Abstract: For several decades, ethologists and comparative psychologists have been using a linguistic terminology to discuss complex communicative abilities in animals, with a particular focus on sound combinatorial rules. One historical example is the possible syntactic ability of songbirds. More recently, context-dependent call combinations have been described in nonhuman primates. This time, the detailed observational and experimental data gathered in this area has even drawn the attention of linguists and has given rise to studies highlighting the relevance of linguistic tools for the study of nonhuman primate communication systems. However, the parallels that can be drawn between humans’, birds’ and nonhuman primates’ verbal/vocal combinations still remain the topic of intense debate possibly because mismatches between the terminologies used have confounding effects. The question is: can we go beyond the traditional dichotomy between phonological and lexical syntax to characterize the diversity of sound combinatorial rules found in animals? Here, we will adopt a two-step approach in order to discuss: (1) what forms sound combination takes in animals, based on structural and functional criteria and when it may or not be appropriate to use linguistic terms; (2) why sound combination may have evolved in some species more than others. We will notably illustrate our arguments with recent findings in some cooperative breeding birds and guenons, where cases of meaningful sound compositionality have been recently described.

Keywords: sounds combinations, verbal combinations, animal communication

1. Introduction

The communicative abilities of animals have received considerable research interest over the years. One likely contributing factor is that animal communication constitutes a particularly fruitful substrate for comparative
analyses with human language. Language, in its full blown form is a product of a variety of communicative and socio-cognitive capacities[11]. This complexity, together with its relatively short window of emergence[12,3], has led to the suggestion that, as with other complex biological phenomena, language might have evolved from pre-existing capacities and structures initially serving other functions[3–5]. One way to shed light on the potential evolutionary path leading to the emergence of language is to decompose it into several core features and to explore their presence and role in non-human animals from various taxa[6,7]. This approach, using animals from distinct and phylogenetically distant taxa is particularly relevant in helping disentangle the relative influence of various factors involved in the evolution of complex communicative abilities culminating in human language.

One relevant historical example is that of capacities for vocal learning, which have been described in some birds (e.g. parrots, starlings and mockingbirds) and mammalian species (e.g. some bats and marine mammals)[8,9] in which individuals are able to acquire new vocalisations. Several studies provide convincing evidence for the importance of social learning and auditory feedback for call and song acquisition[10–13] which are central to language acquisition in humans. In addition, some studies even reported the cultural transmission of vocal dialects in several bird and mammal species[14–16]. Given, the distribution of vocal learning species in distant taxa these capacities are assumed to result from convergent evolution[17,18]. Moreover, the study of the neural substrates and genetics of species that possess vocal learning capacities, compared with humans and with non-learning species (e.g. gull, dove, suboscines) has generated hypotheses for the emergence of vocal learning[18–21]. For instance, comparative studies of the structure and expression patterns of the now famous FoxP2 gene, have shed light on its potential role in the emergence of vocal learning in the human lineage: using data from comparative studies and clinical studies in human, some authors proposed that ancient neural functions of FoxP2 have been co-opted to subserve aspects of vocal communication, and notably vocal learning, in several species including humans[22–25]. This example demonstrates the utility of broad comparative studies to clarify particular aspects of the evolution of communication. However, the study of FOXP2 in isolation is not sufficient and other factors are likely involved and have to be explored regarding vocal learning in animals[22]. For instance, some
so-called ‘non-learners’ have recently been shown to display significant abilities in vocal plasticity (e.g. elephant\[26\], goat\[27\], marmoset\[28\], guenons [29–31], gibbons\[32,33\]) questioning the relevance of the traditional learner/non-learner dichotomy.

Several other capacities involved in language have been studied and described to various extents in non-human species, including many non-human primates. For instance, various species display the capacity to produce intentional signals\[34–36\] or semantic-like signals (also termed referential signals i.e. that refer to an external object of the word)\[35,37–43\]. Also, vocal exchanges in some species are strictly organised and display “conversation-like” properties (i.e. based on call overlap avoidance and turn-taking between exchanging partners\[44,45\]). Furthermore, the capacity to combine sounds into complex structures has long been a topic of much contention\[46\]. Indeed, combinations of vocal units have been extensively reported in animals from various taxa, including numerous bird species (e.g. winter wren (Troglodytes troglodytes)\[47\], Bengalese finches (Loncura striata)\[48\], mockingbirds\[49\], European starlings (Sturnus vulgaris)\[50\], several species of chickadees\[51,52\], and blue-throated Hummingbirds (Lampornis clemenciae)\[53\]) as well as mammals such as rock hyraxes (Procavia capensis)\[54\], several species of bats (i.e. mustached bats (Pteronotus parnellii)\[55\]; free-tailed bats (Tadarida brasiliensis)\[56,57\]; sac-winged bats (Saccopteryx bilineata)\[58\]), whales (humpback whales (Megaptera novaeangliae)\[59,60\]; Killer whales (Orcinus orca)\[61,62\]; pilot whales (Globicephala sp.)\[63\]; sperm whales (Physeter macrocephalus)\[64\]) and non-human primates (e.g cotton-top tamarins (Saguinus Oedipus)\[65\]; gorillas (Gorilla sp.)\[66\]; red-bellied titi monkeys (Callicebus moloch)\[67\]). Sound combinations occur in wide diversity of contexts, such as alarm contexts\[68–70\], socio-positive interactions\[71–73\], mate attraction or territorial defence\[54,74–76\] and can take various forms. Some animals, for example, can merge acoustic units (i.e. basic element consisting of a continuous mark on a sonogram, also termed notes in birds) into combined calls (e.g. consisting of several units merged linearly with little to no silence between them, also termed motifs in birds). Furthermore, calls (simple and/or combined ones) can then also be combined into higher-order call sequences (i.e. series of calls uttered in sequence and separated by a silent interval always shorter than silent gaps between sequences, also termed songs in birds)\[56,71,77\].
Combinatoriality is central to language, and drawing parallels between sound combinations in animals and the combinatorial systems of language is tempting. Language relies on the combination of sounds (phonemes, see Table 1 for a definition) into larger units (morphemes and words) which are themselves combined into larger utterances (sentences)[78]. However, useful comparisons are hard to achieve partly because, contrary to most ethological definitions given for sound combinations, linguistic definitions often heavily rely on functional aspects and include element’s meaning or grammatical function as a way to characterise them[78–80].

Language’s generativity (i.e. capacity to generate an infinite number of ends using finite means) is a product of dual articulation[81,82], which allows combination at two distinct layers: phonology and morphosyntax[83] (Table 1). Phonology corresponds to the combination of meaningless sounds (i.e. phonemes) into meaningful elements (i.e. morphemes and monomorphemic words). Simply put, a phoneme is a sound which, when added, deleted or used to replace another sound in a word, creates a phonemic contrast changing the meaning of the word. For example, in English the sounds /k/ and /b/ are phonemes as they differentiate the words ‘cat’ and ‘bat’. Two words that differ only by one phoneme are termed ‘minimal pair’. Morphosyntax (Table 1), corresponds to the second layer of combination, in which meaningful elements (morphemes and words) are combined into larger structures whose meaning depends on the elements composing them and their order. Some words consist of only one morpheme (i.e. monomorphemic words such as ‘happy’) but morphemes can also be combined together into polymorphemic words. For example, the word “happy” can be combined with the suffix “ness” to create the polymorphemic word “happiness” or with the prefix “un” to create ‘unhappy’. Finally, in phrases, morphemes are combined according to grammatical rules[84]. These rules are a key contributor to language’s generativity: with a finite number of rules it is possible to generate, using a finite number of elements, an infinite number of structures among which rules distinguish well-formed (or grammatical) syntactic structures from ill-formed (or non-grammatical) syntactic structures[78,85].
**Table 1: Main linguistic concepts and definitions to be used in this chapter**

<table>
<thead>
<tr>
<th>Concept</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phoneme</strong></td>
<td>Smallest meaning differentiating sounds in a language[^78] i.e. meaningless sounds that allow to differentiate between two words.</td>
</tr>
<tr>
<td><strong>Morpheme</strong></td>
<td>A minimal unit of meaning or grammatical function[^78].</td>
</tr>
<tr>
<td><strong>Phonology</strong></td>
<td>Combinatorial layer of language in which phonemes are combined to form morphemes and words[^78,83].</td>
</tr>
<tr>
<td><strong>Morphosyntax</strong></td>
<td>Combinatorial layer of language in which meaningful elements are combined into larger structures whose meaning depends on the elements composing them and their order. Morphosyntax includes both morphology, where morphemes can be combined into more complex structures (i.e. polymorphemic words), and syntax, where mono- and polymorphemic words are combined into sentences[^85,86].</td>
</tr>
<tr>
<td><strong>Dual articulation</strong></td>
<td>Characteristic of language whereby speech can be analysed at two complementary levels: phonology and morphosyntax. Duality of patterning has been characterised as a design-feature of language which is partly responsible for language’s virtually infinite generativity[^81,87].</td>
</tr>
<tr>
<td><strong>Scalar implicatures</strong></td>
<td>Linguistic concept related to pragmatic inference. The core idea is that the utterance of a sentence S implicates the falsity of stronger alternatives (i.e. more informative ones) as for any stronger alternative S’ to S, a cooperative speaker would have used S’ rather than S if s/he believed S’ to be true[^88–90]. <em>e.g.</em> the sentence S “some of my trees are oaks” implies that not all my tree are oaks as, if all were, I would have used the more informative S’ sentence “all of my trees are oaks”.</td>
</tr>
</tbody>
</table>

Peter Marler[^91], proposed to differentiate between animal combinatorial structures depending on their organisation and the likely meaning of their components. To this end, he used terms borrowed from linguistics and distinguished two main types of organization: phonological syntax (or phonocoding) and lexical syntax (or lexicoding). He defined phonological syntax as the concatenation of sounds without independent information content and which are not used singularly, or meaningful sounds that lose their original content when combined. He defined lexical syntax as the level at which meaningful elements are combined. Whilst Marler borrowed terms from linguistics, several important differences with human phonology and morphosyntax remain. First, the concept of ‘meaning differentiation”
(i.e. change in meaning of the whole resulting from change in one of its meaningless elements) is absent from Marler’s definition of phonological syntax. In addition, contrarily to phonology, Marler’s phonological syntax allows the combination of meaningful units. Finally, the definition he proposed for lexical syntax – though closer to morphosyntax than phonological syntax is to phonology - differs strikingly from its ‘human’ counterpart as the importance of elements’ order on combined call meaning is overlooked while the order of words and morphemes are of central importance in most human utterances.

Recently, the study of combinatorial systems in animals has received renewed interest from researchers from various fields including ethology, linguistics and psychology\[46,86,88\]. A multitude of studies have illuminated new perspectives and subsequently given rise to the development of interdisciplinary work\[88,92,93\] as well as prompting controlled experiments investigating combinatorial structures found in the natural communication of animals and their relevance to receivers\[69,70,94–96\]. Within this framework, some authors have questioned the relevance of the definitions proposed by Marler notably because the joint use of the terms ‘phonological’ and ‘syntax’, which correspond to very distinct linguistic concepts, is misleading\[86\]. We concur and forward that further reflection is now necessary in order to develop an accurate terminology to characterise the structure and functional aspects of sound combination in animals. In addition to providing the groundwork facilitating understanding of animal combinatorial systems, this will best serve comparative analyses with language.

This chapter therefore aims to review comparative work on sound combinations in humans and animals, with two intended outcomes:

- to propose a basis for future interdisciplinary work aiming to develop a more appropriate terminology, and shed light on some potentially fruitful prospects for future studies of sound combination.
- To elucidate evolutionarily relevant factors likely to have influenced the development of combinatorial communication systems.

We will firstly turn our attention to the variety of combinatorial structures found in primates and birds. We will (1) examine possible bases to define rudimentary parallels with sound combination in language and (2) review
recent empirical studies providing convincing evidence for combinatorial capacities parallel to language in animals. In the third part of this chapter we will focus on the recent advances brought by the use of formal linguistic analyses on animal communication systems. Finally, in line with the comparative rationale adopted at the beginning of this chapter we will build on the taxonomic diversity of examples described as a way to formulate potential hypotheses regarding why combinatorial systems emerge.

2. Combinatorial systems: diversity and terminology

This section reviews various animal combinatorial systems and evaluates the terminology used to characterize them. Given the key role played by meaning in combinatorial systems, particularly when comparing combinatorial structures with those in human language, providing a clarification is important. We will use the term meaning in a form approaching Gricean natural meaning[97,98], i.e. as the significance/information that receivers derive from a signal and its context (because of its regular association with a given event, individual or object), without assuming emitter’s intention to inform others[99,100].

2.1 Parallels with phonology?

Many bird species rely on the combination of apparently meaningless units into larger structures. However, studies describing such systems often lack information on the contextual correlates of the combinatorial variants emitted[49,101]. Furthermore, experiments testing a potential intrinsic meaning of single units (or changes in meaning accompanying changes in the type or order of units combined) are often missing[48,102,103]. Some experiments in songbirds have shown that receivers’ reaction can be influenced by unit diversity[104], fine acoustic structure[105], or by the simultaneous modification of several frequency and temporal parameters[106], suggesting that information (about caller’s quality, or identity) is conveyed. Nevertheless, as frequently suggested by behavioural observations, such modifications, as well as changes in unit’s type or order, do not seem to alter the main function or “semantic content” of the sequence (i.e. mostly social bonding, mate attraction and/or territorial defence in the case of songbirds[47,49,74,86,104]). As a result, any parallel with the phonological layer of language is lim-
ited at best and subsequently it has been argued that such systems may be better described in terms of “phonetic patterning”, that relates to the physical properties of sounds but does not characterise sounds as meaning-differentiating\[86\].

Previous work in non-human primates, particularly gibbons has suggested possible additional parallels with phonological organisation seen in language\[107\]. An observational study on white-handed gibbons (Hylobates lar), for example, indicated that their communication, like that of songbirds, relies on the combination of apparently meaningless units into sequences. However as far as we know, in contrast to songbirds, gibbons give two types of sequences that are associated with strikingly distinct contexts and functions: one is produced routinely in the morning, while the other functions to signal the presence of a predator\[107\]. In both contexts, these sequences are given in duets during which two partners produce song in a coordinated way but the organisation of sequences differs between morning duets and predatory ones. More precisely, morning and predatory duets differ in three ways: (1) in the proportion of one type of note (the “hoo” note, with on average 100 vs 10 “hoo” notes introducing predatory induced songs and morning duets respectively), (2) in the order of motifs involved (female-specific calls is given later and answered slower by her male partner in predatory contexts) and (3) in the presence of two note types as (i.e. ‘learning-wa’ notes are globally absent from predator-induced songs while ‘sharp wow’ notes are absent from morning duets). Moreover, natural observations indicate that wild individuals react differently to the distinct sequences suggesting that the structuring of the signal encodes information. Further experimental work is now required to clarify how. Playback experiments comparing receivers’ reaction to natural sequences and artificial stimuli in which the order, proportion, and type of notes given are manipulated will be particularly necessary to identify what receivers use to discriminate between sequence types. In addition, further clarifications about whether notes’ acoustic structure varies between contexts and whether they possess an intrinsic meaning (notably ‘learning wa’ and ‘sharp wow’ notes) could be obtained using acoustic analyses and playbacks. Such information would help determine the nature of the system (i.e. showing parallels with phonology, morphology or neither) and may
also shed light on the possible cognitive processes underlying communication in this species.

Thus, although previous studies on birds and primates reviewed above match some of the criteria used to define phonological combination in language, none of them did so fully, primarily because the demonstration of changes in message according to sequence organisation (i.e. meaning-differentiation) was lacking or because the intrinsic meaning of notes was unclear. We propose that convincing evidence for parallels with phonology in animals would require: (1) a combination involving units that are not associated to any particular behavioural context (hence from which receivers could not individually extract specific information about the environment, or caller’s behaviour). (2) that the combination (or addition) of given “meaningless” units in a given order creates a signal which can be reliably associated with one (or several) external events or indeed a caller’s behaviour(s)\[96] and critically (3) that changes in unit order or composition triggers changes in signals’ content. Finally, to parallel in a rudimentary way the productivity of language, we would also expect such systems to involve the reuse of units across distinct types of utterances.

2.2 Parallels with morphology?

The second layer of language, morphosyntax, relies on the combination of meaningful sounds into larger structures whose meaning depends on their components and organisation. Several studies have described vocalisations composed of apparently meaningful calls but here, again, the parallel with the morphosyntactic organisation of language is not always clear. A series of studies investigating gorilla communication has described a potential combinatorial system in a great ape species\[66,108\]. Both mountain and Western gorillas possess a graded repertoire composed of five main types of close calls. Each type of call can be given alone or combined with every other close call unit in non-random ways. The authors analysed the contextual correlates of emission for three types of units and their most common combinations: atonal grunts (A1), short tonal grunts (T2) and grumbles (T4) as well as A1-T4 and T2-T4 combinations. The results show that, while T4 is given more in foraging contexts, in particular when there is no individual less than 5 meters from the emitter, A1 and T2 are associated with resting
contexts, notably when other individuals are around the emitter (i.e. <5m) and do not differ in their context of emission. A1-T4 and T2-T4 combinations are given in the same context as A1 and T2 calls but, in contrast to single calls, combinations are associated strongly with vocal exchanges. These results suggest that in this system T4 units, which may serve as a « localisation » call due to their longer duration, can be added to A1 or T2 units (whose « normal » context of emission is thus respected) during vocal exchanges. However, whether combinations triggers changes in the information content of the calls remains unknown, because receiver’s reaction to single and combined units have not been tested and, more importantly, because the contextual correlates of the vast majority of combinations given by gorillas (more than 150 different types\[66\]) have not been investigated yet. In addition, the role of repetition and call order in combined vocalisations that seem to vary greatly remains poorly understood\[108\].

A series of studies on the alarm call system of male putty-nosed monkeys also revealed an intriguing system which relies on the combination of calls that appear to carry meaning\[109–113\]. Indeed, male putty-nosed monkeys use two distinct loud calls « Pyow » and « Hack ». A first series of studies using natural observations, playbacks and predator presentation experiments suggested that sequences of « Pyows » were regularly given to leopards, while sequences of « Hacks » as well as transitional Hack series (i.e. several Hacks followed by several Pyows) were common responses to crowned-hawk eagles. Interestingly, Pyow-Hack sequences (i.e. 1–4 Pyows followed by 1–4 Hacks) reliably trigger movement (both natural sequences and sequences artificially composed of calls given in other contexts). The relationship between the apparent meaning of Pyow-Hack sequences and their components has raised questions and four main interpretations have been proposed\[86,112\]: (1) a phonological interpretation in which Pyow and Hack would work as « phonemes » i.e. allowing differentiation of meaning of single units (Pyow, Hack) and of their joint use (Pyow-Hack sequence)\[86\], (2) an idiomatic interpretation in which the original –compositional- meaning of Pyow-Hack sequence was blurred, similarly to human idioms (e.g. it’s raining dogs and cats)\[86,112,114\] and two more « semantic » interpretations: (3) one in which Pyow and Hack would respectively carry the meaning « move on the ground » and « move in the air » while Pyow-Hack sequence would carry a combined general meaning « we move, let’s go »
as putty-nosed monkeys occupy various strata at a time and can travel on
the ground as well as in the canopy\textsuperscript{[86]} and (4) an interpretation based on
weak meanings from Pyow (i.e. underspecified, general alarm) and Hack
(i.e. non-ground movement or high arousal depending on the analysis) and
inferences based on the pragmatic principles of competition and influence
of contextual cues\textsuperscript{[114]}. Now, further investigation of the possible mental
representations triggered by conspecific calls as well as putty-nosed mon-
keys’ capacities to handle and understand combinatorial structures more
generally are necessary to determine which of these interpretation (or oth-
ers) is most plausible.

The examples reviewed above show that combining meaningful calls into
larger structures (either combined calls or call sequences) is not sufficient to
offer a robust parallel with the morphosyntactic organisation of language.
In particular, we argue that to be considered as a rudimentary parallel with
morphosyntax, a system would obviously need to (1) involve the combina-
tion of vocal units, from which receivers can individually extract informa-
tion, into a larger structure. It would also need that the information content
changes depending on and reflects (2) the units merged together and their
respective content and (3) rules for unit combination (i.e. systematic order
of combination and consistent alteration of the information conveyed by
signal).

3. Focus on promising examples: the cases of babblers
and guenons

3.1 Parallels with phonology

To date and to our knowledge, only one study documenting note combina-
tions in chestnut-crowned babblers has provided convincing evidence for
a parallel with the phonological layer of language. In this study, Engesser
et al.\textsuperscript{[96]} combined natural observations, acoustic analyses and playback of
natural and artificially recombined sounds in chestnut-crowned babblers.
These cooperatively breeding birds living in arid areas of South-Western
Australia possess a vocal repertoire of discrete calls, most of which are com-
posed of apparently meaningless notes. Critically, some notes are reused
across call types, such as the ‘A’ and ‘B’ notes that can be combined together
into an ‘AB’ structure during flight (i.e. flight call) and a ‘BAB’ structure
during nestling provisioning (i.e. prompt call) (Figure 1). Acoustic analyses showed no difference in notes’ structure between call types.

Figure 1: Spectrogram of double-element flight call (i.e. F1 F2) and triple-element prompt call (i.e. P1 P2 P3) of adult chestnut-crowned babblers. Figure reproduced from Engesser et al. (2015) [94].

Receivers’ reaction did not differ between natural and artificial stimuli (i.e. artificial flight calls created by deleting the first ‘B’ unit of a prompt call and artificial prompt calls created by adding a ‘B’ unit to a flight call) within a call type. In addition, the broadcast of single ‘B’ units and artificial ‘CAB’ stimuli (‘C’ being a call element naturally given in combination with other notes by chestnut babblers) triggered surprised reactions that differed from those obtained by the broadcast of flight or prompt calls. These additional testing conditions thus ruled out a possible ‘priming effect’ of a ‘B’ element as well as responses being driven by superstructure effects [96]. Thus, the flight call/ prompt call complex in CCBs seems to match the three key criteria needed to draw a parallel with phonology i.e. (1) a combinations of ‘meaningless’ elements into (2) a structure meaningful to receivers and (3) which meaning changes if elements change order and presence. Indeed, the authors argue this example represents a rudimentary form of phonemic contrasts. Given several other calls in the repertoire of these birds form call pairs (i.e. two calls given in distinct contexts and that differ only by one note) [115], future studies will be needed to determine whether they also make use of a similar combinatorial mechanism. Tackling this question would be of particular importance as this would also help shed light on the productivity of the system (i.e. the extent to which notes are reused across the repertoire to create various types of utterances).
3.2 Parallels with morphosyntax?

The combination of meaningful calls, via a mechanism resembling morphosyntax can involve the merging of sounds into combined calls as well as the combination of sounds (separated by silent gaps) into call sequences. In light of this, we therefore propose to differentiate between combined calls and call sequences. Such a distinction is advantageous for several reasons. Firstly, because this distinction falls in line with the traditional distinction forged between words and phrases (and between calls and call sequences in animals). Secondly, it may serve to facilitate analyses, notably during preliminary phases of investigation. Finally, it may also be more realistic, as different underlying capacities, such as working memory requirements may be required to perform and interpret combinations of meaningful elements at these two levels in animals.

In addition, the combination of morphemes can involve two types of elements: bounded morphemes that are always used in conjunction with others (e.g. suffixes), and free morphemes that constitute monomorphemic words when used alone. An analogous form of the combination of these two types of morphemes in animals would correspond to the merging of one individual call (that can be used alone) with a vocal unit that is never given by itself and the merging of two individual calls respectively. Interestingly, evidence for both types of call combination (i.e. using individual calls or calls that are never used alone) have been recently reported in the literature of one cooperatively breeding bird (southern pied babbler, Turdoides bicolor) and of two species of guenons: Campbell’s and Diana monkeys (see Table 2 for a summary).

Southern pied babblers

A recent study highlighted a combination mechanism in the alarm calls of the southern pied babbler, a cooperatively breeding bird living in the arid areas of South-Africa. These birds emit an alert call with a broadband structure to low-urgency threats and a tonal, repetitive recruitment call in non-alarm contexts to attract group members to a new location (e.g. roosting or foraging). Critically, pied babblers sequentially combine alert and recruitment calls when encountering and mobbing terrestrial predators (Figure 2, Table 2).
Using a playback experiment, the authors tested the combinatorial structure of the mobbing sequence and its relevance to receivers by comparing wild, but habituated, pied babblers’ reaction to the broadcast of natural and artificially created stimuli (i.e. mobbing sequences created by combining alert and recruitment calls and single-call stimuli extracted from natural mobbing sequences). Subjects’ reaction to natural and artificial stimuli did not differ, which demonstrated that mobbing sequences consisted in the linear merging of alarm and recruitment calls and thus confirmed their combinatorial nature. Also, the distinct reactions given to the three call types presented (i.e. alert call, recruitment call and mobbing sequence) demonstrated the relevance of these calls to receivers. This study thus satisfies the three criteria proposed for parallels with linguistic morphosyntax in a non-human animal i.e. individually meaningful calls, combined into a meaningful structure whose meaning reflects that of the elements involved. Interestingly, receivers’ reaction to mobbing sequences exceeded the sum of reactions to their components (i.e. higher attentiveness and quicker approach). This suggests that, in this case, the combination of two elements did not simply lead to an addition of their meanings but potentially gave rise to a ‘richer’ meaning (i.e. ‘mobbing a predator’), that is related to, yet goes beyond, the meaning of its parts\[84,118\].

**Diana monkeys**

Other studies, focussing on the communication of two cercopithecids, Diana and Campbell’s monkeys, revealed meaningful combinatorial systems that could offer rudimentary parallels with the morphosyntactic organisation of language. These arboreal primates live sympatrically in the dense primary
forests of West Africa\cite{119,120} and their communication, which relies almost exclusively on sex-specific vocal signals, has been studied intensively over the past decades (e.g.\cite{29,30,40,121–128}). In Diana monkeys, females possess four main types of social calls: H, L, R and A. The first three calls are associated with distinct contextual valences for the caller (very positive social context, neutral to mildly positive context and socio-negative or mildly dangerous context respectively). The last call (A) is given in a broad range of contexts and strongly signals caller’s identity\cite{29,121}. Each of these calls can be given alone or in combination according to the following pattern: a contextual unit (i.e. H, L or R) merged with an arched unit (i.e. A) (Figure 3).

Figure 3: combined calls of female Diana monkeys. (a) HA call (socio-positive contexts), (b) LA (neutral to positive contexts) and (c) RA calls (negative contexts and mild danger).

To verify the combinatorial structure (i.e. whether apparently combined calls consisted in the merging of individual calls) and test the relevance to receivers of distinct combined calls, Coye et al.\cite{95} conducted a playback experiment on females from a wild habituated group of Diana monkeys. In particular, to determine the relevance to receivers of the contextual unit, they compared subjects’ reaction to the stimuli created by merging L or R units (i.e. relating to distinct contexts) with an A call from a group member (i.e. LA\textsubscript{Group} and RA\textsubscript{Group} stimuli). To determine whether A calls allowed receivers to identify the caller, they compared subject’s reaction to stimuli created by merging an R unit with either A calls from group members or A calls from females in a neighbouring group (i.e. RA\textsubscript{Group} and RA\textsubscript{Neighbour} stimuli). The change of one unit systematically triggered predictable changes
in receivers’ reaction. The results strongly suggest that the contact call system of female Diana monkeys relies on a combinatorial operation through which two independent calls are combined into a larger structure whose information content reflects its components. Hence, female Diana monkey’s contact call system matches the three criteria we proposed and may offer a parallel with morphosyntax.\[95\]

**Campbell’s monkeys**

Similarly to Diana monkeys, adult Campbell’s monkeys possess a sex-specific vocal repertoire as females’ communication relies mostly on social calls while males give mainly alarm calls\[116–117\]. A notable example from female Campbell’s monkeys involves the merging of a low-pitched trill (resembling Diana monkey’s L call), which can also be used alone and varies with caller’s emotional state\[129\], with an arched unit that strongly signals a caller’s identity and social affiliation\[122,130\] (i.e. resembling Diana monkey’s A call). However, contrarily to Diana monkeys, the second unit combined (i.e. identity-rich arch) by Campbell’s monkeys is never used alone, suggesting a mechanism more akin to suffixation. Playbacks verifying the combinatorial structure of complex calls remain to be performed.

**Intriguingly, male Campbell’s monkeys also use a combinatorial system resembling suffixation in their alarm calls (Figure 4, Table 2)**\[131,132\]. More precisely, in Campbell’s monkeys, males possess two urgent alarm calls, Krak and Hok. While the first generally signals the presence of an urgent ground danger (i.e. classically a leopard in the Taï National park), the latter signals urgent aerial dangers (i.e. classically an eagle)\[133\]. These calls can also be combined with a unique ‘oo’ unit to create Krak-oo and Hok-oo calls (Figure 4). While the ‘oo’ unit is never used alone, its addition to Krak or Hok calls seems to reduce the danger signalled given Krak-oo and Hok-oo calls signal a general disturbance (e.g. a duiker) and an aerial danger of lesser urgency (e.g. a fight in an associated group of red colobus)\[40\].
Figure 4: Sonograms of the calls of adult male Campbell’s monkeys. (a) Krak (urgent ground danger), (b) Krak-oo (non-urgent general disturbance), (c) Hok (urgent aerial danger) and (d) Hok-oo calls (non-urgent aerial danger). Figure reproduced from Coye et al., 2015[92]. On sonograms (b) and (d), the black arrow signals the position of the ‘oo’ unit.

Natural observations were complemented by a playback experiment aiming to verify the combinatorial nature of Krak/Krak-oo calls in this alarm system[94]. The authors analysed the reaction of Diana monkeys (which react to the distinct alarm calls of Campbell’s monkeys with their own referential alarm calls[134]) to broadcasted natural and artificially recombined Krak and Krak-oo calls created by deleting the ‘oo’ part of a Krak-oo or by adding an ‘oo’ part to a Krak call respectively. Subjects’ reaction to Krak and Krak-oo calls reflected their distinct levels of urgency, regardless of their origin (i.e. natural or artificially created). Statistical analysis suggested that, although subtle changes in acoustic structure of the Krak part were perceived by receivers (possibly as a result of caller’s emotional state at the time of calling), the presence or absence of a suffix was the main factor leading subject’s reaction. Thus, in addition to confirming the combinatorial nature of Krak-oo calls (i.e. which result from the linear merging of a Krak call with an ‘oo’ unit), this experiment demonstrated that changes in call structure triggered predictable changes in receiver’s reactions and...
confirmed the biological relevance of the addition of an ‘oo’ unit to decrease the urgency of Krak calls\[^{94}\]. Here again, we can thus conclude that the three criteria proposed to define rudimentary parallels with morphosyntax in animals are met.

The long-term study of Campbell’s monkeys’ communication also revealed another type of combination in the alarm calls of males that are given in long sequences, the organisation of which appears to vary with the context\[^{40}\]. Although this system does not possess the complexity of syntactic structures occurring in language, the type of calls involved as well as the position of some call types in the sequence seem to obey non-random rules and may well be meaningful to receivers\[^{40,133}\]. Krak-oo calls, which signal general alerts, are found in most (but not all) alarm sequences given by males\[^{124}\]. In addition, several regularities have been described, as not only sequence composition varies according to the context but also to the order of calls given and their rhythm of emission. Distinct call types can be added to Krak-oo sequences, such as Krak and Hok calls (that appear at the beginning of a sequence), depending on the type of danger, in particular the type of predator (leopard and eagle) encountered\[^{40}\]. In addition, the urgency of the situation (e.g. visual vs auditory detection of the predator) influences the speed of call delivery of Krak-oo calls in the sequence\[^{135}\] while the speed of call delivery for Hok calls (when an eagle is detected) relates to a male’s willingness to attack the predator\[^{135}\]. Boom calls (i.e. another call type) are always given in pairs and trigger group gathering and movement when produced singly. However, when Booms are followed by other calls, they signal non-predatory events and the calls following them vary with the context. For example, a Krak-oo sequence follows Booms when a large branch or tree is falling down\[^{40,136}\]. The insertion of Hok-oo calls, systematically between Boom and Krak-oo series, to these “tree-falling” sequences (i.e. Booms-Krack-oons) occurs during inter-group encounters with neighbours (i.e. Booms-Hok-oons-Krack-oons)\[^{40}\]. A first playback study has investigated the ‘non-predatory’ modification of a sequences’ message through the addition of Boom calls by comparing receivers’ reaction to natural predator-deterring sequences preceded or not by Boom calls\[^{137}\]. Now, several studies will be required to experimentally verify the other interesting patterns of organisation derived from observational data.
Table 2: Summary of the main characteristics of the combined calls in the species under the focus of this chapter.

<table>
<thead>
<tr>
<th>Species</th>
<th>Structure of combined vocalizations</th>
<th>Meaningful elements?</th>
<th>Link between the meaning of units &amp; combined structures</th>
<th>Mechanism for meaning differentiation</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chestnut-crowned babbler (Pomatostomus ruficeps)</td>
<td>Calls composed of several notes</td>
<td>No</td>
<td>No</td>
<td>Rudimentary form of ‘phonemic contrast’</td>
<td>[94, 155, 157]</td>
</tr>
<tr>
<td>Southern pied babbler (Turdoides bicolor)</td>
<td>Combination of individual calls into call sequence</td>
<td>Yes</td>
<td>Yes</td>
<td>Rudimentary form of ‘morphosyntax’: combination of single calls</td>
<td>[129, 153–154]</td>
</tr>
<tr>
<td>Diana monkey (Cercopithecus diana)</td>
<td>Combination of individual calls into larger calls</td>
<td>Yes</td>
<td>Yes</td>
<td>Rudimentary form of morphosyntax: combination of single calls</td>
<td>[29, 93, 113, 115]</td>
</tr>
<tr>
<td>Campbell’s monkey (Cercopithecus campbelli)</td>
<td>Combination of alarm calls with a call unit never used alone. Combination of calls in sequences</td>
<td>Yes</td>
<td>Yes</td>
<td>Rudimentary form of morphosyntax: combination of calls with a ‘proto-suffix’</td>
<td>[40, 92, 118]</td>
</tr>
</tbody>
</table>

The experimental results presented in this section demonstrate that some animal species combine meaningful structures in non-random ways to create richer signals (i.e. conveying more complex information) or to diversify the messages conveyed with only a limited number of distinct calls (Table 2). In each example described in this section, the three criteria we proposed to classify calls as phonological or morphosyntactic structures were met. Additional testing will clearly be necessary to further our understanding of the relevant mechanisms underlying such combinatoriality and subsequent changes in meaning. Notably, it will be necessary to repli-
cate the recombination experiments on other calls of Campbell’s monkeys (i.e. males Hok/Hok-oo calls) and Diana monkeys (e.g. HA calls) to determine the pervasiveness of sound combinations in these species. Experiments manipulating the order of units in combined calls and call sequences will also be required to fully determine how such changes alter the information extracted by receivers.

4. Formal linguistic analysis of combinatorial systems:

Whilst ethologists have relied on linguistics as a source of inspiration for years, more recently linguists have also begun to systematically compare and contrast animal and human communication systems applying methods from formal linguistics (i.e. posing rules to define formally a ‘lexicon’, a ‘syntax’ and ‘semantics’ for a given system)\(^{[88,92]}\). Among other primates, the vocal systems of Campbell’s and Diana monkeys have been subjected to such analyses in studies by Schlenker and colleagues. The authors reanalysed existing data on these guenons providing complementary investigations to the ethological approach\(^{[88,92,138]}\).

A first study on Campbell’s monkeys, focussed on the possible semantic content of Krak, Hok and their ‘suffixed’ versions. It compared models built using methodologies from the field of formal semantics to shed light on the possible meanings of these calls and on the mechanism by which the addition of an ‘oo’ unit alters the meaning of the call ‘stems’ (i.e. Krak and Hok). Authors specifically focused on the distinct calling patterns of males from two populations of Campbell’s monkeys in Ivory Coast (Taï National Park) and Sierra Leone (Tiwai island)\(^{[88]}\). Crowned-hawk eagles are present in both areas and leopards still being present in Taï, but absent from Tiwai for as long as thirty years\(^{[139]}\). Importantly, while Hok functions to signal the presence of an eagle in both populations, Krak is used primarily to signal the presence of a leopard in Taï but it has the distribution of a general alarm call on Tiwai (i.e. given to a broad range of disturbances including falling trees and eagles)\(^{[40,124,140]}\). To determine which ‘semantic’ explanation best captured the patterns observed, the authors systematically tested the predictions of two models against the data. The first model posits the hypothesis that in both populations Krak and Hok calls have the same “innate” meanings (i.e. Krak: general disturbance; Hok: aerial predator)
and that the addition of an ‘oo’ unit decreases the urgency of the innate meaning of both calls (i.e. Krak-oo: general and less urgent disturbance; Hok-oo less urgent aerial disturbance). Finally, this model hypothesises that, while Krak-oo is derived from the innate meaning of Krak in both areas (i.e. the Krak ‘root’ of Krak-oos kept its original meaning), the ‘lexical entry’ for Krak in Taï has changed to ‘leopard-related disturbance’. The second model proposes an alternative hypothesis to explain the pattern described: the innate meaning of Krak and Hok calls is the same in both populations (i.e. Krak: general disturbance; Hok: aerial predator) and it holds in both unsuffixed and suffixed calls. But while both Krak-oo (i.e. non-urgent danger) and Hok (i.e. aerial predator) are specific, Krak has a rather broad meaning (i.e. general alarm call). The second model thus proposes that the competition between more specific calls and Krak calls may lead to the strengthening of the meaning of Krak in a mechanism akin to scalar implicatures[88,92] (see Table 1 for a definition). Specifically, when a male gives Krak calls, a receiver might infer that there is a non-weak and non-aerial disturbance as the call given is not Krak-oo nor Hok. Hence, the meaning of Krak calls can be strengthened from ‘general urgent disturbance’ into ‘dangerous non-aerial predator’. In Taï the presence of leopards led to the strengthening of the meaning of Krak calls as ‘dangerous non-aerial predators’ but not in Tiwai where the absence of ground predators prevented it. From this, the authors concluded that the second model was more parsimonious and more likely to describe the associated ‘meanings’ of calls in the call system of Campbell’s monkeys than the first one[88].

A second study led by Schlenker and collaborators proposed to analyse the communication of female Diana monkeys, using both a statistical analysis of transition probabilities between units and call types and a formal semantic analysis of utterances based on their context of emission[93]. Again, the authors proposed two alternative competing hypotheses to describe the system observed. The first hypothesis proposed that combined calls consisted of two simple calls given in close succession (i.e. maximized adjacency hypothesis). Contrarily, the second hypothesis proposed that combined calls (i.e. HA, LA and RA calls) resulted from the combination of two units that were subsequently used as one call (i.e. combined calls hypothesis). To determine which hypothesis was the more likely, the authors developed two corresponding models (e.g. putative ‘rules’ of call use...
describing the observed patterns) and compared them. This work showed that treating ‘combined’ calls as sequences of simple calls given in close succession failed to account for their distribution in sequences. The most parsimonious model was obtained under the ‘combined call hypothesis’ (i.e. ‘combined calls result from the combination of vocal units and are used as one call) as the alternative hypothesis (i.e. maximized adjacency hypothesis) would need to be supplemented by phonological complexity in order to account for the data with respect to maximal sequence length and call repetition[93].

Other recent articles by the same authors offer analyses of the calling systems of additional species using similar methods (e.g. black-fronted titi monkeys and putty-nosed monkeys)[92,114]. The results obtained converged with field observations and these articles are key in not only generating testable hypotheses but also confirming the relevance of using linguistic methodologies to analyse combinatorial systems in non-human animals. We argue these studies bring key additional support to our findings while adopting different, yet complimentary methodological approaches. Indeed, although previous studies had also described non-random patterns of transitions between elements comprising vocal sequences produced by animals (e.g. marine mammals[141], bats[57], birds[48,53,142]), they failed to take into account the meaning and relevance to receivers of sequence organisation and composition. For instance, Kershenbaum and collaborators[80] analysed the vocal sequences produced by animals from several taxa (i.e. killer and pilot whales, rock hyraxes, Bengalese finches, Carolina chickadee, free-tailed bats and orangutans) using various transition models of increasing complexity to determine which one matched best the transition between elements in the sequences recorded. Such studies are very informative regarding the possible evolution of sequence complexity in animals and may participate in bridging the gap between human language and animal communication[80]. However in language, combination is relevant only because it is meaningful[84]. The work reviewed in this chapter highlights the need to include a more systematic analysis of animal sequence structure, meaning and composition and its relevance to receivers, in particular, when one aims to undertake a comparative approach with human language. We believe that the current progress on animal combinatorial abilities, together with future developments in complementary methodological approaches and appropria-
ate terminology, will pave the way to a more comprehensive understanding of the evolution of sound combination in animals. In the final section we will discuss the evolutionary insights such comparative data can provide on the drivers of the emergence of combinatorial abilities.

5. Evolutionary relevant insights from animal combinatorial systems?

The topic of language origins is frequently accompanied by heated debates over the analogous (i.e. convergent evolution) or homologous (i.e. inherited from a common ancestor) nature of some parallel features of language described in non-human primates, including combinatorial abilities (e.g. [3,6,143–148]). However, we forward that this is not the most pressing question, because we can learn a lot from the study of animal communication regardless of its shared or distinct evolutionary history with language. Indeed, if language is a unique communication system, it is also clearly the product of a gradual evolutionary process and, in this regard, it does not differ from other animal communication systems. Thus, in our opinion, a more important question to tackle would be – what pressures drove the evolution of combinatorial abilities?

Social complexity

Social life is often viewed as a major driver of communicative complexity and this hypothesis has been supported by empirical studies highlighting a positive relationship between indexes of social complexity and signal diversity for both social and alarm calls[47,149–152]. The description and testing of combinatorial systems in animals suggests that sound combination may allow the diversification of a species’ repertoire using a limited number of signals. Interestingly, two studies comparing closely related species of non-human primates and herpestidaes reported a correlation between the complexity of a species’ social life and the presence, diversity and frequency of use of combinatorial structures[153,154]. In line with this, each species in which meaningful sound combinations were described has also been reported to reside in a complex and strongly bonded social group[30,155–158]. These observations support the idea that increased needs for complex communication
resulting from social complexity might have played an important role in the emergence of combinatorial capacities in animals.

Phonatory limits

Another possible factor leading to the emergence of combinatorial capacities results from the phonatory limits that some species face, notably in non-human primates. Work from computational modelling provides relevant additional insight here. Nowak and colleagues\textsuperscript{[159]} modelled scenarios for the emergence and propagation of certain language features in a population, such as arbitrary signals, sound combinations and grammatical rules. Nowak et al. proposed that combinatorics would emerge after a communication system reaches a threshold number of signals above which the addition of new signals (because they would be likely to resemble existing ones) ultimately increases the error risk due to mis-comprehension. In this case, the combination of sounds would allow a continued increase in a language’s fitness (through addition of new signals) without increasing the risk of ambiguous information transfer. This rationale relies on the hypothesis that a species is capable of increasing its repertoire via the acoustic diversification of signals in the first place. We propose that it is also valid in species with limited capacities of vocal production but that in this case, the first limit to signal diversification might be the species’ lack of vocal plasticity rather than the breadth of the existing repertoire. This hypothesis is supported by the fact that all the species in which sound combination has been shown to play a meaning-differentiating role display limited capacities for vocal production\textsuperscript{[118,121,130,156,160]}. Further studies investigating the presence of meaningful call combinations in species characterised by various levels of social complexity as well as distinct capacities for vocal learning are key to testing these hypotheses with more extensive empirical data.

Habitat and constraints on communication

Finally, habitat has often been proposed as a factor influencing species’ communication. In particular, it has been proposed that dense habitats, which impose constraints on sound propagation and visual access to others, may favour the emergence of discrete communication signals (i.e. as opposed
to communication systems of graded signals, whose acoustic structures form a continuum without distinct boundaries between call types\(^{[161-164]}\). In dense habitats, discrete signals would allow more robust communication and prevent ambiguities resulting from poor visual access to others\(^{[65,163,165]}\). Sound combination may therefore benefit animals through the production of more efficient communication signals. This is, for instance, the case in female Diana monkeys, whose combined calls linearly convey information about a caller’s emotional state and identity. Here, females concatenate signals sequentially which might have already evolved to ensure maximal efficiency of information transfer (e.g. calls with more salient identity cues or with an improved acoustic adaptation to propagation constraints). This organisation allows information to be temporally segregated creating richer signals without increasing ambiguity due to the accumulation of information. Interestingly, a combinatorial system resembling that of Diana monkeys has been described in the graded contact calls of desert-living banded mongoose. More precisely, banded mongooses use a contact call composed of two segments, given in three distinct contexts: when the caller is digging, searching and moving\(^{[166]}\). The first segment relates strongly to a caller’s identity and remains identical in the three contexts. A playback experiment confirmed that between-caller variations in the identity segment were relevant to receivers\(^{[167]}\). The second segment has a graded structure and varies with caller’s activity: when the caller is digging the segment is absent (or very short), its duration increases when the caller is searching, and reaches its maximal duration (together with more pronounced harmonics) when the caller is moving. Thus, here again, the use of combinatoriality seems to increase the information content of calls while maintaining a low level of ambiguity.

Importantly, banded mongoose live in an open habitat but lack visual access to conspecifics because their foraging strategy constrains them to face the ground most of the time\(^{[168,169]}\). Although literature traditionally pitted species with graded and discrete repertoires against each other, multiple concerns with the relevance of this dichotomy have been raised, notably because of evidence for subtle gradation in the communication of ‘discrete’ species as well as evidence for categorical perception of graded signals by receivers\(^{[162,170-173]}\). Taken together, these observations suggest that, more than habitat-based propagation constraints, the lack of visual access to
others may be important in influencing the emergence of short-distance combined social call structures that convey complementary information about a caller’s identity, activity and localisation. Finally, this hypothesis can be aligned with the theoretical work of Nowak and collaborators. Indeed, if combination has emerged to limit the risk of ambiguous communication, the inability to disambiguate the context of calling or caller’s identity using visual cues (e.g. due to habitat constraints or foraging strategy) is a possible additional factor triggering its emergence. To investigate the potential relative impacts of habitat density and actual visual access to others on the development of combinatorial capacities, we would need to extend the comparison to other species whose visual access to others contradicts the predictions that could be proposed by simply looking at their habitat density (e.g. other species than banded mongoose with poor visual access in spite of an open habitat).

The hypotheses proposed above shed further light on the factors involved in the evolution of language and other communication systems and have largely resulted from data provided by only a few species where receivers have been experimentally documented to process and use combinatorial structures. Various additional examples also exist that have not yet been completely described and are likely to fit the definitions we have previously used (i.e. in section 1). In particular, several systems in which apparently meaningful calls are combined into larger structures whose context of emission reflects that of its parts have been described in wedged-capped capuchins, cotton-top tamarins, female Campbell’s monkeys, red-bellied titi monkeys, black-fronted titi monkeys, red-capped mangabeys, bonobos and chimpanzees as well as in non-primate species such as Japanese great tits, banded mongooses and meerkats. The diversity of species in which meaningful sound combinations have been documented (e.g. birds and mammals, including primates and herpestidae) suggests that it may be an evolutionary solution to deal with communicative demands and the comparative study of these species is central to test any hypotheses regarding the potential drivers promoting the emergence of combinatoriality.
6. Conclusion: Towards a more comprehensive approach of combinatorial abilities

The study of animal combinatorial abilities appears to be a promising research area, with a number of avenues open to exploration. In this chapter, we restrict discussion to examples exhibiting more or less marked parallels with language. However, a large number of other combinatorial systems in animals remain to be investigated, among which some have been described but remain only partially understood (e.g. gorillas, putty-nosed monkeys, gibbons, rock hyraxes, mustached bats[54,55,66,92,107–109,114]). In addition, some animal combinatorial systems may differ strikingly from language in their organisation and underlying mechanisms facilitating information transfer. For instance, in some systems the diversity of units (e.g. in some song-birds[104]) or the proportion of various units (e.g. in bonobos[176,179]) seem to play a role in meaning generation.

Joint efforts from linguists, psychologists and ethologists is clearly necessary to provide a unified and more relevant framework. One possible first step would be to develop a terminology suitable to describe the vast diversity of combinatorial systems found through the animal kingdom. Indeed, whilst some rare examples developed in this chapter can be captured by pre-set definitions, it seems clear that a number of sound combination systems will not. However, even (if not especially) in those cases, the use of strict definitions is essential. This is important firstly to provide a clearer view of the diversity and complexity of combinatorial organisations in the animal kingdom. Secondly, and perhaps more pertinently, because the study of varied examples relating to potentially meaningful, non-random and contextually flexible combination patterns may be an important step to further understand the biological relevance of vocal combination in animals and its evolution(s).

The rationale adopted to build the two definitions we propose for parallels with phonology and morphosyntax (section 1) could be generalised to develop a more suitable terminology for alternative systems of sound combination. Notably, the definitions we proposed involve three components: (1) whether the vocal units combined possess an intrinsic meaning, (2) whether (and how) the meaning of their combination reflects the meaning of individual elements (if they possess one) and (3) which rules (if any)
best describe the mechanism for meaning-differentiation between combined utterances with distinct functions. In addition, the third component of the definitions would also allow us to capture systems that differ strongly from language, such as those relying on the proportion of one call type, or on the diversity of units involved. As such, this three-component structure could be further expanded to characterise the variety of animal combination systems described while offering a systematic basis for interspecific comparison.

In addition, it may be useful that authors specialised in various taxa (e.g. ornithologists, primatologists, marine biologists) and disciplines (e.g. ethologists, philosophers, linguists) readdress questions pertaining to the nature of meaning as this central question may be approached from various directions (e.g. can regional ‘dialectal’ variations of a song also be considered as changes in meaning?). In any case, future studies focussing on diverse combinatorial systems, including systems that differ strongly from language, are likely to be fruitful. Indeed, understanding the organisation and evolution of systems that differ strongly from ours will, if anything, bring insights into the various evolutionary paths that the human lineage did not follow and may be a relevant source of information to identify important turning points in our “history”.

This chapter focussed mostly on studies that relied on the simultaneous use of natural observations of calling contexts and experiments. These two complementary approaches are essential to investigating the combinatorial structure (i.e. transferability of units in combination) of complex utterances and the relevance to receivers of changes in meaning as a result of changes in the combinatorial structure. We argue that this is a key first question to tackle in order to provide a comprehensive description of animal communication systems. However, beyond the combination of phonemes into words and words into sentences, language relies on a set of rules that allow interlocutors to produce and understand completely novel utterances\[84,87,159\]. The cognitive mechanisms underlying our processing of rules and our capacity to generalise them are at the very base of language generativity\[81,87\].

Now that some animal communication systems have been characterised in terms of their basic structure and meaning-differentiating mechanisms, the next step should be dedicated to clarifying the cognitive mechanisms underlying their combinatoriality. In particular, it will be important to determine whether animals perceive combined utterances as a mosaic of
elements whose message can be inferred from the elements and their relationships (i.e. as compound signals) or as unique elements whose combinatorial nature is only structural. For instance, can Campbell’s monkeys learn the meaning of Krak, Krak-oo, Hok and Hok-oo calls independently or do they learn the meaning of Krak, Hok and the alteration of meaning associated with the presence of an oo unit? And would they be capable to generalise that “rule”? Previous studies have shown that some non-human primates possess particularly sophisticated social cognition skills involving a hierarchically structured representational knowledge of social relationships, governed by rules and involving causal inference – a likely result of their complex social life. In addition, studies based on experimental tasks suggest that some animals possess, to some extent, capacities to parse combinatorial and sequential artificial structures or rules. For instance non-human primates of several species have been shown to learn sequential lists of items, and to compute probabilities of occurrence, and dependencies between syllables or letters. Cotton-top tamarins have the capacity to acquire general ‘rules’ of structuring such as simple pFSA grammar and starlings, as well as language-trained animals (apes, dolphins and parrots), even acquired more complex rules such as recursive structures or generative languages. In line with this, a particularly interesting perspective would be to compare the experimental performances in laboratory settings (e.g. assessing a species’ capacity to acquire and handle various artificial grammars) of species using combinatorial structures to various extents and determine whether their performances at parsing artificially constructed structures correlates with their natural tendency to use sound combination during communication. Finally, studies clarifying the ontogeny and acquisition mechanisms of vocal repertoires involving combinatorial structures would be important to complete our knowledge, especially in species that seem to rely on meaningless notes, as the arbitrariness of combinations should be based on a mandatory learning phase.

To conclude, we can say that there now exists a growing array of species that rely on sound combinations and as such provides an intriguing starting point for investigations into the evolution and emergence of these abilities. When reviewing combinatorial structures across species, however, it becomes clear that questions associated with meaning and information conveyed by animal signals are central and should be taken into account in
the development of an appropriate terminology to describe sound combinations in animals. Finally we actively encourage interdisciplinary research uniting linguists, ethologists, psychologists and anthropologists, to build a unified framework and to further explore the links between human language and animal communication.

References


[93] Veselinovic, D., Candiotti, A. and Lemasson, A. (Under revision). Female Diana monkeus (Cercopithecus diana) have complex calls.


