Biol Philos (2017) 32:779–796
https://doi.org/10.1007/s10539-017-9606-y

How language *couldn't* have evolved: a critical examination of Berwick and Chomsky's theory of language evolution

Ronald J. Planer[1]

Received: 11 June 2017 / Accepted: 22 November 2017 / Published online: 29 November 2017

Abstract This article examines some recent work by Berwick and Chomsky as presented in their book *Why Only Us? Language and Evolution* (2015). As I understand them, Berwick and Chomsky's overarching purpose is to explain how human language could have arisen in so short an evolutionary period. After articulating their strategy, I argue that they fall far short of reaching this goal. A co-evolutionary scenario linking the mechanisms that realize the language system, both with one other and with cognitive mechanisms capable of exploiting linguistic expressions, is surely unavoidable. And yet this is precisely what Berwick and Chomsky in effect rule out.

Keywords Evolution of language · Symbolic behavior · Rapid evolution · Syntax · Merge · Hierarchical structure · Massive modularity

Introduction

Can we ever know how human language evolved? There is certainly no shortage of skeptics, but an increasing number of theorists are more optimistic. And with good cause: we now possess a range of plausible hypotheses which make contact with empirical evidence in one way or another. This is not to say that there is wide-reaching consensus in the area of language evolution, however; far from it. Indeed, theorists often conceive of their evolutionary target—human language—in importantly different ways. But this is to be expected of any scientific field still in its infancy.

✉ Ronald J. Planer
Ronald.Planer@anu.edu.au

[1] School of Philosophy, The Australian National University, Canberra, Australia

The purpose of this article is to critically examine some recent work on the evolution of language by Berwick and Chomsky (henceforth "B&C") as set out in their book, *Why Only Us? Language and Evolution* (2015).[1] B&C's treatment of this topic is rich in detail, bringing together ideas from fields as diverse as archeology, evolutionary developmental biology, and computer science, in addition to the various branches of linguistics proper. This is a very welcome contribution as many theorists have felt that the views of Chomsky and his followers in this area have been underspecified.

As I will explain, the real puzzle of language evolution in B&C's view is how language could have evolved in so short a period of time. B&C's reasons for thinking language evolved in this way will be set out below; for now, suffice it to say that they think "symbolic behavior" is the only reliable indicator of language and that the archeological record tells us such behavior evolved only very recently in our line.

Borrowing some terminology from philosophy, we can think of B&C's main objective as that of providing a good "how-possibly" explanation (Resnik 1991) for the sudden appearance of language. A how-possibly explanation for some phenomenon is an account of how that thing could have occurred which remains speculative in its empirical details but exemplifies other explanatory virtues such as completeness, coherence, fruitfulness, testability and so forth. Like explanations more generally, how-possibly explanations can be better or worse; a good how-possibly explanation will exemplify these other explanatory virtues to a high degree; a bad one, to a low degree.

As I understand B&C, their strategy involves two main steps. The first of these is to argue that the mechanism underlying human language syntax could have arisen in a single evolutionary event. The second is to motivate the idea that all, or virtually all, of the machinery for implementing the rest of the human language system could have been present in the human brain before the evolution of syntax. I say "virtually all" here because I read B&C as wanting to allow that there may have been a very small number of minor changes to our language system following the emergence of syntax. Thus, with the sudden appearance of syntax came human language, or at least a version of language requiring no more than a few "finishing touches".

I am skeptical of both steps, but I will here limit my criticism to the second. I will argue that, even if we grant B&C's quite controversial way of thinking about language, it is extremely doubtful that virtually everything else needed for language might have already been in place. Or put another way, I will argue that there are key pieces of neurocomputational machinery needed for the realization of language whose evolution prior to Merge would be deeply mysterious. If I am correct, then B&C's how-possibly explanation for the sudden evolution of language is far from a good one to say the least.

A quick caveat before getting started: I take B&C's characterization of the human language system to provide us with a functional analysis of that system (Cummins 1983). This form of analysis involves taking a complex capacity and breaking it down into a number of related sub-capacities each of which is simpler

[1] All page references for B&C will refer to this book unless otherwise noted.

than the capacity being analyzed. These sub-capacities are then conceived of as properties of sub-systems making up the larger containing system. Functional analysis might in turn be applied to these sub-capacities until we reach a level of specificity appropriate to our current theoretical aims. When the capacity being analyzed is a high-level one, functional analysis remains largely neutral on issues of physical realization while at the same time making definite claims about internal structure or organization.

I thus take functional analysis to be an importantly different exercise from that of simply describing some capacity of a target system in an abstract way. I raise this contrast because Chomsky and his followers sometimes talk as though the latter is all they are interested in doing; they are merely interested in providing an abstract characterization of what our (linguistic) brains are capable of which is wholly non-committal about the internal structure or organization of the brain. I think there is plenty in B&C's work to suggest that this cannot be all they are up to; as will become clear below, they think, for example, that it makes perfect sense to ask where in the brain the mechanism of syntax is located and how it is physically realized. But more significantly, it is very hard to see how B&C's characterization of language could have any bearing on evolutionary questions were it to say nothing about how the brain is internally structured (Ippedico, in preparation). For why should we prefer their characterization to some other one when thinking about the evolution of language if both provide equally accurate descriptions of our linguistic capacities? The fact that theirs might be "simpler" is irrelevant if it is also the less faithful of the two in terms of internal structure.

Language and the language system

B&C conceive of language as an infinite set of hierarchically-structured mental expressions each of which has a definite semantic and sensorimotor interpretation. Let me explain.

It is obvious what it means to say that this set of expressions is infinite: it has no upper bound. And the implications of this feature are also easily grasped; since the human brain is finite, this set must be generated using recursion of some kind. The idea of hierarchical structure is more complicated. By this, B&C have in mind that the constituents of a given linguistic expression are ordered by something like a *structural distance* relation. One—perhaps the main—reason to accept the psychological reality of such structure is that it would seem indispensable to systematizing and predicting our semantic intuitions. Consider, for example, the sentence *instinctively, birds that fly swim* (an example of B&C's). English speakers take this sentence to mean that birds that fly are instinctive swimmers, this despite the fact that the word *instinctively* is actually closer to *fly* than it is to *swim* in terms of linear distance. Moreover, this pattern generalizes: *instinctively* will be interpreted as modifying *swim* in the sentence, *instinctively, birds that fly and sing swim*, in the sentence, *instinctively*, *birds that fly and swim and lay eggs swim*, and so on, implying that *instinctively* and *swim* can be separated by an arbitrary linear distance. Hence, in order to capture this regularity, it seems that we must invoke some notion of structural distance, of the number of hierarchical levels that separate two or more constituents.[2]

Implicit in this explanation of our semantic intuitions is the idea that hierarchical structure affects the semantic interpretation of linguistic expressions. How is this supposed to go, exactly? In getting clear on this question, it will help to first set out an answer which B&C do not (and arguably cannot) appeal to. This is that semantic interpretation consists in the production of a particular thought couched in an independent and antecedent linguistic system, one we can denote as a "language* of thought". This is perhaps the orthodox view within linguistics and philosophy. Proponents of such a view have an easy answer to the above question: they can say that hierarchical structure affects semantic interpretation simply by influencing the thought that the linguistic expression produces within one's language* of thought.

This is not at all how B&C conceive of matters. Rather, for them, linguistic expressions *are* thoughts; when I think that the weather will be pleasant tomorrow it is in virtue of tokening a particular linguistic expression that represents the world in this way, *not* in virtue of tokening a sentence in a language* of thought with this content. It is for this reason that B&C often describe language in their sense as our "language of thought" (p. 87). Indeed, they go so far as to say that a language* of thought is "unnecessary" and "quite obscure" (p. 172). This view makes sense given B&C's other theoretical commitments. In particular, as we will see below, B&C hold that language was initially selected for due to its (profoundly) beneficial effects on cognition, not communication. Such a selective story would be difficult to square with the idea that linguistic expressions depend for their meaning on the production of sentences in a language* of thought. For if the meaningfulness of language is parasitic on the existence of an equally-rich language of thought*, then it is hard to see what the (direct cognitive) benefit of language in B&C's sense would be.

Where does this leave us? B&C do not offer a worked-out account of semantic interpretation nor of how hierarchical structure can affect it. This is to some extent understandable; questions pertaining to semantics are among the most difficult ones linguists and philosophers continue to face. But it is unfortunate, nonetheless, as it makes core parts of their overall theory difficult to evaluate. For the purposes of this article, I propose to understand semantic interpretation in something like the following way: the semantic interpretation of a complex linguistic expression is determined by some portion of its computational role within the human conceptual system ("conceptual system" is B&C's terminology for the "system of thought" [p. 66]). The qualifiers "complex" and "some portion of" are important here; the former allows that the semantic interpretation of simple expressions (lexical concepts) might be fixed in some other way, while the latter allows that only a select subset of the expression's computational effects is determinative of the expression's semantic interpretation (one might, for example, maintain that it is only those inferences which hold in virtue of the expression's constituent structure and the

[2] Here the idea would be that INSTINCTIVELY modifies SWIM in each of the corresponding hierarchical mental expressions because it is closer to SWIM in terms of structural distance than it is to any other constituent it might modify. (Here and below I follow the convention of using small caps to name concepts.)

mode of composition of those constituents that are relevant). We can then say that hierarchical structure affects semantic interpretation by causally influencing that part of the expression's computational role within the conceptual system that is determinative of the expression's semantic interpretation. The satisfaction of this condition implies computational machinery that is sensitive to hierarchical structure and is capable of reacting differently to different forms of hierarchical structure. Such machinery would allow two expressions—say, [MAX [LEFT [MARY ALONE]]] and [MAX [[LEFT MARY] ALONE]],[3] to differ in their semantic interpretation despite being composed out of the very same parts.

So much, then, for semantic interpretation; what about the sensorimotor side of things? We can again take interpretation to be a matter of computational role, though this time within the human sensorimotor system. More precisely, we can say that the sensorimotor interpretation of a linguistic expression is determined by (possibly some subset of) that expression's effects on motor planning and behavior. Hierarchical structure is relevant here only insofar as it potentially shapes how we externalize a linguistic expression. Unlike in the case of semantic interpretation, where each hierarchically-structured linguistic expression maps to a unique semantic interpretation, more than one hierarchically-structured linguistic expression can map to the same sensorimotor interpretation, ultimately giving rise to ambiguous utterances. (There are no ambiguous thoughts, by contrast.)

B&C think of language in this sense as the product of a certain computational system installed in the human mind-brain; hence, for them, the problem of explaining the evolution of language reduces to the problem of explaining how this computational system evolved. They conceive of this system as being composed out of the following components: (1) a lexicon; (2) the computational operation Merge; (3) a workspace; (4) an interface with the conceptual system; and finally (5) an interface with the sensorimotor system.

The lexicon consists of the individual's stock of "word-like atomic elements" (p. 40). According to B&C, these concepts are fundamentally distinct from the type of concepts possessed by other animals in that the former do not depend for their meaning on correlations with external states of the world. They write:

> What we understand to be a person, a tree, water, and so on, consistently turns out to be the creation of what seventeenth-century investigators called the "cognoscitive powers," which provide us with rich mechanisms to refer to the outside world from intricate perspectives" (p. 85).

These concepts form the raw material for the computational operation Merge, the "star" of B&C's theory of language evolution. Merge takes any two linguistic expressions X and Y, where these can be either simple or complex, and combines them to form a third, complex expression of the form G = [X Y]. This new expression is stored at a workspace location where it remains computationally accessible to Merge. Because Merge can also read from this workspace at a later time, Merge can be composed with itself an arbitrary number of times—it can be

[3] I will use square brackets to illustrate hierarchical structure; "[X [Y Z]]" represents an expression in which X is superordinate to Y and Z.

recursively applied—thereby generating an infinity of hierarchically-structured linguistic expressions.[4]

This leaves the two interfaces. The interface with the conceptual system "relates hierarchical structures built by Merge to systems of reasoning, inference, planning, and the like" (p. 140), that is, to cognitive mechanisms. B&C do not explicitly say as much, but it is obvious that this interface must also relate conceptual-system expressions and/or processes with Merge; were this not the case, it is unclear how linguistic expressions relevant to ongoing cognition and communication could be constructed. The interface with the sensorimotor system relates these same expressions to systems involved in vocal and/or motor production, "thus externalizing internal computations and thought" (p. 66). Again, while B&C do not acknowledge the point, it is clear that this interface must also enable influence to run in the other direction; language comprehension would not be possible otherwise.

The evolution of language

For B&C, the main challenge we face in connection with language evolution is to explain how language could have evolved in a so rapid a manner. According to them, the communication systems of other animals are "radically different from human language in structure and function" (p. 63). Moreover, they think the same is true with respect to the communication systems of extinct non-human hominins. Hence, human language is without a precursor on their view.

B&C thus take it that human language evolved only recently, sometime after the division of our line into *Homo Sapiens* and Neanderthals ($\sim$ 400–500 kya [Green et al. 2006]). How recently? According to B&C, the only persuasive evidence we have for the possession of language is human "symbolic behavior"—the production of personal ornaments, abstract and figurative art, music, and so on. They think that the first unambiguous signs of such behavior are very recent indeed; here they cite the shell beads and engraved pieces of ochre dated to $\sim$ 80 kya from Blombos Cave in South Africa (d'Errico et al. 2005). Moreover, they do not see any precursor to our symbolic capacities before this time in the archeological record. Following archeologist Ian Tattersall, they see this as just another example of a more general tendency in human cognitive and technological evolution. B&C tell us, "… we do not see any kind of 'gradualism' in new tool technologies or innovations like fire, shelters, or figurative art," but rather a "pattern of stasis followed by innovative jumps" (p. 39).

B&C also claim that the human language system appears to have undergone very little, if any, evolution since it first arose. The main idea here is simply that all modern humans are equally capable of acquiring and using any human language. If so, then the language system must have "finished" evolving by the time our ancestors left Africa. Given the dates for colonization of the Old World and

[4] To see how this goes, consider the linguistic expression [X [X [X Y]]]. This structure would be built as follows: first X and Y would be Merged, yielding G = [X Y]; then X and G would be Merged, yielding H = [X G] = [X [X Y]]; finally, X and H would be Merged, yielding I = [X H] = [X [X [X Y]]].

Australia, it is now believed that this exodus occurred no later than ~ 60 kya (McEvoy et al. 2011). Hence, if language only appeared some 80 kya, one might think there was just not enough time for the language system to undergo a significant change.

Thus, in B&C's view, language must have evolved within the sliver of time defined by these constraints. This line of reasoning might be resisted, of course. In particular, one might think that B&C's criterion for identifying language—archeological evidence of symbolic behavior—is too strong. For one thing, as has been pointed out by Dediu and Levinson (2013), we know of populations of humans for whom we have a dearth of evidence of symbolic behavior and yet, given how recently these humans lived, it cannot be seriously doubted that they lacked language.[5] (They cite the case of colonizers of the Americas who would have lived just some 12 kya.) In addition, though admittedly more speculatively, one might think that archeological evidence of robust technological-cum-cultural traditions (Stout and Chaminade 2012; Planer 2017a, b) and of economic transitions (Sterelny 2016) lying in the much deeper past are reliable indicators of linguistic advances. In short, then, I think a powerful case can indeed be mounted against B&C here. But as it is not the aim of this article to make that case, suffice it to say that many theorists, myself included, will already part ways with B&C and have ample reason to do so. This is one significant take-home of the present work.

Setting this skepticism aside, let us now turn to B&C's strategy for explaining how language might have evolved so suddenly. According to them, developments within linguistic theory over the last several decades strongly suggest that human language syntax can be reduced to the operation Merge. This view contrasts sharply with early descriptions of syntax which portrayed it as consisting in a large set of ordered, complex, transformational rules. Moreover, Merge is a very simple operation; as explained above, all it does is take two expressions and combine them to form a third, complex expression that is accessible to further computation.

Given its simplicity, B&C think that Merge "may have arisen from something as straightforward as a slight rewiring of the brain" (p. 79). They offer one concrete proposal for the neurobiological realization of Merge, though it is intended mainly to convey a sense of the type of neurobiological change they have in mind here: on this proposal, Merge resulted from the completion of a fiber-ring tract connecting Broca's area and Broadmann's area 45, both in the frontal lobe, with Wernicke's area in the temporal lobe. They see the completion of this tract in a single step as on par with the sorts one-step phenotypic changes evolutionary-developmental biology has taught us sometimes occur; for example, the one-step evolution of a new beak morphology in Galapagos finches or of a new spine morphology in stickleback fish.[6] And like the latter set of cases, B&C claim that the completion of this tract may have involved no more than a single mutation to a gene-regulatory region: "[a] small genomic change in a growth factor for one of the fibers, along with proper fiber tract guidance" (p. 164).

[5] I thank Richard Moore for drawing this work to my attention.

[6] See Gerhart and Kirschner (2007) for an excellent discussion.

That is the gist of how B&C intend to explain the evolution of human language syntax. As I see it, the other part of their strategy is to motivate the idea that everything else needed for language might have already been present in the brain. They are far less successful in this endeavor.

B&C devote the lion's share of their attention to the machinery implementing the realization of the interface with the sensorimotor system. In their view, we have already learned most if not all of what it is possible for us to know regarding the nature and evolution of this machinery. This knowledge has been gained through a combination of work in neurobiology, developmental and comparative biology, and genetics dealing with species capable of complex vocal learning and production. What this work has revealed is that these capacities are underlain by highly similar neurobiological machinery in animals as distantly related as birds and mammals and that this machinery is built using highly similar and/or the same developmental and genetic "programs". B&C take this to show that:

> ... the "toolkit" for building vocal learning might consist of a (conserved) package of perhaps 100–200 or so gene specializations no matter what the species that can be "booted up" quickly—and so evolved relatively rapidly. (p. 45)

Of course, none of this tells us *why* our ancestors might have evolved the requisite machinery for learning and producing complex vocal behavior prior to some form of spoken language, though. B&C do not offer an answer to this question.

This is one of the less serious problems for B&C on my view. As several theorists have suggested, one possibility is that our ancestors "sang" before they spoke (Mithen 2005; Dunbar 2012; Killin 2017). More precisely, the idea here would be that machinery for producing and learning complex vocal behavior first evolved in connection with "musical" vocalization. Music-like vocalization may have played an important role in signaling to mates or competitors just as it does in songbirds (Mithen 2005). Alternatively, it may have helped to promote and maintain within-group social bonds. This may have taken the form of singing together, strengthening social bonds through joint activity, or singing to one another, thereby triggering the release of biomolecular messengers conducive to prosocial feeling.[7] Indeed, all of these factors may have been at work. Importantly, however, it is another thing to claim that this machinery might have reached a state of "language readiness" in the sense that no further evolution, or only very little evolution, would have been necessary in order for it to support fully-modern human (vocal) language. This is the claim that B&C need to go through here and it is at best unclear whether it is plausible. (Theorists open to a gradual, incremental scenario for language evolution, by contrast, might simply claim that this machinery served as a rough starting point for the evolution of the machinery underlying fully-modern vocal language.)

Moving on to the "word-like atomic elements" (p. 112): B&C do little more than simply assert that they might have existed prior to syntax. As I read them, they think they are under no real obligation to motivate this claim because the nature and

[7] For a discussion of the effect of music on the production of neurotransmitters, hormones, and other biomolecular messengers, see Gangrade (2012).

therefore evolution of these concepts remains so obscure. They tell us, "by all accounts, the origin of mind-dependent word-like elements remains a big mystery—for everyone, us included" (p. 149); that "no one has any idea" (p. 86) as to how they might have evolved. So, why *not* think they evolved prior to syntax? It may well be that no one has a firm grip on the nature of these concepts, but it is worth noting that at least one promising line of thought may be off-limits to B&C (whether it is in fact off-limits turns on whether they allow a pre-linguistic form of syntax governing mental expressions; more on this below [see "Consumer machinery of merged expressions" section]). This is that some form of syntax was critical to the evolution of concepts that did not depend upon correlations with their referents. A version of this idea is presented by Terrence Deacon in his book *A Symbolic Species* (1998). He tells us:

> Some sort of regimented combinatorial organization is a logical necessity for any system of symbolic [i.e., non-iconic and non-indexical/correlational] reference. Without an explicit syntactic framework and an implicit interpretative mapping, it is possible neither to produce unambiguous symbolic information nor to acquire symbols in the first place. Because symbolic reference is inherently systematic, there can be no symbolization without systematic relationships (p. 100)

Deacon develops this idea within the context of Pierce's general theory of signs and connectionist models of human cognition, but it is possible to unpack his main insights without committing to these frameworks (Planer, in preparation). In any case, I propose to let this issue slide; it is again a relatively minor problem compared to the ones we shall get to shortly.

For B&C, the interface with the conceptual system is a more integral part of the language system than the sensorimotor interface. They say, "there is a significant asymmetry between the two interfaces, with the [interface with the conceptual system]—the link to systems of thought and action—having primacy" (p. 71). It is thus shocking that they devote so little attention to the machinery that is supposed to implement this component. As far as I can tell, the closest they actually come to addressing this issue is to tell us that the brain likely operates in accordance with "principles of efficient computation" (p. 71). How is this relevant? What I take B&C to be saying is that however Merge interacts with the conceptual system, that arrangement is computationally efficient one; it fulfills the necessary information-processing tasks while minimizing energy expenditure. But this does nothing to motivate the idea that that the machinery implementing this arrangement may have existed before the evolution of Merge. Indeed, if anything, the assumption of computational efficiency or optimality would seem to actually work *against* B&C at this point. For not only must they claim that this machinery was already in place; they must claim that it was already optimized for carrying out functions it did not possess until Merge arose!

I will have more to say regarding the interface with the conceptual system below; for now, let us take stock. For B&C, the hardest part of the problem of language evolution is to explain how language might have evolved so rapidly. Their explanation has two parts: one is supposed to convince us that the mechanism of human language syntax might have arisen in a single step; the other that all, or virtually all, of the other ingredients needed for language might have already been on hand. Thus, the appearance of syntax led to the creation of a "linguistic phenotype" in single evolutionary step which was either already identical with our modern phenotype or else only superficially different (therefore requiring only a few "tweaks"). As B&C put it elsewhere in an article with colleagues:

> … the language faculty is an extremely recent acquisition in our linage and it was acquired not in the context of slow, gradual modification of preexisting systems under natural selection but in a single, rapid, emergent event that built upon those prior systems but was not predicted by them (Bolhuis et al. 2014, p. 4).

Mysterious machinery

In this section, I argue that there are several substantial pieces of neurocomputational machinery needed to realize language in B&C's sense whose evolution prior to Merge is extremely implausible. If I am correct, then B&C's way of trying to explain the sudden evolution of language will turn out to be a non-starter.

Managing merge

It is obvious that Merge builds different linguistic expressions in different cognitive and communicative contexts. Moreover, these expressions tend to be relevant to ongoing cognition and communication. How does this work?

B&C make much of the fact that Merge is so computationally simple. This makes sense; it is Merge's simplicity, after all, which enables them to argue that human language syntax might have evolved in a single step. But this simplicity comes at a cost. For there is nothing in the operation Merge itself which determines what the operation is to be applied to at a given time. Hence, there must be some mechanism external to Merge which controls the inputs Merge receives at a time and so the complex hierarchical expressions it builds over time.[8]

We might call this system "the Merge Control Unit" or "MCU" for short.[9] Recall that Merge is unrestricted in its operation: every lexical item is available to it as is every previous output of Merge. Accordingly, the MCU must be able to select for processing by Merge any item in the individual's lexicon and any item currently

[8] One might think that this work is done by the semantic "features" posited by the Minimalist program. Put simply, features are like "tags" that attach to lexical items and determine what those expressions (and the complex expressions they make up) can and cannot combine with. If one wishes to appeal to features in this context, then a story must be told about how they came to exist prior to Merge. (This will be more or less difficult depending on what one wishes to say about the forms of mental representation that existed prior to Merge.) But even granting that such a story can be told, it will not obviate the need for a mechanism in the above sense; this is because features merely constrain rather than determine what is combined with what.

[9] To be sure, there are a few ways we might conceive of the MCU and its relation to Merge, but the details do not really matter here.

stored in the workspace. It must, therefore, bear appropriate linkages with lexicon and workspace.

But this is just the beginning. For in order to produce expressions that are relevant to ongoing cognition and communication, the expressions the MCU directs Merge to build at a time must bear some kind of systematic relation to conceptual-system processing at that time. So, the MCU must receive information about the latter. Moreover, if this relation is at all complex, then the internal structure of the MCU will be correspondingly complex; it will either embody or else contain an explicit representation of a set of correspondingly complex computational principles that govern its behavior. Suppose, for example, that the MCU checks to see whether two conceptual-system expressions concern the same entity before directing Merge to combine their contents in some way. Even in cases where the two expressions contain tokens of the same sub-expression, the MCU will still have to confirm that there is a formal match between the sub-expressions. That is, it will have to compute something like the "=" relation.[10] What about cases where there is no such formal overlap between the expressions? It's hard to say in the absence of a more detailed architecture, but the computations involved are unlikely to be trivial. The MCU may also have to select specific portions of these expressions for processing by Merge, or perhaps select some expression which is part of neither conceptual-system expression but which bears an important semantic connection to the latter expressions. And so on.

Given this all this external and (presumably) internal structure, I do not see how a mechanism even coarsely resembling the MCU might have evolved prior to Merge. Or to put the point in a more general way: There is undoubtedly a range of complex computational tasks surrounding the production of linguistic expressions that are relevant to ongoing cognition and communication. If these tasks are not carried out by the mechanism of human language syntax itself, then some other mechanism will have to pick up the slack. That's fine; but then we are owed an account of how and why a mechanism of the latter sort might have been assembled in advance of the mechanism of syntax. I doubt that a plausible account along these lines can be provided. It would seem that the two must have co-evolved.

Consumer machinery of merged expressions

Recall that, on B&C's view, each linguistic expression has a definite semantic interpretation which is determined in accordance with its hierarchical structure. Earlier I proposed that we understand the semantic interpretation of such a linguistic expression to be fixed by some portion of that expression's computational role within the conceptual system. Appealing to computational role to spell out the content or meaning of a mental representation is familiar move from the philosophy of mind and cognitive science.[11] This was offered as an alternative to the idea that semantic interpretation consists in the production of a sentence in an independent

[10] I am here assuming the existence of complex expressions prior to Merge. More on this assumption below.

[11] The most seminal article in this area is perhaps Ned Blocks' "Conceptual role semantics" (1998).

and antecedent language of thought. As I explained there, B&C explicitly reject the latter as a model of semantic interpretation, and with good reason, given their other theoretical views. Perhaps there is some other way for them to unpack semantic interpretation which coheres with their other commitments; but if there is, it is far from obvious, and so I leave that job to B&C. For now I will assume that semantic interpretation must be understood in something like the above way.

If linguistic expressions can differ in their semantic interpretation in virtue of how they are hierarchically organized, then the conceptual system must contain computational machinery which is sensitive to hierarchical structure and which is caused to behave differently in the face of different forms of hierarchical structure. This point was made earlier. What we now want to know is this: How might such machinery have evolved prior to the evolution of Merge?

It is very hard to see how such an evolutionary scenario might go, to put it mildly. This is because, assuming B&C are correct, there simply wouldn't have been any expressions with hierarchical structure for the brain to process until Merge arose. What sorts of expressions would there have been, exactly? B&C are unclear on this matter. One would expect them to say that our pre-Merge brains computed expressions having linear syntax only; what Merge did was introduce hierarchical syntax to our brains. The main rationale for such a view is that linear syntax is needed for all but the most primitive forms of computation. More specifically, without linear syntax, it is not possible to effect an analytical decomposition of referents, and without the latter ability, one is generally limited to the use of lookup tables to store knowledge of the world. Having one's algorithmic options limited in this way profoundly constrains computational power (Gallistel and King 2011; Planer, under review). And yet, some of the things B&C say make it sound like they think complex expressions of any sort did not exist before Merge.[12] For example, at one point they tell us, "[c]oncatenation requires Merge or some similar operation along with order and some principle to erase structure …" (p. 91). Taken at face value, this would mean that our pre-Merge brain was limited to the processing of wholly atomic expressions. Note how radical such a position would be: it would imply that the behavior of all other animals—indeed, the behavior of our very recent hominin ancestors!—can be explained without at any point adverting to complex expressions. But more to the present point: Since atomic expressions can have no syntax, there would have been no syntactic properties for our brain's computational machinery to be sensitive to. The idea that this machinery was nevertheless pre-adapted to be systematically affected by hierarchical syntax—which is inherently more complicated than linear syntax—is absurd.

Let us ignore such remarks from B&C, then, and assume the existence of expressions having linear syntax prior to Merge.[13] Still, I think there is a real worry. The worry is this: Why should we think that computational machinery that evolved

[12] An expression can be complex—can have parts—without having linear syntax, but not vice versa. (See Gallistel and King 2011, Ch. 5, for a nice overview.) .

[13] Importantly, this is not to assume the prior existence of an independent language of thought that was as expressively rich as language in B&C's sense. (For one thing, we are here assuming that the expressions which existed prior to Merge had linear syntax only. This antecedent representational system might have been limited in all sorts of other ways too.).

for processing expressions with a linear syntax only would show systematic sensitivity in the face of hierarchical structure? No further evolution on the part of our brain's computational machinery was required? Surely, B&C must provide us with some story at this point.

So far we have been concerned simply with the evolution of computational machinery which is capable of distinguishing expressions based upon their hierarchical properties. But we are owed a good deal more than this, actually. As I mentioned earlier, according to B&C Merge was not selected for upon its evolutionary debut because it improved our communicative abilities; this, they think, came only later with the cultural creation of natural languages once Merge had spread to a sufficiently large number of individuals. Rather, they hold that it was a direct cognitive benefit produced by Merge which led to Merge's being initially selected for. Hence, it follows that our brain's computational machinery must have been capable of harnessing the hierarchical syntax introduced by Merge, of exploiting it in some way, at the time Merge arose.[14] Or to put the point another way: the act of "relat[ing] hierarchical structures built by Merge to systems of reasoning, inference, planning, and the like" (p. 140) must have produced a cognitive payoff.[15]

This is not at all trivial. To see this, consider a (toy) inference mechanism that works as follows: it takes as input an expression of the form X R Y, where X = Y = {Alex, Ben, Carla, …} and R = {Hits, Hugs}; if R = Hits, the mechanism outputs the expression X Is_Dangerous; else, it outputs the expression X Is_Friendly.[16] Now suppose that each x in X refers to a unique individual, Hits to the act of hitting, Hugs to the act of hugging, Is_Dangerous to the property of being dangerous, and Is_Friendly to the property of being friendly. Finally, suppose that if one individual hits another, the former is dangerous, whereas if they hug another, they are friendly. Given this, and given that the system containing this mechanism benefits from knowing who is dangerous and who is friendly, this mechanism might prove quite useful. For embedded within its operation is the knowledge that people who hit others are dangerous while those who hug others are friendly.

This inference mechanism exploits the linear syntax of its input expressions in the generation of useful outputs; that is, it exploits the fact that the first part (i.e., the first atomic/unstructured element) of the expression encodes the identity of the person who performed the act; the second part, the identity of the act; the third part,

[14] One might be thinking at this point, "Regardless of what B&C say, couldn't one hold that Merge was selected for due to its facilitating communication?" It is far from clear how this might go, given B&C's picture of the language system and its relation to the rest of the mind. It would be too large a digression to delve into this issue here, however. Suffice it to say that any new communicative sophistication on the part of some Merge-possessing sender would appear to be lost on a receiver unless that receiver (1) also possessed Merge, and (2) possessed cognitive mechanisms capable of exploiting the Merged expressions which the sender's communicative behavior cause him (i.e., the receiver) to construct. But to assume (2) just is to assume cognitive machinery that would have upgraded cognition independently of communication. (Note also that (1) is not a trivial assumption: it presupposes that Merge appeared simultaneously in two or more habitually interacting individuals (e.g., a pair of siblings).)

[15] To be clear, this act of relating would run from Merge (or better: the workspace) to the conceptual system via the interface with the latter.

[16] I use an "_" here to mean that the conjoined words represent a single, unstructured concept.

the identity of the recipient of the act. What would happen if we suddenly started feeding this mechanism input expressions with a different syntax? One possibility is that it would fail to operate at all; it might do nothing. But even assuming that the mechanism *was* cued into action, the result would almost certainly be nonsense. (This will be a familiar point to anyone who is ever attempted to design a computer chip.) The mechanism works—it is useful—precisely because it exploits specific structural features of the set of input expressions it was designed to operate on. Suddenly change that structure, and the mechanism becomes useless.

All of this applies to hierarchical syntax in spades. Imagine feeding hierarchically-structured expressions built by Merge to this mechanism. Even if these expressions functioned to describe similar states of affairs (i.e., dyadic episodes of hitting and hugging), there is no reason to expect that useful outputs would be generated. If expressions like [[The [Tallest Girl]] [Hit Ben]] or [Carla [Hugged [The [Oldest Boy]]]] managed to cue this inference mechanism into action at all, they would not tap the knowledge embedded within it.[17]

The main guiding thought behind all of these remarks is this: powerful and adaptive computational systems are delicately arranged physical systems; the format of the expressions that are computed over and the machinery that effects the computation are designed to fit hand-in-glove. They are co-adapted to one another. In the words of Gallistel and King, such systems are "created by careful engineering" (p. 57, 2011). In the case of manmade computational systems, it is we who effect this careful engineering, while in the case of evolved computational systems, it is co-evolutionary processes governed by natural selection (Planer, under review).

Cortical circuitry

The plausibility of B&C's claim that the machinery implementing the interface with the conceptual system may have been present before Merge crucially depends upon the architecture of the human conceptual system. Unfortunately, we still do not know what that architecture is like.

According to one influential line of thought, the conceptual system consists entirely of "mental modules" (Cosmides and Tooby 1992; Sperber 1996, 2001; Mithen 1996; Carruthers 2002, 2006). There is a real issue here with how the notion of a module is supposed to be taken. What is clear to everyone is that these modules cannot be modules in the sense articulated by Jerry Fodor in his seminal work, *The Modularity of Mind* (1983). Fodor used nine features to characterize a module: domain specificity, mandatory operation, limited central accessibility, fast processing, informational encapsulation, "shallow" outputs, fixed neural architecture, characteristic and specific breakdown patterns; and characteristic ontogenetic pace and sequencing. In the case of some of these features, it is simply incoherent to attribute them to conceptual modules; it makes no sense to say, for example, that

[17] If the point is still not clear, perhaps the following analogy will help: Imagine a standard addition chip which has been designed to operate on expressions that encode numbers using the binary encoding system. Now imagine feeding this chip a new set of expressions which, while still encoding numbers, encode them using some other encoding scheme. The chip will produce nonsense.

conceptual modules are fast (relative to what?); while in the case of others, the problem is not one of incoherence but rather implausibility; the idea that conceptual modules might in general be informationally encapsulated, for example, is very difficult to maintain.[18] Thus, proponents of massive modularity offer various non-Fodorian accounts of modularity, such as this one from Peter Carruthers (2006):

> … modules [are] isolable function-specific processing systems, all or most of which are domain specific (in the content sense), whose operations aren't subject to the will, which are associated with specific neural structures (albeit sometimes spatially dispersed ones), and whose internal operations may be inaccessible to the remainder of cognition. (p. 12)

A massively modular architecture for the human conceptual system would represent the worst-case scenario for B&C. Why? Linguistic expressions can feature concepts from any mental domain and so the outputs of each of these autonomous computational systems must be delivered to the vicinity of Merge.[19] The idea that all of the cortical circuitry needed to realize such an arrangement may have been in place prior to Merge cannot be taken seriously; all the more so if that circuitry is computationally efficient or optimal.

It is thus deeply ironic that B&C appear to favor a something like a massively modular picture of the human conceptual system. B&C appeal to experimental work by Hermer-Vazquez et al. (1999) to illustrate the "definite selective advantages" (p. 166) of Merge. The main result of the study was that pre-linguistic children were incapable of integrating geometric and non-geometric information in the context of problem-solving. B&C tell us:

> One explanation for this behaviour, above and beyond sheer memory overload is that language is the *lingua franca* that binds together the different representations from geometric and nongeometric "modules", just as an "inner mental tool" should. Being able to *integrate a variety of perceptual cues and reason about them*—is the animal above or below the rock—would seem to have definite selective advantages. (p. 165–166). [italics mine]

Note that B&C's explanation for why Merge would have been initially selected for simply assumes that the circuitry necessary to bring together contents from the brain's different modules was already in place. Note also that it simply assumes that conceptual-system computational machinery would have been able to exploit these non-domain-specific expressions (B&C tell us we would have been able to "reason about them" [p. 166], that is, the states of affairs represented by the non-domain-specific expressions). This assumption is already problematic for the sorts of reasons discussed in the previous subsection (i.e., general, hardware compatibility concerns), and even more problematic when coupled with the view that the human

[18] See Carruthers (2006) for a discussion.

[19] I say "the vicinity of Merge" because one might think that these systems deliver outputs to a mechanism controlling the operation of Merge—our MCU from above—rather than Merge itself. (Here I am assuming that the MCU and Merge would themselves be in the same vicinity, as is surely plausible given the brain's tendency to "save wire" [Striedter 2005].) This would be compatible with Merge having been directly connected to one or more of these modules when Merge first appeared.

conceptual system is massively modular. For on the latter view, when Merge first arose the brain would have been exhausted by modules for processing domain-specific information of one sort or another. What benefit might result from, say, presenting a module which has been specialized for processing biological information with an expression integrating content from the biological and some other mental domain? One might think that, at best, the module would ignore the non-biological part of the expression and simply process the biological content. This would do nothing to improve cognition, however.[20] These problems are all very worrying as the above is the only concrete suggestion B&C make regarding the initial adaptive benefit of Merge.

So much for then for massive modularity. Looking to the opposite end of the spectrum, the architecture of the human conceptual system might have been a *massively integrated* one prior to Merge. More precisely, it might have been that there was a place in the brain where "everything came together", something like a "global workspace" in Bernard Baars' sense (Baars 1997). We can conceive of this location as something like a central memory system where each subsystem of the conceptual system wrote to and read from. This would represent the best-case scenario for B&C from a cortical wiring perspective. For then all B&C need to claim is that Merge arose bearing appropriate connections to this central memory store. On this proposal, the appearance of Merge would have involved slightly more than a "slight rewiring of the brain", but not much more. The novel circuitry needed to functionally integrate Merge with the central memory system might then be explained by appeal to normal processes of neural plasticity (or so one might reasonably argue).

There is obviously a large intermediate range between these two extremes. The thing to note is this: to the extent that it is at all plausible to think the machinery implementing the interface with the conceptual system was present before Merge, it is plausible only for a very restricted portion of this range. Basically, unless there was already a place in the brain where "everything came together", the integration of Merge with the conceptual system would have required very considerable changes in neuronal connectivity, far more than that which might occur in a single evolutionary step or even several such steps. This is a very strong empirical assumption for one's theory to be hostage to.[21] One might object at this point that I am underestimating the power of small genomic changes to effect large-scale wiring changes in the brain. Specifically, a small genomic change controlling the rate of cell division in some part of the brain can result in a large increase in the size of that region. The enlarged region will then project a correspondingly larger number of new axons to other brain regions while simultaneously providing a correspondingly larger number of new anchor sites for axons coming into the region (Deacon 1990; Striedter 2005). The result would be an abundance of novel connections for development to shape. While all of this may be true enough, such a process cannot be relevant in the present context. For the brains of archaic humans (who would

[20] Because of the expression's novel format, another possibility is that it might simply fail to cue the module into action at all; it might prove epiphenomenal. Alternatively, it might cause the module to operate but to produce some non-adaptive output.

[21] And an ironic place for B&C to wind up, it's worth noting, given their apparent sympathy for a massive modularity picture of the mind.

have lacked language, according to B&C) already overlapped the normal brain-size range of modern humans.[22, 23]

Conclusion

In this article, we had a close look at B&C's how-possibly explanation for the rapid emergence of language. I argued that, even granting B&C's very contentious way of understanding language, their explanation is seriously lacking. More specifically, I argued that there are several key pieces of neurocomputational machinery needed to realize the language system in B&C's sense whose evolution prior to Merge is extremely doubtful. If so, then B&C's idea that the sudden appearance of Merge caused the sudden appearance of language is unpersuasive. A scenario on which these other pieces of machinery co-evolved with the mechanism of human language syntax, and with cognitive mechanisms capable of exploiting linguistic expressions, seems inescapable. But this is precisely the kind of scenario which B&C are committed to denying.

Acknowledgements I am grateful to Matt Spike and Kim Sterelny for conversation regarding this article and to Bruno Ippedico for comments on an earlier draft. I would also like to thank Richard Moore and an anonymous reviewer for this journal for their helpful feedback.

References

Baars BJ (1997) In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. J Conscious Stud 4(4):292–309

Berwick RC, Chomsky N (2015) Why only us: language and evolution. MIT press, Cambridge

Block N (1998) Conceptual role semantics. In: Craig E (ed) Routledge encyclopedia of philosophy. Routledge, Abingdon, pp 242–256

Bolhuis JJ, Tattersall I, Chomsky N, Berwick R (2014) How could language have evolved? PLoS Biol 12(8):e1001934

Carruthers P (2002) Modularity, language, and the flexibility of thought. Behav Brain Sci 25(6):705–719

Carruthers P (2006) The architecture of the mind. Oxford University Press, Oxford

Cosmides L, Tooby J (1992) Cognitive adaptations for social exchange. Adapt Mind Evol Psychol Gener Cult 163:163–228

Cummins R (1983) The Nature of Psychological Explanation. Bradford/MITPress, Cambridge

Deacon TW (1990) Rethinking mammalian brain evolution. Am Zool 30(3):629–705

Deacon TW (1998) The symbolic species: the co-evolution of language and the brain. WW Norton & Company, New York

Dediu D, Levinson SC (2013) On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. Front Psychol 4:397. https://doi.org/10.3389/fpsyg.2013.00397

d'Errico F et al (2005) Nassarius kraussianus shell beads from Blombos Cave: evidence for symbolic behaviour in the Middle Stone Age. J Hum Evol 48(1):3–24

Dunbar RIM (2012) On the evolutionary function of song and dance. In: Bannan N (ed) Music, lang, and human evolution. Oxford University Press, Oxford, pp 201–214

[22] I thank Kim Sterelny for bringing this point to my attention.

[23] Indeed, on B&C's view, even anatomically modern humans would have lacked language for over 200,000 years. But such humans—as the nomenclature is meant to indicate—were physically just like us, brain size included.

Gallistel CR, King AP (2011) Memory and the computational brain: Why cognitive science will transform neuroscience, vol 6. Wiley, New York

Gangrade A (2012) The effect of music on the production of neurotransmitters, hormones, cytokines, and peptides: a review. Music Med 4(1):40–43

Gerhart J, Kirschner M (2007) The theory of facilitated variation. Proc Natl Acad Sci 104(suppl 1):8582–8589

Green RE et al (2006) Analysis of one million base pairs of Neanderthal DNA. Nature 444(7117):330–336

Hermer-Vazquez L, Spelke ES, Katsnelson AS (1999) Sources of flexibility in human cognition: dual-task studies of space and language. Cogn Psychol 39(1):3–36

Ippedico B Symptoms of a married linguistics (**In preparation**)

Killin A (2017) Plio-Pleistocene foundations of hominin musicality: coevolution of cognition, sociality, and music. Biol Theory 12(4):1–14

McEvoy BP et al (2011) Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. Genome Res 21(6):821–829

Mithen S (2005) The singing Neanderthals: the origins of music, language, mind, and body. Harvard University Press, Cambridge

Mithen S (1996) The prehistory of the mind: a search for the origins of art, science and religion. Thames and Hudson, London and New York

Planer R (2017a) Protolanguage Might Have Evolved Before Ostensive Communication. Biol Theory 12(2):72–84

Planer R (2017b) Talking about tools: did early pleistocene hominins have a protolanguage?. Biol Theory 12(4):211–221

Planer R Evolving languages of thought (**Under review**)

Planer R Symbols, symbolic cognition, and symbolic behavior (**In preparation**)

Resnik DB (1991) How-possibly explanations in biology. Acta Biotheor 39(2):141–149

Sperber D (1996) Explaining culture: a naturalistic approach, vol 323. Blackwell, Oxford

Sperber D (2001) In defense of massive modularity. In: Dupoux E (ed) Language, brain and cognitive development: essays honor Jacques Mehler. MIT Press, Cambridge, pp 47-57

Sterelny K (2016) Cumulative cultural evolution and the origins of language. Biol Theory 11(3):173–186

Stout D, Chaminade T (2012) Stone tools, language and the brain in human evolution. Phil Trans R Soc B 367(1585):75–87

Striedter GF (2005) Principles of brain evolution. Sinauer, Sunderland