JAMES R. HURFORD

the Origins
of Grammar

Language in the Light of Evolution

THE ORIGINS OF GRAMMAR

James R. Hurford

Contents

Detailed Contents

Introduction to Part I: Twin Evolutionary Platforms—Animal Song and Human Symbols

Before complex expressions with symbolic meaning could get off the ground, there had to be some facility for producing the complex expressions themselves, even if these were not yet semantically interpreted. Birds sing in combinations of notes, but the individual notes don't mean anything. A very complex series of notes, such as a nightingale's, only conveys a message of sexual attractiveness or a threat to rival male birds. So birdsong has syntax, but no compositional semantics. It is the same with complex whale songs. Despite this major difference from human language, we can learn some good lessons from closer study of birds' and whales' songs. They show a control of phrasal structure, often quite complex. The songs also suggest that quantitative constraints on the length and phrasal complexity of songs cannot be naturally separated from their structure. This foreshadows a conclusion about how the human language faculty evolved as a composite of permanent mental structure and inherent limits on its use in real-time performance.

Also before complex expressions with symbolic meaning could get off the ground, there had to be some facility for learning and using simple symbols, arbitrary pairings of form and meaning. I argue, contrary to views often expressed, for some continuity between ape cries and human vocalized words. There was a transition from innate involuntary vocalizations to learned voluntary ones. This was a biological development to greater behavioural plasticity in response to a changing environment. The biological change in the make-up of individuals was accompanied by the development in social groups of shared conventions relating signals to their meanings. One pathway by which this growth of shared social norms happened capitalized on sound symbolism and synaesthesia. Later, initially iconic form-meaning mappings became stylized to arbitrary conventions by processes which it is possible to investigate with

modern experiments. With the growth of a learned lexicon, the meanings denoted by the developed symbols were sharpened, and previously unthinkable thoughts became accessible.

Thus, the two chapters in this part survey the situation before any semblance of modern grammar was present, exploring the possibility of non-human antecedents for control of complex syntax and of unitary symbols, proto-words. These two chapters deal respectively with pre-human semantically uninterpreted syntax and early human pre-syntactic use of symbols.

Animal Syntax? Implications for Language as Behaviour

The chapter heading poses a question, and I will answer it mostly negatively. Some wild communicative behaviour is reasonably described as having syntactic organization. But only some wild animal syntax provides a possible evolutionary basis for complex human syntax, and then only by analogy rather than homology. That is, we can find some hierarchical phrase-like syntactic organization in species distantly related to humans (e.g. birds), but not in our closer relatives (e.g. apes). The chapter is not, however, a wild goose chase. It serves (I hope) a positive end by clarifying the object of our search. Non-linguists find linguists' discourse about syntax quite impenetrable, and the chapter tries to explain some theoretical points that cannot be ignored when considering any evolutionary story of the origins of human syntax. Using a survey of animal syntactic abilities as a vehicle, it will introduce and discuss some basic analytic tools applicable to both human and non-human capacities. These include such topics as semantic compositionality (as opposed to mere combinatorial structure), the competence/performance distinction, the hierarchical structuring of behaviour and the relation of overt behaviour to neural mechanisms. A special tool originating in linguistics, Formal Language Theory (FLT), will come in for particular scrutiny. This body of theory is one of the most disciplined and mathematical areas of syntactic theorizing. FLT is on firmer formal ground than descriptive syntactic theories, giving an accumulation of solid results which will stand the test of time. These are formal, not empirical, results, in effect mathematical proofs. Some may question the empirical applicability of Formal Language Theory to human language. It does

give us a precise yardstick by which to compare the syntax of animal songs and human language. It will become clear to what extent any syntactic ability at all can be attributed to songbirds and some whales.

The first section below will, after a survey of candidates among animals, reinforce the point that animals in the wild indeed do not have any significant semantically interpreted syntax. The second and third sections will examine how much, or how little, the non-semantic syntactic abilities of animals can tell us. I will illustrate with facts about the songs of birds and whales,[1] fascinating in themselves. To make the proper comparison between these songs and human syntax, it is necessary to introduce some key concepts underlying the analysis of syntax in humans. Discussing these key concepts in a context away from the common presuppositions of linguistics allows us to reconsider their appropriateness to human language, and to suggest some re-orientation of them. So some of this chapter is theoretical and terminological ground-clearing, spiced up with interesting data from animals.

Syntax, at its most basic, is putting things together. Of course, 'putting things together' is a metaphor, but a significantly insightful one. Syntactic spoken language is not *literally* a putting together in the sense in which bricks are put together to make a wall, or fruit and sugar are put together to make jam. Speech is serial behaviour, but serial behaviours differ in the complexity of control of putting things together. Breathing is basic and can be described as putting certain routines of muscular contraction together in a prolonged sequence. Walking is a bit more complex, and the way strides are put together involves more volition and sensory feedback from the surroundings. All animals put actions together in serial behaviour. Indeed that is a defining characteristic of animals, who seem to have some 'anima',[2] dictating the order of their movements. In all animals many action sequences are instinctive, somehow programmed into the genome, without any shaping by the environment in the individual's lifetime. Quite a lot of animals also learn motor sequences, used for practical purposes of survival. With most sequences of actions carried out by animals, the environment provides constant feedback about the state reached and prompts the animal for its next step. For example a gorilla picks, packages, and eats nettles in a somewhat complex way (Byrne and Byrne 1991; Byrne 1995). All through this systematic behaviour the animal is getting feedback in

---

[1] Birdsong and whale songs are typical enough to make my general points, and space prohibits discussion of gibbons and other singing species.

[2] The etymology of *animal* reflects a narrowing of Aristotle's concept of anima or $\psi \upsilon \chi \eta$, which he saw as the essence of all living things, including plants. Aristotle's *anima* is often translated as *soul*, but he did not regard it as a non-physical substance.

the form of the current state of the nettles, whether they are (1) still growing undisturbed in the earth, (2) with stalk held tightly in the gorilla's right hand, (3) stripped of their leaves, held in the left hand, or (4) leaves folded into a package and ready to pop into the mouth. There are millions of such examples, wherever an animal is dealing systematically with its environment. Much rarer are learned routines of serial behaviour not scaffolded throughout the sequence by feedback from the environment. During the singing of a nightingale's song, there are no external landmarks guiding it to its next note. All the landmarks are within, held in the animal's memorized plan of the whole complex routine. Most, and maybe all, such complex 'unguided' routines are communicative, giving information to conspecifics. Although all complex serial behaviour has a kind of 'syntax' or 'grammar', I will restrict the term 'syntax' in the rest of this work to complex, unguided communicative routines. No doubt, a specialized facility for syntax in this narrow sense evolved out of a facility for serial behaviour more generally.

A fancier term for 'putting things together' is **combinatorial**. Music has combinatorial syntax, because it involves putting notes together in strictly defined ways. Different musical traditions are roughly like different languages, in the sense that they define different rules for combining their elementary constituents—notes for music, and words for language. Dances, the tango, the waltz, the Scottish country dance Strip-the-Willow, each have their own syntax: ways of putting the elementary moves together into an approved sequence. The syntax of such human activities tends to be normative, hence the use of 'approved' here. But behaviour can be syntactically organized without the influence of any norms made explicit in the social group, as we will see in this chapter when discussing the structured songs of birds and whales. (This does not, of course, mean that syntactic organization cannot be influenced by the behaviour of others, through learning.)

Peter Marler (1998) distinguishes between **phonological syntax** and **lexical syntax**. In its broad sense of putting things together, syntax applies to phonology. Phonology puts phonemes together to make structured syllables. Each language has its own phonological syntax, or sound pattern. The units put together in phonology don't mean anything. The English /p/ phoneme, on its own, carries no meaning. Nor does any other phoneme. And it follows that the syllables put together out of phonemes can't mean anything that is any function of the meanings of the phonemes, because they have no meanings. *Cat* does not mean what it means because of any meanings inherent in its three phonemes /k/, /a/, and /t/. Phonological syntax is the systematic putting of meaningless things together into larger units. Birdsong, whale song and gibbon song all exhibit phonological syntax, and I will discuss two of these in the third section

below. It is possible that some phonological syntactic ability developed in our species independent of meaning, which is why I devote space to these complex non-human songs.

Lexical syntax, or lexicoding, as Marler calls it, is the kind of putting things together where the elements mean something, and the whole assembly means something which is a reflection of the meanings of the parts. This is **compositionality**. Complex meanings are expressed by putting together smaller meaningful units. As Marler summarizes it, 'Natural lexicoding appears to be a purely human phenomenon. The only animals that do anything remotely similar have been tutored by humans' (Marler 1998, p. 11). In order to be clear that this is indeed the case, the first section of this chapter will look at some challenges to Marler's assertion that have surfaced since he wrote. With some tiny reservations, Marler's assertion stands. (Marler mentioned animals tutored by humans. We will come to them in a later chapter.)

I will weave into the second and third sections of this chapter an introduction to Formal Language Theory. On its own, such an introduction might seem both dry and unmotivated. But the Formal Language Theory approach to repertoires of complex meaningless songs[3] turns out to give a useful way of classifying the overt characteristics of song repertoires. The approach also draws out some differences and similarities between these animal songs and human languages that push us to question some of our common assumptions about human language.

1.1 Wild animals have no semantically compositional syntax

This section describes some non-starters as candidates for evolutionary analogues or homologues of human semantically compositional syntax. In brief, no close analogues or homologues are to be found in wild animal communication systems. But surveying cases that show, or might appear to show, some compositionality can clarify what exactly we are looking for.

1.1.1 Bees and ants evolve simple innate compositional systems

Insects are only very distantly related to humans. But even some insects put elements together in a semantically composed signal. Parts of the signal are

---

[3] The songs are not wholly meaningless, of course, or the animals would not sing them. I mean that the songs do not convey referential meanings by combining the meanings of their elementary parts. One way of putting this is to say that the songs have pragmatic, but not semantic, significance.

combined to express a message which is a function of the meanings of the parts. These communication systems are (1) extremely simple, comprising only two meaningful elements, (2) extremely limited in the domain to which they apply—location of food or a good hive site, and (3) innate. These simple systems are adaptive, enhancing the survival chances of the animals. How far can nature go in engineering a genetically fixed semantically compositional system? The insect systems seem to be the limit. There are no more complex innate systems in nature. Without **learning**, a semantically compositional system cannot evolve beyond the narrowest limits we see in a few insects. So we have an important conclusion here already. Highly complex semantically compositional systems need to be learned. Now I'll briefly survey what we know about the unlearned insect systems. In their way, they are impressive, but impressive in a totally different way from the wonders of human language, which has evidently taken a different evolutionary course.

The honeybee, *Apis mellifera*, provides a well known example of animal communication. Surprisingly, for an animal genetically so far distant from us, bees use a simple, but arguably semantically compositional, system.[4] They signal the location of food relative to the hive by a vector with two components, a distance component and a direction component. Distance is signalled in analogue fashion by the duration of the 'waggle' dance—the longer the dance, the farther away is the food. And direction is signalled by the angle to the vertical of the waggle dance: this corresponds to the angle relative to the sun's position in which the food lies. Thus a fairly precise location is described in terms of two components and each component is signalled by a separate aspect of the overall signal. The receiving bees may possibly be said in some sense to 'compose' the location from its elements, direction and distance.

The question arises, however, whether this description is our own anthropomorphic account of their behaviour. The bee observing the dance no doubt registers somehow the two components of the signal, and responds systematically to both, by flying a certain distance in a certain direction. And then, of course, it gets to roughly the right place. But it does not follow that the bee has in its brain any representation of the place it is going to before it actually gets there. If I give you precise latitude and longitude specifications of a place, you can consult a map and know what place I am talking about.

---

[4] The summary of bee communication given here is basic and omits many fascinating details of the variety between species, and the scope of their responses to different environmental conditions. For a highly informative and readable account, see Lindauer (1961). Other significant works are von Frisch (1923a, 1923b, 1967, 1974); Riley et al. (2005).

Or, more familiarly, if I say 'You know, the pub two hundred yards south of here', you will identify what I mean, and we can talk about it, without either of us necessarily flying off there. There is some evidence that bees can do this as well.[5] Gould (1986) showed that bees could find their way directly to a feeder station when released at a novel site away from the hive, and construed this as evidence that the bees were computing the new route by reference to a cognitive map. The term 'cognitive map' requires some unpacking. For Gould, it was consistent with 'landmark map', and his bees could be taken to be finding their way by reference to familiar landmarks. It is accepted that bees use landmarks in their navigation. On the basis of more carefully controlled experiments, Dyer (1991) argues, however, that the bees acquire 'route-based memories' but not cognitive maps. Dyer released his experimental bees in a site, a quarry, from where they could not see landmarks visible from the hive. On release from the quarry, they tended to fly off on a compass bearing close to that on which they would have flown from the hive, that is in a wrong direction. Dyer concludes that his 'results suggest that honey bees do not have the "mental maps" posited by Gould (1986), or any other mechanism to compute novel short cuts between familiar sites that are not in view of each other' (p. 245). Nevertheless, it is clear that signalling bees do base their performances on a computation of several factors. 'Fully experienced bees orient their dances on cloudy days by drawing upon an accurate memory of the sun's entire course relative to familiar features of the terrain' (Dyer and Dickinson 1994, p. 4471). More recently, and using hi-tech radar equipment, Menzel et al. (2005) were able to track the entire flights of bees. They concluded:

Several operations must be at the disposal of the animal: (i) associations of headings and distance measures toward the hive with a large number of landmarks all around the hive that are recognized from different directions; (ii) shift of motivation (flight to hive or feeder); (iii) reference to the outbound vector components of the route from hive to feeder; and (iv) addition and subtraction of the heading and distance components for at least two conditions, those that would lead directly back to the hive and those that lead from the hive to the feeder. It is difficult to imagine that these operations can be done without reference to vectors that relate locations to each other and, thus, make up a map.   (Menzel et al. 2005, p. 3045)

---

[5] Reznikova (2007) cites Dyer (1991): 'In the experiments of Dyer (1991), bees left the hive when the returning scout indicated that the food was beside a lake. However they did not leave the hive when they were informed that food was near the middle of the lake. Thus, honey bees appear to interpret the meaning of the dance—possibly by identifying the potential location of food, and then decide whether it is worth making the journey'. Unfortunately, this passage is not actually to be found in the cited article by Dyer, so the lake story must have come from somewhere else.

All these navigational experiments involve observing the flights taken by bees, and are not directly about what is signalled in the honeybee waggle dance. Thus the compositional nature of the dance signal itself is not directly investigated. But the evidence for quite rich navigational abilities makes it seem unlikely that the response to the dance by bees already familiar with the landscape is entirely robot-like, following two instructions simultaneously, flying for a certain distance in a certain direction. On the other hand, inexperienced bees, who have not become familiar with the local topology, can do nothing but follow the two components of the message conveyed by the waggle dance, insofar as the landscape allows them. On the evidence, the processing of the signal by experienced bees seems likely to be somewhat analogous to what happens when a human understands a phrase such as *two hundred yards south-west of here*, even when a straight-line walk to that location is not possible, because of the street layout. The human, if he already knows the locality, can make a mental journey by putting the two elements of meaning together, and perhaps never take the actual physical journey. The bee is almost as clever (in this very limited domain), but not quite. Von Frisch (1967) reviews experiments in which bees had to go around an obstacle such as a large ridge to get to their food, thus making a two-leg trip with an angle in it. On returning, their dance signalled the real compass direction of the food (which was not a direction of either leg of their flight) and the actual distance flown, around the obstacle. This shows impressive ability to integrate two flown angles, and the distances flown at those angles, into a single angle. But the location signalled was technically false, being further away from the hive (in a straight line) than the actual food source. One can see this as a simple evolutionary solution to the problem of signalling location over a restricted communication channel. The bee receiving the signal goes in the direction signalled, as best she can, for the distance signalled. Signalling a complex two-leg journey would be more of a challenge.[6] This is a case where the bees' private computational capacity, allowing them to do complex path integration, outstrips what they can communicate publicly. The given message is a simple synopsis of their more complex experience.

In later experiments, it was found that bees could be tricked into believing that they had flown longer distances than they actually had. Srinivasan et al. (2000) trained bees to fly, on either their outward or their inward flight, through a tube painted with many closely-packed colours. After flying through

---

[6] Even humans asking for directions in a strange town find it hard to remember oral instructions with more than about three legs.

such a tube, bees signalled distances much longer than the actual distances flown. Following this up, De Marco and Menzel (2005) made bees take a 90° detour through a painted tube to get to their food. Once these bees had arrived at the food source they took a diagonal shortcut back to the hive, presumably relying on landmarks. The experimenters watched the signalling behaviour of the returning bees. They found that the bees signalled the direction of the shortcut route to the food, figured out from their return journey, but the perceived long distance flown through the tube on their outward journey. On this evidence, bees can separate out two factors of their experience, the length (sometimes misperceived) of their outward flight, and the direction of their return flight. And they code these separate aspects of their experience into the waggle dance. This is compositional coding, but of course in an extremely limited domain, and is not learned behaviour.

Bees have an accurate sense of time and anticipate the movement of the sun across the sky as the day proceeds (Lindauer 1961; Dyer and Dickinson 1996; Dickinson and Dyer 1996). Bees who have received a message in the morning about the direction of food can be kept in the hive for a few hours, and when they are released later in the afternoon they compensate for the movement of the sun during the time they were cooped up. For example, if the waggle dance at noon signals due south, and the bees are released immediately, they fly off directly towards the sun;[7] but if after receiving that same signal at noon they are not released until 3.00 p.m., they don't fly directly towards the sun, but about 45° to the left of it. Thus the code is interpreted with some contextual 'pragmatic' input, namely the time elapsed since reception of the message. This is a lesson that simply having a code is not enough for practical communication. The information conveyed in a code is supplemented, even in such a simple system as honeybee dancing, by contextual information.[8] (Fascinatingly, Lindauer also reports experiments in which bees who had been accustomed to the movement of the sun in one global hemisphere (i.e. left-to-right in the northern and right-to-left in the southern) were shifted overnight to the other hemisphere. The originally transported bees did not adapt, but their descendants, after 43 days, did make the correct new adjustment, interpreting the direction aspect of the dance in the new appropriate way. See Lindauer (1961, pp. 116–26) and Kalmus (1956).

---

[7] in the northern hemisphere.
[8] Humans who leave a message on a door saying 'Back in an hour' seem oblivious of the importance to the receiver of such contextual information about when the message was written.

Some species of ants, socially organized like honeybees, also show evidence of semantically compositional signalling (Reznikova and Ryabko 1986; Reznikova 2007). It seems that ants communicate by contact with each other with their antennae. In controlled experiments, scout ants reported the location of food to teams of forager ants, who reliably followed the directions given by the scout through a series of T-junctions in a maze. There was individual variation: not all ants were very good at transmitting such information. In the case of the effective ant signallers, the evidence for compositional signalling is indirect. That is, the research has not 'decoded' the signals given by the ants into their component meaningful parts, as von Frisch did with the honeybees. Rather, the experimenters carefully controlled the **amount** of information, measured in bits as defined by Information Theory (Shannon and Weaver 1963). Each turn taken at a T-junction in the maze counted as one bit of information. In complex cases, it was possible for the food to be located at a point six turns into the maze from the entrance. Not surprisingly, a correlation was found between the complexity of the message in bits (i.e. number of turns in the maze), and the time taken by ants to convey it.[9] More significantly, where there were regular patterns in the message to be conveyed, such as a succession of turns in the same direction (e.g. Right-Right-Right-Right-Right, or Left-Left-Left-Left-Left), the time taken to convey such messages was shorter than in the case of less regularly structured messages, such as Right-Left-Left-Right-Left. This, as the authors point out, is evidence of data compression.

One way in which data compression can be achieved is with some kind of compositional coding, where one element of the code systematically denotes the way in which the data is to be compressed. For example, we can imagine (although we don't know exactly) that a message such as Right-Right-Right-Right-Right was compressed by the signalling ant into the equivalent of 'All-Right' or 'Only-Right'. A less regularly structured message could not be compressed in this way, assuming obvious intuitions about what is 'regular structuring'. We must remember that the natural environment of ants in the wild is unlikely to present them with routes so neatly defined as a series of T-junctions in a lab maze. But the correlation between regularity in the message, measured in information bits, and duration of the signalling episode needs some explanation. The data raise the possibility that these ants have a semantically compositional (albeit very simple) code.

---

[9] Three species of ant showed such behaviour in these experiments, *Formica pratensis, F. sanguinea* and *F. rufa*. (There are over 11,000 species of ant.)

However, the data also support another interpretation, which is that the ant signals are entirely holophrastic. That is, the ants may just have the equivalent of a lexicon, a lookup table in which each separate mapping from a meaning to a form is stored, with no general rules for constructing the signals from meaningful subparts. (This presupposes that the number of conveyable messages is finite, and presumably small.) The observed correlation between short signals and repetitively structured messages (e.g. Right-Right-Right-Right-Right) may come about through some tendency to associate such meanings with short signals, holophrastically. Information Theory tells us that more efficient communication is achieved if the most frequent messages are coded as the shortest signals. This fact is illustrated by several well-known phenomena, including Zipf's Law inversely correlating word frequency with word length, and Morse Code, in which the commonest English letter, E, is signalled by the shortest possible dot-dash sequence, namely a single dot. The messages to be conveyed by the ants in these experiments did not vary significantly in frequency, so Information Theoretic efficiency of coding is probably not a driving force here. But there might be something salient about such repetitively structured meanings to ant brains which makes them assign them shorter signals. The fact of signal compression in itself does not necessarily imply compositionality in the code. Morse Code, for example, is not semantically compositional in its mappings from dots and dashes to letters: the letters of the alphabet are not treated as bundles of features, with each feature signalled by something in the code. Incidentally, humans find it easier to remember sequences of digits, such as telephone numbers, if they contain repetitions; 666 1000 is much easier to remember than 657 3925.

These several species of bees and ants may have converged in their evolution on a common principle for efficient information transmission, applying it in very limited ways, and in very narrow domains. These insect encoding and decoding systems are probably wholly innate. (This is not to deny that bees, at least, can learn to apply the messages of the signals appropriately in the context of their local landscape.) We are interested in syntactic systems with a much more significant element of learning and with much wider expressive range.

1.1.2 *Combining territorial and sexual messages*

Birds' songs typically express either a courtship or a territorial message—'Welcome, ladies', or 'Keep away, gents'. Can these two messages be combined into a single composite song? If so, could this ability to compose songs be a remote beginning of more complex semantically compositional syntax?

Chaffinches, unlike ants and bees, learn their songs to some extent. The characteristic chaffinch song is quite complex, as we will see later. It can be divided into two main parts, an initial 'trill' and a final 'flourish'. The whole signal serves a dual function, acting both as a territorial challenge to other males and a way of attracting females. Using experimentally manipulated playback calls in the field, Leitão and Riebel (2003, p. 164) found that 'Males showed the closest approach to songs with a relatively short flourish. . . . These were the songs found less attractive by females tested previously (Riebel and Slater 1998) with the same stimuli'. In other words, if the flourish part of the song is short, males will tend to come a lot closer to other males than if the song has a longer flourish. It would be an oversimplification to say that the trill is a territorial challenge to rival males while the flourish functions to attract females, but undoubtedly the two parts of the song do tend somewhat to emphasize these respective functions.

Dual function calls that serve both a territorial and a courtship function are common in nature. But it is not so common that different features of the call can be teased apart and analysed as serving the different functions. Another example is the coqui frog, named after the two parts of its simple call, a low note followed by a higher note (the reverse of a cuckoo call, and higher pitched overall). Here again, it seems that a separate meaning can be assigned to each separate part of the call, each serving a different function. 'Acoustic playback experiments with calling males in their natural habitat and two-choice orientation experiments with females indicate that males and females of the neotropical tree frog *Eleutherodactylus coqui* respond to different notes in the two-note call of the male' (Narins and Capranica 1976, p. 378). 'In the Puerto Rican "Co Qui" treefrog, *Eleutherodactylus coqui*, the duration of the first note "Co", is critical in eliciting male territorial behavior, while the spectral content of the second note, "Qui", is crucial in eliciting positive phonotaxic responses from females' (Feng et al. 1990). The low 'Co' part of the call tends to serve a territorial function, while the higher 'Qui' part of the call tends to serve a courtship function.

Are these chaffinch and frog calls candidates for semantic compositionality, with the meaning of the whole call being formed by a combination of the meanings of its parts? No. The two meanings, territorial challenge and courtship invitation, are incompatible, and directed at different receivers. In the coqui frog, in fact, the male and female brains are tuned differently to be sensitive to the different parts of the call (Narins and Capranica 1976), so it is possible that neither male nor female actually hears the whole call, let alone puts its parts together. The parts of the chaffinch call cannot be combined in the way that distance and direction, for example, can be combined to yield location. The

closest to a compositional interpretation would be that the whole call conveys a conjunction of the meanings of the components.

1.1.3 Combinatorial, but not compositional, monkey and bird calls

Monkeys are more closely related to us than the insects, birds, and frogs that we have considered so far. Can we see any signs of semantically composed messages in monkeys? Klaus Zuberbühler is a leading investigator of this question. My conclusion from his work, surveyed below, is that some monkey communication is at the margins of semantic compositionality, expressing nothing even as complex as *hit Bill*. Likewise, there is no firm evidence of semantic compositionality in bird calls.

Arnold and Zuberbühler (2006) describe a call system used by putty-nosed monkeys in which different call elements are strung together. These monkeys only have two elementary (i.e. unitary) signals in their repertoire, labelled 'pyow' and 'hack'. They also have the ability to combine the elementary 'pyow' and 'hack' signals into longer sequences. This combinatorial power gives ways of expressing more than two meanings. So 'pyow' roughly means LEOPARD, 'hack' roughly means EAGLE, and 'pyow-hack' seems to mean LET'S GO, and so on. Note that the meaning LET'S GO is not a function, in any natural sense, of LEOPARD and EAGLE. This, then, is a (very small) **combinatorial** system, but it is not obviously semantically **compositional**, because in the case of the 'pyow-hack' the meaning of the whole is not a function of the meanings of the parts.

Arnold and Zuberbühler write, very carefully, 'Our findings indicate that non-human primates can combine calls into higher-order sequences that have a particular meaning'. There are two ways to interpret the data. One interpretation is that the internally represented meaning of 'pyow-hack' in the receiving monkey's mind has nothing to do with eagles or leopards, and that it invokes instead some separate notion of imminent travel. In this case the 'particular meaning' that the researchers mention is not a function of the meanings of the basic calls combined, and so the 'pyow-hack' call of the putty-nosed monkeys is not semantically compositional. This would be a case of animals overcoming the limits of their repertoire of individual calls by combining them, but not in any way reflecting the composition of the meanings expressed.

The other interpretation of the data, perhaps more plausible, is that 'pyow-hack' conjures up in the receiver's mind both concepts, EAGLE and LEOPARD, and the monkey takes appropriate action. In this case, the call is, in the simplest sense, compositional, expressing a **conjunction** of the meanings of its parts, that is EAGLE & LEOPARD. In a later paper (Arnold and Zuberbühler 2008),

somewhat extended data is described, with responses to longer series of pyows and hacks. Series combining pyows and hacks again elicited travel. Here the authors use the title 'Meaningful call combinations in a non-human primate'. This is again careful: the call combinations are meaningful, but whether they are interpreted compositionally remains an open question.[10]

A similar point can be made about another case carefully observed, and carefully discussed, by Klaus Zuberbühler (2002). This is more problematic, because part of the story involves the responses of one species, Diana monkeys, to the alarm calls of another species, Campbell's monkeys. Campbell's monkeys have specific alarm calls for leopards and eagles, and Diana monkeys respond to these by giving their own different alarm calls for these predators. There is some interest in first discussing the significance of the calls to the Campbell's monkeys alone. Zuberbühler writes 'In addition to the two alarm calls, male Campbell's monkeys possess another type of loud call, a brief and low-pitched "boom" vocalization. . . . This call type is given in pairs separated by some seconds of silence and typically precedes an alarm call series by about 25 s. Boom-introduced alarm call series are given to a number of disturbances, such as a falling tree or large breaking branch, the far-away alarm calls of a neighbouring group, or a distant predator. Common to these contexts is the lack of direct threat in each, unlike when callers are surprised by a close predator' (2002, p. 294). The responses of Campbell's monkeys to these boom-introduced calls are not described, but if they are like the responses of the Diana monkeys (to the Campbell's calls), the Campbell's monkeys show little or no alarm on hearing a pair of booms followed about 25 seconds later by what sounds like a regular alarm call. The booms could be interpreted as in some sense negating, or qualifying, the normal meaning of the alarm call, just as the English expressions *maybe* or *not-to-worry-about* might modify a shout of 'Police coming!' This is the strongest interpretation one can put on the facts. The 20-second delay between the booms and the alarm call is problematic, as it does not suggest composition of a unitary message. One would expect a unitary communicative utterance consisting of several parts to be produced with little or no delay between the parts (unlike the slow stately progress of whale songs.) The contexts in which the boom-introduced calls occur, as Zuberbühler describes them, can possibly be thought of as semantically composite, for example something like THREAT + DISTANT, but

---

[10] Another interesting fact is that in these studies female receiving monkeys only responded to the calls of 'their own' males, so this is not a case of a group-wide code. Also, Anderson (2008a, p. 800) has an identical take to mine on the 'pyow-hack' data.

the do-nothing responses cannot be seen as any obvious function of the panic reactions induced by the plain alarm calls.[11]

More recently, a team including Zuberbühler (Ouattara et al. 2009) have found more complex behaviour among wild Campbell's monkeys. Besides the 'boom' (B) call, they distinguished five different types of 'hack', which they labelled 'krak' (K), 'hok' (H), 'krak-oo' (K_+), 'hok-oo' (H_+) and 'wak-oo' (W_+). Their observations are worth reporting at length as they are the most complex yet seen in wild primates, and have some syntax, though it is not semantically compositional.

The different call sequences were not randomly assembled but ordered in specific ways, with entire sequences serving as units to build more complicated sequences. As mentioned, pairs of booms alone instigate group movements toward the calling male, while K_+ series functioned as general alarm calls. If combined, the resulting sequence carried an entirely different meaning, by referring to falling wood. In all cases, the booms preceded the K_+ series. We also found that another sequence, the H_+ series, could be added to boom-K_+ sequences, something that callers did when detecting a neighboring group. H_+ series were never given by themselves. . . .

These call combinations were not random, but the product of a number of principles, which governed how semantic content was obtained. We found five main principles that governed these relationships. First, callers produced sequences composed of calls that already carried narrow meanings (e.g., K = leopard; H = crowned eagle). In these instances, sequence and call meanings were identical. Second, callers produced meaningful sequences, but used calls with unspecific meanings (e.g., K_+ = predator). Third, callers combined two meaningful sequences into a more complex one with a different meaning (e.g., B + K_+ = falling wood). Fourth, callers added meaningless calls to an already meaningful sequence and, in doing so, changed its meaning (e.g., B + K_++ H_+ = neighbors). Fifth, callers added meaningful calls to an already meaningful sequence and, in doing so, refined its meaning (e.g. K + K_+ = leopard; W + K_+ = crowned eagle). We also found regularities in terms of call order. Boom calls, indicative of a nonpredation context, always preceded any other call types. H′ and K calls, indicators of crowned eagles or leopards, were always produced early in the sequence and were relatively more numerous if the level of threat was high.

<div align="right">(Ouattara et al. 2009, p. 22029)</div>

These monkeys do produce systematically formed call-sequences, so, like birds, they have some combinatorial syntax. The sequences are meaningful, apparently referential, but the meanings of the sequences are not functions of the

---

[11] For sure, one can always think of **some** function getting from one concept to another, but it won't necessarily be a very natural function. This assumes, of course (what else can we assume?) that what is a 'natural' function for a monkey is also at least somewhat natural for us human investigators.

meanings of the parts, so the syntax is not semantically compositional. What could be happening here is that there is a felt need to express more meanings than can (for some reason) be expressed by an inventory of four one-unit calls 'boom', 'krak', 'hok', and 'wak'. The monkeys cannot produce any further one-unit distinct calls, so they resort to making new signals by concatenating what they have. The meanings expressed are all of the same level of concreteness— LEOPARD, EAGLE, NEIGHBOURS, TREE-FALLING—and not in any hierarchical relation with each other, so a compositional system would not be appropriate. This is pure speculation, and not very convincing, at that, but maybe other similar examples will be found that shed some light on this case. It seems unlikely that Campbell's monkeys are the only species with such behaviour. We need more empirical field research.

Moving on to birds, the dominant consensus in the birdsong literature is that songs are meaningful in the sense that they function to attract mates or defend territory. The great variety in some birdsong repertoires is interpreted as impressive display, or versatile echoing of rival songs. Very few authors claim any compositional semantics for birdsong. Exceptions to this general trend are Hailman et al. (1985), writing about the black-capped chickadee, and Smith (1972), on its close relative, the Carolina chickadee.

These preliminary discoveries of S. T. Smith obviously do not specify referents of note-types completely, but they do suggest that the locomotory signals have something to do with such acts as take-off, landing, flight, perching, and change of direction.

(Hailman et al. 1985, p. 221)

S. T. Smith (1972) went on to make preliminary identification of note-types with specific 'messages' about locomotion, and noted that the combination of these notes in calls encoded a combination of their separate messages. She also pointed out that note-types are commonly repeated within a call, which suggests that the repetitions encode intensive aspects of the basic message of note-types.   (Hailman et al. 1985, p. 191)

Hailman et al. have no hesitation in writing about the 'referents' of the various note-types, of which there are just four in the black-capped chickadee. The last quotation above is a clear statement of compositionality, but it has not, to my knowledge, resurfaced in the literature. At most, the kind of compositionality involved expresses a conjunction of the meanings of the basic notes. For example, if note '**A**' signals something to do with take-off, and '**B**' signals something to do with change of direction, then the sequence **AB** might signal something to do with take-off and with change of direction. This is like the well-known child example 'Mommy sock', meaning something to do with Mummy and with a sock. It is the simplest form of compositionality. As Hailman et al. (1985) concede: 'Unlike written words made recombinantly from their

component letters, calls strung into bouts have no evident higher level of structure such as the grammar of human sentences' (p. 221).

In sum, there is no compelling evidence for any semantically compositional learned signalling in wild animals. Even if the problematic cases that have been mentioned are held to be strictly compositional, they are of limited scope, and provide only a slight platform upon which the impressive human capacity for compositionality might have evolved.

1.2 Non-compositional syntax in animals: its possible relevance

Some wild animals do produce syntactically complex behaviour, in semantically uninterpreted 'songs'. In such songs, although they are structurally somewhat complex, the meaning of a whole signal is not in any way composed as a function of the meanings of its elementary parts. How might signals which don't express any complex meaning be relevant to the evolution of human language? A number of writers, dating back several centuries, have seen in this behaviour the beginnings of human syntax. For these authors, the link lies in the sheer syntactic complexity of the songs. In this section and the next I survey these animal systems, and extract some general lessons about how to conceive of such pure syntactic abilities within biological organisms.

One evolutionary possibility is that after the chimp/human split the ancestors of humans developed somewhat syntactically complex songs like birds or gibbons, initially with no systematic combining of the meanings of the elements to convey some perhaps complex message (even if the elements had some meanings, which they might not have had). This is in fact a venerable idea. Rousseau and Darwin believed it, and so did Otto Jespersen, a renowned early twentieth-century linguist. These all saw music, in some form, as a pre-existing kind of syntactically complex expressive behaviour from which referentially meaningful language later evolved. The function of such complex songs was purely for display, to attract sex partners, they suggested (Darwin 1871; Jespersen 1922). The idea was of a separate syntactic ability, used for composing seductively complex songs—that is songs which were seductive purely by virtue of their complexity, and not by virtue of any semantic content, because they had none (apart from 'come mate with me'). For birdsong,

The evidence from the laboratory data is highly consistent and shows that, when females are exposed to large repertoires, they display higher levels of sexual arousal

than when they hear small repertoires (e.g. Catchpole et al. 1986; Lampe and Saetre 1995; Searcy and Marler 1981) Field data however are not as straightfor-ward. . . . [However] in the great reed warbler *Acrocephalus arundinaceus*. . . cuckolded males had smaller song repertoires than their cuckolders (Hasselquist, Bensch, and T. von Schantz, 1996).   (Gil and Slater 2000, p. 319)

The hypothesis of an early-evolved syntactic, specifically musical, ability, predating the exaptation of syntactically structured songs for propositional semantic purposes by humans, is explicitly argued by Fitch (2005, p. 16). 'The many similarities between music and language mean that, as an evolutionary intermediate, music really would be halfway to language, and would provide a suitable intermediate scaffold for the evolution of intentionally meaningful speech'.[12] Mithen (2005, 2009) has argued for a closely related view, involving coevolution of the human musical and linguistic capacities; see also Molnar-Szakacs and Overy (2006) who emphasize a common neural substrate for music and language, and similar hierarchical structure. Fitch points out that the function of such song need not be for sexual attraction, but could also have a role in promoting group cohesion, or could be used by mothers to calm their young. Cross and Woodruff (2009, pp. 77–8) also stress the functions of music in 'the management of social relationships, particularly in situations of social uncertainty'. For birds with extremely large repertoires, such as the nightingale, it has been pointed out that sexual attraction is an implausible function, as females are unlikely to spend time listening to over a hundred songs, just to be impressed by the male's versatility. In this case, a territory-marking function may be more likely, but the question remains whether rival males need to be told in so many different ways to get lost.

   Music has complex syntax, but the meaning of a whole tune is not put together from the elementary meanings of each note or phrase; and music certainly does not refer to outside objects or events (though it may iconically evoke them). It is possible that some purely syntactic capacity, possibly used for display, or to enhance group cohesion, or to claim territory, evolved in parallel with private, somewhat complex, conceptual mental representations. (Here 'syntactic' simply means 'exploiting combinatorial possibilities, given a set of elementary forms'.) Then, according to this hypothesis, at some later stage the conceptual and syntactic abilities got combined to give complex semantically compositional syntax. The syntax-from-song hypothesis has been seriously argued by serious people, so I will give it a fair hearing in this chapter. I do not think that pre-existing complex song can be the whole story of how

---

[12] See also an essay by Fitch at http://languagelog.ldc.upenn.edu/nll/?p=1136.

humans got complex syntax. But it may be some part of the story. How large that part is cannot be argued, given present evidence.

Command of a range of different complex songs may have served a mnemonic function when they finally began to carry some semantic content. Sometimes you have to repeat a sentence to yourself before you really understand what it means. The ability to repeat it before fully understanding it involves some capacity for holding a (somewhat) meaningless, but nevertheless structured, string in your head.

One intriguing similarity between the songs of many bird species and human utterances in conversation is that they are of roughly the same duration, between two and about ten seconds. A bird will sing one song from its repertoire, lasting, say, about five seconds, and then wait for a similar period, during which a territorial rival may sing its responding song, often identical or similar (Todt and Hultsch 1998, p. 488). Thus a kind of discourse exists with the same temporal dimensions as human casual conversation. (But whalesong is an entirely different matter, with individual songs lasting up to half an hour; this conceivably is connected to the greater distances over which songs transmitted through water can carry.)

A striking difference between bird repertoires and human languages illustrates the unproductivity of bird syntax: 'The composition of vocal repertoires reveals a basic principle in most songbirds: The sizes of element-type repertoires are larger than the sizes of their song-type repertoires' (Hultsch et al. 1999, p. 91). This assertion is surprising to a linguist if one equates element-types with words and song-types with sentences. This fact is also stated by Todt (2004, p. 202) and Bhattacharya et al. (2007, p. 2), and is borne out by the examples I will discuss here. Podos et al. (1992) devised a method to put identification of song-types, and hence song repertoire sizes, on a firmer objective footing. They introduced a concept of 'minimal unit of production', MUP for short. An MUP is typically an individual note, but can be a sequence of notes if these notes always occur together in the same order. Then one can count the MUP repertoire size and the song repertoire size of any bird. Using this method, Peters et al. (2000) quantified the MUP repertoire and song repertoire sizes of five geographically separate groups of song sparrows. In all cases the MUP repertoire sizes were greater than the song repertoire sizes, by factors of about six or seven.

Much depends, of course, on how you count song-types. Hailman et al. (1985) studied chickadee (*Parus atricapillus*) 'calls' (most of which more recent researchers would classify as 'songs'). They counted 362 different 'call-types' composed from a basic vocabulary of four notes. This includes one-note, that is non-combinatorial, calls, and calls with different numbers of repetitions of

the component notes, which other researchers would classify as belonging to the same song-type. Counting only songs in which notes are combined and counting repetitions of the same note as one, the number of distinct songs comprising over 99 per cent of the repertoire comes, by my reckoning, to just four, the same as the basic vocabulary. A spectacular example of a bird's failure to exploit syntactic combinatorial possibilities is provided by the brown thrasher (*Toxostoma rufum*). This bird is reported as being at the extreme of vocal virtuosity, having 'a repertoire of several thousand different types of song' (Brenowitz and Kroodsma 1996, p. 287). The original students of this bird's repertoire (Kroodsma and Parker 1977) report that each distinct song type is in fact a repetition of a distinct syllable type. There is not, apparently, any combination of one syllable type with another in the same song. So this bird has an estimated vocabulary in the thousands, and its song repertoire is in fact no larger than its vocabulary. This extreme example illustrates a general point that whatever syntax can be found in bird repertoires, they do not take advantage of its combinatorial possibilities. An analogy from English orthography would be a repertoire of, say, five words which happen to use all 26 letters of the alphabet. Given so many letters, and some possibility of combining them, why restrict the combinations to less than the number of letters? Why not make up and use more words? In human languages, the inventory of phonemes is always orders of magnitude smaller than the vocabulary size; and the vocabulary size is always orders of magnitude smaller than the number of possible sentences. Birdsong is thus strikingly different in this respect.

Conceivably, an ability for complex song provided an evolutionary basis for human **phonological syntax**, but no basis, or only a slight basis, for the semantically interpreted syntax of whole sentences. '[P]honology (sound structure), the rules for ordering sounds, and perhaps the prosody (in the sense that it involves control of frequency, timing, and amplitude) are the levels at which birdsong can be most usefully compared with language' (Doupe and Kuhl 1999, p. 573). MacNeilage (2008, pp. 303–8) also finds suggestive parallels between the serial organization of birdsong and human phonological syntax. A complementary part of the story, and perhaps the whole story, of how we got complex semantically compositional syntax is that it evolved on a platform of complex conceptual representations, plus some natural principles of the communication of information. These last complementary ideas are not for now but for later chapters.[13]

---

[13] The evolutionary contribution of pre-existing song-like syntax to modern semantically interpreted syntax is bound up with a debate between advocates of two different possible routes to modern syntax, an 'analytic' route and a 'synthetic' route. This debate will be the topic of a later chapter.

Pure uninterpreted syntax is not found in communication systems in the recent human lineage. The closest species to us displaying such asemantic song are gibbons. The absence of any complex songlike behaviour in great apes is not necessarily a problem. Complex song occurs many times in nature, in subsets of classes and families. Many, but not all, bird species have complex song. Among oscine birds, chaffinches have complex songs, but crows do not. Some hummingbirds have complex song (Ficken et al. 2000), while others don't. Among whales and dolphins, humpback whales have the most complex song. Among primates, only gibbons have complex songs. Complex song, it appears has evolved separately several times. So it could have evolved separately in humans after the chimp/human split.

Learned vocal behaviour also cross-cuts phylogenetic classifications, and so has probably also evolved independently several times. There is a close correlation between complexity of song and the degree to which the song is learned. If we can class human speech with song, humans have 'songs' that are both complex and learned. Despite the great genetic distance between songbirds and humans, and despite the large differences in their brain structure (e.g. birds do not have a many-layered cortex like mammals), there are significant similarities in the neural circuitry used for the production and learning of vocalizations. Jarvis (2004a, 2004b, 2007) argues for a hypothesis that 'vocal learning birds—songbirds, parrots, and hummingbirds—and humans have comparable specialized forebrain regions that are not found in their close vocal non-learning relatives' Jarvis (2007, p. 35). To argue this, Jarvis has to depend on a number of hypothesized functional equivalences of parts among the anatomically different brains involved (of parrots, songbirds, hummingbirds, and humans). He gives a long list of evidence that lesions in equivalent places in these brains produce functionally similar deficits in the respective species (2007). In similar vein, Doupe and Kuhl (1999, p. 567) summarize a broad survey thus: 'Although some features of birdsong and speech are clearly not analogous, such as the capacity of language for meaning, abstraction, and flexible associations, there are striking similarities in how sensory experience is internalized and used to shape vocal outputs, and how learning is enhanced during a critical period of development. Similar neural mechanisms may therefore be involved'. They also cite lesion and stimulation studies which bring out the similarities among learners, and their differences from non-learners. The relevant areas are areas of higher control:

Both songbirds and humans have high-level forebrain areas that control the preexisting hierarchical pathways for vocal motor control..., whereas nonlearners do not. There are no neocortical sites in monkeys from which vocalization can be elicited

by stimulation nor whose ablation affects calls (Ploog 1981). In striking contrast, in humans the entire perisylvian cortical area as well as posterior parieto-temporal cortex is critical for speech production, as shown by both stimulation and lesion studies.

(Doupe and Kuhl 1999, p. 599)

This again suggests convergent evolution by somewhat different kinds of brain onto a common working solution to the problem of vocal learning.

Complex signals of wild animals are only partly learned, or not at all; in all species, there is a very hefty innate component. Without assigning percentages to innate and learned components, it is clear that the parallel between human language and animal songs is not strong on this point. Commitment to a nativist and syntactocentric view of language can lead to an emphasis on parallels between birdsong and language:

Certainly, little or no overlap occurs in the details of the development of speech in children and of song in birds. Equally obvious, however, is the remarkable similarity of these two processes at only a modest level of abstraction. ... We should have little hesitation in seeing both processes as essentially similar, as the working out of a species' developmental program in biologically guided maturation. In other words, nestlings and babies both grow up in a specific way, determined in its essence by the fact that they are birds and humans, respectively.   (Anderson 2004, p. 165)

What this view underemphasizes is the massive functional (semantic) difference between birdsong and language, accompanied by an equally great difference in structural complexity, differences that Anderson elsewhere acknowledges. Additionally, a remarkable difference between nestlings and babies growing up and learning their language is the fact that birds do not learn their song incrementally through a process of discourse with their parents (or other group members). Birds store the patterns they hear as nestlings, and then only later, sometimes as much as eight months later, start to produce their own songs.[14]

In birdsong, there is also some evidence of voluntary control. 'We found that chaffinches (*Fringilla coelebs*) in noisier areas (i.e., close to waterfalls and torrents) sang longer bouts of the same song type before switching to a new type, suggesting that they use increased serial redundancy to get the message across in noisy conditions' (Brumm and Slater 2006a, p. 475). In another study, Brumm and Slater (2006b) found that zebra finches sang louder when the receiving female was further away, and draw a superficial parallel with humans raising their voices. However, they suggest that 'this behaviour can be

---

[14] See Fehér et al. (2009) for an interesting recent study in which zebra finches developed a wild song type, over three or four generations, by iterated learning starting from birds who had had no model to imitate.

accounted for by simple proximate mechanisms rather than by the cognitive abilities that have been thought necessary in humans' (p. 699).

To recap, it is worth looking at complex song in species not closely related to humans because of the possibility of a parallel independent evolution adapting to similar functions, and involving similar brain mechanisms. If this happened, then some time after the human/chimp split, our ancestors developed a capacity for complex musical or song-like behaviour that was later recruited for the expression of complex meanings. Perhaps it did happen. Some empirical light could be shed on this question by testing the susceptibility of apes and monkeys to various sequences with music-like structure.

1.3 Formal Language Theory for the birds, and matters arising

So far, I have only mentioned that birdsong and whalesong can be syntactically 'complex'. But how complex is 'complex'? In the rest of this chapter, we get to grips with a way of comparing meaningless syntax across species. It will emerge that despite big quantitative differences between animal song and human language, the more complex animal songs do have some similarities with language. Apart from the obvious lack of compositional, and referential, semantics, these songs are not qualitatively, but only quantitatively, different in their basic combinatorial structure.[15]

If we are seriously to compare human syntax and the complex songs of animals, we need some common scale by which to measure each of them. Formal Language Theory provides a scale which is in some ways suitable. The cross-disciplinary exercise of applying this body of theory to animal songs will reveal some of the serious theoretical issues that arise when applying the tools of one trade to data from another. One conclusion will be that applying this scale shows that human languages are not just head and shoulders above animal songs in syntactic complexity, but (to continue the metaphor) head, shoulders, trunk, and legs above them. The familiar assertion of a huge gap between humans and non-humans is thus reinforced. But it is good to have a non-impressionistic way of justifying this common assertion, and Formal Language Theory provides a tool for this. The other main conclusion to arise from this exercise is that certain issues which have been contentious in theorizing about

---

[15] This is not to deny that certain semantico-syntactic, or pragmatico-syntactic features of human language are absent from animal song (see Chapters 3 and 4). I assume that these features were superimposed on any basic syntactic structure if and when it was recruited for expressing complex meanings.