

Evolutionary roads to syntax

Klaus Zuberbühler

Date of deposit	10/05/2019
Document version	Author's accepted manuscript
Access rights	Copyright © 2019, Elsevier Ltd on behalf of The Association for the Study of Animal Behaviour. This work is made available online in accordance with the publisher's policies. This is the author created, accepted version manuscript following peer review and may differ slightly from the final published version.
Citation for published version	Zuberbuhler, K. (2019). Evolutionary roads to syntax. <i>Animal Behaviour</i> , 151, 259-265.
Link to published version	https://doi.org/10.1016/j.anbehav.2019.03.006

Full metadata for this item is available in St Andrews Research Repository at: <https://research-repository.st-andrews.ac.uk/>

St Andrews Research Repository

1 **Evolutionary roads to syntax**

2

3 **Abstract**

4 Syntax is habitually named as what sets human language apart from other
5 communication systems, but how did it evolve? Comparative research on animal
6 behaviour has contributed in important ways, with mainly three sets of data. First,
7 animals have been subjected to artificial grammar tasks, based on the hypothesis that
8 human syntax has evolved through advanced *computational capacity*. In these
9 experiments humans generally outperform animals, but there are questions about
10 validity, as experimental stimuli are (deliberately) kept devoid of semantic content.
11 Second, animal communication has been compared in terms of the *surface structures*
12 with the aim of developing a typology of animal syntax, based on the hypothesis that
13 syntax is an evolutionary solution to the constraints of small signal repertoires. A
14 wide range of combinatorial phenomena has been described, mainly in non-human
15 primates, but there is little support for the hypothesis that syntax has emerged due to
16 repertoire size constraints. A third way of studying the evolution of syntax is to
17 compare how animals perceive and communicate about external events, the mental
18 deep structure of syntax. Human syntax is closely aligned with how we perceive
19 events in terms of agency, action, and patience, each with subsidiary functions. The
20 event perception hypothesis has been least explored in animals and requires a serious
21 research programme.

22

23 Theories of syntax

24 Studying the evolution of language is notoriously difficult. Neither brains nor
25 behaviours fossilise, such that the archaeological record can only offer little insight
26 into much of what makes humans unique. The comparative approach has turned out to
27 be a viable alternative, by which the behaviours and cognition of closely related
28 animal species are compared in order to draw inferences about the past evolutionary
29 history. The assumption is that behaviour, and its underlying cognitive governance,
30 has a heritable, genetic basis that can be traced phylogenetically. If the topic concerns
31 a human-specific trait, such as syntax, non-human primates naturally play a key role
32 in such endeavours, something that is also reflected in this opinion piece.

33 What is syntax and how did it evolve? The standard dictionary entry for syntax is
34 something like “a set of rules, principles and processes that determine how sentences
35 are formed from words and phrases in a language”. For evolutionary studies,
36 however, this definition is unsatisfactory because it presupposes linguistic units, i.e.,
37 words, phrases, sentences, which are themselves not available to animals. It is
38 possible, however, to modify the standard syntax definition by using functional
39 placeholders (word = meaningful unit; sentence = utterance that conveys a statement,
40 question, exclamation, or command). Hence, syntax is *the set of rules, principles and*
41 *processes that determine how statements, questions, exclamations or commands are*
42 *formed from meaningful units.*

43 How did syntax evolve? The prevailing view, at least amongst biologically trained
44 scholars, is that the evolution of syntax was a gradual process to the effect that its
45 evolutionary history can be reconstructed by comparative evidence. Three lines of
46 investigation have produced relevant data for evolutionary considerations: syntax as a

47 computational capacity (the computational hypothesis), syntax as way to evade
48 repertoire constraints (the surface structure hypothesis), and syntax as a reflection of
49 event perception (the deep structure hypothesis).

50 Syntax as computational capacity

51 A first road to study the evolution syntax derives from theoretical linguistics, which
52 seeks to describe language in terms of formal, artificial grammars with increasing
53 complexity (the ‘Chomsky hierarchy’) using computer science tools (Chomsky,
54 1956). The assumption is that computationally simple syntax, such as finite state
55 grammar, requires fewer computational operations and thus fewer cognitive resources
56 than complex syntax, such as phrase structure grammar. In finite state grammar, the
57 meaning of a sentence emerges from taking into account the relations of adjacent
58 words, i.e. decisions are taken serially. In natural languages the finite state grammar
59 cannot explain the entire range of phenomena, mainly because there are also
60 dependencies between non-adjacent words, requiring more complex phrase structure
61 grammar (Chomsky, 1957).

62

63 In behaviour experiments, subjects are exposed to stimulus sequences that comply
64 with (or are in breach of) the grammar under investigation. The prediction is that
65 successful processing enables a subject to perceive syntactic violations, measured by
66 increased attention (Fitch & Hauser, 2004). This reasoning is analogous to how
67 linguists examine natural grammars, by asking native speakers to make
68 grammaticality judgements (Chomsky 1957). For example, the sentence “colourless
69 green ideas sleep furiously” is typically judged as grammatical, despite the fact that it
70 is semantically nonsensical. In artificial grammar research, however, ‘sentences’ are

71 usually represented by sequences of tones, speech sounds or vocalisations that do or
72 do not comply with the respective grammar under study. The main conclusion from
73 this research has been that only humans can deal with complex artificial grammars
74 (Fitch & Hauser, 2004; Wilson, Smith, & Petkov, 2015) due to the limited
75 computational power of animal brains (Friederici, 2004), but see (Gentner, Fenn,
76 Margoliash, & Nusbaum, 2006; van Heijningen, de Visser, Zuidema, & ten Cate,
77 2009). The evolution of syntax, in this view, is a direct consequence of the evolution
78 of computational power required for syntactic processing.

79

80 One issue with artificial grammar research is that stimulus sequences are usually
81 meaningless simple tones. This is a deliberate choice so that the syntactic apparatus
82 can be investigated in its pure state, uncontaminated by semantics. Although the logic
83 is pertinent, there are questions about the ‘ecological’ validity of this approach.
84 ‘Colourful green ideas sleep furiously’ may be nonsensical but the sentence is still
85 composed of meaningful units, which may trigger processing in different brain areas
86 than processing of meaningless tone sequences. One debate therefore is whether
87 artificial grammar experiments reveal something relevant for evolutionary theories of
88 syntax or whether they are more informative regarding acoustic pattern recognition
89 (Hochmann, Azadpour, & Mehler, 2008; Zuberbuhler, 2018). Brain imaging studies
90 would provide valuable input towards this question.

91

92 Syntax in surface structures

93

94 Another influential hypothesis is that syntax evolves as soon as lexicons reach their
95 limits, because of memory or production limits: “...natural selection can only favour

96 the emergence of syntax if the number of required signals exceeds a threshold value”
97 (Nowak, Plotkin, & Jansen, 2000). Although intuitively appealing, the hypothesis is
98 difficult to test because it presumes species-specific thresholds, but there is no theory
99 as to how they could be determined. Nevertheless the hypothesis predicts that, in
100 closely related species, syntax is only present in species that have reached the
101 threshold, i.e. the ones with larger repertoires.

102

103 *A typology of syntax in animal communication*

104 There is a long ethological tradition of studying the surface features of animal
105 communication, i.e., the way species combine elements of their signal repertoires into
106 sequences. Pioneering were studies of birdsong that have revealed, for example, that
107 syntax plays a role in social interactions (Marler & Peters, 1988). Birdsong functions
108 to attract mates and keep rivals away and, as such, mainly contains information about
109 caller identity (Catchpole & Slater, 1995). More recently, research on bird syntax has
110 shifted towards the question of how meaning is conveyed by combinations of signals
111 that carry their own meaning, with relevant work on babblers (Engesser, Ridley, &
112 Townsend, 2016) and Japanese tits ((Engesser et al., 2016; Suzuki, Wheatcroft, &
113 Griesser, 2016). For mammals, studies exist on rock hyraxes (Kershenbaum, Ilany,
114 Blaustein, & Geffen, 2012) and various primates (Crockford & Boesch, 2005;
115 Hedwig, Mundry, Robbins, & Boesch, 2015; Schamberg, Cheney, Clay, Hohmann, &
116 Seyfarth, 2016; Zuberbuhler, 2018).

117

118 In primates, early studies reported syntactic structures for example in Cebus and
119 squirrel monkey calls (Newman, Katzlieblich, Talmageriggs, & Symmes, 1978;
120 Robinson, 1984). More recently, combinatorial calling has been found in various

121 primate alarm and contact calls (alarms: Diana monkeys (Stephan & Zuberbuhler,
122 2008), Campbell's monkeys (Lemasson, Ouattara, Bouchet, & Zuberbuehler, 2010;
123 Ouattara, Lemasson, & Zuberbuhler, 2009a; Zuberbühler, 2002), King Colobus
124 (Schel, Tranquilli, & Zuberbuhler, 2009); contact calls: Diana monkeys, Campbell's
125 monkeys (Candiotti, Zuberbühler, & Lemasson, 2012; Coye, Ouattara, Arlet,
126 Lemasson, & Zuberbuhler, 2018; Coye, Ouattara, Zuberbuhler, & Lemasson, 2015;
127 Coye, Zuberbuhler, & Lemasson, 2016)).

128

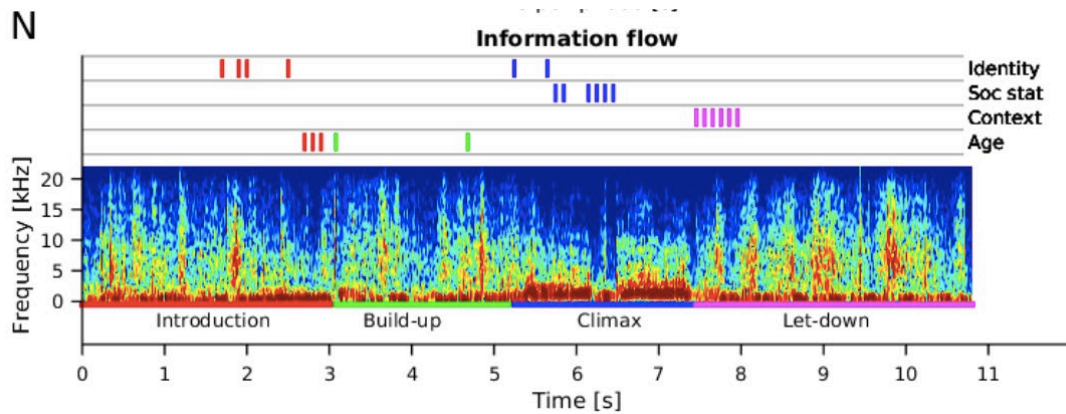
129 Putty-nosed monkeys have been particularly well studied, with males producing two
130 alarm call types, pyows and hacks, either singly or in combination. Series of pyows
131 are given mainly to terrestrial disturbances (e.g. leopards) and series of hacks to aerial
132 dangers (e.g. crowned eagles) (Arnold & Zuberbuhler, 2006). In addition, males can
133 combine both calls into 'pyow-hack' sequences, which carry a different meaning
134 (travel) unlinked to the meanings of the component calls (Arnold & Zuberbühler,
135 2006). The composed meaning appears to reside in the pyow-hack transition,
136 regardless of the number of pyows and hacks (Arnold & Zuberbühler, 2012).

137 Listeners respond to the combinatorial features, by perceiving the combination as a
138 meaningful unit, which is different from its components (Arnold & Zuberbühler,
139 2008).

140

141 Although these bird and monkey studies are relevant, an evolutionarily informed
142 theory of syntax necessarily requires data from our closest relatives, the apes. Gibbon
143 song has long been of interest (Demars & Goustard, 1972), although structural
144 changes were not usually linked with changes in meaning. More recently, it was
145 found that white-handed gibbon songs were produced both as duets and to predators;

146 snakes, clouded leopards and humans. Crucially, predator and non-predator songs
147 were assembled in different ways, albeit from the same basic song units, with
148 syntactic differences particularly visible during the early parts of a song, with
149 indications that recipients discriminated between the different song types (Clarke,
150 Reichard, & Zuberbuhler, 2006). For chimpanzees, an early study found also evidence
151 for regular use of call combinations but no clear semantic effects (Crockford &
152 Boesch, 2005). Some progress has been made with a study on the syntax of pant hoot
153 call utterances. Using machine learning and automated feature extraction, the study
154 produced evidence for encoding of age, rank, identity and context, across the four
155 phases (fig. 1), (Fedurek, Zuberbühler, & Dahl, 2016). The introduction and build-up
156 phases are low amplitude signals and contained mainly caller identity information,
157 suggesting they are directed at nearby individuals. The climax phase, in contrast, is
158 acoustically conspicuous and high-amplitude and contained information on both
159 identity and social status (low vs. high rank), presumably targeting faraway group
160 members and neighbouring groups. This is a relevant finding because, in
161 chimpanzees, decisions about whether to engage in intergroup conflict are largely
162 based on attending to neighbouring pant hoot vocalisations (Herbinger, Papworth,
163 Boesch, & Zuberbuehler, 2009; Wilson, Hauser, & Wrangham, 2001; Wilson et al.,
164 2014; Wilson et al., 2012; Wilson, Hauser, & Wrangham, 2007). Finally, pant hoots
165 usually end with low-amplitude, let-down units, which inform nearby group members
166 about the caller's forthcoming behavioural intentions (feeding vs. travelling). Callers
167 can omit one or several of the four phases, allowing them to target specific audiences
168 with specific information (fig. 1).



169
 170 Figure 1. Four phases of chimpanzee pant hoot vocalisations: introduction, build-up,
 171 climax, let-down. Adults usually produce pant hoot phases in this order, although one
 172 or several phases can be omitted. Different phases convey different sets of
 173 information, as indicated by the top information flow panel (reprinted from (Fedurek,
 174 Zuberbühler, & Dahl, 2016) licensed under a Creative Commons Attribution 4.0
 175 International License).

176 Another line of research in great apes is on the combined use of different modalities.
 177 Here, call-gesture combinations were generally rare and only used in social
 178 interactions of very positive or very negative connotations. Gestures were often added
 179 when vocal utterances failed to achieve a desired social goal, an expression of
 180 underlying persistence (Hobaiter, Byrne, & Zuberbuhler, 2017). For bonobos,
 181 (Schamberg et al., 2016) demonstrated natural call combinations in a wild population,
 182 whereas for captive groups there was evidence for call/gesture combinations, notably
 183 to disambiguate meaning (Genty, Clay, Hobaiter, & Zuberbühler, 2014). These
 184 studies demonstrate that call/call and call/gesture combinations exist and as such
 185 provide the groundwork for further research on the evolution of signal combinations
 186 in our closest living relative. Also relevant is the finding that when encountering food,
 187 bonobos produced sequences of call types that depended on the perceived quality of
 188 food (Clay & Zuberbuhler, 2009). Playback experiments confirmed that listeners were

189 able to attend to the different sequences and make predictions about what type of food
190 the caller has found (Clay & Zuberbühler, 2011). For gorillas, finally, grunts,
191 grumbles and hums can be given singly or in combinations (Harcourt & Stewart,
192 1996; Harcourt, Stewart, & Hauser, 1993; Salmi, Hammerschmidt, & Doran-Sheehy,
193 2013; Seyfarth, Cheney, Harcourt, & Stewart, 1994; Stewart & Harcourt, 1994).
194 When produced in isolation, grunts were given by individuals resting in close
195 proximity of each other, whereas grumbles were given during foraging. When
196 produced in combination, grumbles appeared to lose their foraging meaning,
197 suggesting that call combinations have less to do with augmenting semantics but to
198 mark social roles during communicative interactions (Hedwig et al., 2015).

199 The studies reviewed so far have revealed a bewildering range of combinatorial
200 structures that can be grouped as follows (Zuberbühler, 2018). First, in *merged units*
201 callers combine vocal structures, mainly to convey identity and event information
202 (Diana monkeys (Candiotti et al., 2012; Coye et al., 2016); Campbell’s monkeys
203 (Candiotti et al., 2012; Coye et al., 2018)). Related to this is *suffixation*, as found in
204 male Campbell’s monkey alarm calls. Here, callers add acoustically invariable ‘oo’-
205 units to three distinct alarm calls, to indicate that danger is non-urgent (Ouattara,
206 Lemasson, & Zuberbuehler, 2009b). At the call sequence level, there are examples of
207 *permutations*, i.e., ordered call deliveries, as found in alarm calling of male
208 Campbell’s and, as discussed before, male putty-nosed monkeys (Arnold &
209 Zuberbühler, 2006; Ouattara, Lemasson, & Zuberbuehler, 2009a). Another line of
210 inquiry has been on New World monkeys in Brazil. Black-fronted titi monkeys
211 produce alarm call sequences to refer to both predator type and location (Cäsar,
212 Zuberbühler, Young, & Byrne, 2013), although meaning is encoded by a *probabilistic*
213 (*stochastic*) rather than a categorical mechanism (Berthet et al submitted). Finally,

214 there is evidence for meaning being conveyed by utterances of varying lengths,
215 further assembled into more complex sequences, a *numeric* and seemingly
216 *hierarchical* structure (Diana monkeys (Zuberbühler, 2000), Campbell’s monkeys
217 (Lemasson et al., 2010), black-and-white Colobus monkeys (Schel, Candiotti, &
218 Zuberbühler, 2010).

219 Syntax in deep structure

220 The studies reviewed so far have revealed little about any underlying cognitive
221 processes and it is even possible that, what appears as syntax, is not linked to any
222 interesting mental processing. For example, syntactic regularities in signal sequences
223 could emerge merely by accident due to physiological constraints (e.g. structural
224 changes due to increasing exhaustion), semantic constraints (e.g. responding to X may
225 warrant some call types but not others), pragmatic constraints (e.g. more urgent calls
226 may be produced before less urgent calls), or on-going changes in the environmental
227 context triggering calls (P Schlenker and E Chemla, pers. comm.).

228

229 In humans, however, syntax is tightly linked with how events in the external world are
230 perceived, structured and mentally represented. In particular, humans have a natural
231 propensity to decompose events into actors, actions, and patients to the effect that
232 there is a curious correspondence between the components of natural events and the
233 grammatical functions of language. Sentences are structured in that they contain
234 agents (doer, cause, experiencer), actions (what), patients, targets or beneficiaries (to
235 whom) who experience the action or state of affair (e.g. “the eagle attacked the
236 monkey”). Arguments usually have additional components, such as the manner (how)
237 by which an action is carried out or the instrument used (with what) for this purpose.

238 Additionally, arguments can contain information about location (where), origin or
239 direction (from - to where), or time (when) an action was, is or will be carried out
240 (e.g. “the eagle attacked the monkey *from above*”).

241

242 Another useful description of events is in terms of predication (“*the eagle*
243 *attacked...*”), modification (“*the large eagle...*”) and coordination
244 (“*eagles and leopards*”) (Townsend, Engesser, Stoll, Zuberbuhler, & Bickel, 2018).
245 Languages have means to express these event features in ways to make them evident
246 to listeners, usually with specific syntactic functions. For example, to syntactically
247 distinguish an agent from a patient, some languages use phonological case marking
248 while others use word order.

249

250 The hypothesis here is that, during human evolution, these event-bound cognitive
251 universals (agents, patients, actions, manners, etc.) have become externalised and
252 assimilated into the communication system. This hypothesis is supported by work on
253 Nicaraguan sign language, which has shown that deaf children will gradually and
254 without specific tutoring develop syntactic structures in spontaneous sign language
255 that enables them to encode the core components of an event, rather than referring to
256 entire events holistically (Senghas, Kita, & Ozyurek, 2004). Modern humans, in other
257 words, have a natural propensity to mark the key components of external events with
258 (arbitrary) syntactic features.

259

260 How do animals perceive natural external events? There is evidence from artificial
261 language studies that marine mammals can be trained to discriminate agents from
262 patients (Herman, Richards, & Wolz, 1984). In natural communication, a study on

263 chimpanzee vocal behaviour has found acoustic differences in screams given in
264 different social roles, i.e. when the caller was the actor or the patient in an aggressive
265 act, which was discriminated by others (Slocombe, Kaller, Call, & Zuberbuehler,
266 2010; Slocombe & Zuberbuehler, 2005).

267

268 Human event perception, however, is vastly more complex than marking agents and
269 patients. Complex event perception is likely to have evolved first, possibly due to
270 increasingly complex social systems and associated brain enlargements. Syntax, in
271 this view, is a mere by-product of perceiving external events in decomposed ways and
272 of the ability to mark these components with communication signals. The human road
273 to syntax may have built on this predisposition, completed with the advent of
274 unprecedented vocal control, allowing event perception to become linguistically
275 encoded with grammatical functions.

276

277 Current issues

278

279 *Syntax without precursors*

280 A particularly contentious on-going debate is whether animal data can contribute in
281 meaningful ways to questions about syntax evolution in humans. One argument is that
282 studies of animal communication are irrelevant, because the only relevant property of
283 human syntax is its generative, hierarchical nature, for which there is no evidence in
284 animal communication. Cognitively, the argument goes, this is achieved by a single
285 mental operation, merge, which takes two syntactic elements and assembles them to
286 form a set (Bolhuis, Tattersall, Chomsky, & Berwick, 2014)(Bolhuis, 2017; Bolhuis,

287 Beckers, Huybregts, Berwick, & Everaert, 2018; Bolhuis et al., 2014; Townsend et
288 al., 2018).

289

290 The ‘merge’ view of language, however, is not universally accepted, even amongst
291 linguists. For example, much of ordinary language use is based on accessing
292 prefabricated phrases from a vast memory stock. Although the retrieved utterances
293 may be analysed in terms of syntactic structure, language users simply retrieve them
294 ‘wholesale’ to fit into appropriate slots (Townsend et al., 2018). Prefabricated
295 expressions account for up to half of all phrases used in conversations (Van Lancker-
296 Sidtis & Rallon, 2004), suggesting that evolutionary investigations of syntax should
297 also focus on non-generative, non-hierarchical combinatorial systems, as frequently
298 seen in animal communication.

299

300 *Varieties of merge*

301 An evolutionarily more fruitful proposal has been to distinguish between different
302 levels of ‘merge’, with increasing generative capacity (Rizzi, 2016). According to
303 this, 0-merge systems operate only with individual items from the lexicon. In fact, this
304 has been the default view of animal communication for decades, i.e., that animal
305 signals function as holistic units without any recourse to combinations (Hauser,
306 2000). 1-merge systems, next, have combinatorial properties insofar as they allow for
307 the formation of two-unit expressions, although the system then stops, with no
308 recursive procedures (i.e., word-word merges). Following this are 2-merge systems
309 that allow for recursion insofar as merged expressions (e.g. word-word or ‘phrase’)
310 can enter new merges, with its own components, but this requires more memory
311 capacity. Thus, 2-merge systems can potentially generate an unlimited set of

312 expressions (word-phrase merges) and are thus truly generative. Finally, 3-merge
313 systems are characterised by the ability to merge already merged expressions (phrase-
314 phrase merges), which requires further memory capacity. Sentence formation in
315 human language requires a 3-merge system, as subjects and predicates consist of
316 merged expressions. Current evidence suggests that animal calling goes beyond 0-
317 merge systems but stops at 1-merge systems, without any recursive applications.

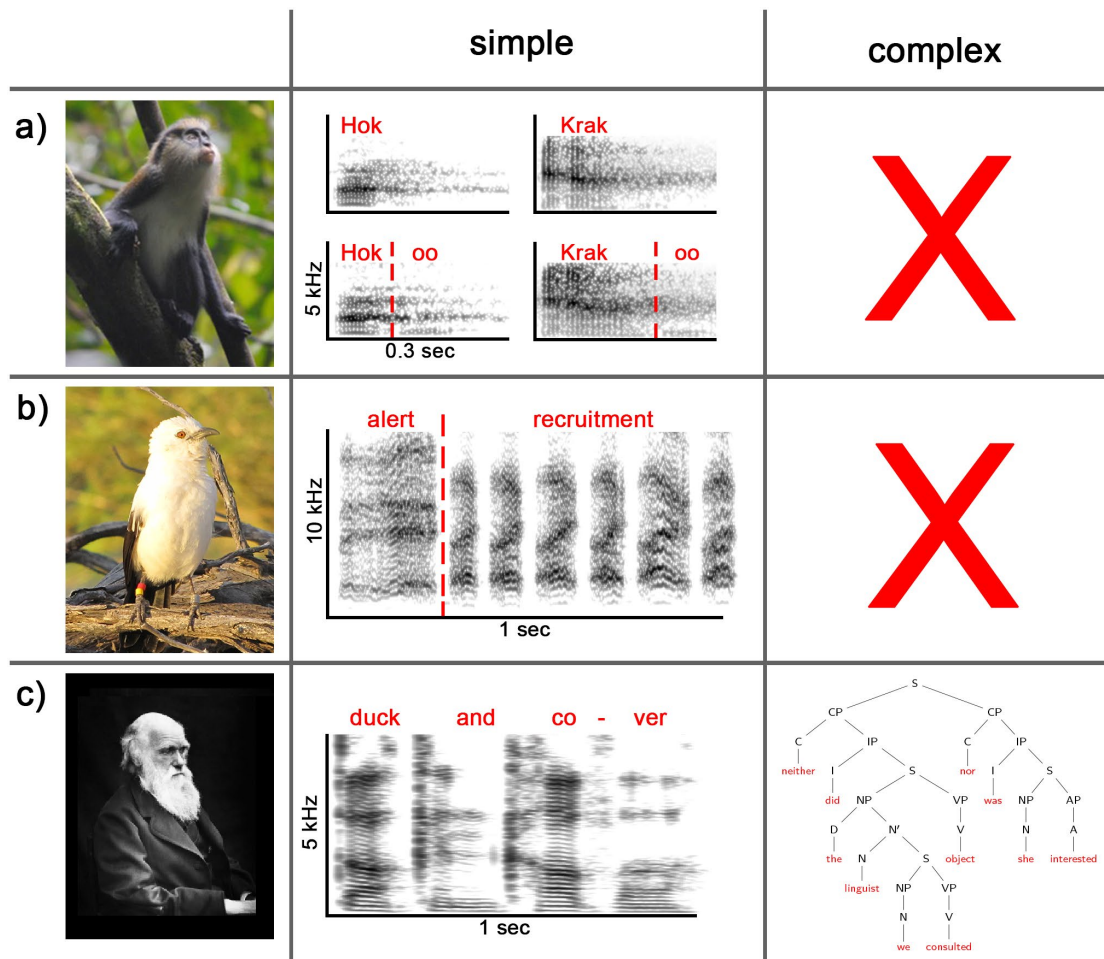
318

319 *Compositionality*

320 Are humans thus unique in having higher-level hierarchical syntax to generate
321 meaning? Most definitions of human language require compositionality, that is, that
322 simple expressions are used to build more complex expressions, whose meaning is
323 determined by the meanings of the constituent simple expressions *and* the rule that
324 combines them. The meaning of the whole is determined by the meaning of its parts
325 and how they are put together, the principle of compositionality.

326

327 In several theory papers, primate call systems have been analysed in such ways,
328 which has led to the conclusion that some systems, particularly Campbell's monkey
329 call suffixation and putty-nosed monkey call permutations, have weak *compositional*
330 properties (fig. 2), a claim with implications for evolutionary theories of language
331 (Schlenker et al., 2014; Schlenker, Chemla, Arnold, & Zuberbuhler, 2016; Schlenker,
332 Chemla, Casar, Ryder, & Zuberbuhler, 2017; Schlenker, Chemla, et al., 2016a, 2016b;
333 Schlenker, Chemla, & Zuberbuhler, 2016).



334

335 Figure 2. Examples of simple and complex compositionality in animals and humans.

336 a) Male Campbell's monkeys produce 'krak' alarms (to leopards) and 'hok' alarms (to

337 eagles), but both calls can also be merged with an '-oo' suffix to generate 'krak-oo'

338 (to a range of disturbances) and 'hok-oo' (to non-ground disturbances) (Ouattara,

339 Lemasson, & Zuberbuhler, 2009b). In playback experiments, suffixation has shown to

340 be meaningful to listeners, suggesting that it is an evolved communication function

341 (Coye et al., 2015). This system may qualify as limited compositionality, as the

342 meanings of krak-oo and hok-oo are directly derived from the meanings of krak/hok

343 plus the meaning of—oo. b) Compositionality in birds: Pied babblers produce 'alert'

344 calls in response to unexpected but low-urgency threats and 'recruitment' calls when

345 recruiting conspecifics to new foraging sites (Engesser, 2016; Engesser et al., 2016).

346 When encountering a terrestrial threat that requires recruiting group members (in the
347 form of mobbing), pied babblers combine the two calls into a larger structure, and
348 playback experiments have indicated that receivers process the call combination
349 compositionally by linking the meaning of the independent parts. c) Compositionality
350 in humans: humans are capable of producing both simple, non-hierarchical
351 compositions (e.g., ‘Duck and cover!’) and complex hierarchical compositions and
352 dependencies. Photo in panel A credited to Erin Kane. Photo in panel B credited to
353 Sabrina Engesser. A, adjective; AP, adjective phrase; C, conjunction; CP, conjunction
354 phrase; D, determiner; I, Inflection-bearing element; IP, inflectional phrase; N, (pro-
355)noun; NP, noun phrase; S, sentence; V, verb; VP, verb phrase (reprinted from
356 (Townsend et al., 2018) under the [Creative Commons Attribution](#) license).

357

358 Conclusion

359 Animal communication research has long worked under the assumption that animal
360 calls are structurally simple, holistic signals that develop under strong genetic control
361 (Snowdon et al., 1992; Snowdon & Hausberger, 1997). As call producers, animals
362 (including primates) were thought to be cognitively unengaged, merely responding
363 with acoustically invariable signals to evolutionarily urgent situations in more or less
364 automated ways (Tomasello, 2008). This point has also been made for great apes,
365 despite the fact that chimpanzees and bonobos have excelled in terms of social
366 cognition and visually based communication (Call & Tomasello, 2008, 2007). More
367 recently, the stance has come under scrutiny, due to a range of empirical
368 developments. First, although primates do not imitate sounds, they have considerable
369 degrees of control over their vocal output, which enables them to refrain from

370 signalling and to modify parts of their vocal repertoire in communicatively functional
371 ways (Lameira, Maddieson, & Zuberbuehler, 2014). Moreover, it has become clear
372 that primate vocal behaviour goes beyond producing single calls to single events, with
373 a steady stream of studies reporting various forms of signal combinations, sometimes
374 even in compositional ways (Zuberbuehler, 2018). As a consequence, research on
375 animal syntax is currently amongst the most productive areas in animal behaviour
376 research, with results being debated across disciplines (Bolhuis et al., 2018;
377 Schlenker, Chemla, et al., 2016b).

378 However, the currently available data do not yet give rise to an empirically informed
379 evolutionary theory of human syntax. Instead, the current literature provides a
380 bewildering diversity of combinatorial systems in animal communication, with no
381 clear evolutionary trends or obvious phylogenetic patterns (Zuberbuehler, 2018).
382 Equally, there is no conceptual agreement in how to integrate the different phenomena
383 into a coherent evolutionary theory of syntax (Kershenbaum et al., 2014; Zuberbuehler,
384 2018).

385 Human syntax is the result of mental processes but this is rarely addressed by animal
386 communication studies. Testing animals with artificial grammars has produced
387 interesting findings, revealing something about the limits of computational capacities,
388 but results are difficult to interpret because stimulus sequences are devoid of semantic
389 relations. Yet "...what distinguishes true language from just collections of uttered
390 words is that the semantic relations among the words are conveyed by syntactic and
391 morphological structure" (Jackendoff, 2007).

392 Future research should focus on how animals, and especially non-human primates,
393 naturally discriminate and mentally represent natural events and whether these

394 representation correspond to the main grammatical functions of human language
395 (actors, patients, descriptions of objects etc.). Data on whether animals perceive
396 events as functionally structured is likely to produce important progress and lead to a
397 better understanding of the evolutionary road to syntax.

398

399 References

400

- 401 Arnold, K., & Zuberbühler, K. (2006). The alarm-calling system of adult male putty-nosed monkeys,
402 *Cercopithecus nictitans martini*. *Animal Behaviour*, *72*, 643-653.
- 403 Arnold, K., & Zuberbühler, K. (2006). Semantic combinations in primate calls. *Nature*, *441*(7091), 303-
404 303.
- 405 Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current*
406 *Biology*, *18*(5), R202-R203. doi: 10.1016/j.cub.2008.01.040
- 407 Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: Compositional or idiomatic
408 expressions? *Brain and Language*, *120*(3), 303-309. doi: 10.1016/j.bandl.2011.10.001
- 409 Bolhuis, J. J. (2017). Making sense of language in the light of evolution. *Mind & Language*, *32*(5), 591-
410 596. doi: 10.1111/mila.12161
- 411 Bolhuis, J. J., Beckers, G. J. L., Huybregts, M. A. C., Berwick, R. C., & Everaert, M. B. H. (2018).
412 Meaningful syntactic structure in songbird vocalizations? *Plos Biology*, *16*(6). doi:
413 10.1371/journal.pbio.2005157
- 414 Bolhuis, J. J., Tattersall, I., Chomsky, N., & Berwick, R. C. (2014). How Could Language Have Evolved?
415 *Plos Biology*, *12*(8). doi: 10.1371/journal.pbio.1001934
- 416 Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in*
417 *Cognitive Sciences*, *12*(5), 187-192.
- 418 Call, J., & Tomasello, M. (Eds.). (2007). *The gestural communication of apes and monkeys*: Taylor &
419 Francis Lea.
- 420 Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012). Context-related call combinations in female
421 Diana monkeys. *Animal Cognition*, *15*(3), 327-339. doi: 10.1007/s10071-011-0456-8
- 422 Cäsar, C., Zuberbühler, K., Young, R. J., & Byrne, R. W. (2013). Titi monkey call sequences vary with
423 predator location and type. *Biology Letters*, *9*(5). doi: 10.1098/rsbl.2013.0535
- 424 Catchpole, C. K., & Slater, P. J. B. (1995). *Bird song: biological themes and variations*. Cambridge:
425 Cambridge Univ. Press.
- 426 Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information*
427 *Theory*, *2*, 113-124. doi: doi:10.1109/TIT.1956.1056813
- 428 Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- 429 Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The Syntax and Meaning of Wild Gibbon Songs.
430 *Plos One*, *1*(1). doi: 10.1371/journal.pone.0000073
- 431 Clay, Z., & Zuberbühler, K. (2009). Food-associated calling sequences in bonobos. *Animal Behaviour*,
432 *77*(6), 1387-1396. doi: Doi 10.1016/J.Anbehav.2009.02.016
- 433 Clay, Z., & Zuberbühler, K. (2011). Bonobos extract meaning from call sequences. *Plos One*, *6*(4). doi:
434 10.1371/journal.pone.0018786
- 435 Coye, C., Ouattara, K., Arlet, M. E., Lemasson, A., & Zuberbühler, K. (2018). Flexible use of simple and
436 combined calls in female Campbell's monkeys. *Animal Behaviour*, *141*, 171-181. doi:
437 10.1016/j.anbehav.2018.05.01
- 438 Coye, C., Ouattara, K., Zuberbühler, K., & Lemasson, A. (2015). Suffixation influences receivers'
439 behaviour in non-human primates. *Proceedings of the Royal Society B-Biological Sciences*, *282*(1807).
440 doi: 10.1098/rspb.2015.0265

441 Coye, C., Zuberbuhler, K., & Lemasson, A. (2016). Morphologically structured vocalizations in female
442 Diana monkeys. *Animal Behaviour*, *115*, 97-105. doi: 10.1016/j.anbehav.2016.03.010

443 Crockford, C., & Boesch, C. (2005). Call combinations in wild chimpanzees. *Behaviour*, *142*, 397-421.

444 Demars, C., & Goustard, M. (1972). STRUCTURE AND DEVELOPMENTAL PATTERN OF SONORAL
445 EMISSIONS OF HYLOBATES HYLOBATES-CONCOLOR. *Bulletin Biologique de la France et de la Belgique*,
446 *106*(3), 177-191.

447 Engesser, S. (2016). *Vocal combinatoinis in the Southern Pied Babbler (Turdoides bicolor) and the*
448 *Chestnut-Crowned Babbler (Pomatostomus ruficeps)*. PhD, Zurich, Zurich.

449 Engesser, S., Ridley, A., & Townsend, S. (2016). Meaningful call combinations and compositional
450 processing in the southern pied babbler. *Proceedings of the National Academy of Sciences*. doi:
451 10.1073/pnas.1600970113

452 Fedurek, P., Zuberbühler, K., & Dahl, C. D. (2016). Sequential information in a great ape utterance.
453 *Scientific Reports*, *6*, 38226.

454 Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a
455 nonhuman primate. *Science*, *303*(5656), 377-380.

456 Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends*
457 *in Cognitive Sciences*, *8*(6), 245-247. doi: 10.1016/j.tics.2004.04.013

458 Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern
459 learning by songbirds. *Nature*, *440*(7088), 1204-1207.

460 Genty, E., Clay, Z., Hobaiter, C., & Zuberbühler, K. (2014). Multi-modal use of a socially directed call in
461 bonobos. *Plos One*, *9*(1). doi: DOI 10.1371/journal.pone.0084738

462 Harcourt, A. H., & Stewart, K. J. (1996). Function and meaning of wild gorilla 'close' calls .2.
463 Correlations with rank and relatedness. *Behaviour*, *133*, 827-845.

464 Harcourt, A. H., Stewart, K. J., & Hauser, M. (1993). Functions of wild gorilla "close" calls: I. Repertoire
465 context, and interspecific comparison. *Behaviour*, *124*, 89-122.

466 Hauser, M. D. (2000). A primate dictionary? Decoding the function and meaning of another species'
467 vocalizations. *Cognitive Science*, *24*(3), 445-475.

468 Hedwig, D., Mundry, R., Robbins, M. M., & Boesch, C. (2015). Contextual correlates of syntactic
469 variation in mountain and western gorilla close-distance vocalizations: Indications for lexical or
470 phonological syntax? *Animal Cognition*, *18*(2), 423-435. doi: 10.1007/s10071-014-0812-6

471 Herbinger, I., Papworth, S., Boesch, C., & Zuberbuehler, K. (2009). Vocal, gestural and locomotor
472 responses of wild chimpanzees to familiar and unfamiliar intruders: a playback study. *Animal*
473 *Behaviour*, *78*(6), 1389-1396. doi: 10.1016/j.anbehav.2009.09.010

474 Herman, L. M., Richards, D. G., & Wolz, J. P. (1984). Comprehension of sentences by bottlenosed
475 dolphins. *Cognition*, *16*(2), 129-219.

476 Hobaiter, C., Byrne, R. W., & Zuberbuhler, K. (2017). Wild chimpanzees' use of single and combined
477 vocal and gestural signals. *Behavioral Ecology and Sociobiology*, *71*(6). doi: 10.1007/s00265-017-2325-
478 1

479 Hochmann, J. R., Azadpour, M., & Mehler, J. (2008). Do humans really learn A(n) B-n artificial
480 grammars from exemplars? *Cognitive Science*, *32*(6), 1021-1036. doi: 10.1080/03640210801897849

481 Jackendoff, R. (2007). *Language, Consciousness, Culture*. Cambridge MA: MIT Press.

482 Kershenbaum, A., Bowles, A. E., Freeberg, T. M., Jin, D. Z. Z., Lameira, A. R., & Bohn, K. (2014). Animal
483 vocal sequences: not the Markov chains we thought they were. *Proceedings of the Royal Society B-*
484 *Biological Sciences*, *281*(1792). doi: 10.1098/rspb.2014.1370

485 Kershenbaum, A., Ilany, A., Blaustein, L., & Geffen, E. (2012). Syntactic structure and geographical
486 dialects in the songs of male rock hyraxes. *Proceedings of the Royal Society B-Biological Sciences*,
487 *279*(1740), 2974-2981. doi: 10.1098/rspb.2012.0322

488 Lameira, A. R., Maddieson, I., & Zuberbuehler, K. (2014). Primate feedstock for the evolution of
489 consonants. *Trends in Cognitive Sciences*, *18*(2), 60-62. doi: 10.1016/j.tics.2013.10.013

490 Lemasson, A., Ouattara, K., Bouchet, H., & Zuberbuehler, K. (2010). Speed of call delivery is related to
491 context and caller identity in Campbell's monkey males. *Naturwissenschaften*, *97*(11), 1023-1027. doi:
492 10.1007/s00114-010-0715-6

493 Marler, P., & Peters, S. (1988). The role of song phonology and syntax in vocal learning preferences in
494 the song sparrow, *Melospiza melodia*. *Ethology*, *77*, 125-149.

495 Newman, J. D., Katzlieblich, A., Talmageriggs, G., & Symmes, D. (1978). SYLLABLE CLASSIFICATION
496 AND SEQUENCING IN TWITTER CALLS OF SQUIRREL-MONKEYS (SAIMIRI-SCIUREUS). *Zeitschrift Fur*
497 *Tierpsychologie-Journal of Comparative Ethology*, *47*(1), 77-88.

498 Nowak, M. A., Plotkin, J. B., & Jansen, V. A. A. (2000). The evolution of syntactic communication.
499 *Nature*, 404(6777), 495-498. doi: 10.1038/35006635

500 Ouattara, K., Lemasson, A., & Zuberbuehler, K. (2009a). Campbell's monkeys concatenate
501 vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences of
502 the United States of America*, 106(51), 22026-22031. doi: 10.1073/pnas.0908118106

503 Ouattara, K., Lemasson, A., & Zuberbuehler, K. (2009b). Campbell's Monkeys Use Affixation to Alter
504 Call Meaning. *Plos One*, 4(11). doi: 10.1371/journal.pone.0007808

505 Ouattara, K., Lemasson, A., & Zuberbuehler, K. (2009a). Campbell's monkeys concatenate vocalizations
506 into context-specific call sequences. *Proceedings of the National Academy of Sciences of the United
507 States of America*, 106(51), 22026-22031.

508 Ouattara, K., Lemasson, A., & Zuberbuehler, K. (2009b). Campbell's Monkeys Use Affixation to Alter Call
509 Meaning. *Plos One*, 4(11).

510 Rizzi, L. (2016). Monkey morpho-syntax and merge-based systems. *Theoretical Linguistics*, 42(1-2),
511 139-145. doi: 10.1515/tl-2016-0006

512 Robinson, J. G. (1984). Syntactic structures in the vocalizations of wedge-capped capuchin monkeys,
513 *Cebus olivaceus*. *Behaviour*, 90, 46-79.

514 Salmi, R., Hammerschmidt, K., & Doran-Sheehy, D. M. (2013). Western Gorilla Vocal Repertoire and
515 Contextual Use of Vocalizations. *Ethology*, 119(10), 831-847. doi: 10.1111/eth.12122

516 Schamberg, I., Cheney, D. L., Clay, Z., Hohmann, G., & Seyfarth, R. M. (2016). Call combinations, vocal
517 exchanges and interparty movement in wild bonobos. *Animal Behaviour*, 122, 109-116. doi:
518 10.1016/j.anbehav.2016.10.003

519 Schel, A. M., Candiotti, A., & Zuberbuehler, K. (2010). Predator-detering alarm call sequences in
520 *Guereza colobus* monkeys are meaningful to conspecifics. *Animal Behaviour*, 80(5), 799-808. doi:
521 10.1016/j.anbehav.2010.07.012

522 Schel, A. M., Tranquilli, S., & Zuberbuehler, K. (2009). The Alarm Call System of Two Species of Black-
523 and-White Colobus Monkeys (*Colobus polykomos* and *Colobus guereza*). *Journal of Comparative
524 Psychology*, 123(2), 136-150.

525 Schlenker, P., Chemla, E., Arnold, K., Lemasson, A., Ouattara, K., Keenan, S., . . . Zuberbuehler, K.
526 (2014). Monkey semantics: two 'dialects' of Campbell's monkey alarm calls. *Linguistics and
527 Philosophy*, 37(6), 439-501. doi: 10.1007/s10988-014-9155-7

528 Schlenker, P., Chemla, E., Arnold, K., & Zuberbuehler, K. (2016). Pyow-hack revisited: Two analyses of
529 Putty-nosed monkey alarm calls. *Lingua*, 171, 1-23. doi: 10.1016/j.lingua.2015.10.002

530 Schlenker, P., Chemla, E., Casar, C., Ryder, R., & Zuberbuehler, K. (2017). Titi semantics: Context and
531 meaning in Titi monkey call sequences. *Natural Language & Linguistic Theory*, 35(1), 271-298. doi:
532 10.1007/s11049-016-9337-9

533 Schlenker, P., Chemla, E., Schel, A. M., Fuller, J., Gautier, J. P., Kuhn, J., . . . Zuberbuehler, K. (2016a).
534 Formal monkey linguistics. *Theoretical Linguistics*, 42(1-2), 1-90. doi: 10.1515/tl-2016-0001

535 Schlenker, P., Chemla, E., Schel, A. M., Fuller, J., Gautier, J. P., Kuhn, J., . . . Zuberbuehler, K. (2016b).
536 Formal monkey linguistics: The debate. *Theoretical Linguistics*, 42(1-2), 173-201. doi: 10.1515/tl-2016-
537 0010

538 Schlenker, P., Chemla, E., & Zuberbuehler, K. (2016). What Do Monkey Calls Mean? *Trends in Cognitive
539 Sciences*, 20(12), 894-904. doi: 10.1016/j.tics.2016.10.004

540 Senghas, A., Kita, S., & Ozyurek, A. (2004). Children creating core properties of language: Evidence
541 from an emerging sign language in Nicaragua. *Science*, 305(5691), 1779-1782.

542 Seyfarth, R. M., Cheney, D. L., Harcourt, A. H., & Stewart, K. J. (1994). The acoustic features of gorilla
543 double grunts and their relation to behavior. *Am. j. Primatol.*, 33(1), 31-50.

544 Slocombe, K. E., Kaller, T., Call, J., & Zuberbuehler, K. (2010). Chimpanzees Extract Social Information
545 from Agonistic Screams. *Plos One*, 5(7). doi: 10.1371/journal.pone.0011473

546 Slocombe, K. E., & Zuberbuehler, K. (2005). Agonistic screams in wild chimpanzees (*Pan troglodytes
547 schweinfurthii*) vary as a function of social role. *Journal of Comparative Psychology*, 119(1), 67-77.

548 Snowdon, C. T., Elowson, A. M., Nishida, T., McGrew, W. C., Marler, P., Pickford, M., & de Waal, F. B.
549 M. (Eds.). (1992). *Ontogeny of primate vocal communication*.

550 Snowdon, C. T., & Hausberger, M. (Eds.). (1997). *Social influences on vocal development*. Cambridge:
551 Cambridge University Press.

552 Stephan, C., & Zuberbuehler, K. (2008). Predation increases acoustic complexity in primate alarm calls.
553 *Biology Letters*, 4(6), 641-644. doi: Doi 10.1098/Rsbl.2008.0488

554 Stewart, K. J., & Harcourt, A. H. (1994). Gorillas' vocalizations during rest periods: signals of impending
555 departure? *Behaviour*, *130*, 29-40.

556 Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for compositional syntax in
557 bird calls. *Nature Communications*, *7*. doi: 10.1038/ncomms10986

558 Tomasello, M. (2008). *Origins of human communication*. Cambridge MA: MIT Press.

559 Townsend, S. W., Engesser, S., Stoll, S., Zuberbühler, K., & Bickel, B. (2018). Compositionality in
560 animals and humans. *Plos Biology*, *16*(8). doi: 10.1371/journal.pbio.2006425

561 van Heijningen, C. A. A., de Visser, J., Zuidema, W., & ten Cate, C. (2009). Simple rules can explain
562 discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the*
563 *National Academy of Sciences of the United States of America*, *106*(48), 20538-20543. doi:
564 10.1073/pnas.0908113106

565 Van Lancker-Sidtis, D., & Rallon, G. (2004). Tracking the incidence of formulaic expressions in
566 everyday speech: methods for classification and verification. *Language & Communication*, *24*(3), 207-
567 240. doi: 10.1016/j.langcom.2004.02.003

568 Wilson, B., Smith, K., & Petkov, C. I. (2015). Mixed-complexity artificial grammar learning in humans
569 and macaque monkeys: evaluating learning strategies. *European Journal of Neuroscience*, *41*(5), 568-
570 578. doi: 10.1111/ejn.12834

571 Wilson, M., Hauser, M., & Wrangham, R. (2001). Does participation in intergroup conflict depend on
572 numerical assessment, range location, or rank for wild chimpanzees? *Animal Behaviour*, *61*(6), 1203-
573 1216.

574 Wilson, M. L., Boesch, C., Fruth, B., Furuichi, T., Gilby, I. C., Hashimoto, C., . . . Wrangham, R. W.
575 (2014). Lethal aggression in *Pan* is better explained by adaptive strategies than human impacts.
576 *Nature*, *513*(7518), 414-+.

577 Wilson, M. L., Boesch, C., Furuichi, T., Gilby, I. C., Hashimoto, C., Hohmann, G., . . . Wrangham, R. W.
578 (2012). Rates of lethal aggression in chimpanzees depend on the number of adult males rather than
579 measures of human disturbance. *American Journal of Physical Anthropology*, *147*, 305-305.

580 Wilson, M. L., Hauser, M. D., & Wrangham, R. W. (2007). Chimpanzees (*Pan troglodytes*) modify
581 grouping and vocal behaviour in response to location-specific risk. *Behaviour*, *144*, 1621-1653.

582 Zuberbühler, K. (2018). Combinatorial capacities in primates. *Current Opinion in Behavioral Sciences*,
583 *21*, 161-169. doi: 10.1016/j.cobeha.2018.03.015

584 Zuberbühler, K. (2000). Referential labelling in Diana monkeys. *Animal Behaviour*, *59*(5), 917-927.

585 Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, *63*, 293-
586 299.

587