have improved the quality of the information transmitted about making and using tools.

This process helps explain our species' impressive tongue dexterity and capacity for vocal mimicry, which are unique among primates. Vocal mimicry is rare in nature, although many bird species who are also excellent vocal mimics have an analogous brain setup.[34] Our neocortical projections, which also innervate the muscles of our jaws,

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

tongues, faces, and vocal chords, could have initially been driven by gestural communication, toolmaking, or any number of other skills. However, once they spread, they could have opened the door to learning vocalizations as a by-product. Vocalizations would have first developed as an addendum to an existing repertoire of communicative gestures, which could have included facial expressions, pointing, and bodily positions. In this scenario, the genetic selection pressures for speech would have been created by a non-speech-based communicative repertoire that gradually favored adding speech elements as a way of freeing the hands and suppressing the need to necessarily look at each other. The cross talk between the tongue and limbs is highlighted at weddings, when inexperienced dancers stick their tongues out while focusing on challenging dance moves.[35]

Whatever the precise evolutionary sequence was for the emergence of gestures, vocalizations, social norms, and tool-use is not crucial since cultural evolution would have created substantial genetic evolutionary pressures for using verbal gestures—speech—to communicate, to free our hands for using tools, throwing weapons, cooking, communicating at night, and maintaining balance while running down prey. Comparative work with other apes shows that our larynx dropped, which lengthened our vocal tracts and expanded our potential phonetic repertoire. Our tongues were freed up to move both side-to-side and up-and-down. We also lost the soft, inflatable air sacs that apes have around their larynx. These changes, however, came with a significant cost. Other mammals, and human infants, can breathe and swallow at the same time and thus can't easily choke on food.[36] Older humans can and do choke.

These genetic evolutionary selection pressures were for communication in general—not only for spoken language—so our ancestors may have maintained gestural languages alongside spoken languages, as many modern hunter-gathers did. In the increasingly cooperative worlds constructed by social norms and reputations, it became relatively more important to communicate and teach others rather than deceive or exploit them. Such pressures explain why humans are peculiar in having our rather small irises set against a white background—the sclera—in our eyes. Anyone watching us can infer where we are looking or whom we are looking at. Experiments with apes and infants show that apes watch the orientation of the heads of others while human infants watch the eyes. This behavior makes it harder for people to hide what they are interested in (a cost), but it is quite useful for actively transmitting cultural information or communicating more generally. Motivated to con-

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

vey cultural information, diverse populations around the globe make use of a somewhat variable set of "pedagogical" cues to facilitate learning. Commonly used cues include eye contact, head movements, and pointing. With infants, people in diverse societies speak in "motherese" (baby talk). With its slow and exaggerated intonations, motherese speeds language learning by providing input that is easier for infants to process.[37]

Recognizing that languages, or at least simpler communicative systems, were coevolving with tools, practices, and social norms helps explain how some of the features of languages might have evolved and why. For example, linguists are often impressed by the hierarchical structure, sequential ordering, and recursion found in modern languages.[38] In a world with only language evolving, it's not obvious where the cognitive abilities that underlie these patterns might have come from. If a genetic mutant with new mental abilities for creating hierarchical linguistic structures emerged in a society having only a simple protolanguage, she wouldn't have anyone with whom to use her advanced abilities. No one lacking her cognitive abilities could understand her or copy her complex patterns.

However, complex tools are also assembled sequentially, hierarchically, and sometimes recursively, just like (most) modern languages. To make a spear, you have to obtain the wood, straighten it, and balance it for throwing. Then, or maybe first, you have to obtain the flint, bone, or obsidian to make the spear tip. This material has to be knapped or carved to make a spear point. Then you need some sinew or resin to attach the spearhead to the shaft. Thus, there are componential parts (objects), actions (knapping), and rules (orderings) for assembling the parts. All of this has to be done with the goal of obtaining a good spear, just as sentences are assembled with the goal of communicating. Similarly, food-processing techniques to detoxify seeds, such as nardoo or acorns, involve sequential protocols with components and subcomponents (grinding, leaching, and cooking), which must be done in order. Beyond sequences and hierarchies, some technical skills, like weaving and knitting, demand recursive applications of the same techniques in a manner that could potentially continue without bound.[39]

Here however, unlike language, there is no need for someone else to have the same abilities to get the benefits. It's directly beneficial to the individual to be able to organize and make a more complex tool. And the existence of these more complex tools will put an even greater selective pressure on those who cannot make them. These cognitive abilities

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

can get redeployed for organizing and transmitting communicative repertoires, or perhaps in transmitting norms and practices about kinship systems. Consistent with this, when humans compete with primates in sequential learning tasks involving manual actions, we achieve superiority only when the task involves a hierarchical structure.[40]

This idea has been explored in computer simulations in which neural networks had to retain a "genetically" evolved ability to learn complex sequences in a nonlinguistic task (like nardoo processing) while at the same time facing the challenge of learning a grammar. The neural networks could evolve genetically to be better learners of the grammar, and the grammars could evolve culturally to be more easily learned by these networks. The results are threefold:

1. Genetically evolving to learn nonlinguistic sequences improved grammar learning—so, synergies exist between tools and language.
2. But having to remain a good nonlinguistic sequence learner (a toolmaker) inhibited genetic evolution's ability to improve grammar learning—so, no genes specifically *for* grammar learning are favored.
3. Nevertheless, neural networks got much better at learning grammars because *the grammars* evolved culturally to be readily learnable by the existing neural networks.[41] In short, cultural evolution made the grammars more learnable.

Genes and Brains for Learning Complex Sequences

Researchers may have isolated a gene that was favored during human evolution to improve our procedural or sequence learning abilities. The gene, *FOXP2*, which is located on human chromosome 7, codes for a protein that influences brain development, particularly in regions related to procedural and motor learning. For 170 million years, this gene appears to have remained relatively fixed across many species; however, since we split with chimpanzees only 5–10 million years ago, there have been two changes in our copy of the gene. Mutations in *FOXP2* influence both grammar and sequence learning in humans, and motor-skill learning in mice. The indirect pathways that connect genes to mental abilities usually make firm links difficult, but it's plausible that *FOXP2* was favored by cumulative cultural evolution to improve our procedural, motor, or sequential learning abilities, for learning complex culturally emerging protocols for toolmaking. Once in place, these abilities could have been harnessed in our communicative repertoire to permit

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

more-complex grammatical constructions. Or perhaps the sequence was the other way around.[42]

Consistent with this, research applying brain-imaging techniques shows that using both language and tools (manual actions) involve overlapping brain regions. In fact, if you focus only on regions that are specific to either tool-use or language-use, you are left with few differences. The differences that do emerge seem trivial, such as language activating the auditory regions and tool-use tapping motor regions. Brain regions once thought to be specific to grammar now appear to be used for tasks requiring sequential or hierarchical structuring and are tapped for complex manual procedures. Similarly, learning to make both sounds and tools tap the same imitative areas, so vocal imitation is not something totally distinct from other kinds of imitation in the brain.[43]

Even now, culture continues to be the primary driver of human genetic evolution, and the cultural evolution of different linguistic features in different populations may be continuing to alter gene frequencies. One suggestive bit of evidence indicates that the emergence of nontonal languages may have created conditions favoring the spread of two different genetic variants. Many languages overlay pitch or pitch change on top of a set of sounds to distinguish different words. These are tonal languages. In Mandarin Chinese, for example, the sound /ma/ can mean "mom," horse," "hemp," or "scold" depending on the intonations or tonal contours. Nontonal languages, such as English or Spanish (or any European language), only use pitch or intonation at the level of the phrase to convey emphasis, feeling, or mood. The two genes are variants of *ASPM* and *MCPH*, and influence brain growth and development, and these have begun spreading under natural selection in the last 50,000 years or so. Both new variants of these genes are associated with nontonal languages: that is, speakers of nontonal languages tend to have the new variants.

Of course, the history of the spread of human populations means that many elements of genes and languages will be correlated, because they so often move together. Consequently, gene-language correlations are rampant and don't generally point to the action of natural selection. But in this case, it appears that historical relationships, linguistic similarities, and spatial proximity cannot explain the gene-language correlation. And so far, these variants cannot be traced to any other psychological or cognitive differences, such as intelligence, brain size, social skills, or schizophrenia.

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

Nevertheless, to be clear, *any human child* raised anywhere can learn the local language, but natural selection may be tinkering with the genetics of just how easy or hard it is, which depends on the enduring features of local languages.[44]

Culture, Cooperation, and Why Actions Speak Louder Than Words

The broad view I've presented in this chapter challenges many claims—old and new—that the emergence of language in our species' evolutionary lineage was THE Rubicon that, once crossed, set us apart from the rest of nature. With the emergence of language, so the story goes, cultural transmission became possible and the problem of cooperation was solved, basically by reputational gossip. While there is no doubt that language plays an immensely important role for our species, there are three main problems with this common view and its overemphasis on language. First, this approach fails to recognize that quite a bit of cultural transmission and cultural evolution is possible without language. Cultural information about tool manufacture, fire making, dangerous animals, edible plants, cooking, and diet (food choice) can all be acquired, at least to some nontrivial degree, without language. Even social norms, such as those related to food sharing, can be transmitted without language. A little way down the culture-gene coevolutionary road, but long before full-blown language, more controlled pointing, facial expressions, and improvised pantomime could have provided communicative tools for cultural transmission, as these still do when people don't share a common tongue. Anything we might call language probably arrived on the scene well after cultural evolution was underway, and had already assembled simple communicative repertoires.

Second, language itself is a culturally evolved product, so it can't *cause* culture. Of course, language can and does greatly facilitate cultural transmission—the flow of cultural information—and opens up whole new pathways, such as the transmission of stories, labeled categories, and songs. But these new pathways were bootstrapped by nonlinguistic forms of cultural evolution, just as the much later spread of writing and literacy was bootstrapped from language to open up new cultural evolutionary avenues.

Finally, language has at its core a rather serious cooperative dilemma: lying, deception, and exaggeration. Lying with language is cheap, at least

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

in the short term, and is a potentially powerful way to exploit and manipulate others. The more complex a communicative system is, the easier it is to lie or shade the truth, and get away with it.[45] If this cooperative dilemma is not addressed, the evolution of language, whether by genetic or cultural evolution, is rather limited. The reason is straightforward. If others are using language to trick or deceive me, I can avoid this by not believing anyone or even by not listening to them at all. If, to avoid being manipulated, everyone stops listening, then there's no reason to try to communicate. Language will go away or remain limited to those situations where deception or manipulation are too difficult.

Thus, for complex communicative repertoires to evolve in the first place, this cooperative dilemma has to have already been at least partially solved. Therefore, language can't be the big solution to human cooperation. Of course, once culturally evolved social norms for honest communication emerge and spread, then language can potentially expand the sphere of cooperation and exchange by helping spread gossip about norm violators and by making the acquisition of social norms faster and more accurate. Many arguments regarding how language solves human cooperation assume—without acknowledging it—that culture already exists because they assume reputations exist and that these provide a policing mechanism for honesty. This presupposes both reputations and social norms about honesty, which are both culturally transmitted and highly variable across societies.

In addition to social norms against lying, the cooperative dilemma of language is further mitigated by our cultural learning abilities in two ways. First, transmitting reputational information does not require language and can effectively be done with nothing more than pointing and facial expressions. For example, I can point at Bill lounging under a shady tree while we are rebuilding his house and make a disgust face or shame display. This will transmit to you my views of Bill's actions. So, the transmission of reputation isn't reliant solely on language. Second, although reputations are subject to manipulation, cultural learners already have cognitive abilities that would have mitigated the impacts of deception by self-interested others. For example, Steve might try to convince you that his rival, John, is a scoundrel so that Steve can gain an advantage on John (suppose John is honorable). However, if you use conformist transmission in acquiring your beliefs about John, you would take input from many people in the community, most of whom would not be John's rival. This would allow you to discard Steve's disinformation and converge on a more accurate version of John's reputa-

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

tion. Thus, our cultural learning mechanisms help prevent lying and deception from destroying the value of reputations.[46]

Nevertheless, the emergence of language would still have created opportunities for individuals, especially those who are successful and prestigious, to manipulate others into doing or believing things that would benefit the manipulator. For example, a manipulator who wants to poison his competitors might try, using language, to spread the idea that blue mushrooms are delicious and nutritious, even though he believes they are mildly toxic. The potential for such manipulative cultural transmission would have created a selection pressure for learners to look for what I call *CRedibility Enhancing Displays* (CREDs). CREDs are actions that a person would be unlikely to perform if he or she believes something different from their verbally stated beliefs or if they prefer something different from their stated preferences. Using CREDs provides learners with a kind of partial immune system, or filter, against manipulators who would exploit the cheap cultural-transmission channel provided by language. A good CRED for actually believing that blue mushrooms are delicious and nutritious is to eat a lot of blue mushrooms and feed them to one's children. If a learner observes a potential model doing this, he should then be willing to weight the model's statements about the nutritional value of blue mushrooms more heavily in forming his own beliefs.

Although research is just beginning, experimental work with children and adults suggests that CREDs are important for the cultural transmission of many beliefs and practices, including food preferences, altruistic behaviors, and supernatural or counterintuitive beliefs.[47] CREDs, for example, help explain why martyrdom can be a powerful force in spreading religious faiths. Anyone willing to die for their supernatural beliefs probably actually holds those beliefs and is not (or was not) engaged in some trickery or manipulation. CREDs also help explain why religious leaders take vows of poverty and celibacy, because these CREDs make them more potent transmitters of the faith. Overall, CREDs provide a partial immunity to the opportunities for exploitation opened up by language, though of course the prevalence of con artists, phony TV evangelists, and other hucksters attests to the persistent problem created by our reliance on cultural learning through language.

Our use of CREDs is why actions speak louder than words and contributes to suppressing self-interested attempts to manipulate or exploit cultural transmission.

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms

Summary

Let me close this chapter with six take-home points:

1. Languages are packages of cultural adaptations for communication. Just as with technologies and social norms, once cultural evolution is unleashed, it will begin generating increasingly complex communicative repertoires adapted to local contexts. We have seen that cultural evolution can even assemble sign and whistle languages or alter the sonority of words to fit the recurrent acoustic environments faced by different speech communities.
2. These repertoires will evolve culturally to be learnable, and this learnability will create shared features among languages as they exploit diverse nonlinguistic aspects of our innate psychology. However, the details of this process will depend on who has to learn the languages—only children or people of all ages.
3. Like toolkits, the size and interconnectedness of populations favors culturally evolving and sustaining larger vocabularies, more phonemes, shorter words, and certain kinds of more complex grammatical tools, like subordinating conjunctions.
4. These patterns combined with anthropological and historical evidence from diverse languages suggests that widely spoken modern languages, like English, probably look quite different from the languages that dominated our species' evolutionary history, just as our technologies and institutions look rather different from those of small-scale societies.
5. Acquiring languages alters aspects of our psychology and endows individuals with new cognitive abilities.
6. As a recurrent product of cultural evolution, languages have produced powerful genetic selection pressures over the course of human evolution, shaping our bodies and brains for speech and cooperative communication. This process continues.
7. Understanding the evolution of language requires seating it within the broader context of the culture-gene coevolution of tools, skills, and social norms.

In chapter 14, we'll see some of the myriad ways in which culture shapes our brains and biology, even when it doesn't mess with our genes. Cultural evolution is a type of biological evolution.

This content downloaded from 47.230.201.5 on Sat, 08 Aug 2026 19:01:37 UTC
All use subject to https://about.jstor.org/terms