Some Third Factor Limits on Merge*

Sandiway Fong, University of Arizona, draft 3/29/2021

The device we call the brain is a marvelous organ, endowing us with the capacity for symbolic thought, language and reasoning far beyond what other animals have exhibited. We possess a *linearly-scaled* primate brain according to Fonseca-Azevedo & Herculano-Houzel (2012), both our neuron density and our neural wetware being unremarkable for primates. (Astrocytes references.) However, with about 86 billion neurons, we possess the biggest primate brain. But we cannot directly attribute our cognitive capacity to sheer brain size.[1]

Although the above is true, this marvel is not the computational powerhouse that we might assume. The biological unit of computation, the neuron, possesses a slow communication mechanism, a signal requires around a millisecond to cross the chemical synaptic gap, and after certain electrical pulse trains, a synapse might require up to 140-150ms to recover (cite Testa-Silva et al. 2014). Although (much faster) direct electrical synapses exist in our nervous system, e.g. they can be found in the retina, slow chemical synapses predominate in the human brain.[2] There is also evidence that the brain does not maximize sensory capacity, which suggests the computational brain is the weak link (or bottleneck) in the chain from external stimulus to thought (and response). For example, we know our eyes are capable of both incredible sensitivity, i.e. single photon level (Tinsley et al. 2016), and resolution, achieving peak acuity of 77 cycles/degree (Curcio et al. 1990), all unnecessary for scene analysis. Even an eagle only possesses eyesight about 3 times better than humans, yet arguably, the eagle requires far better resolution. Human olfactory thresholds can be of the order of parts per billion (ppb) (Wackermannová et al., 2016). Our eardrums can detect vibrations smaller than the diameter of a hydrogen atom (Fletcher & Munson, 1933). In case after case, the brain does not make use of the full resolution of available sensory inputs. Perhaps the answer is that it cannot, as a slow organic system, it does not possess the necessary bandwidth, and therefore, it must (selectively) throw away much of the signal. The idea that this pressure for efficiency also

---

* Acknowledgments: Noam Chomsky inspired and repeatedly encouraged me to write this paper. He also provided much cogent discussion; however, all mistakes and misconceptions of the general enterprise contained here are solely mine. Thanks go also to Riny Huybregts and Masayuki Oishi for valuable comments. *List of acknowledgments go here.*

[1] Gorillas and orangutans are primates having brains 40% the size of ours. Remarkably, humans born with only 1/2 the normal number of neurons can function nearly normally, e.g. the celebrated case of Jonathan Keleher who is entirely missing a cerebellum (containing about 50% or more of our neural matter). (See his NPR interview, cite https://www.npr.org/sections/health-shots/2015/03/16/392789753/a-man-s-incomplete-brain-reveals-cerebellum-s-role-in-thought-and-emotion.) There is also much evidence that the cerebellum contributes substantially to (non-motor) language processing, see (Mariën et al. 2013). Evaluating the total number of neurons, the African elephant has us beat (cite Herculano-Houzel et al. 2014b). Even when limited to the cerebral cortex, certain kinds of whales also outpoint us (cite Mortensen et al. 2014).

[2] Some have theorized that computation is inside the cell, where signals are electrical, see (Trettenbrein 2016) for a summary and discussion relevant to cognition. Outside of the cell, chemical communication is clearly the bottleneck. It is highly improbable that UG, including the workspace and LEX, is primarily implemented entirely inside a single cell.

pertains to both data and computation in language, born out of biological limitations, was termed *The Third Factor* by Chomsky (2006).[3]

Recently, in the Minimalist Program (Chomsky 2019 & *forthcoming*), Chomsky has laid out (seemingly conflicting) conditions that a theory of language must meet in order to qualify as a genuine explanation. A theory must only make use of mechanisms simple enough to have plausibly evolved, yet rich enough that the burden on the language learner is not impractical given primary linguistic data. The phenomenon of language is complex enough to have resisted the best efforts of statistical and deep learning methods without the data and computational resource limits that all organic systems must respect.[4]

Chomsky has argued for a theory, UG, the core of which consists of the simplest possible combinatorial operation, binary set Merge, generating the linguistic expression of thought from atoms of lexical input (LEX). Binary Merge (BM), applied recursively, is the minimum operation necessary to obtain a discrete infinity of hierarchical structures: represented abstractly as {α, β} formed from α (resp. β) = the result of BM or α (resp. β) ∈ LEX.[5] If another maximally simple operation, that of linear concatenation, is unavailable to the core, we have an account for the case of structural dependency being preferred over linear proximity, as exhibited by all human languages. A system of externalization (EXT), separate from the core, maps linguistic expressions into sound (and other possible modalities, e.g. sign). Due to organic sensory-motor limitations, linear order must be imposed on linguistic expressions, subject to language variation. A separate system of interpretation (INT) reads hierarchical structure without regard to basic word order differences between languages.

There is no simpler possible mechanism capable of discrete hierarchical infinity than BM. Therefore, in terms of evolutionary plausibility, BM represents the optimal solution. In terms of biological computation, BM will also be optimal provided it minimally burdens available resources, i.e. with respect to working memory size and combinatorial options. To take a simple example, consider the BM possibilities in (1a–b) for atoms *a*, *b*, and *c* drawn from LEX.

(1) a. a b c ⇒ {a, b} c ⇒ {{a, b}, c}
    b. a b c ⇒ {a, b} *a* *b* c ⇒ {{a, b}, c} {*a, b*} *a* *b* c

{{a, b}, c} is formed in both derivations, taking two steps (⇒). However, (1a) is minimal with respect to working memory size (note: Chomsky uses the term *workspace*). One way to characterize the complexity of the workspace is to count the number of terms.[6] Call this measure workspace size (WSS). (1a) is preferred as WSS increases by exactly 1 on each step, whereas in (1b), WSS first increments by 3, then 5. An increment of at least 1 is required as BM always constructs a new phrase (a workspace object). In (1b), BM preserves its inputs into the

---

[3] According to Chomsky (2006), the three factors are: (1) genetic endowment, (2) experience, and (3) general principles not specific to the faculty of language. The third factor includes principles governing acquisition and principles of efficient computation, the latter one being the central topic of this paper.
[4] Deep learning systems can perform remarkably poorly outside of the narrow band that constitutes their training data. See (Fong, *ms*.) for example.
[5] The calculations and results in this paper generalize to *n*-ary merge, $n \geq 2$. Binary merge is merely the base case.
[6] LEX atoms are terms. The output of BM is a term.

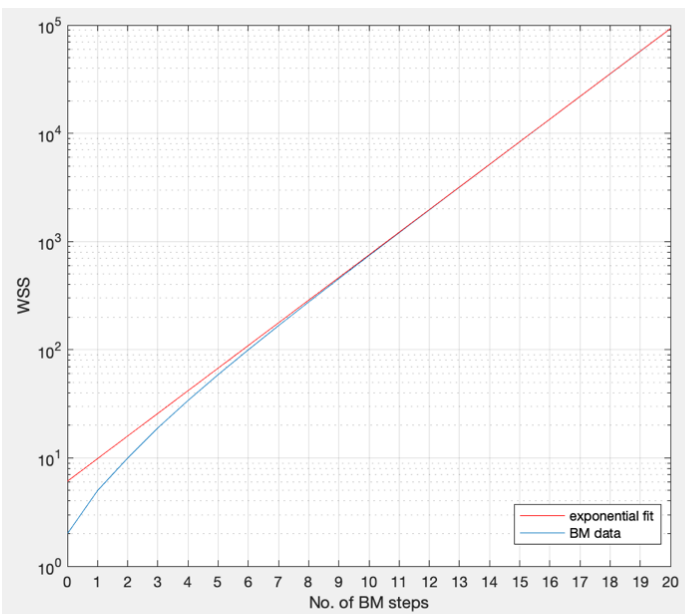
next cycle, whereas in (1a), the inputs for one merge cycle are not carried forwards. Chomsky observes that (1a) is to be preferred to (1b) on empirical grounds, as (1b) essentially permits copies to be replicated anywhere in structure (but unattested in language).

We can also fashion a resource restriction argument for (1a) over (1b). In terms of the combinatorics, given *n* LEX items, we observe that (1a) must complete within *n-1* steps with WSS 2*n*-1 terms. Generally, given *n* atoms, there are an exponential number of ways to combine them, but all of them must terminate within *n*-1 steps because BM minimally reduces the number of workspace objects by 1. The case for *n* = 4 is illustrated in (2a-b) below.[7]

(2) a. a b c d ⇒ {a, b} c d ⇒ {{a, b}, c} d ⇒ {{{a, b}, c}, d}
b. a b c d ⇒ {a, b} c d ⇒ {a, b} {c, d} ⇒ {{a, b}, {c, d}}

In contrast to (1a), (1b) actually diverges, i.e. it is always possible to perform more merges. The WSS grows exponentially in the worst case, as shown in (3).[8] Exponential growth is to be avoided for all systems, including organic ones, as it quickly outgrows any fixed resource.

(3)



[7] (2a-b) illustrate the basic patterns. Each case expands combinatorially given choice of inputs to BM at each stage. Suppose we have *n* WS items. Then there are n(n-1)/2 distinct choices for BM. Generally, given *n* items from LEX, there will be $n!(n-1)!/2^{n-1}$ distinct *derivations*. Since $c^n < n!$ for constant *c* and sufficiently large *n*, the number of possible derivations grows faster than exponential.

[8] Suppose there are *k* LEX atoms, k ≥ 2 (needed to activate BM). Then WSS at step 0 is WSS(0) = *k*. Generally, WSS(*n*) = WSS(*n*-1) + # terms in the worst possible BM, i.e. maximize the # terms, let us call this quantity m(*n*), n ≥ 1. For the worst case, let m(0) = m(-1) = 1 (base cases: # terms in a LEX atom; -1 used solely to simplify the recurrence relation). Then m(*n*) = m(*n-1*) + m(*n-2*) + 1, i.e. the syntactic object with the maximum number of terms can be formed by BM picking the maximums from the previous two steps. As a concrete example, consider the last step in (1b). The worst case next step is simply to pick {{a, b}, c} and {a, b} as input to BM. Notice the recurrence relation for function *m* parallels the recurrence relation for the Fibonacci series, and therefore, has a similar closed-form solution.

Chomsky (2019) defines MERGE over a workspace WS consisting of objects P and Q and X_1, ..., X_n as follows:

(4) MERGE(P, Q, WS) = [{P, Q}, X_1, ..., X_n] = WS’ (Chomsky, 2019)

WS objects X_1, ..., X_n not directly involved in BM must not (mysteriously) vanish, which rules out operations that meet the tightest resource restriction, respected by (4), by virtue of omission, e.g. as in (5a–d):

(5) a. a b c $\underline{d}$ $\underline{e}$ ⇒ {a, b} a b c
b. a b c d $\underline{e}$ ⇒ {a, b} a c d
c. a b $\underline{c}$ $\underline{d}$ $\underline{e}$ ⇒ {a, b} a b
d. a b $\underline{c}$ $\underline{d}$ $\underline{e}$ $\underline{f}$ ⇒ {a, b} a b

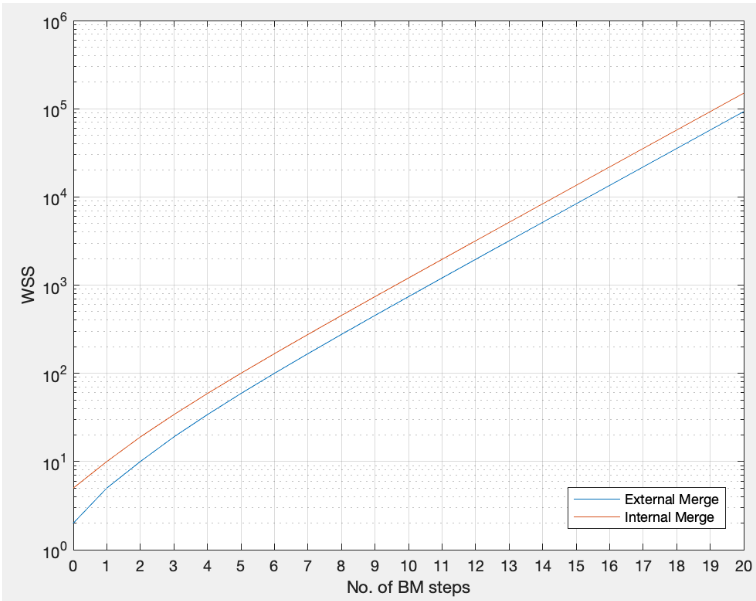
In (5a), *d* and *e* (underlined) are removed to make room for the continuation of both *a* and *b*, so that WSS still increments by exactly 1. Similarly, in (5b), one of the inputs to BM, viz. *a*, is smuggled through to the next round at the cost of losing *e*. Finally, (5c–d) illustrate (illicit) overachievers in terms of workspace management, WSS staying unchanged and decreasing by 1, respectively. The *nothing-lost* condition in (4), in conjunction with the minimal WSS increment limit, means that WS objects can only directly participate in recursive BM exactly once, as previously illustrated in (1a) (except for the case of Internal Merge to be addressed below). If previously selected WS objects cannot propagate forwards in workspace history, BM has no access to earlier stages of Merge computation. Thus, Chomsky observes that Merge-based computation is strictly Markovian. As discussed above, general access to history results in combinatorial explosion.

Another ubiquitous property of language is displacement, i.e. phrases may appear in surface locations other than where they are interpreted. In the discussion so far, we have been assuming that BM applies to distinct objects in the workspace, call this External Merge (EM). Chomsky’s definition in (4) also covers the case of Internal Merge (IM), in which P is a term of Q, Q a workspace object, and P and Q distinct. How does IM impact WSS? Consider (6a-b) below:

(6) a. {a, b} c ⇒ {a, {a, b}} c
b. {a, b} c ⇒ {a, {a, b}} {a, b} c

In (6a), Q is {a, b} and P is a sub-term, *a*, of Q. IM creates a copy of *a*, BM’ing it to Q. The unaffected workspace object *c* passes through, as required by (4). WSS increments by 2, one more than we had previously since BM adds a new phrase and the implicit copy adds another term. When Minimal Search (MS) is factored in, WSS returns to being incremented by 1, as the higher copy of *a* blocks the lower copy from being accessed. (We return to consider MS below.) (6b) illustrates the case in which general access to history is permitted. As with (1b), (6b) also diverges with the same exponential behavior as shown in (7) below.[9] Therefore, resource limitations also result in this option being excluded.

(7)



Next, let us turn to conditions for parallelism in this BM model. Consider the bare argument structure for a transitive sentence, as shown in (8) below.[10]

(8) {EA, {v*, {R, IA}}}
(R = verbal root; IA/EA = internal (external) argument (*resp.*); v* = verbalizer v licensing EA.)

EA and *IA* are phrases, possibly with complex, nested structure, e.g. cf. (9b–c) with (9a).

(9) a. men
b. the few men in the room
c. the few men that were in the room at that time

[9] See note 8 for the general explanation. Only the initial conditions are different because we need to BM two workspace objects first (to minimally form {a, b}) before IM can apply. Observe the IM worst case analysis obtains when we select (for raising) the largest subterm of the largest workspace object. But that subterm must have been the largest workspace object two cycles ago. IM itself adds one term (the new largest object). Therefore, m(*n*) = m(*n*-1) + m(*n*-2) + 1 at step *n*. For example, the largest subterm of {a, {a, b}} is {a, b}, but IM applied to {a, b} in the previous round (by raising *a*).

[10] Note that (8) only relates to bare argument structure, further merges will be required to form a convergent derivation for a complete sentence. For example, IM may apply to (8), e.g. in the case of *wh*-object questions, to the edge of the v* phase, forming {*wh*-IA, {EA, {v*, {R, *wh*-IA}}}}. Further displacement required to the edge of INFL (inflectional) and certain C (complementizer) heads, e.g. interrogative C_Q, are also not described here.

EM is the only available option for merging EA with {v*, {R, IA}}. The Extension condition, implicit in (4), prevents us from inserting material inside already-constructed phrases.[11] Nothing in principle prevents the workspace from being logically partitioned into sub-workspaces in which computation can (asynchronously) proceed in parallel. These distinct sub-threads of computation only need to be synchronized when the phrases come together, i.e. at {XP, YP} for the argument structure case, or <XP, YP> for Pair Merge in the case of adjunct phrases such as *in the room* in (9b-c) and *at that time* in (9c). For cases involving sequences of phrases, e.g. as in *John, Bill, and Mary ran for the stairs, the escalator, and the elevator* (*respectively*), each conjunct may be independently formed prior to synchronization.[12] Would this limited kind of parallelism be permissible given resource restrictions? At the level of the conscious mind, there is ample evidence the brain is a *mono*-processor, employing inefficient task-switching to cope with this bottleneck, see, for example, the literature summary in (Strobach et al. 2018). However, at the subconscious processing level, we know parallel processing exists in the brain. For example, in the case of vision, different functional units operate in parallel on the same signal in humans, as well as other primates (Breitmeyer 1992). Recent fMRI studies have suggested there may be something of the order of 50-60 independent brain processes active simultaneously on visuo-motor tasks, e.g. (Georgiou 2014). Suppose, then, that multiple BM functional units are available to the language faculty. If so, it seems these BM units must observe non-interference, i.e. they are free to operate on the workspace so long as they do not overlap their inputs. Consider (10a-b):

(10) a. a b c d ⇒ {a, b} {a, c} d
b. a b c d ⇒ {a, b} {a, b} c d

In (10a), we have two BM units operating simultaneously, one selecting *a* and *b*, the other *a* and *c*. This is the limiting case of (illicit) Parallel Merge, in which *a* participates in the creation of two different workspace objects, viz. {a, b} and {a, c}. By simply observing non-interference, Parallel Merge cannot be computed. In (10b), we have two BM units, both simultaneously selecting *a* and *b*. Again, the output of (10b) is ruled out by non-interference.[13]

Contrast the application of parallelism above with the situation in (11a-b) below:

(11) a. {a, b} c ⇒ {a, {a, b}} c
b. {a, b} c ⇒ {a, b} {a, c}

In each of (11a-b), only one new workspace object is formed; therefore, they cannot be a case of parallelism. (11a) (= 6a) is the familiar case of IM, but in (11b), we have an example of Sideways Merge, sometimes used to analyze adjunct control and parasitic gap constructions (Hornstein 2009). The difference in (11b) is that the selected subterm of {a, b}, viz. *a*, undergoes

---

[11] For example, we cannot first form {*men*, {v*, {R, IA}}}, then "tuck in" both *the* and *few* to obtain {*the few men*, {v*, {R, IA}}} (structural details for the phrase *the few men* not shown). The Extension condition is a sub-case of the Non-Tampering Condition (NTC), i.e. no modification to already-merged structure is permitted.
[12] In Chomsky (*forthcoming*), conjunct synchronization is carried out via a FormSequence operation that imposes similarity constraints on the conjuncts.
[13] Incidentally, the two cases of BM parallelism in (10a-b) increase WSS by 3 and 4, respectively.

BM with another workspace object, *c*, rather than simply extending {a, b}.[14] In either case, WSS increments by 2. Chomsky observes that IM can be assimilated to the same minimal WSS increment-by-1 story for EM, provided WSS is sensitive to Minimal Search (MS). Under MS, the lower copy of *a* becomes inaccessible, shielded from access by the higher copy that c-commands it. Assuming there is no reason for MS to expand to apply across the entire workspace, (11b) can be eliminated by the minimal WSS increment argument, as the two *a*'s are not in a position to shield one another from access.[15] Unfortunately, this story runs into some trouble as IM is generally *not* invariant with respect to the minimal WSS increment property.

Consider (12a-b) below:

(12) a. $\{a, \{a, b\}\}\ c \Rightarrow \{\{a, b\}, \{\underline{a}, \{a, b\}\}\ c$

b. $\{\{a, b\}, \{\underline{a}, \{a, b\}\}\ c \Rightarrow \{\{a, \{a, b\}\}, \{\{a, b\}, \{a, \{a, b\}\}\}\ c$

In (12a), IM has been applied to the output of (11a). The higher copy of {a, b} shields the lower one from access. Although {a, b} has been raised over *a̲*, there is no c-command relation between the highest two *a*'s in the structure. Therefore, *a̲* is accessible. Applying IM once more, to the output of (12a) by raising {a, {a, b}}, we obtain (12b). Although {a, {a, b}} shields its lower counterpart from access, WSS has incremented by 3, as illustrated in (13b-c).[16] Also, in (12a), the WSS has incremented by 2, as illustrated in (13a-b).

(13) a. $\{_1\, a_2, \{_3\, a, b_4\}\}\ c_5$` (WSS: 5)

b. $\{_1 \{_2\, a_3, b_4\}, \{_5\, a_6, \{a, b\}\}\ c_7$ (WSS: 7)

c. $\{_1 \{_2\, a_3, \{_4\, a, b_5\}\}, \{_6 \{_7\, a_8, b_9\}, \{a, \{a, b\}\}\}\}\ c_{10}$ (WSS: 10)

Should "resource-busting" moves such as in (13a-c) always be illicit? Chomsky (p.c.) opines that this kind of IM may be permissible in just one case, when the movement is terminal, i.e. feeds no more operations. Note that (12b) involves the extraction of a phrase, {{a, {a, b}}, that already had a sub-term, {a, b}, extracted earlier in (12a). Remnant Movement (RM) in Germanic is an example of this kind of iterated movement:

(14) $\{VP, \{C, \{\ldots \{INFL, \{\underline{IA}, \{EA, \{v^*, \{R, \underline{IA}\}\}\}\}\}\}\}\}$

[14] Note this operation does not violate the Extension Condition (or the NTC).

[15] It can be argued that MS operates only within a workspace object (and not the entire workspace) because MS applies within IM, and IM itself is limited to a single object (assuming Sideways Merge is not a valid extension of IM).

[16] In (13a-c), subscripts are used to identify and count the relevant accessible terms.

In (14), VP = {R, IA} is extracted to the edge of the clause after IA is first scrambled.[17] (14) is licit in German, see (Müller 1996) for a general discussion of various constraints on RM.[18] We leave a detailed exploration of this idea for future work.

Next, let us consider two cases of vacuous, unbounded movement, characterized in (15a-b) below:

(15) a. $\{\underline{a}, b\} \Rightarrow \{\underline{a}, \{a, b\}\} \Rightarrow \{\underline{a}, \{a, \{a, b\}\}\} \Rightarrow \{\underline{a}, \{a, \{a, \{a, b\}\}\}\} \Rightarrow \ldots$

b. $\{a, \underline{b}\} \Rightarrow \{b, \{\underline{a}, b\}\} \Rightarrow \{a, \{\underline{b}, \{a, b\}\}\} \Rightarrow \{b, \{\underline{a}, \{b, \{a, b\}\}\}\} \Rightarrow \ldots$

In (15a), we iterate the same IM operation, repeatedly raising *a* (indicated by underling), and in (15b), we raise *a* then *b* repeatedly (cycling between the two underlined items). We assume that infinite loop behavior of this kind is both empirically unnecessary in the grammaticalization of thought and computationally inefficient.[19] In (15a), since the highest copy of *a* shields all of the lower *a*'s, each IM operation here obeys the minimal increment condition. Note that the same is true (for highest *a* and *b*) in the case of (15b). Then, how can we rule out (15a-b)?

First, let us consider a naïve solution to the problem before describing the proposed approach. In (15a) above, MS selects the visible copy of *a* at each step. The repeated pattern π_1 is the singular sequence <*a*>. In (15b), MS repeats the pattern π_2 = <*a*, *b*>. to read as *first b, then a*. In an infinite loop, these patterns are repeated endlessly. A suitable filter can then be defined simply as in (16), i.e. no pattern repetition is permitted at all.[20]

(16) $*\pi\pi$

Although the description in (16) is conceptually simple, its implementation requires us to add a new memory device since we need to record and recognize a previously seen pattern π. Needless to say, conceptions of this sort require extraordinary justification and should be avoided if possible. Perhaps the answer lies within the multiple functional unit architecture for BM. Recall the principle of non-interference introduced earlier to rule out cases of parallelism. In (13a), the same BM operation is repeated over and over again. Suppose that non-interference between invocations of the BM units extends over (a short period of) time. Roughly speaking, the same IM operation, i.e. raising *a* in (14a), back-to-back, so to speak,

[17] See also note 10 for a brief description of some of the details elided in (14). Spellout in EXT also makes a significant contribution, e.g. the verbal complex must amalgamate at root C for V2 languages, as head movement is no longer formulable, as suggested in (Chomsky; *forthcoming*). Finally, the copy of IA in the fronted VP must also not be spelled out.

[18] The data seems rather complex and context-dependent. For example, Riny Huybregts (p.c.) has observed that embedded root clauses (exhibiting V2) is possible in Dutch provided the higher clause is also a root clause (but not when the higher clause is verb-final).

[19] One possible view is that the recursive system freely overgenerates in this fashion, but the computed structures are rejected by the interpretative component (INT). A less resource-intensive view is that infinite loops of this sort are simply not generated at all.

[20] (15a-b) represent a simplified and incomplete picture of infinite loops. Other kinds of infinite loops may contain complex patterning not detected by (16), e.g. repeated extraction of the workspace object from the previous step, as in $\{a, b\} \Rightarrow \{b, \underline{\{a, b\}}\} \Rightarrow \{\{a, b\}, \underline{\{b, \{a, b\}\}}\} \Rightarrow \{\{b, \{a, b\}\}, \{a, b\}, \{b, \{a, b\}\}\}\}$ (underlining used to highlight the object from the previous round). For simplicity of presentation, we set these aside.

cannot be performed. Note this checking can come for free, i.e. it does not require a special memory device, provided multiple BM units are available and simultaneously involved in the workspace (as assumed earlier). Broadly speaking, this kind of parallelism seems feasible and subject to non-interference (also assumed earlier).[21] In the case of (15b), we have to assume that the repetition of IM of *b* and IM of *a* can be independently detected before it fades away.

Finally, let us consider the nature of the interface between core UG and INT from the perspective of communication bandwidth, also a third factor consideration. Communication between organic (and artificial) systems is generally a bottleneck.[22] Suppose INT can dip into UG whenever it wants (so to speak) and grab whatever is there for interpretation. Two questions arise. 1) At any given time, the WS may consist of a number of syntactic objects. Does INT transfer the whole WS? The minimalist answer should be no. Then the question becomes: which objects does INT pick and how does it search the WS efficiently? Phases, i.e. full clausal and argument structure objects such as (8), are the natural candidates for WS objects that INT should be able to read, see Chomsky (1998). Therefore, having only Phases accessible to INT is a way of efficiently limiting that search. 2) Suppose INT can dip into UG at the highest possible time resolution, i.e. after each merge operation. This would mean that INT, on seeing each move, would have *de facto* access to the history of Merge computation, therefore violating the Markovian assumption mentioned earlier, access that Merge itself does not have. Crucially, INT does not require this access either.[23]

References

Breitmeyer, B. 1992. Applications of parallel processing in vision / (Vol. 86, Advances in Psychology). North-Holland.

Chomsky, N.A. 1998. Minimalist Inquiries: The Framework. *MIT Occasional Papers in Linguistics* 15.

Chomsky, N.A. 2006. Three Factors in Language Design, *Linguistic Inquiry*, Vol. 36, No. 1, 1–22.

[21] Here is a tentative sketch of the idea. At the beginning of this paper, it was mentioned that chemical synapses cannot fire indefinitely at high frequency. A refractory period is necessary for the neuron to recover. Given the availability of multiple BM units, the task of the next iteration for a (potentially infinite) loop will be taken up by a *fresh* BM unit. If the current and fresh BM unit overlap in time, a matter of a small amount of persistence, the inputs may be compared (as independently needed anyway for non-overlapping inputs). If the inputs are found to be sufficiently similar, the computation can be halted. Therefore, computational resources are conserved.

[22] In organic systems such as in some mammalian brains, the two hemispheres of the brain do not freely communicate. Instead the brain is bandwidth-limited by the corpus callosum, the major interface that connects the two. For example, visual processing proceeds in both contralateral hemispheres. Perception experiments suggest that hemispheric communication is dispreferred, e.g. the motion quartet paradigm in which vertical motion is more likely to be perceived (intrahemispheric computation) vs. horizonal motion (hemispheric communication required), see Genç et al. (2011).

[23] To be more precise, INT reads argument structure and computes the copy/repetition possibilities without needing to know whether constituents have been IM'ed or EM'ed. See Chomsky (forthcoming) for details.

Chomsky, N.A. 2019. *The UCLA lecture: Lecture 4*. Transcript available at https://ling.auf.net/lingbuzz/005485; original video: https://www.youtube.com/playlist?list=PLa6MU-5gBvQkjDhBz_LsI8pU6B7AjPgU5

Chomsky, N.A. *forthcoming*. Genuine Explanation.

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human Photoreceptor Topography. *The Journal of Comparative Neurology*, 292, 497–523.

Fletcher, H. and Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *The Journal of the Acoustical Society of America*, 5, 82–108.

Fong, S. (ms.) *Adversarial Testing of Statistical Parsers: the Case of Temporary Ambiguities*.

Fonseca-Azevedo & Herculano-Houzel. 2012. Metabolic constraint imposes tradeoff between body size and number of brain neurons in human evolution. In *Proceedings of the National Academy of Sciences of the United States of America* , Vol. 109, No. 45. 18571–18576.

Herculano-Houzel, S., P. R. Manger & J. Kass. 2014. Brain scaling in mammalian evolution as a consequence of concerted and mosaic changes in numbers of neurons and average neuronal cell size, *Frontiers in Neuroanatomy*. Vol. 8, Art. 77. 1–28.

Herculano-Houzel, S., K. Avelino-de-Souza, K. Neves, J. Porfírio, D. Messeder, L. Mattos Feijó, J. Maldonado & P. R. Manger. 2014b. The elephant brain in numbers. *Frontiers in Neuroanatomy*, Vol. 8, Art. 46. 1–9.

Hornstein, N. 2009. *A theory of syntax: minimal operations and universal grammar*. Cambridge University Press.

Genç, E., J. Bergmann, W. Singer & A. Kohler. 2011. Interhemispheric Connections Shape Subjective Experience of Bistable Motion. *Current Biology*. Vol 21. 1494–1499.

Georgiou, H.V. 2014. Estimating the intrinsic dimension in fMRI space via dataset fractal analysis: counting the 'cpu cores' of the human brain. arXiv:1410.7100.

Mariën, P., H. Ackermann, M. Adamaszek, C. H. S. Barwood, A. Beaton, J. Desmond, E. De Witte, A. J. Fawcett, I. Hertrich, M. Küper, M. Leggio, C. Marvel, M. Molinari, B. E. Murdoch, R. I. Nicolson, J. D. Schmahmann & C. J. Stoodley, M. Thürling, D. Timmann, E. Wouters & W. Ziegler. 2014. Consensus Paper: Language and the Cerebellum: an Ongoing Enigma. *Cerebellum*. 13:386–410.

Mortensen, H.S., B. Pakkenberg, M. Dam, R. Dietz, C. Sonne, B. Mikkelsen & N. Eriksen. 2014. Quantitative relationships in delphinid neocortex. *Frontiers in Neuroanatomy*, 26.

Müller, G. 1996. A Constraint on Remnant Movement. *Natural Language and Linguistic Theory*. Vol. 14 355–407.

Strobach, T., M. Wendt & M. Janczyk. 2018. Multitasking: Executive Functioning in Dual-Task and Task Switching Situations. *Frontiers in Psychology*. Vol. 9, Art. 108 1–5.

Testa-Silva, G., M. B. Verhoog, D. Linaro, C. P. J. de Kock, J. C. Baayen, R. M. Meredith, C. I. De Zeeuw, M. Giugliano, H. D. Mansvelder. 2014. High Bandwidth Synaptic Communication and Frequency Tracking in Human Neocortex. *PLOS Biology*. Vol. 12, Issue 11.

Tinsley, J. N., Molodstov, M. U., Prevedel, R., Wartmann, D., Espigulé-Pons, J., Lauwers, M., & Vaziri, A. 2016. Direct detection of a single photon by humans. *Nature Communications*, 7, 12172.

Trettenbrein, P.C. 2016. The Demise of the Synapse As the Locus of Memory: A Looming Paradigm Shift? Frontiers in Systems Neuroscience. Vol. 10, Art. 88 1–7.
Patrick C.
Wackermannová, M., L. Pinc, & L. Jebavý. 2016. Olfactory sensitivity in mammalian species. *Physiological Research*, 65, 369-390.