Journal of Language Evolution, 2017, 114–125
doi: 10.1093/jole/lzx005
Advance Access Publication Date: 7 April 2017
Research article

OXFORD

In the beginning‡

A review of Robert C. Berwick and Noam Chomsky's Why Only Us

Michael Studdert-Kennedy[1,] and Herbert Terrace[2,*,**]**

[1]Haskins Laboratories and [2]Columbia University

*Corresponding author: terrace@columbia.edu
**The authors names are alphabetical, but each contributed equally to this review.
‡This manuscript was accepted for publication shortly after the death of M.S.-K. The final draft was prepared solely by H.T.

Abstract

We review Berwick and Chomsky's *Why Only Us, Language and Evolution*, a book premised on language as an instrument primarily of thought, only secondarily of communication. The authors conclude that a Universal Grammar can be reduced to three biologically isolated components, whose computational system for syntax was the result of a single mutation that occurred about 80,000 years ago. We question that argument because it ignores the origin of words, even though Berwick and Chomsky acknowledge that words evolved before grammar. It also fails to explain what evolutionary problem language uniquely solved (Wallace's question). To answer that question, we review recent discoveries about the ontogeny and phylogeny of words. Ontogenetically, two modes of nonverbal relation between infant and mother begin at or within 6 months of birth that are crucial antecedents of the infant's first words: intersubjectivity and joint attention. Intersubjectivity refers to rhythmic shared affect between infant and caretaker(s) that develop during the first 6 months. When the infant begins to crawl, they begin to attend jointly to environmental objects. Phylogenetically, Hrdy and Bickerton describe aspects of *Homo erectus'* ecology and cognition that facilitated the evolution of words. Hrdy shows how cooperative breeding established trust between infant and caretakers, laying the groundwork for a community of mutual trust among adults. Bickerton shows how 'confrontational scavenging' led to displaced reference, whereby an individual communicated the nature of a dead animal and its location to members of the group that could not see it. Thus, both phylogenetically and ontogenetically, the original function of language was primarily an instrument of communication. Rejecting Berwick and Chomsky's answer to Wallace's question that syntax afforded better planning and inference, we endorse Bickerton's view that language enabled speakers to refer to objects not immediately present. Thus arose context-free mental representations, unique to human language and thought.

Key words: evolution of language; displaced reference; evolution of words; protolanguage; Wallace's question.

Here at last is a book that explains Noam Chomsky's views on the evolution of language. We have long known that he sees little room for the gradualism of natural selection in the evolution of syntax, attributing it rather to a slight rewiring of the brain by a minor chance mutation (Chomsky 2010). Now in four occasionally overlapping essays Robert Berwick, a computer scientist, and Chomsky (Berwick and Chomsky 2016) reveal the reasoning behind these speculations.

The review is divided into three sections. In the first, we lay out the book's argument. In the second, we develop a critique and some counterarguments. In the third, we propose a partial alternative scenario, speculating on the origins of words.

1. The argument

Every normal human child can learn to speak and understand any natural human language. No other ape can do this. Language is specific to humans and '... as far as we know, apart from pathology, uniform in the human population' (54). Why then are there so many different languages? How do they differ and what do they have in common? What is the common core of language that emerges in every child?

The answers begin to take shape if we recognize that language is primarily an instrument of thought, only secondarily a means of communication. From this it follows that language is a property of the individual, not the group, as Saussure, Bloomfield, and others supposed, and therefore, as Chomsky was among the first to recognize, open to evolutionary study.

What then evolved, how, when, and why? 'Why' is '... the dilemma that plagued the Darwinian explanation of language evolution from the start' (2), sometimes called 'Wallace's problem' after Alfred Russel Wallace, the co-discoverer of natural selection.

Wallace could see no problem solved by language that could not be solved without it. Berwick and Chomsky remind us that generative theory has been interested in these questions for more than 60 years but had little to say because the grammars of the day '... were so complex that it was clear at the time that they could not possibly be evolvable' (2).

What the grammar needed was radical simplification, a narrower language phenotype, and this is what 60 years of research has achieved with the Minimalist program (Chomsky 1995). 'We can now effectively use a "divide and conquer" strategy to carve the difficult evolutionary problem of "language" into three parts...: (1) an internal computational system that builds hierarchically structured expressions with systematic interpretations at the interfaces with two other internal systems, namely, (2) a sensorimotor system for externalization as production or parsing and (3) a conceptual system for inference, interpretation, planning and the organization of action – what is informally called "thought"' (11). Externalization includes '... aspects of language such as word-formation (morphology) and its relationships to language's sound systems (phonology and phonetics), readjustment in output to ease memory load during production, and prosody' (11). Thus 'divide and conquer' separates syntax and symbolic thought from behavior and their expression in speech.

What languages have in common, then, are parts (1) and (3), their capacity to generate thought by assembling hierarchical syntactic structures. Where they differ is largely (perhaps entirely) in Part 2, the sensorimotor system for externalization. According to the Minimalist Program, the generative mechanism of syntax is a single recursive operator, Merge. Its iterative function is to combine syntactic objects (words, phrases, clauses), forming sets of unordered 'computational atoms, word-like, but not words' (90), so as to form sentence-like thoughts. Importantly, these syntactic combinations have hierarchical structure, admitting indefinitely long phrasal and clausal embedding and structural dependencies, but no order. Linear order of words emerges from the secondary process of externalization in speech or sign. Thus, the sets of syntactic objects created by Merge are hierarchical, unordered, abstract, and amodal.

What then are the 'computational atoms' on which Merge operates? They are lexical concepts, word-like inasmuch as they are discrete units of meaning, but not words because they have no phonological structure and no physical form outside the brain. Other animals may have concepts or categories for physical, 'mind-independent' aspects of the world. Human concepts, by contrast, are 'mind-dependent'—'inward ideas' (83). Berwick and Chomsky roundly reject the '... reference relation in the sense of Frege, Peirce, Tarski, Quine, and contemporary philosophy of language and mind' (85), but they offer no alternative. On the contrary, '... the origin of human specific concepts and 'the atoms of computation' that Merge uses remains for us a mystery—as it is for other contemporary writers such as Bickerton (2014). Elsewhere, Berwick and Chomsky claim that Bickerton 'shrugs his shoulders' at the problem (149).

Setting aside the mystery of computational atoms, how did Merge itself arise? We should be clear, first of all, that natural selection is basically a sieve; it cannot give rise to evolutionary novelties, new bodily forms, new neural structures, or modes of action; all it can do is to pass on, or not pass on, whatever is presented to it.

Evolutionary novelties arise in two main ways. They arise from exposure to new conditions, as in '… the Tibetan ability to thrive at high altitudes … or the ability to digest lactose past childhood in dairy farming cultures …' (27). Second, they arise from chance mutations that happen to introduce, say, a change in the timing of regulatory action by a gene, or new cells that yield a sharp discontinuity in function, such as the light sensitive cell and its shadowing pigment cell from which the vertebrate eye evolved. Berwick and Chomsky argue that Merge arose from just such a chance discontinuity and radiated relatively rapidly.

Indeed, the paleoanthropological evidence suggests that symbolic behavior may have first appeared sometime between the first anatomically modern humans in Southern Africa about 200,000 years ago and the diaspora out of Africa about 60,000 years ago. The minor mutation that caused Merge occurred sometime in that interval in one or two members of a small *Homo sapiens* group. 'If there was no externalization [that is, if no one was yet speaking] then Merge would be … just like any other "internal" trait that boosted selective advantage internally, by means of better planning, inference, and the like' (164.) This, incidentally, is Berwick and Chomsky's solution to Wallace's problem: Merge increased mental efficiency. The intellectual gain was strong enough for Merge to radiate through the group by natural selection and to establish the core of Universal Grammar (UG) before the diaspora from Africa.

The third and final component to evolve was externalization, that is, vocal learning and speech. Here Berwick and Chomsky rest their case on the 'revival' of 'Darwin's Caruso theory' (4) by which human males began courting females, as songbirds do, by singing. Continued singing strengthened and perfected the vocal organs, leading to speech and, as the brain grew larger, to language. Recent work has indeed revealed the same genes acting in the same way in vocal learning species (songbirds, parrots, humming birds, humans), but not in nonvocal learners (doves, quails, macaques) (42). Songbirds and humans have evidently converged on the same genes and homologous circuits as their common ancestor hundreds of millions of years ago. 'In other words, the "toolkit" for building vocal learning might consist of a (conserved) package of perhaps 100-200 gene specializations … that can be "booted up" quickly – and so evolved relatively rapidly' (45). On the other hand, Berwick and Chomsky remark elsewhere, '… externalization may not have evolved at all; rather, it might have been a process of problem solving using existing cognitive capacities found in other animals' (83). Be that as it may, all this meshes neatly with the notions that (1) language evolved rapidly and (2) syntax and the conceptual system evolved before and independently of the capacity for speech.

In conclusion:

A very strong thesis, called the Strong Minimalist Thesis (SMT), is that the generative process is optimal: the principles of language are determined by efficient computation … [L]anguage is something like a snowflake, assuming its particular form by virtue of laws of nature – in this case principles of computational efficiency … Insofar as [the SMT] is correct, the evolution of language will reduce to the emergence of Merge, the evolution of conceptual atoms of the lexicon, the linkage to conceptual systems and the mode of externalization … Note that there is no room in this picture for any precursors to language – say a language-like system with only short sentences. There is no rationale for positing such a system: to go from seven-word sentences to the discrete infinity of human language requires the same recursive procedure as to go from zero to infinity, and there is of course no direct evidence for such protolanguages. Similar observations hold for language acquisition, despite appearances … (71–2).

2. Some counterarguments

We turn now to a critique of *Why Only Us*. We wish that the title had ended with a question mark, indicating a request for an answer rather than the answer itself. Nonetheless, we should say at the outset that we agree with, or at least do not dispute, much of Berwick and Chomsky's story. For one thing, we agree that language in adults may be primarily an instrument of thought, even though, in our view, thought evolved from the spoken word. For another, although (or perhaps because!) we are not syntacticians, we accept that Merge radically simplifies syntax and may achieve optimal (i.e. least effort) efficiency of computation. Here we respect the undoubted authority not only of Berwick and Chomsky themselves but also of Bickerton (2014) and Jackendoff (2002) who propose parallel alternatives. Finally, we accept the likelihood that language evolution culminated rapidly in the past 100,000 years or so, even if it began very much earlier than Berwick and Chomsky suppose.

The central flaw of the book is what the authors regard as the key to its success, their strategy of 'divide and conquer'. By separating syntax and the conceptual system from 'externalization', that is, from behavior in the form of spoken words, they separate themselves

from a plausible account of the origin of words and of how speakers come to combine them. Indeed, they implicitly acknowledge this by postulating mental precursors for spoken words, 'computational atoms, word-like, but not words', that have all the properties of words, except that no one ever speaks them. This tortuous argument stems from Berwick and Chomsky's fear of 'behavior', the *bête noire* of generative theory. Indeed, they speculate that the 'modern conception [of language as communication] … derives from lingering behaviorist tendencies which have little merit' (102).

By inventing mental computational atoms to substitute for spoken words, Berwick and Chomsky ignore two facts, central to evolutionary theory. First, the force driving all evolution by natural selection is behavior. '[C]hanges in behavior generate new selection forces which modify the structures involved. Many, if not most acquisitions of new structures in the course of evolution can be ascribed to selection forces exerted by newly acquired behaviors … Behavior thus plays an important role as the pacemaker of evolutionary change' (Mayr 1982: 612). Berwick and Chomsky do implicitly acknowledge the role of behavior in, for example, evolution of 'the Tibetan ability to thrive at high altitudes where there's less oxygen [and] the ability to digest lactose past childhood in dairy farming cultures' (26, 27), but they assign it no role in the evolution of language, and so no role for natural selection.

The behavior that set the pace for language was speech, spoken words without which combinatorial syntax, or Merge, would have had nothing to combine. Chance favors the prepared context. Neither the light sensitive cell nor Merge could have passed the natural selection sieve, had they not occurred in a behavioral context that favored them. But Berwick and Chomsky have nothing whatever to say about the behavioral or the social context in which Merge appeared.

The second fact ignored by Berwick and Chomsky is that in evolution behavior builds brains, not brains behavior. Multi-cellular organisms were behaving and thriving millions of years before their behavior brought brains into existence. Consider a few familiar examples, cited by Bickerton (2014). Surely no one supposes that bats began to feed on moths because a chance mutation granted them echolocation. Just as our blind learn to tap their paths along a sidewalk, bats who discovered flying food learned to whistle for their supper, and a suite of chance mutations that might once have vanished into what Berwick and Chomsky refer to as 'a stochastic gravity well' (22) granted them echolocation. Or think of beavers. We shall never know how their first dam came to be built, but it does not seem likely to have come from a lucky dam-building mutation. More likely is the accident of a tree, perhaps felled by their own bark-eating habits, falling across a stream, soon then blocked by branches, twigs, and silt, offering shelter from predators and, eventually, as the beavers developed the interior of the dam, dry spaces for eating, sleeping, and tending the young. Here, utterly novel architecture and a subtly thinking brain sprang from inventive behavior prompted by the contingencies of beaver life.

We could go on indefinitely with examples of behavior initiating evolutionary novelty and presumably corresponding development of supporting neural structures. But before we leave the topic, let us briefly mention sign language, cited several times by Berwick and Chomsky as evidence of the amodality of language. Given their attention to birdsong, Berwick and Chomsky evidently assume that speech was the original modality of externalization, and they term sign language an 'invention' (83). Much research over the past 40 years has demonstrated the plasticity of the brain in response to novel input. Studies of the production and comprehension of spoken and signed language have found that both forms of language engage the same areas of the brain in some linguistic tasks, but different areas in others (Emmorey et al. 2007, 2014). Such work does indeed argue for the amodality of language, but it also illustrates a new mode of behavior differentially shaping the brain.

Berwick and Chomsky's aversion to vocal behavior is further evident in their treatment of externalization. They introduce the topic with 'Darwin's Caruso theory' (4) in which speech evolved through males courting females by song. Whatever weight Darwin's arguments carried in 1871, they no longer carry today. Consider the species (e.g. crickets, frogs, bats, songbirds, marine mammals) in whom courting by sound has arisen. All these species live in visually obscure surroundings (grass, reeds, darkness, leafy trees, murky waters), so that males and females cannot easily find each other. In oscine birds the solutions were bright plumage and distinctive song in male birds. Thus, song and other forms of vocal courtship in these species probably began as localization calls. Sexual selection would have come into play as competing males elaborated their calls into rival songs, as instruments of courtship and territoriality. We have no reason to believe that early human males and females had difficulty in finding each other or in claiming territory, and so no reason to posit song as the precursor of speech. Other reasons, including the limiting of early speech to males, also argue against the theory (Fitch 2013). And, we may add, as an answer to Wallace's question, speech as courtship is peculiarly feeble.

Nonetheless, as Berwick and Chomsky explain, recent work has found interesting genetic analogies among vocal learning species. What these studies have not revealed, however, and a topic about which Berwick and Chomsky have nothing to say (although it was of some concern to Lenneberg (1967: Chapter 1) is how and why the neuroanatomy of the human vocal apparatus evolved. Berwick and Chomsky remark, citing Tattersall (1998), that the vocal tract had already taken a form adequate for speech over 500,000 years ago (64), and they cite Fitch (2010) as claiming that the hominid vocal perceptual and production apparatus was 'vocal ready', but say nothing about how or when the central neural structures and dense peripheral innervation of tongue, lips, velum, and larynx that activate and coordinate their movements might have arisen. There seems to have been little comparative work on primate vocal neuroanatomy, but Mu and Sanders (2010) report that '... the innervation of the human tongue has specializations not reported in other non-human primates. These specializations appear to allow for fine motor control of tongue shape' (777). How many generations and how many genes it took for this subtle machinery to evolve, we do not know, but it surely called for more than the songbird 'toolkit' invoked by Berwick and Chomsky. We take up this matter in Section 6.

A final error in Berwick and Chomsky's story is their solution to Wallace's problem. They evidently do not recognize that evolutionary changes respond to *specific*, not general, problems. They propose that Merge was '... just like any other internal trait that boosted selective advantage internally, by means of better planning, inference, and the like'. But this is far too broad. Planning for what? And why was Merge the solution to planning rather than, say, an increase in short-term memory? What specific problem did *H. sapiens* have that language alone could solve? That is the question that Wallace was asking and that Berwick and Chomsky do not grasp.

3. To the brink of syntax: an alternative scenario

Our alternative scenario begins where Berwick and Chomsky end, with spoken words, symbols for human concepts. We take up three disparate strands of evidence to trace speculative answers to 'questions we will never answer' (Lowentin 1998), because we agree with Berwick and Chomsky that 'even a speculative outline might lead to productive lines of inquiry' (158). In each strand, evolutionary novelties emerge from changes in behavior that alter the conditions of natural selection and help to set hominids along the path to language.

First, in 'The Ontogeny of Words', we draw on ontogeny to suggest how infant–mother relations enhance mutual understanding in which vocal sounds are associated with immediately present events: here-and-now people, animals, and events of shared interest. Next, in 'An Answer to Wallace's Question', we turn to two recent phylogenetic accounts of the origin of symbolic communication in *Homo erectus*. Hrdy (2009) describes the emergence of 'emotionally modern humans' through cooperative breeding that established trust between an infant and diverse caretakers, laying the background for a community of trust and mutual understanding among adults. Bickerton (2014) proposes that context-free concepts and the words that symbolize them had their origin in the displaced reference to absent objects and events demanded by thousands of years of confrontational scavenging. Finally, in 'Differentiation of the Vocal Apparatus', we sketch an account of how a growing vocabulary and neuroanatomical differentiation of the vocal apparatus may have co-evolved in an interactive spiral to yield the store of phonetic contrasts from which every language builds its lexicon.

4. The ontogeny of words

In 1965, Chomsky (1965) proposed what came to be known as the Language Acquisition Device (LAD) to explain how children learn language. The LAD is a hypothetical, innate component of the brain that enables a child to discover the grammar of the language she hears. Without a LAD, children would be unable to learn grammar because of the 'poverty of the stimulus' to which they are exposed. But a LAD would only be useful for children who had already learned some vocabulary. It has nothing to say about how children learn words.

If language were simply a matter of biology, we might expect an infant to produce words by maturation, just as we would expect her to walk without any external input. Indeed, walking and talking are similar in that both activities seem to develop without instruction. However, a child raised in silent isolation might very well learn to walk, but would certainly not learn to talk, because children learn to talk only in conversation with others.

Ironically, Chomsky himself saw this nearly 30 years ago. In what was perhaps an unguarded moment, he remarked that, to learn a language, an infant needs 'triggering events'.

> *. . . a stimulating loving environment in which their natural capacities will flourish. A child that is raised in an orphanage . . . may be very restricted in his abilities. In fact*, it may not learn language properly
> *(emphasis added) (Chomsky 1988)*.

How can we study that 'stimulating loving environment'? The obvious place to look is the relation between an infant and her mother. That has been a major focus of developmental psychologists during the last 40 years. They have discovered two significant modes of mother-infant interaction during the infant's first year, *before* she actually begins to speak. Like other nonhuman primates, human infants form a strong attachment to their caregivers at birth. That is evident in their tendency to cling and cry during fearful situations. But, in addition, infants have unique capacities to share their emotional and cognitive experiences with their caregivers, which are the first steps toward language. Shortly after birth, the human infant begins to develop a bond based on reciprocal expressions of affect. Toward the end of her first year, she is able to share her mother's attention to external objects. The first process is called intersubjectivity; the second, joint attention. Both processes are nonverbal and uniquely human.

4.1 Intersubjectivity

Intersubjectivity grows from an infant's physical relation with her mother. Among primates, only humans cradle their infants, not only because the mother has no body hair to which the infant can cling, but also because newborn infants are the least developed of all primates. The volume of the infant brain is approximately 25 per cent of its adult size; in chimpanzees, 45 per cent. Similarly, the human skeletal system is poorly developed. As a result, an infant cannot locomote, and has to be cradled for 6 months. An important benefit of cradling is the proximity of the infant's and the mother's eyes, allowing them to share each other's affect and gaze, one of many quirks of evolution that laid the groundwork for language. In compensation, as it were, for the infant's lack of mobility, infant and mother can observe and anticipate each other's behavior during cradling (Trevarthen 1993).

That has been shown in many experiments in which 3- to 4-month-old infants and their mothers were video recorded. In a typical study, the infant sits in a high chair across from her mother, or on her mother's lap (Murray and Trevarthen 1985). Separate video cameras record the infant's and the mother's behavior, after which the recordings are synchronized. Independent observers then analyze those recordings, at normal and slow speeds, for temporal patterns of each individual's behavior: smiling, vocalizing, moving their bodies, expressing anger, rejection, and other modes of affect (Trevarthen and Aitken 2001).

Such analyses have shown that infant and mother coordinate their affect and activities and predict each other's behavior. For example, a micro-analysis of vocalization measured the onset and offset of a mother and her infant's vocalizations and pauses (Beebe et al. 1988). On average, mother and infant matched the duration of their pauses. That is, before taking a new turn, each partner paused for a duration that roughly matched the other's most recent pause. The bi-directional contingent relation between the mother's and the infant's vocalizations prompted Beebe et al. and others to refer to such exchanges as 'proto-conversations'. That interpretation seems justified because the infant and the mother alternated their utterances, as adult speakers and listeners do in real conversations (Stern et al. 1975)

Bi-directional contingent relations between mother and infant are not restricted to vocalization, however. Other studies analyzing videotapes of face-to-face communication between infant and mother obtained significant correlations between the mother's attentiveness and the infant's smiling and cooing (Lavelli and Fogel 2013) and other expressions of affect between mother and infant (Beebe et al. 2016).

Developmental psychologists refer to the close temporal correlation between infant and mother's affect and behavior as dyadic to highlight the fact that the coordination of those events contains more information than individual analyses alone. Dyadic relations suffice until the infant begins to crawl and to explore objects in her environment. Beginning at about 6 months, triadic relations develop between the infant, her mother, and objects of mutual interest. Those relations facilitate joint attention.

4.2 Joint attention

While playing with an infant, it is commonplace for a mother to engage the infant's interest in an object by looking at that object, waiting for the infant to gaze at it, and then look back at her and smile. That sequence is an example of joint attention. It provides the first instance in which an infant and another person share the contents of their minds, in this example, knowing that each one saw a particular object. Significantly, joint attention, a nonverbal process, is evident before the infant learns her first words.

Joint attention is crucial for word learning. Consider, for example, a mother teaching her infant that the name

of the object they are playing with is a 'doll'. Without joint attention, 'doll' might refer to any other item in the room, a chair, a fan, a shoe, a dog (Premack 1986). However, once the 'common ground' of joint attention is achieved, it is easy for the child to identify the object as a 'doll' (Wilkes-Gibbs and Clark 1992).

Joint attention is more complicated than shared gaze, a phylogenetically older ability, in which two individuals simply look at the same object. To appreciate the difference, imagine that you turn your head toward a passing car and that your friend does the same. Unless you have some way of communicating what you saw, you have no way of knowing if you both saw the same thing. That is why, in the previous example, smiling after shared gaze is important. It is a nonverbal way of saying 'I saw what you see'. Thus, for joint attention to mean shared experience, it is important for one person to engage in a communicative act after the other person looks at the object in question. To indicate sharing, children often smile, or point, or literally offer the object to their caretaker (Liebal et al. 2013)

Joint attention not only facilitates the acquisition of words but also predicts the size of a child's vocabulary at 24 months and at older ages (Meltzoff and Brooks 2008). The higher the rate of joint attention at 12 months, the larger a child's subsequent vocabulary (Morales et al. 2000). Joint attention is also significant because the words a child learns are declarative and, in that sense, part of a conversation. After the mother says, 'doll', the child might reply 'doll' to indicate that she saw it. Such exchanges appear to be the only way children learn their first words.

Once a child learns to use words, the influence of joint attention may vary. How much, is an open question. In non-Western cultures (e.g., in Mozambique) higher expressive vocabularies have been reported in urban than in rural areas (Mastin and Vogt 2016). More information is needed, however, to determine if that is a reflection of differences in the measurement of joint attention and/or of estimates of vocabulary by urban and rural mothers. But possible differences in expressive vocabularies should not detract from the contribution of joint attention to a child's discovery that objects, people, and actions have names and that names are used conversationally. That discovery is crucial for the development of language.

4.3 From intersubjectivity and joint attention to words

Berwick and Chomsky's disregard of intersubjectivity and joint attention follows directly from their lack of concern for the nature and the origin of words. To be sure, they postulate 'computational atoms, word-like, but not words' as units of thought, the neural precursors of words, but they have nothing to say about the nature of words themselves. For example, they argue that, 'words are radically different from anything in animal communication systems' (90), yet they refer to 'words' being used in an experiment on ape language (148ff).[1] Certainly, they explain why Nim's 'two-*word* utterances' are not syntactical (148, emphasis added) but they ignore the fact that none of Nim's utterances even qualified as words in the first place. Nim's utterances were always imperative, never declarative (Terrace 1985, 2013). Declarations imply a conversational use of language, a topic that Berwick and Chomsky avoid completely. Imperatives are uni-directional, require no response and form a minuscule fraction of a child's vocabulary. Indeed, language would never have evolved if children learned only to use words as imperatives.

Declaratives are bi-directional and typically occur in conversation between a speaker and a listener who take turns talking. All languages are conversational and children could not learn language without conversation. Why do Berwick and Chomsky avoid the topic? Presumably because, as observed earlier, generative theory systematically excludes behavior from a role in the evolution of language.

5. An answer to Wallace's question

Nearly 150 years ago, Wallace wondered why a human has ' . . . a large and well developed brain quite disproportionate to his actual requirement' (Wallace 1870: 342). He could see no problem solved by language that could not be solved without it, that is, no problem to which natural selection might have picked a solution leading to language. That concern conflicted with his belief that 'all nature can be explained' by the principles of natural selection (131).

In his book, *More Than Nature Needs*, Derek Bickerton (2014) proposes a frankly speculative answer to Wallace's question by turning not to apes, but to man's hominid ancestors. Although Berwick and Chomsky refer to *More Than Nature Needs*, the only book, so far as we know, to have directly addressed Wallace's question, they appear to have ignored much

1 Berwick and Chomsky refer to Project Nim, as a 'well-known attempt to "teach" a chimpanzee (to use) sign language' that was performed by 'researchers at Columbia'. But they do not cite the actual authors of the study (Terrace et al. 1979; Terrace, 1979).

of its contents. Berwick and Chomsky not only dismiss Bickerton's idea of protolanguage but they mention neither his theory of concepts nor his solution to the problem of their origin. Instead, they observe that the origin of concepts '. . . is entirely obscure, posing a very serious problem for the evolution of human cognitive capacities, language in particular' (90), and that, 'Words and concepts appear to be similar . . . (and) . . . seem unique to human language and thought and have to be accounted for somehow in the study of their evolution. How, no one has any idea' (86).

Bickerton, however, does have an idea, an idea that not only answers Wallace's question, but also accounts for the origin of human concepts and words. Let us be clear at the outset where Berwick and Chomsky's 'mystery' lies. The problem, first described by Saussure (1916/1959), is that words symbolize not objects in the world, but our mental representations, or concepts, of those objects. How did such verbal concepts arise? We can this question by recalling the design feature of language listed by Hockett and Altmann (1968) as 'displacement', or 'displaced reference', and defined as 'the ability to talk about things that are not physically present'. Such an ability requires the speaker–listener to have a mental representation, or concept, of the object talked about. An evolutionary account of the origin of such concepts raises the question: What were the ecological conditions among early prehumans that might have required them to communicate about things that were not physically present?

To address this question, Bickerton conjures up the nutritional needs, ecological conditions, and cognitive capacities of *H. erectus*, a species that evolved about 2 million years ago. As compared to *Homo habilis*, its immediate ancestor, the brain size of *H. erectus* was significantly larger. Because of extreme climate change, *H. erectus* lived in a relatively dry environment of open grassland. Instead of obtaining fruit trees, food had to be obtained from fauna (antelope, zebra, deer) and megafauna (elephant, rhinoceros, hippopotamus).

Like other animals, *H. erectus* presumably spent most of their waking hours looking for food, eating it, or resting after eating it, so that, apart from predators, food was the most likely referent of their communicative acts. To satisfy the caloric needs of its larger brain, *H. erectus* required meat as its primary source of food. Although they had tools to butcher dead fauna, *H. erectus* had no weapons to kill them. Instead, they first had to find dead animals and then recruit absent helpers to help in the butchering and in warding off other scavengers. Such 'confrontational scavenging' became essential to the survival of *H. erectus* and its growing brain.

Bickerton's theory has two parts; one factual, the other conjectural. How do we know that *H. erectus* engaged in scavenging? Fossil bones of megafauna display two types of clue: bite marks from animals that defleshed them and cut marks from hominid tools that cut through their thick hides. Before 2 million years ago, cut marks lie above bite marks, indicating hominid access to those bones only after other animals had scavenged them. After 2 million years ago, bite marks lie above cut marks, indicating that hominids had first access to the bones (Dominguez-Rodrigo et al. 2005). Since many of the animals that hominids butchered were too large to have been hunted, the conclusion that *H. erectus* scavenged is inescapable.

How did a scout, having located a dead animal, recruit distant followers to fend off rival scavengers and help in butchery? Bickerton proposes that *H. erectus*, thanks to newly formed cooperative habits of that species (see below), was already using sounds and/or gestures (but not articulate speech) to refer to objects that were physically present, and perhaps even to footprints or droppings of absent animals. But to recruit followers, the scout had to communicate the nature of the carcass and its location, and that could be done only by displaced reference.

The first 'words', or units of semantic communication, then, necessarily referred to mental entities, representations of absent objects. The form of such communication is a matter of speculation: perhaps mimetic gestures or sounds imitating the nature of the carcass to be scavenged, its location, and the nature of rival scavengers. The vocal modality would have come to prevail, leaving hands and eyes free to go about their more important functions. Transformation from hominid calls and cries to articulate speech may have taken as many as hundreds of thousands of years (Section 6), but proto-words of some form would gradually have come into use. They would have been strung, like beads on a string, relying on the pragmatics of the situation to make sense.

That had two consequences. Proto-words used repeatedly in combination with other proto-words acquired lexical status creating pressure for syntax. As Bickerton states, 'from an evolutionary perspective, it seems obvious that words came first but had only a small subset of the properties of modern words, that their arrival precipitated syntax, and that their subsequent interactions with syntax built the set of modern properties' (105).

For confrontational scavenging to work, *H. erectus* had to engage in an unprecedented degree of cooperation. Bickerton does not comment on the source of

such cooperation but, as Hrdy (2009) has observed, such cooperation was a consequence of the intersubjectivity instilled by cooperative breeding, as practiced by *H. erectus*, the only early hominid to engage in that practice. Compared to apes, whose mothers never allow others to care for their young, infants in species that engage in cooperative breeding are cared for and provisioned not only by their mothers but by other members of their group (alloparents). For *H. erectus*, trust in an alloparent's benevolence was a consequence of their contribution to child rearing. A mother would never engage in cooperative breeding and share care of her offspring unless she trusted members of her group.

How did mutual trust evolve among adults of species that used collective breeding, that is, how did *H. erectus* become 'emotionally modern humans'? According to Hrdy, an infant had to learn to share affect not only with her mother but also with other caretakers. Infants that succeeded would obtain more attention from their caretaker than those that did not. The benefit of more attention increased the likelihood that those infants would survive and that, as adults, they would tend to trust their fellows and understand them.

To return to Wallace's question, we propose the following antecedents of language in *H. erectus*: intersubjectivity, joint attention, and conversational communication with arbitrary words. As those antecedents took root, the competitive world of apes in which communication was based on a small number of innate and involuntary utterances, was supplemented by one in which *H. erectus* lived more cooperatively and in which they began to exchange words voluntarily.

Intersubjectivity advanced the emotional development of human infants beyond apes in two important ways. For the first time, an infant learned to care about her affective relation with her mother and, as noted earlier, she also began to engage in proto-conversation. Joint attention is significant in being the first instance in which an infant shares a mental state with another person.

Neither intersubjectivity nor joint attention was sufficient for the development of language. Even though infants engaged in proto-conversations with their mothers, there is no reason to assume that language would follow. The same is true of joint attention. But once joint attention took root, adults would likely converse about mutually interesting events in their environments.

At present, it is not possible to confirm when in our prehistory the sequence of presyntactic stages of language we have suggested occurred. However, given the universal sequence of these stages in human infants, they likely occurred in the order we described. It is difficult to imagine the occurrence of joint attention without a foundation of intersubjectivity, or of vocabulary development without the foundation of joint attention.

Displaced reference answers Wallace's question by showing that language supported confrontational scavenging by referring to objects or events not immediately present to the senses. Only language could do that. Thus, *H. erectus* and their descendants possessed the key to language and thought: mental representations, or proto-concepts, that came to be independent units of thought, free of time and space, and voluntarily accessible.

6. Differentiation of the vocal apparatus

Perhaps the single most implausible remark in Berwick and Chomsky's book is: '... externalization may not have evolved at all; rather it might have been a process of problem solving using existing cognitive capacities found in other animals' (83).

Berwick and Chomsky not only fail to cite these capacities but also do not consider how their remark is scarcely compatible with the facts. To begin with, as one of us (M.S.-K.) has written elsewhere:

In English we readily produce and comfortably understand speech [between pauses] at a rate of 120-180 words/minute or 10-15 phonetic segments/second. (Readers may want to check these numbers by reading a text out loud at a brisk rate for a minute.) If we break the segments down into discrete movements of lips, tongue, velum and larynx, we arrive at a rate of some 15-20 movements/second. By way of comparison, a violinist's tremolo may reach 16Hz and a hummingbird can beat its wings at over 70 Hz. But these are identical repetitive movements of a single organ. Speech, by contrast, engages half a dozen organs (lips, tongue blade/body/root, velum, larynx) in as many different combinations as there are different phonetic segments in the speech stream, all nicely executed within a tolerance of millimeters and milliseconds. In fact, it is precisely the [perfectly coordinated] distribution of action over different articulators that makes the high rate of speech possible (Studdert-Kennedy 2005: 55). For fuller discussion, see Lenneberg (1967: Chapter 3).

What we have then in speech is perhaps the fastest, most highly differentiated and most precise form of sustained motor action in the animal kingdom. That the system emerged by problem solving using existing cognitive capacities found in other animals does not seem likely. Rather, it would seem to have evolved under

pressure from short-term memory constraints on computation in speaking and listening. Notice that we are not dealing with action guided by the environment as in the rapid zig-zagging of a downhill skier or of a warbler threading through the branches and leaves of an orchard to land on a twig. Speech is the *self-generated*, coordinated action of an integral system of several more or less independently moving parts.

Critical to this development was the capacity for vocal imitation, unique among primates to humans. This capacity may itself have evolved from an earlier capacity for facial imitation, also unique to humans and also calling for the integrated action of several moving parts: lips, cheeks, nostrils, eyelids, and brow. Andrew Meltzoff and his colleagues have built a solid body of evidence for infant facial imitation, starting within 72 hours of birth (Meltzoff and Moore 1997). If we accept Merlin Donald's (1991) long and persuasive argument for a culture in *H. erectus* and later hominids in which individuals communicated by mimesis (i.e. by enacting or reenacting actions, events, and feelings), voluntary control, and imitation of facial expression would have been an important component of mimetic representation.

Darwin (1872/1998: 96) observed the effect of facial expression on the quality of vocalizations. Lenneberg (1967) reports on muscles of the human face, lips, and mouth that

'… have a decisive influence upon speech sounds' (34). We know that changes in position of lips, jaw, and teeth in rhesus monkeys affect the spectral structure of vocalizations, as in humans (Hauser et al. 1993). And we know that mirror neurons are activated by communicative mouth actions in macaque monkeys (Ferrari et al. 2003) and, in all likelihood, by speech in humans (Fadiga et al. 2002). Thus, we have a plausible evolutionary route from facial to vocal imitation.

Here we assume that vocal imitation and words evolved together, as neo-Darwinism would assume, by repeated innovative acts, each modestly extending the conditions of selection and gradually building the language niche, as Bickerton (2014) argues. The first referential utterances would have associated whatever grunts, trills, and hoots the hominid vocal tract afforded with some 'mind-independent' object or happening. Eventually, searching for new sounds, hominids would have come upon the syllable or syllable string, formed by rapidly opening and closing the mouth while activating the larynx. Once the habit of vocal reference was established, the search for new sounds and patterns of sound would have entered an interactive evolutionary spiral with differentiation of the vocal tract into its quasi-independent articulators. Berwick and Chomsky tell us that: 'All human languages draw from a fixed, finite inventory, a basic set of articulatory gestures' (55). We propose that this interactive spiral was the source of these gestures.

As the vocabulary of proto words grew, finite articulations forced reuse of gestures and gestural patterns in different proto words. From these patterns, constantly recurring in more and more different contexts, there emerged, to facilitate rapid motor access, the cohesive patterns of gesture that we term segments, the meaningless phonetic segments (phonemes) afforded by what Carré et al. (2017) claim to be our optimally adapted vocal apparatus (Lindblom 1998; de Boer 2005; Oudeyer 2006). Notice here the parallel with the emergence of concepts and words, postulated above. Repeated use of an initially context-bound unit in many different contexts, whether semantic or phonetic, ultimately sets it free for independent, context-free use. Thus, concepts and phonemes may have arisen by the same evolutionary mechanism.

The new capacity for phonetic imitation had further consequences for language evolution. Delay of imitation beyond its original occasion of use would have given rise to short- and long-term phonetic memory. Long-term phonetic memory would have supported displaced reference and the eventual proliferation of words. Both short- and long-term phonetic memory were prerequisite for syntax. Long-term memory was necessary to formulate and understand a syntactically organized utterance. Short-term memory was necessary in speaking to hold upcoming words in premotor store, in listening, to hold words without commitment to meaning, while computing their syntactic relations (cf. Studdert-Kennedy 2000) Thus, the capacity to speak, remember, and repeat words ramified through the hominid mind to undergird the emergence of syntax.

7. Conclusions

Despite its subtitle, *Language and Evolution*, this book has little to say about evolution. Language is a dynamic process of constant behavioral change, both biological and cultural, but Berwick and Chomsky have nothing to say about behavior. By dismissing behavior, they also dismiss natural selection. Confronted with the origin of human concepts and words, they do not shrug their shoulders, as they falsely accuse Bickerton of doing (149). They simply throw up their hands and declare it a 'mystery'. In place of words they postulate computational atoms, static neural entities that no one ever speaks, and to account for syntax, they posit a fluke, a chance mutation in a few individuals living in a social vacuum.

In our alternative scenario, language begins with the conversational exchange of arbitrary words. We have proposed a framework for the origin of words by natural selection that incorporates recent advances in our understanding of their ontogeny and phylogeny (Terrace and Studdert-Kennedy 2015). Starting with Hrdy's concept of 'emotionally modern humans', we have shown how and why they became more cooperative and attentive to each other's needs. Bickerton's protolanguage then showed how the voluntary social factors needed to converse about immediately perceived events led to the development of communication about displaced referents. The steady elaboration of communication about displaced referents led to emergence of context-free concepts, and so, ultimately, of syntactic structures, encompassing our entire world of language and thought. These antecedents of words are necessary components of a theory of the evolution of language, topics that Berwick and Chomsky avoid in their exclusive focus on syntax.

Funding

Preparation of this manuscript was supported by NIH grant MH081153-06.

Acknowledgements

We thank Beatrice Beebe, Michael Lewis, Björn Lindblom, Katherine Nelson, Kathrin Perutz, Robert Remez, Ann Senghas and Charles Yang for their helpful comments.

References

Beebe, B. et al. (1988) 'Vocal Congruence in Mother-infant Play', *Journal of Psycholinguistic Research*, 17: 245–59.

—— et al. (2016) 'A Systems View of Mother-Infant Face-to-Face Communication', *Developmental Psychology*, 52: 556–71.

Berwick, R. C. and Chomsky, N. (2016) *Why Only us*. Cambridge, MA: MIT Press.

Bickerton, D. (2014) *More Than Nature Needs: Language, Mind, and Evolution*. Cambridge, MA: Harvard University Press.

Carré, R. et al. (2017) *Speech Dynamics and Modeling*. The Hague: De Gruyter, in press.

Chomsky, N. (1965) *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N (1988) *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, MA: MIT Press.

Chomsky, N. (1995) *The Minimalist Program*. Cambridge, MA: The MIT Press.

—— (2010) 'Some Simple Evo devo Theses: How True Might They be for Language?' In: Larson, R. K. et al. (eds) *The Evolution of Human Language*, pp. 45–62. Cambridge: Cambridge University Press.

Darwin, C. (1872/1965) *The Expression of Emotions in Man and Animals*. London, UK: J. Murray.

de Boer, B. (2005) 'Evolution of Speech and Its Acquisition', *Adaptive Behavior*, 13: 281–92.

Dominguez-Rodrigo, M. et al. (2005) 'Cutmarked Bones from Pliocene Archeological Sites at Gona, Afar, Ethiopia', *Journal of Human Evolution*, 48: 109–21.

Donald, M. (1991) *Origins of the Modern Mind*. Cambridge, MA: Harvard University Press.

Emmorey, K. et al. (2007) 'Amodal Aspects of Linguistic Design', *Neuroimage*, 36: 202–8.

—— et al. (2014) 'How Sensory-motor Systems Impact the Neural Organization for Language: Direct Contrasts between Spoken and Signed Language', *Frontiers in Psychology*, 5: 1–13.

Fadiga, L., Craighero, L., Buccino, G. and Rizzolatti, G. (2002) 'Speech listening specifically modulates the excitability of tongue muscles: a TMS study', *European Journal of Neuroscience*, 15(2): 399–402.

Ferrari, F., Gallese, V., Rizzolatti, G. and Fogassi, L. (2003) 'Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex'. *European Joiurnal of Neuroscience*, 17(8), 1703–1714.

Fitch, W. T. (2010) *The Evolution of Language*. Cambridge: Cambridge University Press.

—— (2013) 'Musical Protolanguage: Darwin's Theory of Language Evolution'. In: Bolhuis, J. and Everaett, M. (eds) *Birdsong, Speech and Language*, pp. 489–503. Cambridge, MA: MIT Press.

Hauser, M. D. et al. (1993) 'The Role of Articulation in the Production of Rhesus Monkey, Macaca Mulatta, Vocalizations', *Animal Behavior*, 45: 423–33.

Hockett, C. F. and Altmann, S. (1968) 'A note on design features', *Animal Communication: Techniques Of Study and Results of Research*, T. A. Seboek. Bloomington, IN, Indiana University Press: 61–72.

Hrdy, S. B. (2009) *Mothers and Others: The Evolutionary Origins of Mutual Understanding*. Cambridge, MA: Belknap Press of Harvard University Press.

Jackendoff, R. (2002) *Foundations of Language*. New York: Oxford University Press.

Lavelli, M. and Fogel, A. (2013) 'Interdyad Differences in Early Mother–Infant Face-to-Face Communication: Real-Time Dynamics and Developmental Pathways', *Developmental Psychology*, 49/12: 2257–771.

Lenneberg, E. H. (1967) *Biological Foundations of Language*. New York, NY: John Wiley.

Liebal, K. et al. (2013) 'Young Children's Understanding of Cultural Common Ground', *British Journal of Developmental Psychology*, 31: 88–96.

Lindblom, B. (1998) 'Systemic Constraints and Adaptive Change in the Formation of Sound Structure'. In: Hurford, J. R. et al. (eds) *Approaches to the Evolution of Language:*

Social and Cognitive Bases. Cambridge: Cambridge University Press.

Lowentin, R. (1998) 'The Evolution of Cognition: Questions We will Never Answer'. In: Scarborough, D. and Sternberg, S. (eds) *An Invitation to Cognitive Science, Volume 4: Methods, Models, and Conceptual Issues*. Cambridge, MA: MIT Press.

Mastin, J. D. and Vogt, B. A. (2016) 'Infant Engagement and Early Vocabulary Development: A Naturalistic Observation Study of Mozambican Infants from 1;1 to 2;1', *Journal of Child Language*, 43: 235–64.

Mayr, E. (1982) *The Growth of Biological Thought*. Cambridge, MA: Harvard, The Belknap Press.

Meltzoff, A. N. and Brooks, R. (2008) 'Self-experience as a Mechanism for Learning about Others: A Training Study in Social Cognition', *Developmental Psychology*, 44: 1257–65.

—— and Moore, M. K. (1997) 'Explaining Facial Imitation: A Theoretical Model', *Early Development and Parenting*, 6: 179–92.

Morales, M. et al. (2000) 'Responding to Joint Attention Across the 6- Through 24-month Age Period and Early Language Acquisition', *Journal of Applied Developmental Psychology*, 21: 283–98.

Mu, L. and Sanders, I. (2010) 'Human Tongue Neuroanatomy: Nerve Supply and Motor Endplates', *Clinical Anatomy*, 7: 771–91.

Murray, L. and Trevarthen, C. (1985) 'Emotional Regulation of Interactions Between Two-month-olds and Their Mothers'. In: Field, T. M. and Fox, N. A. (eds) *Social Perception in Infants*, pp. 177–97. Norwodd, NJ: Ablex Publishers.

Oudeyer, P. Y. (2006) *Self-organization in the Evolution of Speech*. New York, NY: Oxford University Press.

Premack, D. (1986) *Gavagai*. Cambridge, MA: The MIT Press.

Stern, D. N. et al. (1975) 'Vocalizing in Unison and in Alternation: Two Modes of Communication Within the Mother-Infant Dyad', *Annals of the New York Academy of Sciences*, 263: 89–100.

Saussure, F. (1916/1959) *Course in general linguistics*. London, Duckworth.

Studdert-Kennedy, M. (2000) 'Evolutionary Implications of the Particulate Principle: Imitation and the Dissociation of Phonetic Form from Semantic Form'. In Knight, C. et al. (eds) *The Evolutionary Emergence of Language*, pp. 161–76. Cambridge, UK: Cambridge University Press.

—— (2005) 'How Did Language go Discrete?' In: Tallerman, M. (ed) *Language Origins: Perspectives on Evolution*. New York, NY: Oxford University Press.

Tattersall, I. (1998) 'The Origin of the Human Capacity', *James Arthur Lecture Series*, 68: 1–27.

Terrace, H. (1985) 'In the Beginning was the Name', *American Psychologist*, 40: 1011–28.

—— (2013) 'Becoming Human: Why Two Minds are Better than One'. *Agency and joint attention*, J. Metcalfe and H. Terrace. New York, Oxford University Press: 11–48.

—— et al. (1979) 'Can an Ape Create a Sentence? ', *Science*, 206: 891–902.

Terrace, H. S. (1979) *Nim*. New York, NY: A. Knopf.

—— and Studdert-Kennedy, M. (2015) 'The mystery of language evolution'. *Language Log*. M. Liberman. Phildelphia, PA, University of Pennsylvania.

Trevarthen, C. (1993) 'The Self born in Intersubjectivity: The Psychology of an Infant Communicating'. In: Neisser, U. (ed.) *The Perceived Self: Ecological and Interpersonal Sources of Self-Knowledge*, pp. 121–73. New York: Cambridge University Press.

—— and Aitken, K. J. (2001) 'Infant Intersubjectivity: Research, Theory, Clinical Applications', *Journal of Child Psychology and Psychiatry*, 42: 3–48.

Wallace, A. R. (1870) *Contributions to the Theory of Natural Selection: A Series of Essays*, 2nd edn. London: Macmillan and Company.

Wilkes-Gibbs, D. and Clark, H. H. (1992) 'Coordinating Beliefs in Conversation', *Journal of Memory & Language*, 31: 183–94.