

Babbling Birds

Riny Huybregts

CHESTNUT-CROWNED BABBLERS (*Pomatostomus ruficeps*) produce a bitonal flight call as they approach their nest and a tritonal prompt call when feeding chicks in the nest. Sabrina Engesser et al. have demonstrated that the tones within these calls are perceptibly distinct within calls, perceptibly equivalent across calls, and meaningless in isolation, conveying no functionally relevant information.¹ Both calls are combinations of two naturally occurring notes, *A* and *B*, differentiated by pitch contour: flight call *AB* and prompt call *BAB*. In a series of experiments, Engesser et al. switched the elements between the *AB* and *BAB* combinations. Babblers were still able to discriminate between the two types of calls despite these changes.² A flight call composed from the elements of a prompt call is still interpreted as a flight call and a prompt call made using the elements of a flight call is still heard as a prompt call. When played in isolation, tones *A* and *B* did not elicit any specific response.

At first glance, this capacity may seem comparable to the way humans form meaningful words from meaningless phonemes. The apparent similarities are suggestive. Studying this phenomenon in birds may, in fact, help us understand how such a capacity first evolved in humans. After all, it is this combinatorial capacity, together with the capacity to combine words into phrases, that constitutes the defining trait of human language—sometimes referred to as duality of patterning.³ Upon closer inspection, a series of differences emerges. The organizational principles governing bird calls appear unlike the phonemic organization of words in natural languages, a process that is primarily based on computational efficiency.⁴ In addition, birds lack an operation analogous to the recursive procedure that builds phrases from words. There is no duality of patterning. This should come as no surprise. In human beings, the externalization of language, whether by speaking or signing, is ancillary. What is fundamental is the capacity to merge linguistic elements drawn from a lexicon, in the way that *rabbits* and *run* are merged into a single set as {*rabbits*, *run*}.⁵ In evolutionary terms, this basic property seems to have emerged recently and it seems to have emerged abruptly.⁶ Babblers are compelled to use only one of four bitonal (*AA*, *AB*, *BA*, *BB*) and eight

tritonal (*AAA*, *AAB*, *ABA*, *BAA*, *BBA*, *BAB*, *ABB*, *BBB*) combinations. Their vocalizations are fixed and directly linked to specific stimuli in the bird's immediate natural habitat. A spontaneously composed *BA* flight call, signaling, say, departure from the nest, lies beyond the bird's brain.

No goodbyes.

By way of comparison, consider the combinatorial capacity of a human language. This discussion will be restricted to three phonemic consonants—*p*, *t*, and *k*—and four vocalic phonemes—*i*, *e*, *a*, and *o*—in Dutch consonant-vowel-consonant words. From a total of thirty-six possible combinations, only *tep*, *tit*, *tat*, *tet*, *tek*, *kep*, and *ket* go unrealized. They remain available for future use. Birds must make do with what they have. Babbler calls are unlike human languages in several other respects. In calling out, birds use a finite lexicon and linear order. The basic combinatoric operation is concatenation over sequences of tonal elements. There are no freely generated bird calls. For all we know, flight and prompt calls could be stored in avian memory as units with no need for phonemic composition. What is more, human beings recognize phonemes mainly on the basis of formant transitions, as when the vocal tract vibrates energetically in the passage from a closed to an open vocal tract.⁷ The difference between /*bu:m*/ (“boom”) and /*du:m*/ (“doom”) depends on the formant transitions to the vowel following the consonant rather than the acoustic differences between /*b*/ and /*d*/. Not so for birdcalls.⁸ In English, the bilabial plosive “*p*,” dental plosive “*t*,” and the round back vowel “*o*” form distinctive phonemes that are meaningless in isolation, but produce meaningful words in different arrangements: “*pot*” (/pɒt/), “*top*” (/tɒp/), and “*opt*” (/ɒpt/). Bird brains hear the *A* in *AB* and *BAB* as alike; but the occurrences of /*t*/ in “*pot*” and “*top*” are not perceptibly equivalent for human brains. The phoneme /*t*/ in “*top*” is aspirated, but lacks aspiration in “*pot*” and may be pronounced without audible release. Substituting /*ɔ*/ in “*pot*” for /*ɒ*/ in “*top*” may result in “*top*” being misidentified as “*pop*.” This is due to the acoustics of coarticulation, a characteristic that is typical of human speech.

The word-like elements of a natural language are built from a sequence of syllables composed of phonemic seg-

ments organized in onsets and nuclei. The phonemic transcription of the word “phoneme” as /fəʊni:m/ reveals an internal structure: the onset /f/ and nucleus /əʊ/ in the first syllable, and the onset /n/ and nucleus /i:m/ in the second.⁹ Unlike syntax, phonology is not recursive. There are no syllables nested inside other syllables.¹⁰ Phonology is iterative and allows repeats.¹¹ Even so, a strong generative capacity must be invoked to explain syllable structure.¹² Why does Dutch use *stol*, *stool*, and *stolp*, but not *stoolp*? Evidently, there is no categorical ban on “lp.” The asymmetry must follow from some hierarchical syllable structure, which permits only two tail-end rimals: VV, VC. It follows that *stoolp* is ill-formed because “ool” is ruled out as a rime and “lp” is excluded as a syllable onset, as in **lpot*. The “ol” and “p” of *stolp* satisfy respective conditions on rimal and onset structures, as in “stol” and “pot.”¹³

In contrast to birdcall communication, the normal use of human language is creative and unbounded. There is no nonarbitrary limit to the length of expressions or the depth of embedding—*We all think that John believed that his sister may have thought that...* It is neither determined by stimuli—*You said what?*—nor random—*Hell’s drop cigarettin’ wigwam ...*—but coherent and appropriate—*If elected, I will not serve*.¹⁴ Babbling vocalizations are a limited, fixed, stimulus-bound repertoire of calls that are involuntary and controlled by instinct. The link to specific stimuli means that they are more likely to be governed primarily by conditions of communicative efficiency.

They get the job done.

In this regard, two aspects of babbling calls play a notable role in enhancing communication. First, the most frequently used birdcalls tend to be short, possibly reflecting a principle of least effort for efficient communication.¹⁵ Less is more. Second, distinctive calls are *maximally* distinct. Bigger is better. A bird communicating its whereabouts during flight will need a more frequently used call than a feeding prompt. Flight calls should be shorter than prompt calls.¹⁶ And so they are. This is reflected in their bitonal and tritonal composition, respectively. Maximal distinctiveness entails maximal tonal differences within and across calls: bitonal calls must be *AB* or *BA*, but not *AA* or *BB*; tritonal calls, *ABA* or *BAB*, but not *ABB*, *BBA*, *BAA*, or *AAB*. Only two combinations are optimally scattered: *AB* (flight) and *BAB* (prompt); *BA* (flight) and *ABA* (prompt). The call *BAB*, it should be noted, contains *AB* but begins with a contrasting onset.

In birdcalls, ease of communication seems to prevail over computational efficiency. The reverse is true for natural language.¹⁷ This remains the case even though word length and frequency seem to be in accordance with Zipf’s law, which pegs the frequency of a word inversely to its length.¹⁸ Although data in vocabulary studies can be approximated by a Zipfian distribution,¹⁹ the law itself does not specifically apply to language. Communicative efficiency cannot be inferred from Zipf’s law and conformity to it is

essentially meaningless.²⁰ If rank-frequency distributions deviated from Zipf’s law, such a deviation from predicted behavior would be a positive result conveying something meaningful about word choice.²¹ A recent revision of Zipf’s law by Edward Gibson et al. holds that “[t]he most communicatively efficient code for [word] meanings is one that shortens the most predictable words—not the most frequent words.”²² Not so. Charles Yang et al. has shown that the claim of communicative efficiency is unsupported.²³ The statistical distributions of words in Gibson’s study are replicated by Yang’s stochastic model, which mechanically pairs sounds to their meanings “without any functional considerations.”²⁴ Zipf’s law has been incorrectly interpreted as indicating lexical efficiency. In any case, it has no relevance to the far more important notion of structural efficiency. The basic properties of human language overwhelmingly demonstrate the prevalence of computational over communicative efficiency.²⁵ *Why has John left the room?* Who knows? But why has “why” been dragged to the front of the sentence from its expected grammatical position at the rear? That is a matter of structure dependence.

Bird communication does not share the capacity of human language to freely generate new meanings from meaningless elements. The differences between the two systems are qualitative and abrupt. Since birds lack a recursive operation for the creative use of call vocalizations, evolutionary and comparative biological studies of avian and human communication will always remain a problematic enterprise.²⁶ Still, some significant similarities between human speech and birdsong have recently come to light. The sensorimotor systems for producing language or birdsong require similar linear arrangements of differently organized structures.²⁷ These appear to be derived from transcription factors for convergent neurogenetic organization in analogous brain regions that are involved in auditory–vocal imitation learning, perception, and production.²⁸ It is plausible that this convergence,²⁹ which is absent in both our closest primate relatives and non-vocal learning birds, may have contributed to externalized language in human evolution.³⁰

Riny Huybregts is a retired theoretical linguist who has formerly held positions at the University of Utrecht, Tilburg University, and Leiden University.



1. Sabrina Engesser et al., “[Chestnut-Crowned Babblers Construct Calls from Meaningless, Shared Building Blocks](#),” *Proceedings of the National Academy of Sciences* 116, no. 39 (2019): 19,579–84, doi:10.1073/pnas.1819513116. This paper was based on a presentation at the University of Zurich Inaugural Workshop *The Evolution of Language: Bridging*

- the Natural and Cognitive Sciences* (ISLE, Zurich, March 4–5, 2019).
2. Engesser et al., "[Chestnut-Crowned Babblers](#)," as well as Sabrina Engesser et al., "Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System," *PLoS Biology* 13 (2015): e1002171.
 3. Charles Hockett, "[The Origin of Speech](#)," *Scientific American* 203, no. 3 (1960): 88–96, doi:10.1038/scientificamerican0960-88.
 4. Noam Chomsky, "[Three Factors in Language Design](#)," *Linguistic Inquiry* 36, no. 1 (2005): 1–22, doi:10.1162/0024389052993655; Noam Chomsky, "[What Kind of Creatures Are We? The Dewey Lectures, Lecture I: What Is Language?](#)," *Journal of Philosophy* 11, no. 12 (2013): 645–62, doi:10.5840/jphil2013110121; Noam Chomsky, "[Some Core Contested Concepts](#)," *Journal of Psycholinguistic Research* 44 (2015): 91–104, doi:10.1007/s10936-014-9331-5; Noam Chomsky, "[The Language Capacity: Architecture and Evolution](#)," *Psychonomic Bulletin and Review* 24 (2017): 200–3, doi:10.3758/s13423-016-1078-6; Martin Everaert et al., "[Structures, Not Strings: Linguistics as Part of the Cognitive Sciences](#)," *Trends in Cognitive Sciences* 19, no. 12 (2015): 729–43, doi:10.1016/j.tics.2015.09.008; Robert Berwick and Noam Chomsky, *Why Only Us: Language and Evolution* (Cambridge, MA: MIT Press, 2016). Riny Huybregts, "[Infinite Generation of Language Unreachable from a Stepwise Approach](#)," *Frontiers in Psychology* 10 (2019): 425, doi:10.3389/fpsyg.2019.00425; Robert Berwick and Noam Chomsky, "[The Siege of Paris](#)," *Inference: International Review of Science* 4, no. 3 (2019).
 5. And also to reapply the binary operation recursively to, for example, *saw* and the previous merge result, constructing the hierarchical structure {*saw*, {*rabbits*, *run*}}, and so on indefinitely. Chomsky, "[What Kind of Creatures Are We?](#)"; Berwick and Chomsky, "[The Siege of Paris](#)."
 6. Furthermore, not until the earliest separations of human populations did this novel property become linked to the evolutionarily more ancient sensorimotor system for externalization, speech when possible or sign when necessary. Chomsky, "[What Kind of Creatures Are We?](#)"; Berwick and Chomsky, *Why Only Us*; Noam Chomsky, "[The Language Capacity](#)"; Riny Huybregts, "[Phonemic Clicks and the Mapping Asymmetry: How Language Emerged and Speech Developed](#)," *Neuroscience & Biobehavioral Reviews* 81 (2017): 279–94, doi:10.1016/j.neubiorev.2017.01.041.
 7. As opposed to formant values for sustained phonemes, which are only seldomly reached in fluent speech. Kenneth Stevens, *Acoustic Phonetics* (Cambridge, MA: MIT Press, 1998).
 8. An analogous experiment conducted with humans would be expected to yield systematic confusion and errors in identifying plosive consonants. In fact, error patterns do occur. E.g., substitution of /u:/ in *boom* for /u:/ in *doom* tricks listeners into mistaking "doom" for "boom." Marcel Just et al., "[Acoustic Cues and Psychological Processes in the Perception of Natural Stop Consonants](#)," *Perception and Psychophysics* 24, no. 4 (1978): 327–36, doi:10.3758/bf03204249.
 9. Onsets may be structurally complex and have asymmetric arrangements, e.g., *plot* versus **lpot*.
 10. William Idsardi, "Why Is Phonology Different? No Recursion," in *Language, Syntax and the Natural Sciences*, ed. Ángel Gallego and Roger Martin (Cambridge: Cambridge University Press, 2018), 212–23.
 11. In fact, human language phonotactics can be weakly generated by (sub)regular grammars of the Chomsky hierarchy of rewrite systems, and most of it may be even strictly local Noam Chomsky, "[On Certain Formal Properties of Grammars](#)," *Information and Control* 2, no. 2 (1959): 137–67, doi:10.1016/S0019-9958(59)90362-6; Jeffrey Heinz, "Learning Long-Distance Phonotactics," *Linguistic Inquiry* 41, no. 4 (2010): 623–61, doi:10.1162/LING_a_00015; Jeffrey Heinz and William Idsardi, "[Sentence and Word Complexity](#)," *Science* 333, no. 6,040 (2011): 295–97, doi:10.1126/science.1210358; Jane Chandlee and Jeffrey Heinz, "[Strict Locality and Phonological Maps](#)," *Linguistic Inquiry* 49, no. 1 (2018): 23–60, doi:10.1162/LING_a_00265.
 12. Daniel Kahn, *Syllable-Based Generalizations in English Phonology* (New York: Garland, 1980). Jonathan Kaye, "[Coda Licensing](#)," *Phonology* 7, no. 1 (1990): 301–30, doi:10.1017/s0952675700001214.
 13. An easy explanation is provided by the coda licensing condition of Jonathan Kaye, which requires that postnuclear rimal positions (codas) must be licensed by a following onset. Since the sequence "lp" in **stoolp(en)* is not licensed by a following onset, it cannot have a rimal position. As a consequence, "lp" must be an onset in **stoolp(en)* as it is in **lpot*. Since the onset is illegitimate, the ill-formedness of these "words" receives the same explanation. In brief, words of natural language are hierarchically structured in syllabic arrangements of phonemic segments that conform to conditions of computational resources. Kaye, "[Coda Licensing](#)."
 14. Noam Chomsky, *Cartesian Linguistics: A Chapter in the History of Rationalist Thought* (Cambridge: Cambridge University Press, 2009 [1966]). Chomsky, "[What Kind of Creatures Are We?](#)"
 15. George Zipf, *Human Behavior and the Principle of Least Effort* (New York: Addison-Wesley, 1949); Steven Piantadosi, Harry Tily, and Edward Gibson, "[Word Lengths Are Optimized for Efficient Communication](#)," *Proceedings of the National Academy of Sciences* 108 (2011): 3,526–29, doi:10.1073/pnas.1012551108.
 16. Unitonal calls are communicatively uninformative or not discriminatory enough, and are therefore assumed not to occur. A different outcome involving flight (*AB*) and prompt (*BA*) calls of equal length but with no change to their frequency would instead have shown consistency with Saussurean arbitrariness of human language signs.
 17. Chomsky, "[Three Factors in Language Design](#)." Chomsky, "[What Kind of Creatures Are We?](#)" Chomsky, "[Some Core Contested Concepts](#)." Chomsky, "[The Language Capacity](#)." Everaert et al., "[Structures, Not Strings](#)." Berwick and Chomsky, *Why Only Us*.

18. George Zipf, *The Psychobiology of Language* (London: Routledge, 1936).
19. Charles Yang, *The Price of Linguistic Productivity* (Cambridge, MA: MIT Press, 2016).
20. George Miller, "[Some Effects of Intermittent Silence](#)," *The American Journal of Psychology* 70, no. 2 (1957): 311–14, doi:10.2307/1419346. George A. Miller and Noam Chomsky, "Finitary Models of Language Users," in *Handbook of Mathematical Psychology*, vol. 2, ed. R. Duncan Luce, Robert Bush, and Eugene Galanter (New York: Wiley, 1963), 419–92.
21. Noam Chomsky, "[Review: Langage des machines et langage humain by Vitold Belevitch](#)," *Language* 34, no. 1 (1958): 99–105, doi:10.2307/411281.
22. Piantadosi, Tily, and Gibson, "[Word Lengths Are Optimized](#)," 3,528. It would be interesting to see if their reformulation will overcome the contrary results of classical studies, such as those described by Miller and Chomsky, with demonstrative proof that randomly constructed pseudo-lexicons show different interword statistical dependencies from human lexicons, as indeed they should if the notion of communicative efficiency is to have any clear empirical content. Precisely such a test, an update to Miller's original thought experiment, has been run by Charles Yang et al., who show that lexicons generated without recourse to functional considerations but incorporating phonotactic structure of language and a semantic dimension exhibit the correlational statistical properties of words attributed to communicative efficiency. Miller, "[Some Effects of Intermittent Silence](#)"; Miller and Chomsky, "Finitary Models of Language Users"; Spencer Caplan, Jordan Kodner, and Charles Yang, "[Miller's Monkey Updated: Communicative Efficiency and the Statistics of Words in Natural Language](#)" (to appear in *Cognition*).
23. Caplan, Kodner, and Yang "[Miller's Monkey Updated](#)."
24. Caplan, Kodner, and Yang "[Miller's Monkey Updated](#)."
25. Chomsky, "[What Kind of Creatures Are We?](#)"; Chomsky, "[Some Core Contested Concepts](#)"; Everaert et al., "[Structures, Not Strings](#)"; Berwick and Chomsky, *Why Only Us*.
26. Berwick and Chomsky, *Why Only Us*; Noam Chomsky, "[The Language Capacity](#)." Martin Everaert et al., "[What Is Language and How Could It Have Evolved?](#)" *Trends in Cognitive Sciences* 21, no. 8 (2017): 569–71, doi:10.1016/j.tics.2017.05.007.
27. Robert Berwick et al., "[Songs to Syntax: The Linguistics of Birdsong](#)," *Trends in Cognitive Sciences* 15, no. 3 (2011): 113–21, doi:10.1016/j.tics.2011.01.002; Johan Bolhuis, Kazuo Okanoya, and Constance Scharff, "[Twitter Evolution: Converging Mechanisms in Birdsong and Human Speech](#)," *Nature Review of Neuroscience* 11 (2010): 747–59, doi:10.1038/nrn2931; Allison Doupe and Patricia Kuhl, "[Birdsong and Human Speech: Common Themes and Mechanisms](#)," *Annual Review of Neuroscience* 22 (1999): 567–631, doi:10.1146/annurev.neuro.22.1.567; Dina Lipkind et al., "[Stepwise Acquisition of Vocal Combinatorial Capacity in Songbirds and Human Infants](#)," *Nature* 498, no. 7,352 (2013): 104–8, doi:10.1038/nature12173. Christopher Petkov and Erich Jarvis, "[Birds, Primates, and Spoken Language Origins: Behavioral Phenotypes and Neurobiological Substrates](#)," *Frontiers: Evolutionary Neuroscience* 10 (2012): 1–24, doi:10.3389/fnevo.2012.00012.
28. Andreas Pfenning et al., "[Convergent Transcriptional Specializations in the Brains of Humans and Song-Learning Birds](#)," *Science* 346, no. 6,215 (2014), doi:10.1126/science.1256846.
29. Marc Hauser et al., "[The Mystery of Language Evolution](#)," *Frontiers in Psychology* 5 (2014): 401, doi:10.3389/fpsyg.2014.00401. Berwick and Chomsky, *Why Only Us*.
30. Many thanks go to Bob Berwick, Gerrit Bloothoofd, Martin Everaert, Marc Hauser, Marc van Oostendorp, and Charles Yang for helpful comments, and especially to Noam Chomsky for valuable discussion.