PHILOSOPHICAL TRANSACTIONS B

Syntax and compositionality in animal communication

Klaus Zuberbühler[1,2,3]

[1]Institute of Biology, University of Neuchatel, Rue Emile Argand 11, CH-2000 Neuchatel, Switzerland; [2] School of Psychology and Neuroscience, University of St Andrews, St Andrews, KY16 9JP, Scotland; [3] Centre for the Interdisciplinary Study of Language Evolution, University of Zurich, Plattenstrasse 54, CH-8032 Zurich, Switzerland

Summary

Syntax has been found in animal communication but only humans appear to have generative, hierarchically structured syntax. How did syntax evolve? I discuss three theories of evolutionary transition from animal to human syntax: computational capacity, structural flexibility and event perception. The computation hypothesis is supported by artificial grammar experiments consistently showing that only humans can learn linear stimulus sequences with an underlying hierarchical structure, a possible by-product of computationally powerful large brains. The structural flexibility hypothesis is supported by evidence of meaning-bearing combinatorial and permutational signal sequences in animals, with sometimes compositional features, but no evidence for generativity or hierarchical structure. Again, animals may be constrained by computational limits in short-term memory but possibly also by limits in articulatory control and social cognition. The event categorization hypothesis, finally, posits that humans are cognitively predisposed to analyse natural events by assigning agency and assessing how agents impact on patients, a propensity that is reflected by the basic syntactic units in all languages. Whether animals perceive natural events in the same way is largely unknown, although event perception may provide the cognitive grounding for syntax evolution.

*Author for correspondence (klaus.zuberbuehler@unine.ch).
†Present address: Institute of Biology, University of Neuchatel, Rue Emile Argand 11, Neuchatel, CH-2000, Switzerland

Introduction

The lineages leading to Homo and Pan began to split some 6-8 million years ago [1], which led to differences in morphology, behaviour and cognition between the three surviving species. A striking example is human language, with nothing comparable in the communication systems of chimpanzees, bonobos and other animals. While all primates communicate with species-specific vocal repertoires, only humans have evolved a secondary communication system that is based on sophisticated, socially learned motor control of the vocal apparatus, unlike any other primate [2] [3]. This difference is already visible in human infants before language: from about 4 months of age, humans begin to control their vocal apparatus, first by engaging in playful babbling, then followed by the production of simple word-like structures. At the same time, human infants produce non-linguistic vocalisations for social communication [4], very similar to what has been reported in non-human primates [5]. Human babbling has a social bonding function, possibly evolved in response to cooperative breeding [6], but it also provides the scaffolding for increasing vocal control and competence for subsequent speech production. At around two years of age, infants go beyond labelling simple referents ("dada", "duck") and start producing word combinations [7], a developmental process that ends with the capacity to produce complex utterances with complex syntactic structures, such as "Neither did the linguist we consulted object nor was she interested" (Fig. 1). Such communication behaviour requires considerable computational power, for both signallers and receivers. One particular challenge is the conversion of hierarchically structured social events into linearly structured speech streams. How and why did humans evolve this capacity?

-- Figure 1 --

Syntax and compositionality

A basic mental operation that underlies this communication system is compositionality, or Frege's principle, which states that the meaning of a complex expression is determined by the meanings of its constituents and the rules used to combine them. In human language, compositionality is found both at the morphological and syntactic level, jointly responsible for the vast expressive power of human language. But syntax and morphology are not always compositional. Idioms ("to kick the bucket"= to die; "to spill the beans" = to

reveal a secret), for example, have figurative meanings different from their literal meanings [9], which cannot be derived from the meaning of the constituents. How did humans evolve the capacity for compositionality? Can the evolutionary history of compositionality be traced back to our common ancestor with non-human primates? Animals often combine signals into sequences, so an essential question is whether resulting combinations qualify as compositional or simple additions of meanings. The null hypothesis in this line of investigation is that callers produce sequences of (meaningful) constituents that merely reflect ongoing changes in the environment, a sort of running commentary of how events unfold.

The purpose of this essay is to review some of the animal evidence for syntax and examine, on a case-by-case basis, whether the systems show indications of compositionality at the morphological or syntactic level. The key terminology is summarised in table 1.

-- table 1 --

Evolutionary theories

Syntax as a recent evolutionary event

How did humans evolve their capacity to produce and decode syntactically organised structures? A current debate concerns the onset of the transition towards human syntax, i.e., when brains became language-ready. One group of theories posit that this process began only very recently, after the split between humans and their primate relatives a few million years ago. Here, the argument is that human syntax is the output of a powerful computation device that evolved very recently and without relevant precursors. This device enables speakers to compute complex phrase structures and allows infants to acquire the syntax of their native language without specific instructions and at high speed ([10], p. 25). Once present, it exerts its power onto the vocal output by rendering thoughts into a syntactically organized, linear speech output. One particular prominent proposal within this group is Chomsky's saltation theory, which explains syntax as the output of a brain that has undergone a qualitative change after the split from non-human primates due to a macromutation [11] (pp. 176–184). Although saltation exists as an evolutionary mechanism, examples are

mostly from plants, where gene duplications and gene transfers can lead to rapid spread and speciation [12] [13]. However, there is no evidence that brain evolution has occurred in this way. A possibly more plausible scenario for why humans have syntax as a mental operation is that it was enabled, as a by-product, by recent brain enlargements. Humans have unusually large brains, especially considering their body size and compared to ancestral and current species (H. sapiens:1,408g; P. troglodytes: 406g, G. gorilla 486g, P. pygmaeus 512g; H. habilis 599g, H. erectus 963g; [14]) and it is possible that the extra amount of brain mass enables humans to perform more complex computations than individuals with smaller brains, including short-term memory operations. What the exact amount of brain volume is to experience a transition to syntax, however, is underspecified but could potentially be tested by examining syntactic processing in pathologically small brains, such as in microcephaly. Whatever the underlying evolutionary history, the claim is that human syntax emerged only recently and is fundamentally different from all known animal systems, due to its hierarchical, compositional, recursive mental operations that produce multi-segment structures as the main output [15].

Syntax as an ancient evolutionary event

A second theoretical position is that syntax evolution started much earlier, deep within the primate lineage (or before), and then changed gradually from simple to more complex operations. This hypothesis can be tested with research on non-human primates and other animals to retrace the evolutionary history of syntax in deep time, with distantly related species providing important data to understand the role of ecological pressures [8]. A particularly influential idea is the lexical constraint hypothesis, e.g. [16], the idea that the capacity for syntax has emerged as a response to increasing lexicon sizes [17]. If a species continuously increases the number of signals it will reach a point where further additions become uneconomical compared to combining already existing ones, either due to production or memory limits.

A second, less explored idea is the event perception hypothesis [18]. The hypothesis states that the origins of syntax lie in the way events are perceived and mentally represented by the members of a species. For humans, natural events consist of agents, actions and patients, they possess a temporal structure and they contain

4

dependencies across event sequences. Assignments of intentionality appear to play a key role during this process [19]. An important observation is that these basic components (agents, patient, action) correspond to the major grammatical functions of language, i.e., subjects, predicates, direct or indirect objects, tense markers or conjunctions. In this view, the syntactic structure of human language is little more than a coding system that reflects how (humans) perceive events. Syntax, in other words, is 'out in the world', constantly present in natural events, so an important research question is how non-human primates, particularly great apes, perceive natural events. Results will be important to decide whether syntax-ready brains have evolved before modern humans, and whether the main transition towards language merely consisted of a communicative externalisation of this cognitive propensity.

Comparative research

Artificial grammar studies

One way to compare the syntactic capacities across species is to carry out artificial grammar studies. This is based on the notion that artificial languages can be ranked in terms of structural complexity, such as whether they contain linear or hierarchical rules. The experimental manipulation is to train subjects on sequences of stimuli that comply with a given rule in order to test whether they subsequently perceive violations [20]. So far, the conclusion has been that the processing abilities of animals is limited to linear grammars, while only humans are able to process hierarchically organized grammars [21]. There are exceptions, notably a study on blackbirds that managed to master phrase structure ruled grammar [22], but results have been questioned due to issues with overtraining and other artefacts [23]. Artificial grammar research has also been criticised for low ecological validity [24]. In most studies, experimental stimuli are sequences without relevant content, requiring advanced auditory pattern recognition but no semantic processing. Although a deliberate design feature (due to the goal of studying syntax 'pure' uncontaminated by semantics), syntax and semantics are not divorced in natural systems, casting some doubt on whether artificial grammar studies are suitable to model syntax evolution.

Natural communication studies

Animal song

A second approach to the evolution of syntax has been to study natural communication systems in across species. In several species, song has been analysed in terms of potential syntactic structures, a particularly striking example being the communication behaviour of humpback whales, with males combining song units into phrases and phrases into songs, all driven by social learning, occasionally even across populations [25]. However, neither humpback whale nor bird songs qualify as syntax in a strict sense (table 1), as the basic units do not seem to bear any meaning. Instead, animal song appears to function typically to repel sexual rivals and attract mating partners, a sexually selected behaviour. In primates, singing has been reported in some socially monogamous species, such as gibbons or indris, but here also it appears to function in reproduction and competition between groups [26]. White-handed gibbons, however, also sing to predators, notably tigers, clouded leopards and pythons, with predator songs and duet songs assembled in different ways [27]. Naturalistic observations [27] and playback experiments [28] suggest that recipients discriminate predator and non-predator songs by showing appropriate behavioural responses. But as individual song units are meaningless, gibbon song does not technically qualify as syntax either. Nevertheless, white-handed gibbon song represents a potential example of a compositional system, albeit with no evidence that songs are socially learned.

Permutation syntax

Types of syntax that prevail in the animal communication literature; permutation and combination. In permutations, signals are arranged in ordered ways, whereas in combinations, signals are selected from a repertoire but assembled without order. Permutations have been observed both within call variants and across signal sequences. An example for call variants is Diana monkey contact calls, which consist of a tonal, arched unit that conveys the identity of the caller. Arches differ individually in shape and contextually in tonality, depending on visibility and background noise, suggesting callers have some control over articulation. Importantly, the arched unit can take a prefix, either a trill or a chatter, depending on social circumstances. Trill-arch combinations are given during socio-positive interactions, whereas chatter-arch combinations are

part of socio-negative interactions [29]. Whether such combinations are compositional remains a possibility, pending evidence that the combined meaning goes beyond the meaning of the constituent parts.

A second example is male Campbell's monkeys predator alarm calls [30]. Here, males produce two basic alarm call types ('krak' and 'hok'), both of which can take an acoustically invariable suffix ('-oo') (fig. 2). Suffixed alarm calls have been associated with low urgency situations (e.g. predator far away) whereas un-suffixed ones are given when danger is imminent (e.g. predator visible). In playback experiments, unsuffixed krak-alarms (given to leopards) caused leopard alarm responses in Diana monkeys (a species that often forms mixed groups with Campbell's monkeys) whereas suffixed krak alarms (given to general disturbances) did not [31]. Importantly, artificially edited krak and krak-oo alarm calls caused the same response patterns as unedited, natural krak and krak-oo alarms, suggesting that the -oo unit modified the meaning of the krak stem. The Campbell's monkey suffixation system has been cited as an example of compositionality, at the morphological level (-oo do not have independent semantic function) [32] [33].

-- Figure 2 ---

Permutations are also common at the sequence level. For example, Campbell's males often produce long sequences of calls to different events [34] and one permutation rule concerns the use of boom calls to introduce non-predator related events, which is meaningful to other primates [35]. Another example is chimpanzee long-distance pant hoot vocalisations, a charismatic vocal utterance that consists of four distinct units; introduction, build-up, climax, and let-down (fig. 3). Using machine-learning, it was possible to establish the semantics of the four units, with caller identity mainly encoded by the introduction and climax units, age mainly encoded by the build-up, social status mainly encoded by the climax and behavioural context (travel vs. food) mainly encoded by the let-down unit [31]. Although it appears that all four units can also be given individually, there is no systematic work on the meaning of these individual calls, so it is not possible to decide whether pant-hoots qualify as compositional signals.

-- Figure 3 --

Another well studied permutational system is putty-nosed monkey alarm calls. Here, males produce two types of alarms, pyows and hacks, typically assembled into longer sequences. Roughly speaking, pure pyow series are given to terrestrial disturbances, whereas pure hack series are given to crowned eagles [37]. However, males sometimes produce mixed sequences, i.e., some pyows followed by some hacks, which reliably predicts group movement [38]. In playback experiments, pyow-hack sequences triggered movements by subjects towards the speaker, in contrast to pyow or hack series [39], suggesting that listeners attend to the permutation rule. The variable number of pyows and hacks in natural pyow-hack combinations had no measurable impact on recipient responses [40]. Due to the fact that the apparent meaning of the pyow-hack sequence cannot be inferred from the individual meanings, the system does not qualify as compositional but has been interpreted as idiomatic [41].

Combination syntax

As opposed to permutation, combinatoriality is defined as 'the selection of a given number of elements from a larger number without regard to their arrangement'. A first relevant example is titi monkey alarm calling where adults produce A- and B-calls to aerial and terrestrial predators, respectively [42, 43]. This basic pattern is disrupted if predators are detected in non-standard locations, i.e., a raptor on the ground or a terrestrial predator in the canopy [44]. Then callers produced mixed sequences containing both A and B calls in combinatorial but non-permutational ways [45]. In subsequent playback experiments, it could be shown that sequences recorded under the four experimental conditions (aerial predator ground or canopy, terrestrial predator ground or canopy) elicited adequate behavioural responses (fig. 4, [46]). One specific combinatorial feature, the proportion of B-call bigrams in the sequence, had a particularly strong impact on recipient responses, suggesting it served as a main carrier of meaning.

-- Figure 4 --

For great apes, combinatorial syntax is reported in bonobo food calling. Here, individuals were shown to select from five different call types to produce seemingly unordered combinations. Individual calls appeared

to carry some meaning, insofar as different calls were preferably given to differently perceived food qualities [47].

-- Figure 5 --

In subsequent playbacks, we trained subjects to anticipate highly preferred food (kiwi) in one part of their enclosure, whereas less preferred food (apple) could be found in another part [48]. Following this initial training, we broadcast segments of natural call sequences originally recorded from individuals discovering either kiwi or apple. Although sequences consisted of only four calls (table 2), subjects responded in highly predictable ways, such that sequences originally given to kiwi triggered mainly searching in the kiwi field whereas the opposite was observed with sequences originally given to apples.

-- Table 2 --

Are bonobo food calls compositional? As mentioned, definitions of language are based on the notion that the meaning of a complex expression fully determined by its structure and the meanings of its constituents [49]. If permutation (order) is an essential requirement for 'structure', then neither titi monkey alarm calling nor bonobo food calling qualify as compositional. However, in both systems there is a curious presence of bigrams, which appears to play an organising role. In titi monkeys, the proportion of B-call bigrams explained much of the monkeys' production and response patterns. In bonobos the meaning of the sequence appears to be determined by the same principle: 15 of 17 playback sequences contained bigrams of B, P, PY, Y calls, and the presence of such structures explained to a large degree the subsequent search behaviour. In particular, B and P bigrams elicited a strong kiwi bias (in search efforts), whereas PY and Y bigrams elicit an apple bias. No bigrams in the sequence (N/A) elicited an ambiguous response pattern, either a kiwi or an apple bias (table 3).

-- Table 3 --

But what meaning has been computed by these expressions? As B and P calls are already associated with high quality foods, why do callers not just use pure series of these calls when detecting kiwi? One possibility is that

the meaning of some of the component calls is broader. This is certainly the case for P (peep) calls, which are given to a wide range of events, including non-food ones, that require others' attention [50]. Producing calls in sequences, and particularly producing bigrams, thus allows individuals to generate acoustically conspicuous structures in response to some external events, but not others. Furthermore, great apes in the wild feed on hundreds of items, which differ in a wide range of variables (e.g., greens vs. fruits, large vs. small amounts, high vs. low quality, novel vs. familiar), potentially enabling callers to compose a range of complex expressions from a small number of calls to convey such differences. Also, callers may have reasons to encode individual identity into their utterances when finding food, and call sequences may facilitate this demand. It has been argued that food calling is some sort of other-serving behaviour, part of social reciprocity (Needle et al in prep), suggesting that marking individual identity could be important when providing information about food encounters. Further work is needed to explore these questions, but for now it would be difficult to defend a compositionality interpretation in the bonobo call system.

Conclusion

Is animal communication compositional?

Are humans unique in having the capacity to compose complex expressions from simpler ones? Schlenker et al. [33] have proposed to distinguish between limited (trivial) and genuine (non-trivial) compositionality. Examples in English are "It's humid" (trivial) and "It's very humid" (non-trivial). The difference lies in the fact that for trivial compositions there is no need for semantic operations; separate utterances are merely added into meaningful expressions. In genuine compositions, a semantic operation is active, such that the composite expression cannot be analysed from its components ("it's humid", "very") in a direct way. In several theory papers, primate call systems have been analysed in such ways, which has led to the conclusion that some systems, particularly Campbell's monkey call suffixation and putty-nosed monkey call permutations, have weak compositional properties, a claim with implications for evolutionary theories of language [33, 41, 51-54]. Arguably, however, the evidence for compositionality is better for passerine birds, with empirical evidence for compositional structure in both Japanese tits [55] and southern pied babblers [56, 57]. For example, Japanese tits produce 'alert' calls to predators, 'recruitment' calls to non-dangerous situations and 'alert-recruitment'

sequences to incite mobbing behaviour to a predator. Here, permutation matters since (non-natural) 'recruitment-alert' concatenations do not trigger any responses. Similar findings have been reported in southern pied babblers, suggesting that this type of compositional syntax may be widespread in passerine birds.

Is merge uniquely human?

Related to this is the distinction between expressions with linear and hierarchical structures. Townsend et al. [8] discuss the sentence "Duck and cover!" as an example of simple, non-hierarchical compositionality in contrast to "Neither did the linguist we consulted object nor was she interested" as an example of a complex expression that consists of three hierarchically organised sentences. Not all human syntax is hierarchical but an evolutionary theory of syntax must encompass the entire range of phenomena. Nevertheless, it is interesting that all currently known examples of permutation syntax in animals consist of systems containing only two calls. Future studies may provide different results, but for now animal and human syntax may simply differ in complexity of 'merge' operations, but not in kind [58], with 0-merge system containing only single lexical items with no combinations and 1-merge systems consisting of simple combinations, with no recursion. As said, all animal syntax appears to fall into this category. 2-merge systems, then, allow for recursion with previously merged complexes whereas 3-merge systems allow for merges of two already merged complexes [58]. This requires an individual to keep a merged complex in short-term memory, while processing further mergers. The different degrees of merging, in this view, may simply be due to short-term memory limits, with only larger brained species being capable of carrying out merges of already merged units. Whether or not animals are limited by a capacity for short-term retention of merged complexes (and whether such a capacity is sufficient for recursion) needs to be tested with specific research.

How does syntax evolve?

The saltation hypothesis is unlikely to carry much weight in evolutionary discussions of language origins. Although saltation can happen it is not known to play a role in brain evolution. Syntax as a by-product of enhanced computational capacity is more plausible, especially in light of the proposal that complex phrase structure merges require short-term memory capacities, perhaps only available to humans. Regarding the hypothesis that syntax is an evolutionary response to increasing lexicon demands, no such pattern is visible in the published record of animal syntax. Some of the most interesting examples are from systems that consist only of a small number of calls (Campbell's monkey alarm calls, putty-nosed monkey alarm calls), suggesting that the evolution of further structure would have been possible.

How then did human syntax evolve? One main unresolved problem concerns the transition from non-hierarchical, 2-merge units in animal communication to generative, hierarchically structured, higher-order merges only seen in human language (e.g. [59]). It is possible that this transition actually involved two independent evolutionary events, a capacity for advanced articulatory control and a capacity to perceive others as governed by mental states. Syntax, in this scenario, can only emerge in species that have an urge to communicate perceptions of natural events to others ('Mitteilungsbedürfnis'), which requires some understanding of other minds. Natural events are structured in hierarchical ways, i.e., in terms of actors and patients interacting in intentional ways, but this needs to be loaded upon a linear signal output, which requires a much richer signal repertoire than normally available to non-human primates. The evolution of syntax and compositionality, in this view, can only emerge in species that have evolved the computational resources to perceive, represent and recall natural events in such hierarchical ways and that have a signal system capable of conveying such information.

Future directions

The event perception hypothesis has not been subject to much empirical research, at least in animals. One possibility is that syntax-ready brains have evolved before modern humans, whereas the main transition towards language merely consisted of a communicative externalisation of the cognitive propensity to perceive events in intentional ways (agents, patients). A first question therefore is whether non-human primates

perceive and represent events in human-like ways or whether their perceptions are syntactically unstructured and holistic. If primates perceive events like humans, we expect them to visually explore social scenes in rule-governed ways along the main grammatical functions, for example by looking first or longer at the likely *actor* ('subject': NP) compared to the likely *patient* ('predicate': VP), regardless of the actor's and patient's actual appearance, emotional valence, or social status ('complement': Adjective). Related question are whether primates perceive inter-dependent event sequences as chronologically structured, suggesting that they may be more surprised by unexpected changes in the actor (NP) than unexpected changes in the action (VP), and whether primates are able to communicate such categories to others. In sum, research on the cognition of event perception is likely to contribute in new ways to the emerging science of animal syntax, currently dominated by studies on natural communication systems, towards a more comprehensive theory of language evolution.

Acknowledgments

References

[1] Langergraber, K. E., Prufer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., Inoue, E., Inoue-Muruyama, M., Mitani, J. C., Muller, M. N., et al. 2012 Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *P Natl Acad Sci USA* **109**, 15716-15721. (DOI:Doi 10.1073/Pnas.1211740109).
[2] Lameira, A. R., Maddieson, I. & Zuberbuehler, K. 2014 Primate feedstock for the evolution of consonants. *Trends Cogn Sci* **18**, 60-62. (DOI:10.1016/j.tics.2013.10.013).
[3] Fitch, W. T., de Boer, B., Mathur, N. & Ghazanfar, A. A. 2016 Monkey vocal tracts are speech-ready. *Science Advances* **2**. (DOI:10.1126/sciadv.1600723).
[4] Kersken, V., Zuberbuhler, K. & Gomez, J.-C. 2017 Listeners can extract meaning from non-linguistic infant vocalisations cross-culturally. *Sci Rep-Uk* **7**. (DOI:10.1038/srep41016).
[5] Laporte, M. N. C. & Zuberbuehler, K. 2011 The development of a greeting signal in wild chimpanzees. *Developmental Sci* **14**, 1220-1234. (DOI:10.1111/j.1467-7687.2011.01069.x).
[6] Zuberbühler, K. 2011 Cooperative breeding and the evolution of vocal flexibility. In *Oxford handbook of language evolution* (eds. M. Tallerman & K. Gibson). Oxford, Oxford University Press.
[7] Tomasello, M. 2009 The usage-based theory of language acquisition. *Cambridge Handbook of Child Language*, 69-87.
[8] Townsend, S. W., Engesser, S., Stoll, S., Zuberbuhler, K. & Bickel, B. 2018 Compositionality in animals and humans. *Plos Biology* **16**. (DOI:10.1371/journal.pbio.2006425).

[9] Jackendoff, R. 1997 *The architecture of the language faculty*. Cambridge, MA, MIT Press.

[10] Chomsky, N. 1965 *Aspects of the theory of syntax*. Cambridge, Mass, MIT Press.

[11] Chomsky, N. 2006 *Language and mind*. Cambridge MA, Cambridge University Press.

[12] Serres, M. H., Kerr, A. R., McCormack, T. J. & Riley, M. 2009 Evolution by leaps: gene duplication in bacteria. *Biology Direct* **4**, 46.

[13] Dufresne, F. & Herbert, P. D. N. 1994 Hybridization and origins of polyploidy. *Proceedings: Biological Sciences* **258**, 141–146.

[14] Herculano-Houzela, S. & Kaasb, J. H. 2011 Gorilla and Orangutan Brains Conform to the Primate Cellular Scaling Rules: Implications for Human Evolution. *Brain Behav Evol* **77**, 33-44. (DOI:doi: 10.1159/000322729).

[15] Bolhuis, J. J., Tattersall, I., Chomsky, N. & Berwick, R. C. 2014 How Could Language Have Evolved? *Plos Biology* **12**. (DOI:10.1371/journal.pbio.1001934).

[16] Arnold, K. & Zuberbuhler, K. 2006 Semantic combinations in primate calls. *Nature* **441**, 303-303.

[17] Nowak, M. A., Plotkin, O. B. & Jansen, V. A. A. 2000 The evolution of syntactic communication. *Nature* **404**, 495 - 498.

[18] Zuberbühler, K. 2019 Evolutionary roads to syntax. *Anim Behav* **151**, 259-265.

[19] Dasser, V., Ulbaek, I. & Premack, D. 1989 The perception of intention. *Science* **243**, 365-367.

[20] Friederici, A. D. 2004 Processing local transitions versus long-distance syntactic hierarchies. *Trends Cogn Sci* **8**, 245-247. (DOI:10.1016/j.tics.2004.04.013).

[21] Fitch, W. T. & Hauser, M. D. 2004 Computational constraints on syntactic processing in a nonhuman primate. *Science* **303**, 377-380.

[22] Abe, K. & Watanabe, D. 2011 Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience* **14**, 1067-U1173. (DOI:10.1038/nn.2869).

[23] Beckers, G. J. L., Berwick, R. C., Okanoya, K. & Bolhuis, J. J. 2017 What do animals learn in artificial grammar studies? *Neuroscience and Biobehavioral Reviews* **81**, 238-246. (DOI:10.1016/j.neubiorev.2016.12.021).

[24] Zuberbuhler, K. 2018 Combinatorial capacities in primates. *Current Opinion in Behavioral Sciences* **21**, 161-169. (DOI:10.1016/j.cobeha.2018.03.015).

[25] Garland, E. C., Rendell, L., Lamoni, L., Poole, M. M. & Noad, M. J. 2017 Song hybridization events during revolutionary song change provide insights into cultural transmission in humpback whales. *P Natl Acad Sci USA* **114**, 7822-7829. (DOI:10.1073/pnas.1621072114).

[26] Marshall, J. T. & Marshall, E. R. 1976 Gibbons and their territorial songs. *Science* **193**, 235-237.

[27] Clarke, E., Reichard, U. H. & Zuberbühler, K. 2006 The Syntax and Meaning of Wild Gibbon Songs. *Plos One* **1**. (DOI:10.1371/journal.pone.0000073).

[28] Andrieu, J., Malaivijitnond, S., Reichard, U. H. & Zuberbühler, K. in prep White-handed gibbons discriminate context-specific song compositions.

[29] Candiotti, A., Zuberbuehler, K. & Lemasson, A. 2012 Context-related call combinations in female Diana monkeys. *Animal Cognition* **15**, 327-339. (DOI:10.1007/s10071-011-0456-8).

[30] Ouattara, K., Lemasson, A. & Zuberbühler, K. 2009 Campbell's monkeys use affixation to alter call meaning. *Plos One* **4**, e7808.

[31] Coye, C., Ouattara, K., Zuberbuehler, K. & Lemasson, A. 2015 Suffixation influences receivers' behaviour in non-human primates. *P Roy Soc B-Biol Sci* **282**, 50265-50265.

[32] Zuberbühler, K. 2019 Evolutionary roads to syntax. *Anim Behav* **in press**.

[33] Schlenker, P., Chemla, E., Schel, A. M., Fuller, J., Gautier, J. P., Kuhn, J., Veselinovic, D., Arnold, K., Casar, C., Keenan, S., et al. 2016 Formal monkey linguistics. *Theoretical Linguistics* **42**, 1-90. (DOI:10.1515/tl-2016-0001).

[34] Ouattara, K., Lemasson, A. & Zuberbuhler, K. 2009 Campbell's monkeys concatenate vocalizations into context-specific call sequences. *P Natl Acad Sci USA* **106**, 22026-22031.

[35] Zuberbuhler, K. 2002 A syntactic rule in forest monkey communication. *Anim Behav* **63**, 293-299. (DOI:Doi 10.1006/Anbe.2001.1914).

[36] Fedurek, P., Zuberbühler, K. & Dahl, C. D. 2016 Sequential information in a great ape utterance. *Sci Rep-Uk* **6**, 38226.

[37] Arnold, K. & Zuberbuhler, K. 2006 The alarm-calling system of adult male putty-nosed monkeys, Cercopithecus nictitans martini. *Anim Behav* **72**, 643-653.

[38] Arnold, K. & Zuberbühler, K. 2006 Semantic combinations in primate calls. *Nature* **441**, 303-303.

[39] Arnold, K. & Zuberbühler, K. 2008 Meaningful call combinations in a non-human primate. *Curr Biol* **18**, R202-R203. (DOI:10.1016/j.cub.2008.01.040).

[40] Arnold, K. & Zuberbuhler, K. 2012 Call combinations in monkeys: Compositional or idiomatic expressions? *Brain Lang* **120**, 303-309. (DOI:Doi 10.1016/J.Bandl.2011.10.001).

[41] Schlenker, P., Chemla, E. & Zuberbuhler, K. 2016 What Do Monkey Calls Mean? *Trends Cogn Sci* **20**, 894-904. (DOI:10.1016/j.tics.2016.10.004).

[42] Caesar, C., Byrne, R., Young, R. J. & Zuberbuehler, K. 2012 The alarm call system of wild black-fronted titi monkeys, Callicebus nigrifrons. *Behav Ecol Sociobiol* **66**, 653-667. (DOI:10.1007/s00265-011-1313-0).

[43] Caesar, C., Byrne, R. W., Hoppitt, W., Young, R. J. & Zuberbuehler, K. 2012 Evidence for semantic communication in titi monkey alarm calls. *Anim Behav* **84**, 405-411. (DOI:10.1016/j.anbehav.2012.05.010).

[44] Cäsar, C., Zuberbühler, K., Young, R. J. & Byrne, R. W. 2013 Titi monkey call sequences vary with predator location and type. *Biol Letters* **9**. (DOI:10.1098/rsbl.2013.0535).

[45] Berthet, M., Neumann, C., Mesbahi, G., Casar, C. & Zuberbuhler, K. 2018 Contextual encoding in titi monkey alarm call sequences. *Behav Ecol Sociobiol* **72**. (DOI:10.1007/s00265-017-2424-z).

[46] Berthet, M., Mesbahi, G., Pajot, A., Cäsar, C., Neumann, C. & Zuberbühler, K. 2019 Titi monkeys combine alarm calls to create probabilistic meaning. *Science Advances* **5**, eaav3991. (DOI:10.1126/sciadv.aav3991).

[47] Clay, Z. & Zuberbuhler, K. 2009 Food-associated calling sequences in bonobos. *Anim Behav* **77**, 1387-1396. (DOI:Doi 10.1016/J.Anbehav.2009.02.016).

[48] Clay, Z. & Zuberbühler, K. 2011 Bonobos extract meaning from call sequences. *Plos One* **6**. (DOI:10.1371/journal.pone.0018786).

[49] Szabó, Z. G. 2017 Compositionality. In *The Stanford Encyclopedia of Philosophy* (ed. E. N. Zalta), Summer 2017 Edition ed.

[50] Clay, Z., Archbold, J. & Zuberbuhler, K. 2015 Functional flexibility in wild bonobo vocal behaviour. *PeerJ* **3**, e1124-e1124. (DOI:10.7717/peerj.1124).

[51] Schlenker, P., Chemla, E., Arnold, K., Lemasson, A., Ouattara, K., Keenan, S., Stephan, C., Ryder, R. & Zuberbuhler, K. 2014 Monkey semantics: two 'dialects' of Campbell's monkey alarm calls. *Linguistics and Philosophy* **37**, 439-501. (DOI:10.1007/s10988-014-9155-7).

[52] Schlenker, P., Chemla, E., Arnold, K. & Zuberbuhler, K. 2016 Pyow-hack revisited: Two analyses of Putty-nosed monkey alarm calls. *Lingua* **171**, 1-23. (DOI:10.1016/j.lingua.2015.10.002).

[53] Schlenker, P., Chemla, E., Casar, C., Ryder, R. & Zuberbuhler, K. 2017 Titi semantics: Context and meaning in Titi monkey call sequences. *Natural Language & Linguistic Theory* **35**, 271-298. (DOI:10.1007/s11049-016-9337-9).

[54] Schlenker, P., Chemla, E., Schel, A. M., Fuller, J., Gautier, J. P., Kuhn, J., Veselinovic, D., Arnold, K., Casar, C., Keenan, S., et al. 2016 Formal monkey linguistics: The debate. *Theoretical Linguistics* **42**, 173-201. (DOI:10.1515/tl-2016-0010).

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

[56] Engesser, S., Crane, J. M. S., Savage, J. L., Russell, A. F. & Townsend, S. W. 2015 Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System. *Plos Biology* **13**. (DOI:10.1371/journal.pbio.1002171).

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

[58] Rizzi, L. 2016 Monkey morpho-syntax and merge-based systems. *Theoretical Linguistics* **42**, 139-145. (DOI:10.1515/tl-2016-0006).

[59] Bolhuis, J. J., Beckers, G. J. L., Huybregts, M. A. C., Berwick, R. C. & Everaert, M. B. H. 2018 Meaningful syntactic structure in songbird vocalizations? *Plos Biology* **16**. (DOI:10.1371/journal.pbio.2005157).

Please note that references to datasets must also be included in the reference list with DOIs where available.

Tables

Table 1. Key terminology and definitions

Term	Definition
Syntax	A set of principles by which meaning-bearing units can be combined into well-formed complexes
Grammar	A synonym for 'syntax' or an umbrella term encompassing the phonology, morphology and syntax of a language
Duality of patterning	A characteristic of language by which structural organisation is found at two levels: Meaningless phonemes can be combined into morphemes and words whereas (meaningful) morphemes and words can be combined into phrases and sentences
Compositionality	A process by which meaning is determined by the meanings of the constituent parts and the rule that combines them
Merge	A mental operation that takes two syntactic elements and assembles them to form a set
Recursion (in linguistics)	Application of a linguistic rule to the result of the same rule (=nesting)

Table 2. Playback stimuli and receiver foraging efforts in bonobos at Twycross Zoo, England

Signallers		Receivers			
Food	**Sequence**	**Kiwi field**	**Apple field**	**Kiwi bias**	**Apple bias**
Kiwi	B B P B	21.0	2.5	9.4	1.1
Kiwi	B B P B	6.5	2.5	3.6	1.4
Kiwi	B B P PY	28.3	5.0	6.7	1.2
Apple	PY B B PY	20.0	12.0	2.7	1.6
Kiwi	P P PY P	79.0	18.8	5.2	1.2
Kiwi	PY P P P	20.8	1.8	12.6	1.1
Kiwi	P PY PY P	1.3	2.5	1.5	2.9
Kiwi	P P PY Y	16.5	6.5	3.5	1.4
Apple	PY P PY PY	14.8	3.3	5.5	1.2
Apple	Y PY PY P	0.0	15.8	1.0	>12
Apple	Y PY P Y	9.3	2.0	5.7	1.2
Apple	PY PY Y PY	5.8	20.8	1.3	4.6
Apple	Y P Y Y	6.5	14.3	1.5	3.2
Apple	Y Y Y P	3.8	40.8	1.1	11.7
Apple	Py PY Y Y	9.5	9.0	2.1	1.9
Apple	PY Y PY Y	6.5	17.3	1.4	3.7
Apple	PY Y Y Y	2.5	10.3	1.2	5.1

Table 3. the role of bigrams in bonobo food calling

Bigram	Kiwi search (s) median	Apple search (s) (median)	Kiwi bias (median)	Apple bias (median)	N
B	20.5	3.8	5.1	1.3	4
P	20.8	6.5	5.2	1.2	3
PY	5.8	9.0	1.5	2.9	5*
Y	5.2	12.3	1.3	4.2	4*
N/A	9.3	2.0	5.7	1.2	1
N/A	6.5	17.3	1.4	3.7	1

*1 of 15 sequence contained both Y and PY bigrams; N/A no bigrams within the first 4 calls

Tables should be inserted at the end of the document, unless they are provided in another format (e.g. Excel) in which case they should be supplied as a separate file. Please provide all tables in an editable format (rather than embedded as an image).

Figure and table captions

Table 1. Key terminology and definitions

Table 2. Playback stimuli and receiver foraging efforts in bonobos at Twycross Zoo, England

Table 3. the role of bigrams in bonobo food calling

Figure 1. Tree-structure of a complex English expression conveying a sequence of three social events involving people interacting with a linguist, first as a patient (of a consultation), then as an agent (responding to consultation, showing lack of interest). A adjective; AP adjective phrase; C conjunction; CP conjunction phrase; D determiner; I inflection-bearing element; IP inflectional phrase; N (pro)noun; NP noun phrase; S sentence; V verb; VP verb phrase. Reprinted from [8] under the Creative Commons Attribution license.

Figure 2. Suffixation in Campbell's monkey alarm calls: (b) krak alarm calls (given to leopards) have a descending frequency transition (large dashed red arrow) and can take on an optional 'oo' suffix (dashed red oval line) to form krak-oo. (c) hok alarm calls (given to crowned eagles) have no frequency transition (small dashed red arrow) and can take on an optional 'oo' suffix (dashed red oval line) to form hok-oo (f). Reprinted from [30] under the Creative Commons Attribution license.

Figure 3. Spectrogram and information content of chimpanzee pant hoot calls. The top panel refers to the location within the expression where different types of information were most strongly encoded. The bottom panel depicts a spectrogram of a typical chimpanzee pant hoot, featuring all four call units. Reprinted from [36] under the Creative Commons Attribution license.

Figure 4. Proportion of time the listener spent looking upwards depending on the experimental condition. The figure shows raw data (one line per individual), model estimates (black circles) and bootstrapped model estimates (coloured circles, 1,000 bootstraps). Subjects looked more upwards when they were tested with sequences elicited by an aerial predator or by a predator in the canopy. Reprinted with permission from [46].

Figure 5. Assessments of meaning of calls given by bonobos when encountering foods of differently perceived qualities. Reprinted from [47] under the Creative Commons Attribution license.

Figures

Supplementary material

Supplementary material can be used for supporting data sets, movies, figures and tables, and any other supporting material. The main article, however, should stand on its own merit. Where possible, supplementary material should be combined into one Word document or PDF.

Additional Information

During the submission process you will be asked to provide the following information. You do not need to include this in your manuscript. These statements will be available to reviewers and included in the published paper.

Ethics
Research on humans must include a statement detailing ethical approval and informed consent. Research using animals must adhere to local guidelines and state that appropriate ethical approval and licences were obtained. Please read our editorial policies carefully before submission. Please send relevant documentation to the Editorial Office.

Data Accessibility
All manuscripts which report primary data (usually research articles) must include a statement about where the article's supporting data, materials (such as statistical tools, protocols, software) and code can be accessed. As a minimum, sufficient information should be deposited so that published research is fully reproducible, and enough primary data provided so that the results reported can be verified. Please read our data sharing policies carefully before submission. If the data is included in the article's Supplementary Material this should be stated here, for instance: 'The datasets supporting this article have been uploaded as part of the

Supplementary Material.'

Authors' Contributions

All submissions, other than those with a single author, must include an Authors' Contributions section which individually lists the specific contribution of each author. The list of Authors should meet all of the following criteria; 1) substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; 2) drafting the article or revising it critically for important intellectual content; and 3) final approval of the version to be published. All contributors who do not meet all of these criteria should be included in the acknowledgements.

Competing Interests

All manuscripts must include a competing interests section. If you have no competing interests please state *'I/We have no competing interests.'*

Competing interests are defined as those that, through their potential influence on behaviour or content or from perception of such potential influences, could undermine the objectivity, integrity or perceived value of publication. For more details please read our editorial policies. You must disclose any relationships, whether professional or personal, with the Guest Editors of the issue.

Figure 1. Tree-structure of a complex English expression conveying a sequence of three social events involving people interacting with a linguist, first as a patient (of a consultation), then as an agent (responding to consultation, showing lack of interest). A adjective; AP adjective phrase; C conjunction; CP conjunction phrase; D determiner; I inflection-bearing element; IP inflectional phrase; N (pro)noun; NP noun phrase; S sentence; V verb; VP verb phrase. Reprinted from [8] under the Creative Commons Attribution license.