



# Infinite Generation of Language Unreachable From a Stepwise Approach

M. A. C. Huybregts\*

*Utrecht Institute of Linguistics OTS (UIL-OTS), Utrecht University, Utrecht, Netherlands*

Language is commonly thought of as a culturally evolved system of communication rather than a computational system for generating linguistic objects that express thought. Furthermore, language is commonly argued to have gradually evolved from finite proto-language which eventually developed into infinite language of modern humans. Both ideas are typically integrated in accounts that attempt to explain gradual evolution of more complex language from the increasingly strong pressures of communicative needs. Recently some arguments have been presented that the probability of the emergence of infinitely productive languages is increased by communicative pressures. These arguments fail. The question whether decidable languages evolve into infinite language is vacuous since infinite generation is the null hypothesis for a generative procedure. The argument that increasing cardinality leads to infinite language is incoherent since it essentially conflates concepts of performance with notions of competence. Recursive characterization of infinite language is perfectly consistent with finite output. Further, the discussion completely ignores a basic insight that language is not about decidability of weakly generated strings but rather about properties of strongly generated structures. Finally, the plausibility proof that infinite productivity evolves from finite language is false because it confuses (infinite) cardinal numbers with (natural) ordinal numbers. Infinite generation cannot be reached with a stepwise approach.

**Keywords:** evolution, communicative pressures, infinite generation, productivity, computable/decidable, recursive, competence, strong/weak generation

## OPEN ACCESS

### Edited by:

Andrea Moro,  
Istituto Universitario di Studi Superiori  
di Pavia (IUSS), Italy

### Reviewed by:

Cristiano Chesi,  
Istituto Universitario di Studi Superiori  
di Pavia (IUSS), Italy  
Anna Maria Di Sciullo,  
Université du Québec à Montréal,  
Canada

### \*Correspondence:

M. A. C. Huybregts  
m.a.c.huybregts@uu.nl

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 09 January 2018

**Accepted:** 12 February 2019

**Published:** 19 March 2019

### Citation:

Huybregts MAC (2019) Infinite  
Generation of Language Unreachable  
From a Stepwise Approach.  
*Front. Psychol.* 10:425.  
doi: 10.3389/fpsyg.2019.00425

## INTRODUCTION

Language is commonly thought of as a culturally evolved system of communication (Dunbar, 1996, 2017; Tomasello, 2003, 2008; Kirby et al., 2007; Smith and Kirby, 2008; Chater and Christiansen, 2010; Kirby, 2017) rather than a computational system for generating linguistic objects expressive of thought (Chomsky, 2010, 2013, 2017a,b). Furthermore, it is also commonly argued that language has gradually evolved from finite proto-language which eventually developed into infinite language of modern humans (Bickerton, 1990; Pinker and Bloom, 1990; Corballis, 2002; Fitch, 2010; Jackendoff, 2010; Dediu and Levinson, 2013; Christiansen and Chater, 2015; Jackendoff and Wittenberg, 2017). Both ideas are typically integrated in accounts that attempt to explain gradual evolution of more complex language from the increasingly strong pressures of communicative needs (e.g., Smith and Kirby, 2008; Christiansen and Chater, 2015). In a recent article Piantadosi and Fedorenko (2017), P&F hereinafter, argue that “increasing the number of signals in a language increases the probability of languages that have—in fact—infinite cardinality. Thus, across evolutionary time, the productivity of human language could have arisen solely from algorithmic randomness combined with a communicative pressure for a large number of signals.”

Here we will argue that these arguments fail. The question whether decidable languages evolve into infinite language is close to being vacuous since infinite productivity is the null hypothesis for a generative procedure. Decidable sets are equivalent to effectively computable functions (Gödel, 1934), which by default are characterized by infinite productivity and are “obtained in the ideal case where all of the practical restrictions on running time and memory space are removed” (Enderton, 1977, p. 530). Since infinite productivity is the null hypothesis, no communicative pressures are required for an internal system to emerge that conforms to laws of nature and third factor principles. Next, the argument that increasing cardinality leads to infinite language is incoherent since it essentially conflates concepts of *performance* with notions of *competence*. Recursive characterization of infinite language is perfectly consistent with finite output. Further, the discussion completely ignores a basic insight that language is not about decidability of *weakly generated* strings but rather about properties of *strongly generated* structures. Finally, the plausibility proof that infinite productivity evolves from finite language is false because it confuses (infinite) *cardinal* numbers with (natural) *ordinal* numbers. Infinite generation cannot be reached with a stepwise approach.

## INFINITE PRODUCTIVITY OF COMPUTABLE FUNCTIONS

The creative aspect of normal language use has been a nearly constant topic of philosophical/scientific discussion at least since Cartesian times (Descartes, Port-Royal grammarians), and even before (Galileo). Some qualifications are in order, e.g., last century’s structural linguistics, grounded in behaviorist thinking, and recent preoccupation with item- or usage-based learning (Tomasello, 2003, 2008; Culicover and Jackendoff, 2005; Goldberg, 2006; Christiansen and Chater, 2015; Kirby, 2017), grounded in statistics-based domain-general learning from particulars. While the creative use of language is still mysterious, there has been substantial progress in understanding the principles underlying the unboundedness of human language (Chomsky, 2013).

Infinite productivity is the default property of a *computable function* (Turing, 1936; Davis, 1958; Watumull, 2012; Watumull et al., 2014). Consequently, if internal language is a computable function, a generative procedure defined in intension, there cannot be any non-arbitrary limit on its productivity. Imposing a limit requires an arbitrary stipulation, a complication that need be explained. Such explanation is provided by Yang (2016), who gives an explicit and enlightening account of how linguistic productivity is *determined* by his *Tolerance Principle* for arbitrary exceptions in specific domains of language. The alternative, rejecting language as a generative procedure leaves no question to investigate. Not only don’t we know in that case what language is but absent any characterization of the set of well-formed sentences, P&F’s question whether infinite productivity could arise from chance under communicative pressure becomes a mission impossible.

To make this perfectly clear, take one of the simplest formal languages imaginable,  $L = \{x \mid x \in \{a\}^*\}$ , the language consisting of all strings on  $\{a\}$ .  $L$  contains an infinite number of strings, sequences of the symbol  $a$ , which could be recursively generated by a FS grammar  $G(L)$  with two rules only: (i)  $S \rightarrow a S$ , and (ii)  $S \rightarrow a$ . These rules generate  $a, aa, aaa$ , etc. A “simpler” finite language  $L'$  consisting of, say, five strings only,  $L' = \{a, aa, aaa, aaaa, aaaaa\}$ , can be characterized only by a more complex grammar  $G(L')$ , which has no less than nine rules.

- (1)  $S \rightarrow a$
- (2)  $S \rightarrow a S_1$ ,      (3)  $S_1 \rightarrow a$
- (4)  $S \rightarrow a S_2$ ,      (5)  $S_2 \rightarrow a S_1$
- (6)  $S \rightarrow a S_3$ ,      (7)  $S_3 \rightarrow a S_2$
- (8)  $S \rightarrow a S_4$ ,      (9)  $S_4 \rightarrow a S_3$

Grammar  $G(L')$  generates all five strings of  $L'$  and only these, permitting a maximum embedding depth of four. Here complexity of a grammar  $G(L)$  for any language  $L$  is understood as the length  $l(G)$  of the shortest description representing grammar  $G$  as measured by the number of rules relative to the cardinality of language  $L$ : the closest is  $l(G)$  to the number of signals, the more complex the language is. Evidently, finite generation is vastly more complex than generation of an infinite set. Infinite generation is therefore the null hypothesis. True, for finite languages, providing a finite list (or a set of rules that generate one terminal string each) may be simpler than any grammar generating the list’s sentences. But in the example above even this finite list is more complex than the recursive grammar generating the infinite language with the arbitrary limit removed. For natural language the situation turns out to be dramatically more complex. Its generative grammar, a computable function, must therefore be recursive by default. Consequently, finite generation with its vast complexity could not have been an earlier step in evolution of language, precursing the simplicity of infinite generation.

There is no longest sentence in human language, which as a consequence must be infinite in size. Therefore, generating the unbounded array of structures requires a finitely specified recursive procedure. Now P&F “consider only languages  $L$  that are decidable,” but since decidable sets are equivalent to computable functions (Gödel, 1934), the latter ideally characterized by infinite productivity, infinity must be the null hypothesis for decidable sets as well. The conclusion must be that computable functions do not evolve into recursive systems with infinite productivity, because by default they start out being infinitely productive. Since human language is infinite and infinite generation is the null hypothesis, any limit on recursive application would be arbitrary, requiring an explanation. No evolution is required, therefore, and, in fact, as discussed in section Infinite Cardinal Numbers as Unattainable Limits, the idea of a stepwise approach to infinity is incoherent. P&F already presuppose what they set out to prove. The question of gradual evolution as a result of external pressures simply does not arise.

Still, this line of reasoning is not unique but shared along a broad spectrum of cognitive scientists (e.g., Scott-Phillips and Kirby, 2010; Christiansen and Chater, 2015; Kirby, 2017). These three studies are further misguided by hidden assumptions.

Successful iterated-learning from finite collections of items actually happens in modern minds already fully equipped with recursive language/compositional meaning. Consequently, in these experiments “cultural evolution” does not contribute anything that wasn’t already there before (Berwick and Chomsky, 2017). Rather these arguments presuppose what they conclude. Capacity for infinite generation need not be learned. It is something humans start out with in life. A recursive system has evolved, clearly, but its evolution must have been saltational, perhaps as a result of cerebral reorganization of an expanding brain. Language is therefore a biological isolate, a novelty, and emerged suddenly, recently in evolutionary time, and as far as we know with little or no evolutionary change since (Chomsky, 2010, 2017a,b; Berwick and Chomsky, 2011, 2016, 2017; Tattersall, 2012, 2017; Huybregts, 2017). In particular, no group differences have been found since we left Africa, and every child, African or non-African, can grow and develop any language of the community she is born into.

There is some intriguing evidence from computational simulation that suggests that nature would select the simplest stable recursive functions, the ones that produce the successor function (Minsky, 1985). And precisely this may have happened in the evolution of human language, which essentially is such a recursive function. No communicative pressures are needed to explain the urge to merge in language. Language’s computational optimality may justify a bolder but no less cogent or plausible conjecture that “the basic principles of language may be drawn from the domain of (virtual) conceptual necessity” (Chomsky and Watumull, in a book chapter currently in press). Universal grammar would then be inevitably universal for both humans and aliens: absent any external selective pressures on an internal system, nature would converge on the same class of computational procedures that conform essentially to laws of nature and third factor principles only [Roberts, Watumull, and Chomsky, presentation “Universal Grammar” at the International Space Development Conference (ISDC), Los Angeles, 2018].

## INFINITE PRODUCTIVITY AND FINITE OUTPUT

We should be careful, however, not to confuse the *infinite productivity* of the generative procedure (a function in intension) with the *cardinality* of the decidable set of structures it generates (the well-formed formulas defined in extension). To recursively characterize a set, applying an intensional function, is not to produce the set, represented as its output. Watumull et al. (2014) offers a principled discussion of these aspects of recursion theory. Without contradiction, finite outputs can be generated by recursive systems. See **Figure 1** for an illustration of this.

Within the SMT framework of current generative theories (Chomsky, 2000, 2015, 2017c), a more empirically informed case may be the morpho-syntactic constraints on case stacking in Kayardild syntax that effectively restrict the output of the recursive operation to a single embedding at most (Evans and Levinson, 2009). The morpho-syntactic conditions restraining

recursive embedding of possessive noun phrases in German or Dutch constitute another cogent and convincing case. There are two strategies for handling these possessive constructions in standard Dutch, a fully productive strategy that involves possessive agreement between possessor and possessee, and a partially productive one that involves genitive case agreement. The latter is more archaic and restricted to proper names, kinship terms and a few “genitive fossils” in the language. The discussion below is simplified somewhat for convenience (phrases (1a)/(2a) mean “the man’s aunt,” and phrases (1b)/(2b) mean “the man’s aunt’s bike”).

- (1) a. [DP [DP de **man**] **z’n** tante ] (but \* *’s man z’n tante*)  
 b. [DP [DP [DP de man ] z’n **tante**] **d’r** fiets ]
- (2) a. [DP [GenDP *’s* man ] -s tante ] (but\* *de mans tante*).  
 b. [DP [GenDP [GenDP *’s* man ] -s tante ] -s fiets ]

Modulo lexical restrictions on genitive case, each construction permits free recursion. Both structures are characterized by agreement between determiner and determiner phrase, Agree(D, DP): person/number/gender agreement in (1) and Case agreement in (2).

In (2) the case-marked head (recursively) agrees in Case with the case-marked determiner phrase in its specifier position. I.e., genitive case-marker “-s” (cliticized to the noun *man*) agrees with genitive determiner *’s* (contracted form of *des*) in (2a), and there is recursive agreement in (2b), where genitive “-s” of *tantes* agrees with genitive determiner phrase *’s mans*. In (1) possessive agreement features of the determiner recursively agree with its determiner phrase in person, number and gender: *z’n* ~ *de man* and *d’r* ~ *de man z’n tante*.

Interestingly, hybridization leads to a curious asymmetry of recursion, explained on present assumptions.

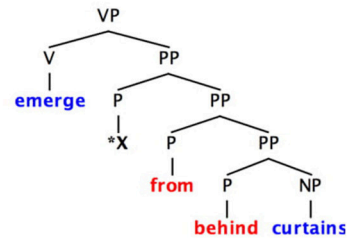
- (3) a. [DP [DP [GenDP *’s* man ] -s **tante**] **d’r** fiets ]  
 b. \* [DP [GenDP [DP de **man**] **z’n** tante ] -s fiets ]

Free recursion is constrained here by an extraneous morphosyntactic factor: genitive case agreement cannot apply (but must apply) in (3b), whereas possessor-agreement must apply (and can apply) in (3a). Both the asymmetry and the constraint (“criterial agreement”) on recursive output receive a principled explanation.

While the limiting conditions on recursive output may receive independent and principled explanations, in the Dutch case labeling requirements on exocentric constructions (Chomsky, 2013, 2015), it is important to stress that these are independent of the recursive computation, and are extraneous to it. The widely publicized Pirahã case (Everett, 2005; Nevins et al., 2009) is no different. The basic operation Merge is part of the faculty of language, the human genetic endowment for language, and applies without exception in humans anywhere, in every language, and in every construction. Recursiveness is a property of the generative procedure applicable to any input, not a feature of its output, which may be arbitrarily constrained by complicating idiosyncratic factors independent of the procedure. The procedure may generate an infinite language but only produce a finite subset of it.

Context-sensitive grammar  $G$  containing rewrite rules (i) through (vii):

- (i)  $VP \rightarrow V PP$
- (ii)  $PP \rightarrow P PP$
- (iii)  $PP \rightarrow P NP$
- (iv)  $V \rightarrow \textit{emerge}$
- (v)  $P \rightarrow \textit{behind} / \textit{---} NP$
- (vi)  $P \rightarrow \textit{from} / \textit{---} \textit{behind}$ .
- (vii)  $NP \rightarrow \textit{curtains}$



A.  $VP \Rightarrow V PP \Rightarrow V P NP \Rightarrow V \textit{behind} NP$

B.  $VP \Rightarrow V PP \Rightarrow V P PP \Rightarrow V P P NP \Rightarrow V P \textit{behind} NP \Rightarrow V \textit{from} \textit{behind} NP$

**FIGURE 1** | Recursive grammar but finite output. CS Phrase Structure Grammar  $G$  runs a recursive program that generates an indefinitely large number of nested prepositional phrases. Still its output is idiosyncratically restricted by context-sensitive rules that allow production of a finite set of linguistic objects only. Rule (ii) applies recursively but  $G$ 's output is strictly finite, (A) [*emerge [behind curtains]*] and (B) [*emerge [from [behind curtains]]*], due to the context-sensitive restrictions of rules (v) and (vi). These must be considered arbitrary lexical constraints relative to the freely applying recursive rule (ii). The phrase marker corresponding to the tree in this figure cannot terminate in a well-formed verb phrase.

Beyond the cases discussed above, there are other conclusive arguments for distinguishing a recursive generative procedure from its (in)finite output. Humans have I-language, a biological system with a computational capacity for generating a discrete infinity of hierarchical structures ("competence"), but their linguistic output ("performance") must necessarily be strictly finite for other, biophysical reasons, quite extraneous to the generative procedure itself. Evidently, "unboundedness is a property of generative competence, not its application in performance" (Watumull et al., 2014).

Unfortunately, P&F do not stand alone in failing to distinguish these significant concepts. E.g., Christiansen and Chater (2015) recently argued for an account of the human ability to process "a limited amount of recursive structure in language" that derives from domain-general constraints on sequence learning, and "does not rely on recursion as a property of grammar." According to Christiansen and Chater, "[T]he ability does not necessitate the postulation of recursive mechanisms to process them." However, their account, building on construction grammar and usage-based approaches to language, leaves the problems unresolved. How are "constructions" and "limited recursive structure" characterized? Recursive characterization is ruled out by them apodictically. It is argued that "language is shaped by the brain" and but there is no answer to the question of why, how or when the brain did evolve a capacity to learn precisely those languages that are UG-compliant. As a result, none of these questions receive a decent answer. It is just not enough to say that what is observable is real (performance), and it is incorrect to assume that what is invisible is unreal (competence). It is the task of the sciences to reduce "complex visibles" to "simple invisibles" (Jean Baptiste Perrin, Nobel laureate in physics, 1926). Clearly, the distinction between the properties of a system and the use of that system cannot be in doubt. *Pace* empiricist prevalence, from finite output alone *nothing* can be inferred for internal language.

The suggestion that a theory that successfully ignores the distinction between competence (internalized system) and

performance (externalized use of the system) should be the null-hypothesis (Christiansen and Chater, 2015; Piantadosi and Fedorenko, 2017) has the story backwards. The notion of an error of performance only makes sense by virtue of the existence of competence. How else could anything be an "error"? Further, there are double dissociations. There are competence effects that have no performative explanation, e.g., structure dependence of rules of grammar (Berwick et al., 2011). On the other hand, there are performative effects that cannot be explained by competence (self-embedding, garden path, blindspots). There may be peripheral impairments that preclude performance but leave competence intact. There are additional asymmetries in performance such as "agreement attraction errors" (Bock et al., 2001, and others). E.g., performance errors like *The key to the cabinets are on the table* occur significantly more frequently than attraction errors like *The keys to the cabinet is available*. To account for these skewed "attraction errors" we need an adequate theory of performance (probability) that presupposes competence (degrees of grammaticalness). Distinguishing between competence and performance is the null hypothesis. Deviation from it needs an argument and a "proof" of explanatory superiority.

## STRONGLY GENERATING THE LANGUAGE PHENOTYPE

The *triple-I* qualification of language (*internal, individual, intensional*) highlights the elementary distinction between a generative procedure ("competence") and its output production ("performance"). Language is *internal* to the *individual* in whose brain it is represented as a function in *intension*, a generative program (Chomsky, 1986). As a result, since recursion is a property of the computational procedure, not a characteristic feature of the program's output, it follows, in particular, that from a set of well-formed formulas nothing can be concluded

about the nature of the generative procedure that generates them. Further, the notion of decidable sets relates to properties of their *weak generation*, the finite procedure that decides for each input if it belongs (“is grammatical”) or does not belong (“is ungrammatical”) to the set. But for the study of *natural* (as opposed to *formal*) language, *decidability* of sets, *weak generative capacity* (WGC), is not even a relevant notion, in sharp contrast with the generative procedure and its *strong generative capacity* (SGC), which is fundamental (Chomsky, 1956, 1965, 2007; Berwick, 1984). Despite much confused parlance otherwise, decidability of grammaticality, relevant for formal language, is hardly an issue in the study of natural language, for which no surface structure characterization can be given.

For natural language, characterization of its unbounded array of hierarchical structures and their properties is fundamental, and proceeds from a finitary generative procedure. We can exemplify these matters, discussing some properties of the basic recursive operation Merge, defined as  $Merge(X, Y) = \{X, Y\}$ , where  $X, Y$  is a lexical item or a phrase  $Z$  that is itself the result of  $Merge(X', Y')$ . Binary  $Merge(X, Y)$  is thus an operation that takes two elements  $X$  and  $Y$  —  $X, Y$  atomic elements or results of a previous Merge application — and constructs a set out of them that contains just these elements as its members, imposing no further structure or ordering arrangement on them (Chomsky, 2015; Everaert et al., 2015). Since recursion is strictly the property of a rule to reapply to its own output, free Merge re-applying to its own output is recursive in this sense. Moreover, Merge must be allowed to operate on multiple syntactic objects in the workspace, applying in parallel to generate, e.g., both the syntactic objects  $\{this, man\}$  and  $\{loves, \{that, woman\}\}$  before merging these to construct a new syntactic object  $\{\{this, man\}, \{loves, \{that, woman\}\}\}$ . Consequently, freely applying Merge, defined by induction,  $\mu(x, y) = \{x, \mu(x', y')\}$ , will be *strongly generative* of structure, allowing for “unbounded expansion of the strongly generated structure.” Furthermore, “by these necessary and sufficient criteria the grammars of all natural languages are recursive” (Watumull et al., 2014). E.g., the structure externalized as “this man loves that woman” is defined by induction and strongly generated by recursive merge applications that cyclically build increasingly complex hierarchical structure:  $\mu(\mu(this, man), \mu(loves, \mu(that, woman))) = GP \Rightarrow \{\{this, man\}, \{loves, \{that, woman\}\}\}$ .

For the study of language, not *grammaticality*, a property of *weak generative adequacy*, but rather *degree of grammaticality*, a property of *strong generative adequacy*, is the linguistically significant concept (Chomsky, 1956, 1965). Formal languages exist by virtue of an *extensional* definition, and any *computational procedure* generating the language will suffice. It is only a *matter of choice*, not an empirical issue. In contrast, natural language cannot be characterized extensionally and are defined only *intensionally*. There is no choice of grammar (I-language) here: only one specific grammar must be selected since biology requires a single specific generative procedure, thus posing an empirical problem.

What is relevant is the nature of the intensional function that generates a discrete infinity of hierarchically structured

expressions, a trait of the language phenotype, not the cardinality of the decidable set it generates. We want to know e.g., why the sentence *John knows a sharper lawyer than Bill* is ambiguous (i.e., between ... *than Bill does* and ... *than Bill is*), but why the sentence *John knows a lawyer who is sharper than Bill* shows constrained ambiguity (i.e., only ... *than Bill is* could be a well-formed continuation). Note that underlying structures *John knows a sharp lawyer* and *John knows a lawyer who is sharp* are semantically equivalent. Therefore, disambiguation has no semantic, pragmatic or communicative causes but must result from properties of the computable program generating their syntactic structures (Everaert et al., 2015).

In fact, comparative constructions are analyzable as a special case of (unbounded) wh-movement of the compared element, and are constrained by conditions on movement (Chomsky, 1977). Examples like *John is richer than (what) you might expect him to be* or *You look much taller than (what) you really are* illustrate this movement theory of comparative deletion. Turning now to the ill-formed sentence, *\*John knows a lawyer who is sharper than Bill does*, we can see that ellipsis has applied to a VP, viz.  $[VP \text{ know a lawyer who is } ec]$ , which already contains a Complex NP Island violation (Chomsky, 1977). The derivation does not converge and non-convergence is explained by restrictions on computational resources, e.g., a phase impenetrability condition on merge-generated phrase structure (Chomsky, 2000). Complex analysis of hierarchical structure (strong generative capacity), not information about class membership (weak generative capacity), provides us with a clear understanding of why ambiguity is constrained, unifying comparative deletion with unbounded movement. Significantly, we do not content ourselves with simply observing that expressions are (un)acceptable. Even if there would be an observationally adequate approximation of acceptability there is no answer as to why they are (un)acceptable. Analogously, we want to know why the stars shine, not merely observe that they shine (Feynman, 1963/2011).

In natural language, a well-defined Merge operation strongly generates a digitally infinite array of hierarchically structured expressions that each receive systematic and determinate interpretations at the cognitive and sensorimotor systems interfacing with internal language. The interface mappings show asymmetry, with the mapping to the thought systems being primary and externalization at the sensorimotor systems ancillary (Everaert et al., 2015, 2017; Huybregts et al., 2016; Chomsky, 2017a,b). Precise insight into the nature of the generative procedure, the properties of the hierarchical structures it generates (Chomsky, 1965, 1981, 2007, 2015, 2017c; Everaert et al., 2015), its growth and development in the language learning child (Crain and Thornton, 2011; Crain, 2012; Piattelli-Palmarini and Berwick, 2012; Yang et al., 2017), development and representation in the human brain (Musso et al., 2003; Moro, 2016; Friederici et al., 2017) as well as its evolutionary origin (Tattersall, 2012, 2017; Berwick and Chomsky, 2016, 2017; Huybregts, 2017) are the primary goals of generative linguistic research, not class membership of individual sentences.

Study of evolution of the language faculty (UG) should therefore be primarily concerned with the emergence of a

specific generative procedure in evolutionary time, and within a narrow evolutionary channel defined by historical contingencies and biophysical constraints (Monod, 1972), not with the probability of an infinitely productive language arising “solely from algorithmic randomness combined with a communicative pressure for a large number of signals.” Communicative efficiency does not govern language though computational efficiency does, unambiguously, ubiquitously, and unexceptionally (Chomsky, 2010, 2013, 2017a,b; Everaert et al., 2015). Algorithmic randomness begs the question of the prior availability of algorithms, computable functions with an inherently infinite productivity. The assertion that infinite human language may have arisen “solely” from these two factors only exacerbates the incongruence and incoherence of the proposal. Evidently, UG is a *species-specific* and *domain-specific* biological system that did not evolve from algorithmic randomness or communicative pressure. For better informed accounts of why language came to stick with only us see (e.g., Tattersall, 2012, 2017; Bolhuis et al., 2014; Hauser et al., 2014; Berwick and Chomsky, 2016, 2017; Chomsky, 2017a,b).

Other linguists, e.g., Bickerton (1990, 2000), Jackendoff (1999, 2002), Jackendoff and Wittenberg (2017), Progovac (2009, 2016) have argued for a proto-language precursor to modern language (but see e.g., Di Sciullo, 2013; Miyagawa and Nobrega, 2015 for opposing views). However, these arguments become irrelevant to the problem of evolution of the language capacity once we realize that human language is recursive. Finite language gradually evolving into infinite language is a logical impossibility (section Infinite Cardinal Numbers as Unattainable Limits below). Conceivably, proto-language may have existed as a system of communication (and probably has), but only as an evolutionary *cul-de-sac*, and not as a precursor to human language. Of course, that leaves open the possibility that surface relics and remnants of proto-language may have been incorporated/integrated into the novel and saltationally evolved language phenotype. But this is a different matter altogether and has no relevance for *evolvability* of recursive language. Finally, it is really a speculation outside the reach of empirical testing to argue that constructions like “step-by-step” or (head-initial) exocentric compounds (French *timbre poste*, *café filtre* and English *scarecrow*, *killjoy*, or *turncoat*) are “fossils of proto-language.” Language does not fossilize. At best, on current understanding, these cases show analytical problems waiting to be resolved under a unified merge-based approach to syntax and morphology.

## INFINITE CARDINAL NUMBERS AS UNATTAINABLE LIMITS

P&F confuse the cardinality of (decidable) sets, defined in extension, with the (infinite) productivity of computable functions, defined in intension. A cardinal number is not a specific number *identifying a member* of the set (i.e., an ordinal number, a property of an *individual member*) but rather a number *characterizing the size* of the set (i.e., a property of the set). For *finite* sets there is a *greatest* number  $n$ , the ordinal

that has no successor in the set, which can be approximated and reached in a finite number of steps. But for an *infinite* set there is no such greatest number: for any ordinal  $n$  there is a least ordinal greater than  $n$  (its *successor* ordinal). Infinite cardinal numbers describe sizes of infinite sets, and may be better thought of as *limit* ordinals.  $\text{Lim}(n) = \aleph_0$ , where *Aleph-zero* denotes the cardinality of the set of natural numbers, can never be reached by adding up finitely many finite numbers.

Natural language, likewise, is infinite, since there is no longest sentence. Recursive merge may expand a bounded range to an unbounded range of output structures, but no finite set of expressions, however large, can reach unboundedness by combining finitely many finite constructions. A stepwise approach cannot reach infinite generation. Communicative pressures for more signals, necessarily finite numbers of them, could never be pressures for an infinite language. Consequently, the generative procedure couldn't possibly have evolved gradually but must have emerged saltationally without any pressures from the external environment and rather as an exaptive side-effect of cerebral reorganization in *Homo*.

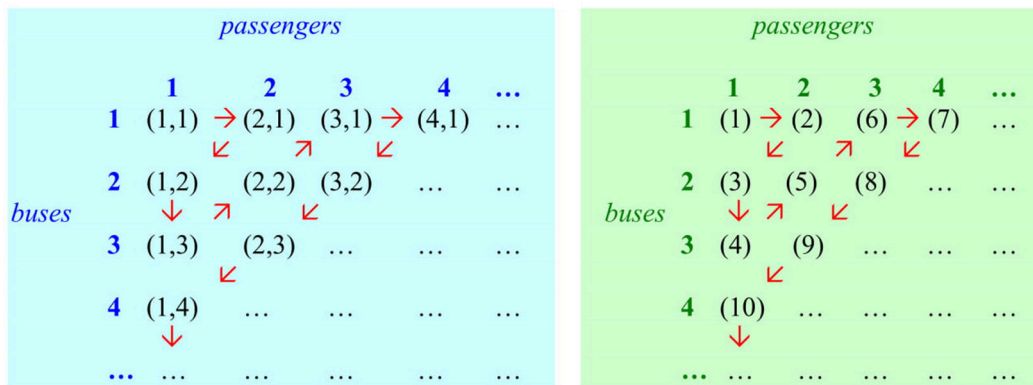
Infinite language has cardinality  $\aleph_0$ . Since cardinality addition and cardinality multiplication do not change cardinality, there is no sense in which infinity can be approximated by a succession of finite additions. But denying precisely that core property lies at the root of P&F's proof of Theorem 2. The proposal is therefore incoherent.

*Theorem 2.* If at least one infinite language has non-zero probability (e.g.,  $\mathbf{P}[\text{crd}(L) = \infty] > 0$ ), then  $\mathbf{P}[\text{crd}(L_p) = \infty \mid \text{crd}(L_p) > B] \rightarrow 1$  as  $B \rightarrow \infty$

The theorem says that “if  $L$  is constrained to contain at least  $B$  strings,  $L$  will be infinitely productive with increasing probability as  $B$  gets large.” Applying the definition of conditional probability  $\mathbf{P}(X|Y) = \mathbf{P}(X \cap Y)/\mathbf{P}(Y)$  to the antecedent of the conditional yields (4).

$$(4) \quad \mathbf{P}[\text{crd}(L_p) = \infty \mid \text{crd}(L_p) > B] = \frac{\mathbf{P}[\text{crd}(L_p) = \infty]}{\mathbf{P}[\text{crd}(L_p) = \infty] + \mathbf{P}[B < \text{crd}(L_p) < \infty]}$$

The second term  $\mathbf{P}[B < \text{crd}(L_p) < \infty]$  in the denominator says that the probability of the size of  $L_p$  staying intermediate between cardinality  $B$  and infinity gets increasingly small as  $B$  increases. The idea is that the bigger  $B$  gets, the closer to *Aleph-zero* the size of  $L_p$  will be. “As  $B$  increases, the second term in the denominator must approach zero, meaning that the fraction goes to 1” (P&F, p. 144). However, the probability never reaches zero since it is impossible to reach, or even approximate, infinite generation from finiteness by a stepwise approach adding finitely many finite signals to a finite class of signals. The “quantum jump” from 999,999,999 to infinity is no less than the “jump” from nine to infinity. This is essentially *Hilbert's Infinite Hotel* argument (Hilbert, 1924/1926), as explained informally by Gamow (1947) and illustrated in **Figure 2**.



**FIGURE 2 |** There's always room in Hilbert's Infinite Hotel. Assume that the hotel is already fully booked with an infinite number of permanent guests, and there are arriving an infinite number of coaches, each with an infinite number of seats. Can the new arrivals be suitably accommodated? Yes, they can. *Unvorstellbar* and yet inconceivably true! Strangely enough, the answer is surprisingly simple. For column *X* (passengers) and row *Y* (buses) there is a one-to-one correspondence between the set of ordered pairs of  $N \times N$  (representing passenger *x* on bus *y*) and the set of integers  $N$  (representing room number).

Pairing function:  $f(x,y) = [(x+y-1)(x+y-2)]/2 + y$ .

Since these sets are equivalent, with the same cardinality,  $N \times N$  passengers can be accommodated perfectly well in a hotel with  $N$  rooms (where  $N$  designates cardinality of the set of positive integers). Diagonal method yields e.g.,  $(1,1) \rightarrow 1, (2,1) \rightarrow 2, (1,2) \rightarrow 3, (1,3) \rightarrow 4, (2,2) \rightarrow 5, (3,1) \rightarrow 6, (4,1) \rightarrow 7, \dots$  For example, passenger 3 on bus 1 is given room 6. Here we assume that the permanent guests have arrived at some earlier time with coach 1.

## CONCLUDING REMARKS

The arguments reviewed make (hidden) assumptions that are questionable or even incorrect. Rather than a system of thought, they take language to be basically a means of communication, governed by communicative rather than computational efficiency. They presume that only gradual rather than saltational evolution of a computable function is realistic. Furthermore, they suppose that the study of language concerns decidability of strings (WGC) rather than the *Basic Principle* of language (SGC), and, finally, they ignore the foundational asymmetry of interface mappings, with the mapping to the cognitive systems (internal language) taking primacy over the mapping to the sensorimotor systems (externalization), both in function and evolutionary time. This way, uniformity and universality (principles of UG) are incorrectly dispreferred to diversity, mutability and complexity (parameters and externalization). These and other issues are regrettably ignored or dismissed, but are of paramount importance and scientific interest in studying evolution of language. Here we have argued that unbounded productivity

of language is the null hypothesis. The complexity of finite language cannot have been a precursor to the simplicity of infinite generation in evolution of language. Furthermore, no stepwise approach to language can reach infinite generation. Therefore, gradual evolution of recursive language must be ruled out. The effort to make that explicitly clear here will hopefully help safeguard generative grammar against potential charges of contributory negligence.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

## ACKNOWLEDGMENTS

This article has benefitted from helpful suggestions and advice from Bob Berwick, Noam Chomsky, Martin Everaert, Charles Yang, and Jeffrey Watumull. Furthermore, I am indebted to both reviewers for critical comments and suggestions for improvement.

## REFERENCES

Berwick, R. C. (1984). Strong generative capacity, weak generative capacity, and modern linguistic theories. *Comput. Linguist.* 10, 189–202.  
 Berwick, R. C., and Chomsky, N. (2011). “The biolinguistic program: the current state of its development,” in *The Biolinguistic Enterprise*, eds A. M. Di Sciullo and C. Boeckx (Oxford: Oxford University Press), 19–41.  
 Berwick, R. C., and Chomsky, N. (2016). *Why Only Us: Language and Evolution*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262034241.001.0001

Berwick, R. C., and Chomsky, N. (2017). Why only us: recent questions and answers. *J. Neurolinguist.* 43, 166–177. doi: 10.1016/j.jneuroling.2016.12.002  
 Berwick, R. C., Pietroski, P., Yankama, B., and Chomsky, N. (2011). The poverty of the stimulus revisited. *Cogn. Sci.* 35:1207. doi: 10.1111/j.1551-6709.2011.01189.x  
 Bickerton, D. (1990). *Language and Species*. Chicago: University of Chicago Press.  
 Bickerton, D. (2000). “How protolanguage became language,” in *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, eds C. Knight, M. Studdert-Kennedy, and J. R. Hurford (Cambridge: Cambridge University Press), 264–284.

- Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., and Schriefers, H. (2001). Some attractions of verb agreement. *Cogn. Psychol.* 43, 83–128. doi: 10.1006/cogp.2001.0753
- Bolhuis, J. J., Tattersall, I., Chomsky, N., and Berwick, R. C. (2014). How could language have evolved? *PLoS Biol.* 12:e1001934. doi: 10.1371/journal.pbio.1001934
- Chater, N., and Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cogn. Sci.* 34, 1131–1157. doi: 10.1111/j.1551-6709.2009.01049.x
- Chomsky, N. (1956). Three models for the description of language. *IRE Trans. Inform. Theor.* IT-2, 113–124. doi: 10.1109/TIT.1956.1056813
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1977). “On wh-movement,” in *Formal Syntax*, eds P. Culicover, T. Wasow, and A. Akmajian (New York, NY: Academic Press), 71–132.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris Publications.
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York, NY: Praeger.
- Chomsky, N. (2000). “Minimalist inquiries: The framework,” in *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, eds R. Martin, D. Michaels, and J. Uriagereka (Cambridge, MA: MIT Press), 89–156.
- Chomsky, N. (2007). Review of Margaret Boden’s “Mind as Machine: A History of Cognitive Science.” *Artif. Intell.* 171, 1094–1103. doi: 10.1016/j.artint.2007.10.005
- Chomsky, N. (2010). “Some simple evo-devo theories about language: how might they be true for language?” in *The Evolution of Human Language: Biolinguistic Perspectives*, eds R. K. Larson, V. Déprez, and H. Yamakido, H. (Cambridge, UK: Cambridge University Press), 45–62.
- Chomsky, N. (2013). What kind of creatures are we? The Dewey lectures. Lecture I: What is language? Lecture II: What can we understand? *J. Philos.* 12, 645–700. doi: 10.5840/jphil2013110121
- Chomsky, N. (2015). “Problems of projection: extensions,” in *Structures, Strategies and Beyond. Studies in Honour of Adriana Belletti*, eds E. Di Domenico, C. Hamann, and S. Matteini, S. (Amsterdam/Philadelphia: Benjamins Publishing Company), 1–16. doi: 10.1075/la.223.01cho
- Chomsky, N. (2017a). Language architecture and its import for evolution. *Neurosci. Biobehav. Rev.* 81(Pt. B), 295–300. doi: 10.1016/j.neubiorev.2017.01.053
- Chomsky, N. (2017b). The Galilean challenge. *Inference: Int. Rev. Sci.* 3.
- Chomsky, N. (2017c). Some Puzzling Foundational Issues (lecture presented at the University of Reading, 11 May 2017).
- Christiansen, M. H., and Chater, N. (2015). The language faculty that wasn’t: a usage-based account of natural language recursion. *Front. Psychol.* 6, 1–18. doi: 10.3389/fpsyg.2015.01182
- Corballis, M. C. (2002). *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.
- Crain, S. (2012). *The Emergence of Meaning*. Cambridge, UK: Cambridge University Press.
- Crain, S., and Thornton, R. (2011). Syntax acquisition. *Wiley Interdiscipl. Rev. Cogn. Sci.* 3, 185–203. doi: 10.1002/wcs.1158
- Culicover, P. W., and Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.
- Davis, M. (1958). *Computability and Unsolvability*. New York, NY: McGraw-Hill Book Company.
- Dediu, D., and Levinson, S. (2013). On the antiquity of language. *Front. Psychol.* 4:397. doi: 10.3389/fpsyg.2013.00397
- Di Sciullo, A.-M. (2013). Exocentric compounds, language and proto-language. *Lang. Inf. Soc.* 20, 1–26.
- Dunbar, R. I. M. (1996). *Grooming, Gossip and the Evolution of Language*. London: Faber and Faber.
- Dunbar, R. I. M. (2017). Group size, vocal grooming and the origins of language. *Psychon. Bull. Rev.* 24, 209–212. doi: 10.3758/s13423-016-1122-6
- Enderston, H. B. (1977). “Elements of Recursion Theory,” in *Handbook of Mathematical Logic*, ed J. Barwise (Amsterdam: North-Holland), 527–66. doi: 10.1016/S0049-237X(08)71114-5
- Evans, N., and Levinson, S. (2009). The myth of language universals. *Behav. Brain Sci.* 32, 429–492. doi: 10.1017/S0140525X099094X
- Everaert, M. B. H., Huybregts, M. A. C., Berwick, R. C., Chomsky, N., Tattersall, I., Moro, A., et al. (2017). What is language and how could it have evolved? *Trends Cogn. Sci.* 21, 569–571. doi: 10.1016/j.tics.2017.05.007
- Everaert, M. B. H., Huybregts, M. A. C., Chomsky, N., Berwick, R. C., and Bolhuis, J. J. (2015). Structures, not strings: linguistics as part of the cognitive sciences. *Trends Cogn. Sci.* 19, 729–743. doi: 10.1016/j.tics.2015.09.008
- Everett, D.L. (2005). Cultural constraints on grammar and cognition in Pirahã: another look at the design features of human language. *Curr. Anthropol.* 46, 621–646. doi: 10.1086/431525
- Feynman, R. P. (1963/2011). *Six Easy Pieces*. New York, NY: Basic Books.
- Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511817779
- Friederici, A. D., Chomsky, N., Berwick, R. C., Moro, A., and Bolhuis, J. J. (2017). Language, mind and brain. *Nat. Human Behav.* 1, 713–722. doi: 10.1038/s41562-017-0184-4
- Gamow, G. (1947). *One Two Three Infinity: Facts and Speculations of Science*. New York, NY: Viking Press.
- Gödel, K. (1934). “On undecidable propositions of formal mathematical systems,” in *Kurt Gödel: Collected Works, Vol. I, Publications 1929–1936*, eds S. Feferman, J. W. Dawson, S. C. Kleene, G. H. Moore, R. M. Solovay, and J. Van Heijenoort (Oxford: Oxford University Press), 346–371.
- Goldberg, A. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., et al. (2014). The mystery of language evolution. *Front. Psychol.* 5:401. doi: 10.3389/fpsyg.2014.00401
- Hilbert, D. (1924/1926). Über das Unendliche. *Mathematische Annalen* 95, 161–190. [Also published in *David Hilbert’s Lectures on the Foundations of Arithmetics and Logic 1917-1933*, (2013), eds W. Ewald and W. Sieg (Heidelberg: Springer-Verlag), 668–756.]
- Huybregts, M. A. C. (2017). Phonemic clicks and the mapping asymmetry: how language emerged and speech developed. *Neurosci. Biobehav. Rev.* 81(Pt. B):279–294. doi: 10.1016/j.neubiorev.2017.01.041
- Huybregts, M. A. C., Berwick, R. C., and Bolhuis, J. J. (2016). The language within. *Science* 352:1286. doi: 10.1126/science.aag0933
- Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends Cogn. Sci.* 3, 272–279. doi: 10.1016/S1364-6613(99)01333-9
- Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Jackendoff, R. (2010). “Your theory of language evolution depends on your theory of language,” in *The Evolution of Human Language: Biolinguistic Perspectives*, eds R. K. Larson, V. Déprez, and H. Yamakido (Cambridge, UK: Cambridge University Press), 63–72.
- Jackendoff, R., and Wittenberg, E. (2017). Linear grammar as a possible stepping-stone in the evolution of language. *Psych. Bull. Rev.* 24, 219–224. doi: 10.3758/s13423-016-1073-y
- Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *Psych. Bull. Rev.* 24, 118–137. doi: 10.3758/s13423-016-1166-7
- Kirby, S., Dowman, M., and Griffiths, T. (2007). Innateness and culture in the evolution of language. *Proc. Nat. Acad. Sci. U. S. A.* 104, 5241–5245. doi: 10.1073/pnas.0608222104
- Minsky, M. (1985). “Why intelligent aliens will be intelligent,” in *Extraterrestrials: Science and Alien Intelligence*, ed E. Regis (Cambridge/New York: Cambridge University Press), 117–128.
- Miyagawa, S., and Nobrega, V. (2015). The precedence of syntax in the rapid emergence of human language in evolution as defined by the integration hypothesis. *Front. Psychol.* 18, 6:271. doi: 10.3389/fpsyg.2015.00271
- Monod, J. (1972). *Chance and Necessity*. New York, NY: Random House (Vintage Books edition).
- Moro, A. (2016). *Impossible Languages*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262034890.001.0001
- Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., et al. (2003). Broca’s area and the language instinct. *Nat. Neurosci.* 6, 774–781. doi: 10.1038/nn1077
- Nevins, A., Pesetsky, D., and Rodrigues, C. (2009). Piraha exceptionality: a reassessment. *Language*. 85, 355–404. doi: 10.1353/lan.0.0107



- Piantadosi, S. T., and Fedorenko, E. (2017). Infinitely productive language can arise from chance under communicative pressure. *J. Lang. Evol.* 2, 141–147. doi: 10.1093/jole/lzw013
- Piattelli-Palmarini, M., and Berwick, R.C. (Eds.). (2012). *Rich Languages From Poor Inputs*, Oxford: Oxford University Press.
- Pinker, S., and Bloom, P. (1990). Natural language and natural selection. *Behav. Brain Sci.* 13, 707–784. doi: 10.1017/S0140525X00081061
- Progovac, L. (2009). “Layering of grammar: vestiges of evolutionary development of syntax in present-day languages,” in *Language Complexity as an Evolving Variable*, eds G. Sampson, D. Gil, and P. Trudgill (Oxford: Oxford University Press), 203–212.
- Progovac, L. (2016). A gradualist scenario for language evolution: precise linguistic reconstruction of early human (and Neandertal) grammars. *Front. Psychol.* 7:1714. doi: 10.3389/fpsyg.2016.01714
- Scott-Phillips, T. C., and Kirby, S. (2010). Language evolution in the laboratory. *Trends Cogn. Sci.* 14, 411–417. doi: 10.1016/j.tics.2010.06.006
- Smith, K., and Kirby, S. (2008). Cultural evolution: implications for understanding the human language faculty and its evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3591–3603. doi: 10.1098/rstb.2008.0145
- Tattersall, I. (2012). *Masters of the Planet: The Search for Our Human Origins*. New York, NY: Palgrave Macmillan.
- Tattersall, I. (2017). How can we detect when language emerged? *Psychon. Bull. Rev.* 24, 64–67. doi: 10.3758/s13423-016-1075-9
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/7551.001.0001
- Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc.* 42, 230–265.
- Watumull, J. (2012). The computability and computational complexity of generativity. *Cambridge Occas. Pap. Linguist.* 6, Article 11:311–329.
- Watumull, J., Hauser, M. D., Roberts, I. G., and Hornstein, N. (2014). On recursion. *Front. Psychol.* 4, 1–7. Article 1017. doi: 10.3389/fpsyg.2013.01017
- Yang, C. (2016). *The Price of Linguistic Productivity*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262035323.001.0001
- Yang, C., Crain, S., Berwick, R. C., Chomsky, N., and Bolhuis, J. J. (2017). The growth of language: Universal Grammar, experience, and principles of computation. *Neurosci. Biobehav. Rev.* 81(Pt B), 103–119. doi: 10.1016/j.neubiorev.2016.12.023

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Huybregts. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.