Anti-Babel:

Three Degrees of Interspecies Comprehension

Philippe Schlenker[a,b], Camille Coye[a], Ambre Salis[a], Shane Steinert-Threlkeld[c], Lucie Ravaux[a], Emmanuel Chemla[d]

Abstract. Recent work seeking to provide a formal account of meaning in animal communication took call form to be arbitrary (Schlenker et al. 2014, 2016b). But a long line of research (from Marler 1955 to Magrath et al. 2020) has shown that this is not always so, especially in birds: there is sometimes widespread convergence among the calls of different species, to the point that a species may understand the calls of another one that is geographically and phylogenetically extremely distant. This suggests that there might be a natural biological code by which the calls of unrelated and unfamiliar species can sometimes be understood without prior exposure. We clarify this possibility by distinguishing among three degrees of interspecies comprehension. In the first degree ('Understand Thy Neighbor'), a species understands some of the calls of a neighboring species because it is exposed to them. In the second degree ('Call Convergence'), a species may understand the calls of an unrelated and unfamiliar species by virtue of Marlerian convergence: a heterospecific call may globally resemble a conspecific call enough to yield understanding. In the third degree ('Featural Interpretation'), a species may use a rule that associates a meaning to a specific acoustic feature – e.g. higher call rate signals greater urgency, higher pitch signals greater arousal, greater noisiness signals greater negativity. This yields a kind of featural compositionality by which a species may understand a heterospecific call that does not globally resemble any familiar call, but still includes the crucial, interpretable feature. There might thus be an entirely new road to the emergence of compositionality in studies of meaning evolution. We lay out possible mechanisms of evolution of these degrees of interspecies comprehension, and isolate predictions that distinguish the third degree from the second.

[a] Institut Jean-Nicod (ENS - EHESS - CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France; PSL Research University.
[b] Department of Linguistics, New York University, New York, USA.
[c] Department of Linguistics, University of Washington, Seattle, WA, USA.
[d] LSCP (ENS - EHESS - CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France; PSL Research University.

1   Introduction

1.1   Babel and Anti-Babel

A long-standing problem in human linguistics might be called the 'Babel Problem': if human language has a strong innate component (as argued by Chomsky (1965) and his followers), why is there more than one human language (e.g. Culicover 2021)? The opposite problem ('Anti-Babel') exists in animal communication, as highlighted by Marler's pioneering work (1955): why do phylogenetically and geographically distant species sometimes display striking levels of signal convergence?

A result of signal convergence is eavesdropping, the phenomenon whereby a species understands some aspects of heterospecific signals. As summarized by Magrath et al. (2020), "over 70 species have been shown through experimental playback to respond to other species' alarm calls". But the most surprising part is that sometimes a species understands the calls of another species that is both phylogenetically and geographically extremely distant, suggesting that there exists a 'biological code' that unrelated species may sometimes share.

Call convergence raises two questions. First, what are the evolutionary mechanisms by which calls with related forms may have related meanings in unrelated species? Second, does call convergence have consequences for the way meaning is represented in the animals' minds? The first question has been amply, if informally, discussed in the ethological literature, from Marler (1955) to Magrath et al. (2020). The second question takes a particular significance in view of recent studies of 'animal semantics'.

1.2   Animal semantics and interspecies comprehension

Recent work seeking to provide a formal account of meaning in animal communication took call form to be arbitrary (Schlenker et al. 2014, 2016b). One natural question in this line of research was how much structure sequences of calls have, and whether they display any traces of compositionality. With some exceptions, the answers coming out of 'animal linguistics' have mostly been deflationary (see also Rizzi 2016, Sauerland 2016). But a long line of research pioneered by Marler (1955) has shown that form is not always arbitrary, especially in birds. The potential existence of a biological code by which species may understand each other raises the question of its cognitive status.

We aim to clarify this question by distinguishing among three degrees of interspecies comprehension.[1] In the first degree ('Understand Thy Neighbor'), a species understands some of the calls of an unrelated but neighboring species because it is exposed to them. In the second degree ('Call Convergence'), a species may understand the calls of an unrelated and unfamiliar species by virtue of Marlerian convergence: a heterospecific call may globally resemble a conspecific call enough to yield understanding. In the third degree ('Featural Interpretation'), a species may know a rule that associates a meaning (a class of situations) to a specific acoustic feature—e.g. higher faster call rate signals a greater urgency, higher pitch signals greater arousal, greater noisiness signals greater negativity. This yields a kind of feature-based compositionality by which a species may understand a heterospecific call that does not globally resemble any familiar call, but still includes the crucial, interpretable feature.

The first and the second degree of interspecies comprehension need not affect the architecture of theories of animal semantics. But if real, the third degree has striking consequences, as it implies that designated acoustic features must sometimes be given a semantics. This opens the possibility that the meaning of certain calls might be derived from the meanings of some of their acoustic features—an entirely new road to compositionality in studies of meaning evolution. A direct consequence might be productivity, as unfamiliar and unrelated calls might then be understood even in the absence of any global similarity with native (or more generally familiar) calls.[2]

---

[1] We set aside the case of comprehension though common descent of the calls, as this case does not raise special problems in the present context.

[2] We write with great caution because we don't think all the crucial experiments have been performed yet, and one of our goals is precisely to propose some crucial experiments. Still, bird specialists routinely write and talk as if the decoding of individual features is at the heart of interspecies comprehension: they assume a version of the third degree. (Two salient examples, discussed below, are Aubin and Brémond 1989 and Aubin 1991, who discuss 'laws of decoding' used to interpret an artificial distress call played back to different bird species.)

We lay out possible mechanisms of evolution of these degrees of interspecies comprehension, and isolate predictions that distinguish the third degree from the second (one consists in pitting overall acoustic similarity against the presence of a crucial acoustic feature to determine whether the latter is interpreted as such; another one consists in testing the productivity of Featural Interpretation). While we definitely do not attempt to settle the debate, we argue that the distinction between the second and the third degree is of utmost importance for two foundational fields: studies of animal semantics, and studies of the evolution of meaning and compositionality. In addition, Featural Interpretation is of independent interest because it is formally close to mechanisms postulated in recent studies of music semantics (e.g. Schlenker 2017, 2022).

1.3   Structure

The rest of this articles is organized as follows. We define and illustrate our three degrees of comprehension in the interlinguistic case first, and then in the interspecies case, highlighting similarities and differences (Section 2). We then survey the empirical landscape, with very clear arguments for our first two degrees, Understand Thy Neighbor and Call Convergence, and far more speculative ones for the third degree, Featural Interpretation (Section 4). We then highlight crucial predictions that could help adjudicate between comprehension through Featural Interpretation and through similarity with familiar calls (Section 5), before drawing some conclusions (Section 6).

2   Three Degrees of Interlinguistic Comprehension

To introduce the main distinctions, we start from human language and consider three cases of interlinguistic comprehension; later sections will develop related concepts for animal languages. Our goals are purely pedagogical: we do not seek to give realistic analyses of linguistic phenomena, but to illustrate in their simplest possible form some mechanisms of meaning production.

2.1   Understand Thy Neighbor

We will consider just two expressions, one with a greeting function, *hello*, and one to request help, *help*. If Ann, a native English speaker, lives in Montreal, a bilingual city, she will have no trouble understanding that the French word *bonjour* has the meaning of *hello* while *à l'aide* has the meaning of *help*. This is of course the very simplest instance of *Understand Thy Neighbor*: Ann will have heard enough French in her environment to learn the meaning of these two common expressions.

Now if Ann knows no other language and encounters a German tourist who tells her: *Hallo*, the acoustic resemblance with *hello* might be enough to allow her to infer that it means *hello*; and maybe similarly if the tourist says *Hilfe* for *help*: even without any knowledge of Germanic historical linguistics, Ann might just tend to interpret foreign words by their acoustic similarity to English words she knows. On the other hand, if Ann hears an Italian tourist say *salve* or *aiuta*, she might not be able to infer that the first means *hello* and the second means *help*.

2.2   Convergence

Still, the expressions *aiuta* and *help* might in some respects be produced in related ways. Since they are requests for help with unspecified addressees that might be far away, they might both be produced with greater loudness than is standard, and they might even be repeated: *aiuta! aiuta!* So even without knowing the meaning of *aiuta*, Ann might be able to use these features—increased loudness, possibly repetition—to infer its meaning.

These truisms raise two questions. First, why do words such as *help* and *aiuta* tend to be produced in related ways with respect to loudness? Second, how can Ann use these commonalities to infer the meaning of words she might not initially know?

To address the first question in a transparent framework, we adopt a simplified version of the Rational Speech Act model and ask what the speaker's options are (see for instance Goodman and Frank 2016, Franke and Jäger 2016, and Bergen et al. 2016). To display a maximally simple contrast between *hello* and *help*, we will assume that the speaker can be in one of two situations: a distress situation w,

and a greeting situation w'. In the first case, she can only utter {help, HELP}, two identical expressions that only differ with respect to loudness; and similarly in greeting situations, with {hello, HELLO}. In other words, we take the meanings of the words to be given and fixed, and only their acoustic modulation to be open to choice.

Of course, one needs to exert oneself a bit more to produce a louder message, hence a slightly greater cost. In greeting situations, the interlocutor is always nearby and hence the message will be heard no matter what. Thus it won't matter whether one uses the loud or of the soft version of the message, and using the loud version won't be worth the effort. But in distress situations, the interlocutor may be far away and the message may be missed; the loud version of the message will have a better chance of being heard, which might justify the additional cost.[3]

A very simple model is displayed in (1)-(2). The assumptions are just that the utility of a (true) message is the probability that it is heard, minus the cost of producing it. A louder message is more costly, and this has benefits in the distress situation but not in the social situation, with the immediate result that *help* is pronounced loud but *hello* isn't. We assume throughout that only true messages can be sent.

(1)   For a world w and a true message m in w, the speaker's utility is given by:

$$U(m \mid w) = p(m, w) - c(m)$$

where $p(m, w)$ is the probability that the message m is heard in the situation w and $c(m)$ is the cost of message m.

(2)   a. Distress situations w
$p(help, w) = .5$
$p(HELP, w) = 1$
$c(help) = .2$
$c(HELP) = .3$

$U(HELP \mid w) = 1 - .3 = .7 > U(help \mid w) = .5 - .2 = .3$

b. Greeting situations w'
$p(hello, w') = p(HELLO, w') = 1$
$c(hello) = .2$
$c(HELLO) = .3$

$U(HELLO \mid w') = 1 - .3 = .7 < U(hello \mid w') = 1 - .2 = .8$

2.3   Featural Interpretation

Of course the reasoning will apply just as well to the Italian counterparts of these expressions, with the result that *salve* and *aiuta* will be pronounced as soft and loud respectively. This, in turn, may help Ann identify the meaning of *aiuta* even if she has never heard this word before. Importantly, however, there are two hypotheses about what might be going on in case she does infer the meaning of the word without prior exposure to it.

One possible hypothesis is that by virtue of its loud pronunciation, *AIUTA* is acoustically similar to *HELP*. This would make this case analogous to that of *hello / hallo*, two words that globally resemble each other despite being from different languages. On this view, Ann applies whatever similarity metric she has to treat *AIUTA* and *HELP* as expressions of her native language. If so, Ann's ability to understand *AIUTA* is a result of linguistic convergence, applied to the use of loudness in some situations but not others, combined with an acoustic similarity metric.

In the case at hand, this account isn't very plausible, however. After all, *AIUTA* really doesn't sound like *HELP*: yes, both expressions are pronounced loudly, but besides that, they have little in common. But an alternative hypothesis presents itself: Ann may reason on the speaker's maximization

---

[3] An additional effect is that the utility differential between being heard and not being heard is probably greater in the distress case, as one has a lot to lose if one doesn't get help. Since our goal is to present the simplest human example that might have an animal counterpart, we disregard this point here.

behavior to infer that they are in the distress world w, and not in the social contact world w'. The reasoning could get off the ground irrespective of which other word is used to express *hello*, and independently of whether there is any global similarity between *AIUTA* and *HELP*.

This reasoning, which is laid out in (3), is straightforward: the addressee reconstructs the speaker's utility maximization behavior described above, and comes to the conclusion, as we have, that the speaker will use a loud word for 'help' and a soft word for 'hello'.

(3) Addressee's reasoning
In view of (2) applied to a language with expressions *<hello>*, *<help>*, unknown foreign counterparts of English *hello*, *help:*

in w the speaker chooses *<HELP>* over *<help>*;
in w' the speaker chooses *<hello>* over *<HELLO>*.

So the addressee takes
<HELP> (in our example, *AIUTA*) to be indicative of w
<hello> (in our example, *salve*) to be indicative of w'.

Rather than global acoustic similarity, it is now a reasoning on the speaker's maximization problem that leads Ann to infer that the Italian equivalent of *help* should be pronounced loudly, and unlike the equivalent of *hello*. Loudness *per se* ends up being a telltale sign that the speaker is in a distress rather than in a greeting situation.

Our three levels of interlinguistic comprehension are now easily illustrated on these highly stylized examples. *Understand Thy Neighbor* is responsible for Ann's understanding of French *à l'aide*, which she has heard in her environment. *Convergence*, motivated by the speaker's maximization behavior, explains why *help* and *aiuta* alike end up being pronounced loud. This *might* be sufficient to produce interlinguistic comprehension if the results are acoustically similar across the two languages. Since this is very dubious in this case, one needs to postulate a third level of interlinguistic comprehension, *Featural Interpretation*: a certain acoustic feature, here loudness, is associated with certain situations but not others. Importantly, the mechanism we outlined to explain this association is based on the addressee's reasoning on the speaker's maximization problem, and is unlikely to have a direct counterpart in the animal realm.

3   Three degrees of interspecies comprehension

When we turn to animal languages, similarities and differences arise with respect to the case of interlinguistic comprehension; we outline them in this section, before discussing the empirical landscape in Section 4.

3.1   *Understand Thy Neighbor*

Understand Thy Neighbor applies in comparable ways in animal communication and in human languages. Numerous cases have been described in which two species have different calls but understand each other, presumably because they live in the same environment and are able to learn the meaning of heterospecific calls. Campbell's monkeys and Diana monkeys diverged approximately 6.5 million years ago (Perelman et al., 2011). They often live in the same trees, and thus they have much experience with each other's calls. While female calls are arguably to some extent preserved in the two species (Coye et al., 2022), males have structurally and acoustically different repertoires across the two species (Ouattara et al., 2009, Zuberbühler, 2000). Still, Diana monkeys understand Campbell's calls (Zuberbühler 2002). In the Tai forest of Ivory Coast, the Campbell's call *krak* call is indicative of leopards. When *krak* calls are played through a hidden loudspeaker, the Dianas respond with their own calls appropriate for ground predators. Cambpell's *hok* calls are indicative of eagle presence, and if they are played back, the Dianas respond with their own raptor calls. In other words, although the Dianas

can't produce the Campbell's calls, they understand some of their meaning.[4] Similarly, Campbell's monkeys understand some Diana calls, and these are even used in field experiments to trigger Campbell's calling behavior (Schlenker et al. 2014). This type of bilingualism isn't restricted to primates. The hornbill (a brightly colored bird with a majestic flight) was shown in field experiments to react appropriately to some Diana monkey calls: it discriminated between eagle Diana calls, which were indicative of a direct threat for the hornbill, and leopard calls, which were not (Rainey et al. 2004a, b; Zuberbühler 2009).

One possible twist relative to interlinguistic communication is that we don't have strong grounds for positing that a given animal treats the calls of different species as anything but mere indices of something that's going on in their surrounding: just like smoke might indicate fire, to a Diana monkey, a Campbell's monkey *hok* call might indicate eagles.[5]

Be that as it may, exposure to the call in one's environment may be sufficient to understand some of its meaning.

3.2   Call Convergence

The second degree of interspecies comprehension is based on the fact that the calls of phylogenetically and geographically distant species may converge. In many (though not all) species, call form and function is thought be innate. When this is the case, convergence must take place over evolutionary times. The existence of such convergence on functional grounds is a classic idea developed in Marler 1955, an article that ended with these words: "many of the sounds used in bird and insect 'language' have not been chosen arbitrarily, but are directly adapted in structure to the function they have to perform". Marler discussed sound properties that facilitated or confused location depending on the call's function: a mobbing call should be easy to locate so conspecifics know where to mob; by contrast, a raptor alarm call (without mobbing) may be more optimal if the caller is hard to locate.

We can illustrate the same logic with loudness. In our human example, we took the convergence in the use of loudness for *help* but not for *hello* to be caused by a rational maximization process. In some cases, its analogue in the animal kingdom would likely be evolutionary in nature:[6] starting from a population that has two soft calls, an alarm call a and a greeting call g, mutants that replace a with its loud version A will gain a fitness advantage (whereas no such advantage will be conferred to mutants to replace g with its loud version G). Here we can use the very same assumptions that were stated in (1) and (2), with the hypothesis that the speaker/sender and addressee/receiver derive the same utility from the (cooperative) interaction.

We will use a standard notion of evolutionary stability, due to Maynard Smith and colleagues. According to the standard definition, strategy I is evolutionarily stable just in case for all alternative strategies J, either (1) the payoff of I against I is greater than the payoff of J against I, or (2) I and J have equal payoffs when played against I, but I has a greater payoff than J when played against J (e.g. Maynard Smith and Price 1973, Skyrms 1996). The basic intuition is that I, the strategy of a resident population, is evolutionarily stable just in case it can successfully resist invasion by mutants J, which is the case if mutants fare less well against residents than the majority residents do against each other

---

[4] Zuberbühler 2002 goes one step further. Campbell's monkeys have a non-predation call, *boom*, produced with air sacs. Diana monkeys have no such call. But they (correctly) understand that when *boom* is prefixed to a Campbell's *krak*-series or *hok*-series, it changes its meaning and the series stops being alarming; under such conditions, Diana males mostly *fail* to reply with their own leopard-appropriate or eagle-appropriate sequences. Thus they understand some of the details of Campbell's call series.

[5] The underlying conceptual distinction is related to that between a cue and signal. Maynard Smith and Harper 2003 write: "We define a 'signal' as any act or structure which alters the behavior of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved." (p. 3) The authors further explain that "the requirement that a signal evolved *because* of its effect on others distinguishes a signal from a 'cue'", which is "any feature of the world, animate or inanimate, that can be used as a guide to future action."

[6] An example that certainly arose over evolutionary times is the loud alarm calls produced by male Campbell's monkeys, which contrast with the soft social calls produced by females (Ouattara et al., 2009, Lemasson & Hausberger, 2011). There are other cases in which animals adapt the loudness of their calls to the environment (Waser & Waser, 1977); these might well be the result of an online maximization process.

(Condition (1)); or mutants and residents fare equally well against residents, but residents fare better against mutants (Condition (2)).[7] It is worth noting that if I is a strict Nash equilibrium, then I is evolutionarily stable because Condition (1) is automatically satisfied; but if I is just a weak Nash equilibrium, this need not be the case and one may have to appeal to Condition (2) to determine whether the equilibrium is evolutionarily stable. Below we will solely appeal to Condition (1).

Let's consider the alarm call *a* first. The resident population uses a soft alarm call *a*, the mutants use the loud version *A* instead. We assume that encounters are symmetric: when two individuals <x, y> interact, in half the cases x is the signaler and y the receiver, and in the other half it's the other way around. Compared to a resident interacting with another resident, a mutant interacting with a resident will yield greater (shared) utility through the use of a loud alarm call. For this reason, the resident 'soft alarm call' strategy will not resist invasion, as Maynard Smith's Condition (1) will be violated. We can make this point concrete by re-using the distribution of utilities and costs defined in (1)-(2) above.

(4) Pay-offs of a communicative interaction when the resident population S(oft) uses *a* and a mutant population L(oud) uses *A* instead, with precisely the situation (= same computations of utility) corresponding to that of *help* in (1)-(2), in a situation of alarm w.

a. Resident S_1 interacting with resident S_2
<S_1's utility, S_2's utility> = <U(help | w), U(help | w)> = <.3, .3>

b. Mutant L interacting with resident S
<L's utility, S's utility>　　　= <.5 (U(help | w)+U(HELP | w)), .5 (U(HELP | w)+U(help | w))>
　　　　　　　　　　　　　　 = <.5 (.3 + .7), .5 (.3 + .7)>
　　　　　　　　　　　　　　 = <.5, .5>

We can now see numerically what we first described qualitatively: Condition (1) of evolutionary stability because a mutant interacting with a resident gets greater utility (namely .5) than a resident interacting with a resident (yielding a utility of .3). The 'soft call' strategy is not evolutionarily stable.

Conversely, the 'loud call' strategy is evolutionarily stable relative to the 'soft call' strategy. In a nutshell, a resident that uses a loud alarm call with another resident will now produce greater (shared) utility than a mutant who uses a soft alarm call with a resident, and thus Maynard Smith's Condition (1) will be satisfied. This is shown with the same numerical example in (5).

(5) Pay-offs of a communicative interaction when the resident population L(oud) uses HELP and a mutant population S(oft) uses help instead, with a situation corresponding to that of *help* in (1)-(2), in a situation of alarm w.

a. Resident L_1 interacting with resident L_2
<L_1's utility, L_2's utility> = <U(HELP | w), U(HELP | w)> = <.7, .7>

b. Mutant S interacting with resident L
 <S's utility, L's utility>　　　= <L's utility, S's utility> (as in (4)b)
　　　　　　　　　　　　　　 = <.5, .5>

By contrast, when we consider the greeting call, the 'soft call' strategy is evolutionarily stable relative to the 'loud call' strategy. Since a loud greeting call has a greater cost but is not more effective than a soft call, a resident using the 'soft call' strategy with another resident produces greater (shared) utility than a mutant using a 'loud call' strategy with a resident. In other words, Maynard Smith's Condition (1) is satisfied; this is illustrated with our numerical example in (6).

(6) Pay-offs of a communicative interaction when the resident population S(oft) uses *g* and a mutant population L(oud) uses *G* instead, with a situation corresponding to that of *hello* in (1)-(2), in a situation of alarm w'.

a. Resident S_1 interacting with resident S_2
<S_1's utility, S_2's utility> = <U(hello | w'), U(hello | w')> = <.8, .8>

---

[7] As Maynard Smith and Price 1973 write, "the relevance of the latter condition [= Condition (2)] is as follows. If in a population adopting strategy I a mutant J arises whose expectation against I is the same as I's expectation against itself, then J will increase by genetic drift until meetings between two J's becomes a common event."

b. Mutant L interacting with resident S

$$\text{<L's utility, S's utility>} = \text{<.5 (U(hello | w')+U(HELLO | w')), .5 (U(hello | w')+U(HELLO | w'))>}$$
$$= \text{<.5 (.8 + .7), .5 (.8 + .7)>}$$
$$= \text{<.75, .75>}$$

Conversely, it can be checked that the 'loud call' strategy for a greeting will not be evolutionarily stable.[8]

In essence, we have just transformed competition among strategies in the head of a single speaker into a related competition, but now among individuals whose fitness depends on these strategies. The outcome is the same: the calls of different species may converge with respect to loudness for essentially the same reason as in the interlinguistic case. Of course loudness is just a motivating example; what matters for our purposes is the logic of the argument, which arises from the differential distribution of utilities and costs, and would thus be applicable to other cases.

As a result, a species x with a call C may be expected to understand the call C' of a phylogenetically and geographically distant species if (i) the calls C and C' of the two species have converged, and (ii) this convergence is enough to make C' similar to C according to whatever acoustic similarity metric species x uses to interpret unknown calls.

We should add that this second degree of interspecies comprehension may in some cases depend on the first one, in the following sense: a native species x may understand unfamiliar calls of a phylogenetically and geographically distant species z because these calls converge with those of x's neighboring species y rather than with x itself. As long as x understand the relevant calls of y, these will count as familiar and similarity to y's calls will yield comprehension by x.

3.3 Featural Interpretation

Call convergence only involved an evolutionary adaptation of the utility-maximizing behavior we posited in the interlinguistic comprehension case (restricted to loudness convergence). Things are different with an addressee's ability to reason on the speaker's maximization problem to associate loudness with some situations but not others (in our example, with alarm but not greeting situations). It would be surprising if such reasoning could be performed by non-human animals. Still, there are empirical arguments for the existence of the corresponding behavior (associating a feature with some situations but not others), which will raise a puzzle about its evolutionary sources.

We will restrict attention to the case of call interpretation, taking call production to be constant across all cases.[9] Formally, the problem can be stated as follows. There are four potential calls to consider, namely <alarm>, <ALARM>, <greeting>, and <GREETING>. Since they involve just two parameters of variation, we can represent them as pairs, namely <a, soft>, <a, loud>, <g, soft>, <g, loud>, where the first component is the base call, and the second its loudness. For Featural Interpretation to arise, receivers should associate the acoustic feature *loud* to alarm situations—in our earlier notation, to the alarm world w and not to the greeting world w'. We will assume the result of the previous section, Call Convergence: we will thus start from a resident population that understands *ALARM* as indicative of alarm situations, and *greeting* as indicative of greeting situations. This is represented with a feature decomposition in (7). The question is how a mutant population could arise in which *loud* is associated with alarm situations.

(7)   Resident population – rules of interpretation
       With two words w = alarm situation and w' = greeting situation, we have:

---

[8] Let's assume that the resident population uses the 'loud call' strategy for a greeting. A resident L_1 interacting with a resident L_2 yields a pair of utilities <L_1's utility, L_2's utility> = <U(HELLO | w'), U(HELLO | w')> = <.7, .7> (as in (2) in the text). By contrast, a mutant S using a 'soft call' greeting strategy interacting with a resident L will yield, as in (6)b (since utilities are shared), <S's utility, L's utility> = <.75, .75>.

[9] While it is usually assumed that call production and call comprehension involve the same cognitive mechanisms, this need not be a given, as there is a striking asymmetry between the two (Seyfarth and Cheney 2010). Call production in primates is highly restricted and in part innately specified. By contrast, call comprehension is far more flexible and gives rise to learning, as suggested by the fact Dianas understand Campbell's calls they can't produce and conversely, as discussed in Section 3.1. Thus it could be that call comprehension involves different cognitive mechanisms from call production.

P(w | <a, loud>) = 1
P(w' | <g, soft>) = 1

There are two salient explanations for the emergence of Featural Interpretation. The simplest one is to assume that it is cognitively easier (less costly) to interpret a call by zeroing in on a single feature than by attending to the entire call. An alternative possibility (possibly a more convincing one) is that a feature-based interpretation for *<a, loud>* in terms of *loud* alone yields a greater benefit because it makes it possible to understand novel calls that are themselves the result of Marlerian convergence.

In order to make things more concrete with a numerical example, we must define the utility derived by the receiver. We will continue to assume that there are just two situations, the alarm situation w (with probability P(w)) and the greeting situation w' (with probability 1-P(w)), and that these exhaust the possibilities. The receiver's utility U_r will thus be a weighted sum of the utilities it derives in w and in w', as stated in (8). We assume that alarm situations are rare and thus that $P(w) < .5$.[10]

(8)　The receiver's utility U_r is the sum of the receiver's utility in w (= the alarm situation) and the receiver's utility in w' (= the greeting situation) weighted by the probability of these situations:
$U_r = P(w) U_r(w) + (1-P(w)) U_r(w')$
with $P(w) < .5$

All the action will be in the alarm situation: non-alarm situations will remain constant (and greetings won't play a role). Thus the total receiver utility U_r will be an increasing function of the receiver's utility $U_r(w)$ derived in alarm situations, with non-alarm situations remaining fixed; we thus write $U_r = f(U_r(w))$, as specified in (9).

(9)　Notation
For notational convenience, we will define $f(x) = P(w) x + (1-P(w)) U_r(w')$. With this notation,
$U_r = f(U_r(w))$
We note that f is an increasing function of its argument (the receiver's global utility increases as its utility in the alarm situation w increases).

One important property of the function f is that the average of f(x) and f(y) is just f(.5x + .5y) (see (32) of Appendix I for details about a more general case):

(10)　$.5f(x) + .5f(y) = f(.5x + .5y)$

This will come in handy when we average the utility obtained by a resident and by a mutant.

❑ *Emergence of Featural Interpretation through interpretive cost*

The simplest line of analysis (maybe not the most convincing one) is to postulate that Featural Interpretation is cognitively easier than call-based interpretation. In other words, it is more economical to interpret <a, loud> by just concentrating on the loudness component than by considering the entire call.

To be concrete, we compare a resident population that relies on call-based interpretation, and mutants that use Featural Interpretation. Residents and mutants derive the very same information from the calls they hear, but mutants do so at lesser cost. As a result, a mutant interacting with a resident produces greater (shared) utility than a resident interacting with a resident. This means that Maynard Smith's Condition (1) is violated and that the mutant strategy invades the resident strategy.

This reasoning is outlined in a very simplified case in (11), with resulting utilities in (12). (Encounters are still symmetric: when two individuals <x, y> interact, in half the cases x is the signaler and y the receiver, and in the other half it's the other way around.) To describe this case in words: we assume that alarm situations always give rise to an alarm call, interpreted in the same way by the resident, which uses call-based interpretation, and by the mutant, which uses feature-based interpretation. As a result, the resident and the mutant alike react appropriately in alarm situations. But the mutant does so at lesser cognitive cost, which gives it an edge over the resident.

---

[10] This will matter below when we assume that an alarm-appropriate reaction is worth adopting only in case the probability of the threat is at least .5. If alarm situations had a probably of .5 or more, this would justify adopting an alarm-appropriate reaction irrespective of whether one has received an alarm signal or not, thus trivializing the problem.

(11) Receiver's utility in an alarm situation w, with the assumption that a conspecific produces <a, loud> with probability 1, i.e. $P(<a, loud> | w) = 1$

Resident (c̲all-based) interpretive strategy:
$P_C(w | <a, loud>) = 1$

Mutant (f̲eature-based) interpretive strategy (= identical to the Resident's interpretive strategy):
$P_F(w | <a, loud>) = 1$

We now assume that there is a cost of $c_C = .2$ to use the call-based strategy and a cost of $c_F = .1$ to use the feature-based strategy. Thus for any receiver i,

$$U_{r\_i}(w \text{ and } <a, loud>) \quad = 1-c_i \text{ iff } P_i(w | <a, loud>) \geq .5;$$
$$= -c_i \text{ otherwise.}$$

Since $P_C(w | <a, loud>) = P_F(w | <a, loud>) = 1$,
$U_{r\_C}(w \text{ and } <a, loud>) = 1-c_C = 1 - .2 = .8$
$U_{r\_F}(w \text{ and } <a, loud>) = 1-c_F = 1 - .1 = .9$

(12) a. Resident C_1 interacting with resident C_2
$U_{r\_C}(w) \quad = P(<a, loud> | w) \, U_{r\_C}(w \text{ and } <a, loud>)$
$\qquad\qquad = U_{r\_C}(w \text{ and } <a, loud>) \qquad\qquad$ (since $P(<a, loud> | w) = 1$)
$\qquad\qquad = .8$

hence

$<C_1\text{'s utility}, C_2\text{'s utility}>$
$= <f(U_{r\_C}(w)), f(U_{r\_C}(w))>$
$= <f(.8), f(.8)>$

b. Mutant F interacting with resident C
$U_{r\_F}(w) \quad = P(<a, loud> | w) \, U_{r\_F}(w \text{ and } <a, loud>)$
$\qquad\qquad = U_{r\_F}(w \text{ and } <a, loud>) \qquad\qquad$ (since $P(<a, loud> | w) = 1$)
$\qquad\qquad = .9$

hence

$<F\text{'s utility}, C\text{'s utility}>$
$= <.5f(U_{r\_C}(w))+.5f(U_{r\_F}(w))), 5f(U_{r\_C}(w))+.5f(U_{r\_F}(w)))>$
$= <f(.5U_{r\_C}(w)+.5U_{r\_F}(w)), f(.5U_{r\_C}(w)+.5U_{r\_F}(w))> \qquad$ (by using the remark in (10))
$= <f(.85), f(.85)>$

Since f is an increasing function, $f(.85) > f(.8)$, and Maynard Smith's Condition (1) is violated: the resident call-based strategy will be invaded by the mutant's feature-based strategy, purely for reasons of cost.

Conversely, it can be checked that the feature-based strategy will be evolutionarily stable relative to the call-based strategy. If the resident relies on feature-based interpretation, a resident interacting with a resident will produce greater (shared) utility and a mutant interacting with a resident, since now the mutant uses the more costly call-based interpretation.[11]

---

[11] The computations in (i) show that Maynard Smith's Condition (1) will now be satisfied.
(i) a. Resident F_1 interacting with resident F_2
$\quad U_{r\_F}(w) = .9 \qquad$ (using the value of $U_{r\_F}(w)$ from (12)a)
$\quad <F_1\text{'s utility}, F_2\text{'s utility}> = <f(U_{r\_F}(w)), f(U_{r\_F}(w))> = <f(.9), f(.9)>$

b. Mutant C interacting with resident F
We note that this is the same computation we already performed in (12)b:
$\quad <C\text{'s utility}, F\text{'s utility}> = <f(.85), f(.85)>$

❑ *Emergence of Featural Interpretation through productivity*

The second salient possibility is that Featural Intepretation emerges because it makes it possible to interpret diverse and novel heterospecific calls that are themselves the result of Marlerian convergence, so that a designated feature is regularly associated with a fixed function.

To keep things maximally simple, we will restrict attention to the interpretive rule for <a, loud>, and we will assume that it does not have repercussions for the interpretation of <g, soft> (greetings will play no role here). So everything will hinge on the maximization the receiver's utility in the alarm world w, namely $U_r(w)$ (since as noted above, the receiver's total utility U_r is an increasing function of its utility in w: $U_r = f(U_r(w))$.

To see the simplest possible case of evolution of Featural Interpretation, we will assume that just two things may happen in the alarm situation w. Either (i) a conspecific produces an alarm call <a, loud>, with probability 1/2, or (ii) a heterospecific neighbor produces an alarm call, with probability 1/2 as well. Heterospecific calls may take a variety of forms <blah, loud> (for different values of *blah*), with the only constant being that all these calls are produced loudly. With respect to (i), residents and mutants have the same interpretive strategy: they correctly interpret the alarm call. It is only with respect to (ii) that they differ: residents cannot interpret <blah, loud> and thus do not gain information from it, whereas mutants, who use Featural Interpretation, interpret the loudness component as indicating alarm.

The numerical details are a bit tedious and can be found in Appendix I. The key result is that a mutant with feature-based interpretation produces more (shared) utility when interacting with a resident with call-based interpretation than a resident does, simply because the mutant understands heterospecific calls but the resident doesn't. This means that the call-based interpretive strategy violates Maynard Smith's Condition (1) and is not evolutionarily stable: it will get invaded by the feature-based interpretive strategy.

3.4 Stepping back

The first degree of interspecies comprehension, Understand Thy Neighbor, is unsurprising. The second degree, based on Call Convergence, is a close evolutionary counterpart of the maximization behavior that leads speakers to pick the acoustic features that optimize call transmission. Convergent evolution in the calls of phylogenetically and geographically distant species can be combined with general similarity metrics across calls to yield interspecies comprehension in the absence of any prior contact.

The third degree of interspecies comprehension is based on an interpretation of individual features. To derive Featural Interpretation in the human case, we assumed that the addressee could reason on the speaker's maximization behavior. Such reasoning is unlikely to be the source of Featural Interpretation in animals, and thus we looked for reasoning-free evolutionary paths that might yield the same result. One rationale is based on the assumption that Featural Interpretation is sometimes less costly than whole call interpretation. Importantly, on this view, one needn't expect that the features that are selected by convergent evolution are also precisely those that give rise to interspecies comprehension; at the very least, a separate argument would be needed to establish this point.

By contrast, an alternative rationale, based on productivity, leads to the expectation that those acoustic features that are selected by evolution in the second degree of interspecies comprehension may be selected as targets of Featural Interpretation in the relevant species in case there are enough neighbors that converged on the same feature. This is certainly not sufficient: these neighbors' calls must be sufficiently diverse that Featural Interpretation is more economical than wholesale acquisition of the neighbors' calls (by an application of Understand Thy Neighbor).

Be that as it may, the third degree of interspecies comprehension is derived in ways that seem unrelated to the addressee's reflection on the speaker's maximization problem in our simplified RSA-like framework.[12]

---

[12] This raises a more general question for future research: under what conditions can some levels of the reasoning recursion in the RSA model be emulated by evolutionary processes?

4   The Empirical Landscape

Having distinguished among three levels of interspecies comprehension, we illustrate their relevance with data from the field of animal communication. *Understand Thy Neighbor* and *Call Convergence* are uncontroversial, and correspondingly discussed briefly below. By contrast, we lay out several possible cases of *Featural Interpretation*, but also lay out weaknesses in the arguments. For all three mechanisms, a detailed literature review can be found in the Supplementary Materials.

Interspecies comprehension is by no means an uncommon phenomenon. In a remarkable survey, Magrath et al. (2020) note that "over 70 species have been shown through experimental playback to respond to other species 'alarm calls", and that "there are even a few examples of lizards responding to bird alarm calls (Vitousek et al. 2007; Ito and Mori 2010; Ito et al. 2013), showing that individuals do not need to have their own alarm calls, or even to produce any vocalizations, to be responsive to the calls of other species". Magrath et al. also sketch different mechanisms by which interspecies comprehension might arise; some calls require learning to be recognized, while others are decoded on first exposure, plausibly because individual features are decoded. As Magrath et al. note, there is probably a trade-off: "Reacting to acoustic features alone allows for response to novel alarm calls on first exposure, but restricts the response to calls with a narrow range of characteristics (Hollén and Radford 2009). By contrast, learning allows individuals to gain information from alarm calls with a broad range of acoustic features, but leaves individuals more vulnerable to predators until they have learned to recognize the calls (Griffin 2004).".

4.1   *Understand Thy Neighbor*

We mentioned in Section 3.1 the case of Campbells' monkeys and Diana monkeys, which live in the same trees and appear to understand each other's calls—not to mention the hornbill, a bird that discriminates among the Diana eagle-related call and the Diana leopard-related call. In principle, learning need not be the only reason there is cross-species understanding in these cases, since call convergence and Featural Interpretation might play a role as well. But Magrath et al. 2020 provide examples in which learning seems to be crucial:

Studies of both geographic and temporal patterns of response to heterospecific alarm calls provide indirect but compelling evidence for learning (…). For example, bonnet macaques, Macaca radiata, respond to playback of alarm calls of only the locally common of two species of langur at any one site (Ramakrishnan and Coss 2000; Fig. 12.3); Himalayan bird communities respond more strongly to familiar than to unfamiliar heterospecific alarm calls, regardless of acoustic similarity (Wheatcroft and Price 2013); and fledgling white-browed scrubwrens, respond to aerial alarm calls of New Holland honeyeaters at a younger age where their territories overlap (Haff and Magrath 2013).

Strikingly, Shriner 1999 directly explores the learning mechanism itself by showing that individual Golden-mantled ground squirrels learn to associate anti-predator responses to a synthetic tone after being exposed to it simultaneously with a predator model.

We provide sample cases in (13) from our own literature review, excluding cases in which a bird species learns the calls of another bird species.

(13)   Sample cases in which the calls of a species (Species 1) are learned by another species (Species 2), with specific arguments that learning is at stake[13]

Species 1 (= Emitter)	Species 2 (= Receiver)	Call type	Call function
Superb starling	Vervet monkey	Alarm call	Predator response

---

[13] Picture credits appear at the end of this piece.

Malagasy paradise flycatcher	Madagascan spiny-tailed iguana	Alarm call & song	Predator response & mate attraction
Malagasy paradise flycatcher	Giant day gecko	Alarm call & song	Predator response & mate attraction
Nilgiri langur, Hanuman langur, and sambar deer	Bonnet macaque	Alarm call	Predator response
Crested coua and Madagascar magpie-robin	Solitary Sahamalaza Sportive Lemur	Alarm call	Predator response
Galapagos mockingbird	Galapagos marine iguana	Alarm calls & song	Predator response & mate attraction
Chimpanzee	Diana monkey	Alarm call	Predator response

In our literature review, we find cases of heterospecific comprehension in diverse taxa (birds, mammals, reptiles), including cases of inter-comprehension across taxa (e.g. Ito & Mori, 2010, Ito et al., 2013). Several non-exclusive mechanisms seem to be at play. It is generally accepted that generalization from species-specific calls sometimes plays a role in reaction to heterospecific calls (Fallow et al., 2011). However, in some cases, evidence suggests that similarity is not the only driver of subjects' reaction, as some species react to the acoustically dissimilar calls of heterospecifics that live in the same area (i.e. sympatric heterospecifics; Hauser, 1988, cited by Magrath 2014, Magrath, 2009, Vitousek et al., 2007, Wheatcroft & Price, 2013). This suggests that other factors may be at play. In several studies, the role of learning is supported by the observation that only individuals living in an area occupied by the target species display appropriate reaction to its calls (Hauser, 1988 cited by Magrath 2014, Magrath & Bennett 2011, Pollock et al., 2017, Ramakrishnan & Coss, 2000, Zuberbühler, 2000). In addition, the role of learning has also been evidenced in experiments showing that animals can learn to associate a heterospecific call (or a synthetic sound) with an external stimulus e.g. the presence of a predator (Dutour et al., 2019, Keen et al., 2019, Magrath & al., 2015, Potvin et al., 2018, Shriner, 1999).

The Supplementary Materials include more detailed information and additional cases, and when relevant reference to arguments that learning is genuinely involved.

4.2   Call Convergence

Following the pioneering work of Peter Marler, the existence of call convergence has been a guiding principle of work on animal communication. Magrath et al. 2020 summarize the historical situation as follows:

In a classic paper on signal design, Peter Marler proposed that the acoustic structure of alarm calls should reflect their function (Marler 1955). Passerine alarm calls given to flying hawks should be difficult to locate, to reduce the risk of attack, so Marler predicted that they should be high-pitched, pure tones, with a gradual onset and ending. By contrast, mobbing calls given to perched raptors should be easily locatable so that others can join in mobbing. They should therefore be abrupt, repetitive, and more broadband than calls given to flying hawks. Consistent with these predictions, many European passerines have "seeet" aerial alarm calls (…) and mobbing calls with exactly these acoustic features (Marler 1955, 1957).

As Magrath et al. further point out, subsequent research supported Marler's insights, in particular by way of playback experiments.

Sample cases from our own literature review appear in (14). We focused on pairs of species that have calls with related functions and related acoustic structures, but are phylogenetically very distant—hence good candidates for Marlerian convergence.

(14)   Sample cases in which the calls two phylogenetically remote species have calls with related functions and related acoustic structures

Species	*Call type*	*Type of acoustic structure*	*Function*	*Habitat*
–Little Penguin –Short-tailed shearwater	Territorial call	Succession of gaps in amplitude and frequency, with a high degree of redundancy	Marking of territory	Burrow nests
–12 species of passerines –3 species of Parrots	Begging call	Sound pressure level, wide band, low dominant frequencies	Feeding by parents	Captive
–Marmoset monkey –Several bird species	Alarm call	Fairly high frequency, long, drawn-out "seet". Gradual beginning and end.	Predator response	Trees
–Large-billed leaf warbler –Frogs of the genus Rana	Song	Short sequences, narrow frequency band, rhythmically separated	Attracting partners	Near torrents

In our review of the literature, we find that call convergence has been described in calls with very distinct functions—e.g. alarm calls, Jurisevic & Sanderson, 1994; begging calls, Jurisevic, 2003 or distress calls, Russ et al., 2003. In some cases, the species whose calls converged share the same habitat (e.g. Jouventin & Aubin, 2000), which may play an important role in shaping the acoustic structure of vocal signals, notably due to background noise or propagation constraints (Jouventin & Aubin, 2000). However, other studies uncover cases of convergence in species whose habitats (or micro-habitats) differ (e.g. Jurisevic, 2003, Vencl, 1977), which suggests that environment is not the only factor at play. Further work expanding our understanding of this phenomenon in more taxa with diverse habitats, social structures and ecological niches will be necessary to clarify the mechanisms underlying call convergence. (Here too, additional details and examples appear in the Supplementary Materials.)

4.3 Featural Interpretation?

While there several discussions of featural interpretation (or 'decoding') in the literature, it it is usually difficult to determine whether understanding of novel calls is due to feature decoding *per se*, or to global acoustic similarity to familiar calls. The best we can do is to mention potential cases of feature decoding from the literature; we will sketch the form of crucial arguments (and alternatives) in Section 5.

To start with a general view, Magrath et al. 2020 mention several cases in which featural interpretation might be involved.

The alarm calls of many species contain acoustic features that are inherently attention-grabbing or frightening, and this may also prompt immediate response to unfamiliar calls (Hirth and McCullough 1977; Morton 1977; Fitch et al. 2002; Rendall et al. 2009). Many mobbing, distress, "panic," or general alarm calls, for example, contain elements that are harsh, abrupt, broadband, or nonlinear. These acoustic features tend to be honest signals of caller distress, increase listener attention and response, and prevent habituation (Manser 2001; Blumstein and Récapet 2009; Townsend and Manser 2011; Blumstein and Chi 2012; Blesdoe and Blumstein 2014; Karp et al. 2014). For example, great-tailed grackles, Quiscalus mexicanus, increase vigilance after playback of novel synthetic calls that include nonlinear features such as abrupt transitions and noise (Slaughter et al. 2013), and western gray kangaroos, Macropus fuliginosus, do not habituate to novel sounds with nonlinear features (Biedenweg et al. 2011).

Two concrete cases are particularly enlightening. Aubin and Brémond 1989 focused on bird distress calls, which are known to evoke interspecific reactions. They proposed that this "results from the use of similar laws of decoding by the species concerned". Specifically, "a simple slope applied to a carrier frequency that corresponds to the acoustic shape of a distress call is sufficient to confer a distress meaning to the signal". Their argument was based on herring gull and starling distress calls. They observed that in both species, these calls involve (among others) (i) a fundamental corresponding to the carrier frequency (CF) and numerous harmonics, (ii) a slow frequency modulation which is "usually descending for the gull and always ascending for the starling", as illustrated in (15).

(15) Distress calls of herring gull (a) and of starling (b) (Aubin and Brémond 1989)
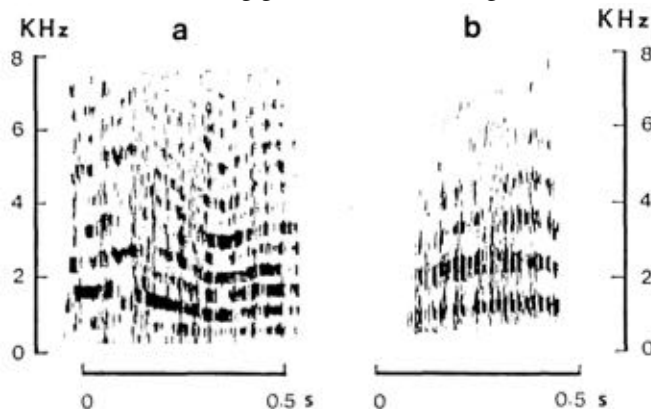
From these natural calls, Aubin and Brémond created synthetic signals without the slope, with a constant fundamental corresponding to what is found in each species, as is illustrated in (16), Signals 1a and 1b. These control signals did not elicit distress-related reactions. By contrast, when these synthetic signals were further manipulated to include a simple ascending or descending slope, reactions were regained: either slope elicited a response in gulls, whereas only an ascending slope elicited a reaction in starlings (counterparts of Signals 1a and 1b with an ascending slope added appear as Signals 2a and 2b in (16)).

(16) Synthetic signals used by Aubin and Brémond 1989
Signal la corresponds with the carrier frequency characteristics of a gull call; signal lb those of a starling call.
Signals 2a and 2b have a slow increasing frequency modulation. 2a was played to gulls; 2b was played to startlings
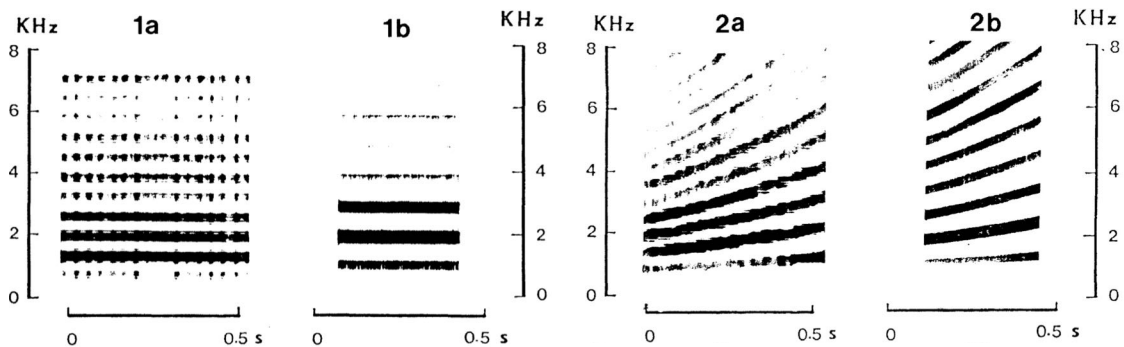
Aubin 1991 went one step further and built on these ideas to synthetize a general distress call that elicits responses in five species of birds (*Larus argentatus*, *L. ridibundus*, *Vanellus vanellus*, *Corvus frugilegus* and *Sturnus vulgaris*). Aubin's synthetic call crucially involved an ascending slope, as seen in (17), which led him to conclude that feature decoding is at stake: "A simple slope (a slow increasing or decreasing frequency modulation) applied to a carrier frequency that follows the acoustic shape of a natural distress call is sufficient to confer to the signal a distress meaning."
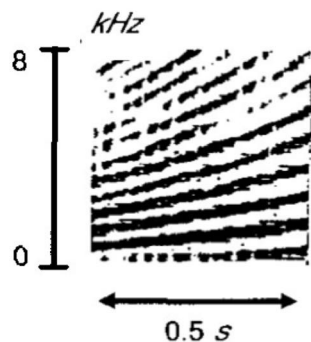
(17) Synthetic distress call used to elicit respnses in five different bird species (Aubin 1991)

As exciting as they may be, these results do not fully disambiguate the underlying mechanism. While feature decoding might be involved, an alternative is that the synthetic distress calls are effective because they globally resemble the natural distress calls of the relevant species.

This problem is arguably solved in a rather different experiment by Blumstein and Récapet 2009. In line with Magrath et al.'s (2020) remark that some features "increase listener attention and response, and prevent habituation", the authors show that a certain acoustic feature added to a natural call changes behavior. As illustrated in (18), Fig. 1, the authors start from a natural marmot alarm call (1st call), and they create two synthetic calls from it: a target call with white noise inserted in the middle of the call (2nd call), and a control call with silence in the middle of the call (3rd call). As can be seen in (18), Fig. 2, the natural call and the control synthetic call had comparable alarm effects in terms of reduction of time devoted to foraging. By contrast, the target call with white noise (non-linearities) had a *greater* alarm effect, suggesting that the acoustic feature added was effective on its own.

(18) Normal, synthetic and control calls used by Blumstein and Récapet 2009 (Fig. 1), with results in terms of foraging behavior (Fig. 2)
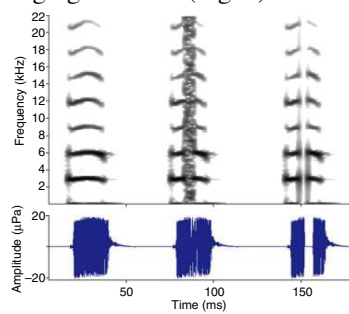

Fig. 1: Spectrograms of the three calls used to construct the playback tracks: normal alarm call (call 1), call including random noise (call 2), call including silence (call 3).
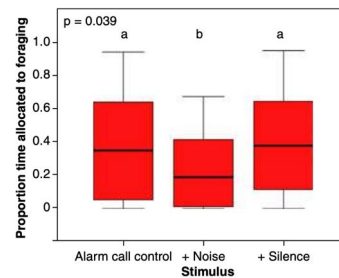

Fig. 2: Response of yellow-bellied marmots in proportion of time spent foraging to three different playbacks (normal alarm calls, calls including noise and calls including silence). Identical letters above box plots (which illustrate the median, inter-quartile range, and 95% confidence intervals) illustrate non-significant (p > 0.05) differences.

Since the target synthetic call triggers a different behavior than both the natural call and the control call (namely a stronger reduction in foraging behavior), it is clear that mere similarity to a known call isn't sufficient to explain the data. This particular objection to Aubin and Brémond 1989 and Aubin 1991 has no force in the present case: it is clear that the inserted white noise is effective on its own. But why is this so? One possibility is that non-linearities are, on their own, interpreted as an alarm signal. As emphasized by Blumstein and Récapet, this would make good evolutionary sense and is compatible with observations in diverse species (involving meerkat alarm calls, some vervet monkey calls, and highly aroused versions of chimpanzee pant-hoots, piglet calls, Japanese macaque calls and domestic dog calls).

But another possibility is that non-linearities just make it harder to habituate to the signal because they make it more unpleasant and irregular. This hypothesis predicts that the addition of non-linearities might decrease habituation in non-alarm calls just like they do in alarm calls. On this view, there is indeed a correlation between the presence of an acoustic feature (non-linearities) and a behavior, but this doesn't show that the acoustic feature has a proper semantics.

Stepping back, both the "universal" bird distress call and the strengthened marmot alarm call are highly evocative of a featural semantics, but they do not prove its existence. In the case of distress calls, decoding of unfamiliar calls may be due to acoustic similarity to familiar calls. In the case of non-linearities added to marmot calls, the acoustic feature on its own has an effect, but it could be because of its physiological effects rather than through feature decoding.

It is clear that further empirical support will be needed to establish the reality of featural interpretation. We conducted a literature review and found several evocative examples, including some listed in (19).

(19) Samples cases of potential Featural Interpretation in animal communication
Details can be found the Supplementary Materials by looking for the species name and line number in the relevant table.

Candidate Feature	Candidate Meaning	Species
amount of noisiness/non-linearities	greater noisiness → greater arousal, greater distress, less affiliative	Multiple species, African elephant, Red junglefowl, Meerkats, Red wolf
frequency (fundamental F0, mean frequency, peak frequency)	higher F0 → greater arousal, more positive emotion, more dangerous,	Multiple species, Bottlenose dolphins, Campbell's monkey, Greater false vampire bat, Red-capped mangabey, Red junglefowl, Weddel seal
call rate	greater call rate → greater arousal/urgency	Multiple species, Weddel seal, Bottlenose dolphin, Humans, Meerkats Red junglefowl
call duration	longer call → greater negativity, danger/urgency, arousal	African elephant, De Brazza monkey, Red-capped mangabey, Campbell's monkey, Weddel seal
call number	more calls → greater arousal/urgency	Multiple species, Bottlenose dolphin, Greater false vampire bat, White-browed scrubwrens

In our literature review, several trends are observable through distinct taxa. Some of the most striking pertain to temporal parameters: an increase in urgency or arousal is generally accompanied by an increase in call duration (although one contradictory observation exists in meerkats alarm calls). In high urgency and arousal contexts, inter-element intervals in call sequences (and multi-element calls) decrease while the number of elements in the signals increases. Interestingly, call rate increases when the emitter's arousal is high, in both positive (e.g. reunion) and negative (e.g. alarm) contexts. Pitch-related parameters also reflect the caller' emotional state in distinct taxa as both parameters relating to the fundamental frequency (e.g. F0, F0max), to formants' frequency and to the overall distribution of energy in the frequency spectrum increase when arousal is high (here again in both positive and negative contexts). Finally, an increase in non-linear phenomena (e.g. frequency jumps) can signal increased arousal while a decrease in pitch saliency is associated with high arousal and/or negative emotions. (Additional details and further examples can be found in the Supplementary Materials.)

Still, for the most part our examples are just suggestive, and involve important difficulties. First, as can be seen in (19), there is some overlap among the acoustic features (e.g. all other things being equal, greater call rate co-occurs with larger number of calls). Second, the description of their effects is currently extremely coarse-grained and gives rise to considerable overlap as well; for instance, all the features listed are associated with greater arousal when their acoustic value increases. Even more importantly, while the similarities found across species makes it likely that these are instances of Marlerian convergence, there are few arguments that they genuinely involve Featural Interpretation; what is currently missing is a hard proof.

5 Featural Interpretation: Arguments and Alternatives

Among our three degrees of interspecies comprehension, Understand Thy Neighbor and Call Convergence are unsurprising and uncontroversial. By contrast, Featural Interpretation is a new interpretive mechanism that requires empirical and theoretical justification. On an empirical level, the data we surveyed from the literature are fascinating but do not clinch the case, and thus we will sketch crucial experiments that might help decide the issue in future research. On a theoretical level, we have written as if there were just two options to interpret novel calls: global similarity to familiar calls, and Featural Interpretation. But similarity metrics can be tweaked so as to give outsized importance to a certain feature, in which case global similarity and Featural Interpretation might converge. However this only holds true when we consider a single feature present in the target call. When combinations of features are considered, Featural Interpretation predicts a productivity that should make it possible to tell animals things they've never been told, in such a way that they understand upon first exposure. This tantalizing possibility is a striking prediction of Featural Interpretation.

5.1 Featural Interpretation I: triggering familiar behaviors with unfamiliar calls

How could one develop a decisive argument for Featural Interpretation? Aubin and Brémond 1989 and Aubin 1991 showed that target birds could interpret novel calls with some crucial acoustic features. The logic of the argument is to show that a familiar behavior (appropriate to a distress call) can be triggered with a call that the target birds have never heard before. But the authors did not disentangle what was due to the feature in question and what was due to global similarity. To disambiguate the findings, one needs to pit the presence of the crucial feature against global similarity.

One possibility to do so is to construct an artificial signal that carries the target feature—for instance, a rising slope—but is otherwise completely different from the signals that the target animals are familiar with. To be concrete, suppose that species x has a call C that can be characterized by 10 feature values, $<F_1, F_2, \ldots, F_{10}>$, with F_1 (for instance, a rising slope) the target feature that is hypothesized to be interpreted. The idea is to create an artificial call C^* that is like C with respect to F_1, but unlike C with respect to the remaining 9 features. Writing $-F$ for the destruction/inversion of a feature F, C^* is thus defined by $<F_1, -F_2, \ldots, -F_{10}>$. On the assumption that reasonable measures of global similarity yield the result that C^* is very dissimilar from C, and also from any other call with a similar function that species x might be familiar with, the key question is whether C^* will trigger the target behavior. If so, we will have the beginning of an argument that Featural Interpretation rather than

global acoustic similarity is at work. One might also need a control condition in which one investigates a call C^{**} produced by destroying F_1 while keeping all other features—in other words, C^{**} is defined by $<-F_1, F_2, \ldots, F_{10}>$. One would predict that despite its overall similarity to C, C^{**} should *fail* to trigger the target behavior due to the absence of the crucial feature F_1.

5.2   *More fine-grained versions of Similarity-based Interpretation*

Our analysis might be too simple-minded, however. We have written as if one only has a choice between standard semantic interpretation on the one hand (combined with a global acoustic similarity metric), and Featural Interpretation on the other. But it could be that semantic interpretation is call-based and standard, but that the acoustic similarity metric gives some features an oversized role.

To be concrete, consider an example in which calls are defined by 4 binary features $<F_1, F_2, F_3, F_4>$ (with the features in $\{0, 1\}$), and suppose that the base call is defined by $C = <1, 1, 1, 1>$. Let us start from the mechanism of Featural Interpretation specified in (20): a call C is interpreted as specified by feature F_1 just in case the first coordinate of C is 1.

(20)   An example of Featural Interpretation
Suppose a call C is defined defined by $<i_1, i_2, i_3, i_4>$, where for $i_1, \ldots, i_4$ are binary features in $\{0, 1\}$. If F_1, corresponding to the first coordinate, is the only interpretable feature, C is interpreted as F_1 if and only if $i_1 = 1$.

Now we consider one target call C and two synthetic calls C^* and C^{**}, defined as in (21). They correspond to the case discussed in the preceding section: C^* shares with C the feature F_1 but is unlike C with respect to the features F_2, F_3, F_4; by contrast, C^{**} is similar to C with respect to F_2, F_3, F_4, but it is similar to C with respect to F_1. Based on Featural Interpretation of F_1, we predict that C^* will trigger the same behavior as C while C^{**} won't, despite the fact that in terms of overall similarity, C^{**} is more similar to C than C^* is.

(21)   C = $<1, 1, 1, 1>$
$C^* = <1, 0, 0, 0>$
$C^{**} = <0, 1, 1, 1>$

Now C^{**} is indeed closer to C than C^* is if one adopts a simple-minded similarity metric, such as this one: to assess the similarity of two calls, take the number of coordinates that they have in common, and divide by the total number of coordinates.[14] With this rule, stated in (34)a, one obtains for C^* and C^{**} the similarities to C given in (34)b,c: C^* has a similarity of .25 with C because the tuple $<1, 0, 0, 0>$ has 1/4th of its coordinates in common with the tuple $<1, 1, 1, 1>$; C* has as similarity of .75 with C because $<0, 1, 1, 1>$ has 3/4th of its coordinates in common with $<1, 1, 1, 1>$.

Notation: If X is a tuple of coordinates, X_i is its ith coordinate (thus C_{1i} is the ith coordinate of C_1).

(22)   If C_1 and C_2 are each defined by 4 binary features in $\{0, 1\}$:
a. $Sim(C_1, C_2)$
= number of coordinates that C_1 and C_2 have in common / total number of coordinates
= $\Sigma_{1 \leq i \leq 4} (1-|C_{1i} - C_{2i}|)/4$
b. $Sim(C, C^*) = 1/4 = .25$
c. $Sim(C, C^{**}) = 3/4 = .75$

But there are further general similarity metrics one could consider. In particular, the target animal's perceptual system might give outsized importance to some features over others. One way to implement this is to assume that the similarity metric gives different weights to different coordinates, as is specified in (23): the similarity between two calls (viewed as tuples) is given by the proportion of features they have in common, but now weighted by different parameters λ_i for different features F_i.

---

[14] See for instance Choi et al. 2010 for a survey of similarity measures over binary vectors. In Choi et al.'s typology, the one we use is the Sokal and Michener measure.

(23) For some family of parameters $(\lambda_i)_{1 \leq i \leq 4}$ such that $\Sigma_{1 \leq i \leq 4} \lambda_i = 1$, if C_1 and C_2 are each defined by 4 binary features in $\{0, 1\}$:

$Sim_\lambda(C_1, C_2) = \Sigma_{1 \leq i \leq 4} \lambda_i(1-|C^1_i - C^2_i|)$

When all the weights are set to 1/4, we obtain the same result as with our earlier similarity metric: only the (unweighted) proportion of features that two calls have in common matters, as stated in (24). But when λ_1 has value 1 while all other weights have value 0, only the first coordinate will matter. As a result, C^* comes out as maximally similar to C, and C^{**} maximally dissimilar to C, as is illustrated in (25).

(24) If for each $i \in \{1, 2, 3, 4\}$, $\lambda_i = 1/4$, we have:
$Sim_\lambda(C_1, C_2) = Sim(C_1, C_2)$

(25) With $\lambda_1 = 1$ and for $i \neq 1$, $\lambda_i = 0$, we have:
a. $Sim_\lambda(C, C) = 1$
b. $Sim_\lambda(C, C^*) = 1$
c. $Sim_\lambda(C, C^{**}) = 0$

In effect, this metric makes it possible to replicate within a similarity-based interpretation a featural interpretation of a single feature, at least in the case at hand.

This broader view of similarity metrics makes it possible to define intermediate cases as well, namely ones in which a feature takes outsized importance without fully obliterating the role of other features. For instance, when the first coordinate accounts for 70% of the weight and each of the other three coordinates each accounts for only 10% of the weight (a case displayed in (26)), we preserve the result that C^* is more similar to C than C^{**} is, thus explaining the role of feature F_1 in the interpretation. But we also obtain the result that C^* is not as similar to C as C itself is, contrary to what was the case in (25)a,b.

(26) With $\lambda_1 = .7$ and for $i \neq 1$, $\lambda_i = .1$, we have:
a. $Sim_\lambda(C, C) = 1$
a. $Sim_\lambda(C, C^*) = .7$
c. $Sim_\lambda(C, C^{**}) = .3$

This more discriminating analysis might have advantages. On a pure feature-based interpretation, we predict that heterospecific calls should be treated *exactly* as conspecific calls as long as they carry the crucial feature. But it is rather implausible that the natives do not distinguish between conspecific and heterospecific calls. Of course one could take this distinction to be non-semantic in nature (as when one recognizes somebody else's voice without thereby using different semantic rules to interpret what they say). But an alternative is that the similarity metric we just outlined is responsible for this behavior, since C^* is very similar to C, but is still not conflated with it.

Stepping back, we initially sketched a way of showing that Featural Interpretation rather than unadorned global acoustic similarity is responsible for the decoding of novel calls. This logic can serve to distinguish Featural Interpretation from simple-minded versions of global similarity, but not from more discriminating ones, which give outsized importance to some features over others. Furthermore, the latter might be at an advantage in case heterospecific calls are treated differently from conspecific calls even when their semantic content is decoded.

5.3   *Featural Interpretation II: triggering new behaviors with new calls*

Interestingly, there still remains an essential difference between Featural Interpretation and similarity-based interpretation. The latter predicts that a novel call may be interpreted, but only with the meaning of a familiar call. By contrast, Featural Interpretation predicts that features arranged in new ways in unfamiliar calls may give rise to new meanings because of 'feature compositionality', the property by which the interpretation of a call is determined by the interpretable features it contains. There are several cases to consider.

First, a novel call C might contain an interpretable feature F that is not present in any of the calls that are familiar to the target animal (whether in those of conspecifics or heterospecifics), and in

principle C might still be interpreted in accordance with the meaning of F. Obviously a similarity-based interpretation wouldn't help in this case.

Second, a novel call C might contain an interpretable feature F_1 that co-occurs with another interpretable feature F_2 whenever it appears in a familiar call. If a novel call C^* contains F_1 but not F_2 and no other interpretable features, we predict the appearance of a novel reaction that needn't correspond to any of the familiar calls.

Third, we could be in a situation in which every familiar call contains a single interpretable feature (different ones for different calls). But a novel call C^* might contain two interpretable features F_1 and F_2 (and no others), say with F_1 normally present in C_1 and F_2 in C_2. In this case too, Featural Interpretation and a similarity-based interpretation make different predictions. Featural Interpretation predicts that C should be interpreted as the conjunction of F_1 and F_2. A similarity-based interpretation predicts that C might be interpreted as one of $\{C_1, C_2\}$ depending on which one 'wins' according to the similarity metric.[15]

To make things concrete, let's assume once again that calls are defined by four binary features with values in $\{0, 1\}$, and let's assume that the familiar calls C_1, C_2 and the synthetic call C^* are defined as in (27).

(27) $C_1 = <1, 0, 0, 0>$
    $C_2 = <0, 1, 1, 0>$
    $C^* = <1, 1, 0, 0>$

On the assumption that positive features F_1 and F_2 and no others are interpreted, the predictions of Featural Interpretation are unambiguous: C_1 should be interpreted as F_1, C_2 as F_2, and C^* as the conjunction of F_1 and F_2; this is laid out in (28). For instance, if F_1 the non-linearity characteristic of alarm (as discussed by Blumstein and Récapet 2009), and F_2 is the slope characteristic of distress (in Aubin and Brémond 1989 and Aubin 1991), Featural Interpretation predicts that C^* should simultaneously convey distress and alarm, even though neither C_1 nor C_2 alone has such a function (C_1 alone should trigger alarm and C_2 should signal distress).

(28) Featural Interpretation for C^*, on the assumption that positive values of features F_1 and F_2 are interpreted (and no others)
    Interpretation of C_1: F_1
    Interpretation of C_2: F_2
    Interpretation of C^*: F_1 & F_2

By contrast, it is clear that, without further elaboration, a similarity-based interpretation will have to select either C_1 or C_2 as being most similar to C^*. In the special case of (27), with the unadorned metric we posited before, just based on the proportion of shared coordinates (as in (22), C^* will be interpreted in the same way as C_1 rather than as C_2, because C^* shares more features with C_1 (namely 3) than with C_2 (namely 2). If C_1 is an alarm call and C_2 is a distress call, the result will be an alarm signal, not a distress signal.

(29) With the similarity metric in (22),
    a. $\text{Sim}(C_1, C^*) = 3/4 = .75$
    b. $\text{Sim}(C_2, C^*) = 2/4 = .5$

One point remains to be discussed, however. What happens if the novel call C^* is equally distant from the known calls C_1 and C_2? It is unclear what is predicted by a similarity-based interpretation in this case. One possibility is that the target bird should fail to react. An alternative is that it should randomize between a C_1-appropriate and a C_2-appropriate behavior. In either case, a similarity-based interpretation will continue to make different predictions from Featural Interpretation, as the latter just predicts a conjunctive meaning. But there is a third conceivable possibility, namely that in the case of a tie between C_1 and C_2, the target bird just conjoins the meaning of C_1 and that of C_2. This would make

---

[15] The marmot experiment discussed in Blumstein and Récapet 2009 can be taken to roughly follow this general logic, with some complications. The authors start from an alarm call with non-linearities, which can be viewed as a feature F_1. By inserting white noise within natural calls, they introduce another version of F_1 so as to produce a novel meaning and behavior, in the sense the receivers are more alarmed by the two copies of F_1 than by a single copy.

the similarity-based analysis harder to distinguish from the featural analysis. This issue is discussed in greater detail in Appendix II.

6   Conclusion

6.1   Results

While the idea of feature-decoding is widespread in the literature on animal communication (especially birds), it is not always distinguished from the phenomenon of Marlerian convergence, the phenomenon whereby calls with similar functions tend to have the same acoustic features. We proposed to clarify the issue by distinguishing between three degrees of intercomprehension: Understand Thy Neighbor, Call Convergence and Featural Interpretation. In the interpersonal case, the counterpart of Call Convergence may be due to a rational process whereby the speaker uses optimal features (e.g. loudness) to express some words based on their function. Featural Interpretation may arise when the addressee reflects on the speaker's maximization problem and infers that certain acoustic features are associated with certain functions. Call Convergence in animals need not arise from rational optimization, but may arise from optimization through evolution. On the other hand, there is no obvious counterpart of the rational process by which Featural Interpretation might arise. There are two salient evolutionary mechanisms that might explain why Featural Interpretation does arise in nature: productivity (Featural Interpretation makes it possible to interpret novel calls, on the assumption that these already converged through a Marlerian process), and possibly cost (if interpretation of a feature is less costly than interpretation of an entire call). Obviously this is only the beginning of an investigation of possible mechanisms.

The literature contains numerous possible instances of feature decoding. But crucial experiments that adjudicate between Featural Interpretation and similarity-based interpretations have rarely, if ever, been conducted. We proposed two schematic experiments that might help decide the issue in future research. One is based on novel calls that trigger the same behavior as an existing call. The key is to pit the presence of the crucial feature against acoustic similarity to show that the former is crucial and the latter is not. The difficulty is that more fine-grained similarity metrics might allow certain similarity-based interpretations to 'imitate' Featural Interpretation in this case. A complementary direction is to create novel calls with combinations of features that are not simultaneously present in any familiar call. The prediction of Featural Interpretation is that these novel calls should be understood on first exposure, and should give rise to a novel behavior thanks to featural compositionality (the property whereby the meaning of a call is computed from the interpretable features it contains and their meanings).

While theoretical at this point, this raises a tantalizing possibility: thanks to featural compositionality, it should be possible to tell animals things they've never been told before, in such a way that they understand on first exposure.[16]

6.2   Open questions

Since our paper is entirely theoretical in nature, the most pressing issue is of course to perform the crucial experiments we described. But there are additional open issues.

First, we did not attempt to survey possible mechanisms by which Featural Interpretation may arise in evolution. We only mentioned that productivity (the ability to understand novel calls that converged through a Marlerian mechanism) might provide an explanation, as might differences in cognitive cost between feature-based and whole-call interpretation. Blumstein and Récapet 2009 evoked a separate mechanism, whereby some acoustic features may have a biological effect on the cognitive system of recipients (see also Magrath et al. 2020); in particular, non-linearities might block habituation. Relatedly, Arnal et al. 2015 argue that the acoustic property of 'roughness' is present in

---

[16] In principle, other compositional systems (non feature-based ones) might give rise to the same result. But systems of animal communication that have been formally analyzed mostly involve the concatenation of propositional meanings (e.g. Schlenker et al. 2016b, Schlenker et al., to appear), hence some limitations on the new meanings that can be produced.

alarm calls because of its physiological effects:[17] it "selectively activates amygdala, involved in danger processing" and "confers a behavioral advantage to react rapidly and efficiently"; we might expect that this gives rise to Featural Interpretation. Numerous further evolutionary mechanisms ought to be considered, possibly different ones for different acoustic features.

Second, we have assumed that when a call C contains two interpretable features F_1 and F_2 (and no others), C is interpreted as the conjunction of F_1 and F_2. In animal linguistics, the term 'trivial compositionality' has been used when two juxtaposed calls C and C' are interpreted as separate utterances, a procedure that yields the conjunction of their meanings (e.g. Schlenker et al. 2016c, Schlenker et al., to appear, Zuberbühler 2020). A conjunctive interpretation of features is in a sense trivial in the same sense, but with an important difference: in our case, F_1 and F_2 appear in the same call and thus cannot be analyzed as being separate utterances. Still, it would be important to ask whether some combinations of features might be interpreted by procedures different from conjunction. This would make good sense in some cases. For instance, a high calling rate applied to a raptor call might conceivably signal that the raptor threat is high, rather than something less specific, to the effect that there is a raptor alarm, and there is *something* very urgent. The more specific meaning might involve a non-conjunctive combination of calls. Whether such cases exist remains to be seen.[18]

Third, to our knowledge no specialist argues that all of animal communication is reducible to feature decoding. Rather, there seem to be two coexisting systems, one based on whole calls (with innate or learned meanings); and another, whose existence has yet to be proven, which is based on Featural Interpretation. While we have sketched mechanisms that could explain how Featural Interpretation can evolve from whole call interpretation, the division of labor between these two systems has yet to be investigated.

Fourth, for the sake of simplicity we took the action to be entirely on the comprehension side. We thus discussed things as if production rules could remain fixed while comprehension rules evolve. This is not inconceivable, as production is often far more rigid and constrained than comprehension is (as shown by the fact that monkey species typically understand calls they cannot produce). But it would make good sense to assume that changes in comprehension rules end up affecting production, which would give rise to far more complicated analyses that the tiny 'proof-of-concepts' models we discussed in this piece.

Finally, Featural Interpretation is strikingly reminiscent of a procedure proposed in recent formal analyses of musical meaning (e.g. Schlenker 2017, 2019, 2022, Migotti 2019, Migotti and Zaradzki 2019, Zaradzki 2021, Rodriguez 2021). The idea was that diverse musical features such as harmonic stability, melodic height, loudness, etc. each give rise to diverse inferential effects, which can be aggregated to yield a full-fledged (if highly underspecified) semantics for music. This connection between musical meaning and animal signals need not be as surprising as it seems. It is an old idea in music cognition that music recycles expressive mechanisms that are found in human voice as well as in animal signals (e.g. Cook 2007; Cross and Woodruff 2008; Bowling et al. 2010, 2012; Huron 2015; Ilie and Thompson 2006, Juslin and Laukka 2003).[19] In fact, following the spirit of Blumstein and Récapet 2009, but transported from marmot alarm calls to human music, Blumstein et al. 2012 show that adding non-linearities to music (noises) increases arousal and decreases valence. A key question for future research is whether it is in fact the same system of Featural Interpretation that underlies music semantics and some animal calls.

---

[17] As Arnal et al. 2015 write, "sounds in this region correspond to amplitude modulations ranging from 30 to 150 Hz and typically induce unpleasant, rough auditory percepts". Note that the fact that certain acoustic features have certain automatic cognitive effects would still have to explained; it might be that this is so *because* of Featural Interpretation, hence a possible 'chicken and egg' problem.

[18] Schlenker et al. 2014 analyzed high calling rate as a separate acoustic feature with a semantics that does not depend on the call that it modifies. But this was admittedly for reasons of formal simplicity, and the authors made no strong claim in this connection.

[19] One salient point of comparison pertains to feature-based encoding in voice and in music. In a review of the literature, Juslin and Laukka 2003 thus argue that emotions are encoded in part by the same acoustic cues in voice and in music. We should add that the relation between animal signals and emotions in music was the topic of a workshop at the music center IRCAM in Paris in 2016 (e.g. https://medias.ircam.fr/x70694c).

Photo credits for (13)

Appendix I. Emergence of Featural Interpretation Through Productivity

In a nutshell, within the first situation described in Section 3.3 (= hetereospecifics produce diverse alarm calls of the form <blah, loud>), a mutant using feature-based interpretation produces more (shared) utility when interacting with a resident with call-based interpretation than a resident does because the mutant understands heterospecific calls but the resident doesn't.

We reiterate in (30) and (31) our assumptions and our notations about receiver utility.

(30) The receiver's utility U_r is the sum of the receiver's utility in w (= the alarm situation) and the receiver's utility in w' (= the greeting situation), weighted by the probability of these situations:
$U_r = P(w) U_r(w) + (1-P(w)) U_r(w')$
with $P(w) < .5$

(31) Notation
For notational convenience, we will define $f(x) = P(w) x + (1-P(w)) U_r(w')$. With this notation,
$U_r = f(U_r(w))$
We note that f is an increasing function of its argument (the receiver's global utility increases as its utility in the alarm situation w increases).

The following is a more general form of the technical remark in (10) in the main text, where we took the value $a = .5$.

(32) Technical remark
Let r and r* be two receivers with utility functions $U_r(\bullet)$ and $U_{r*}(\bullet)$.
If $U_{r*}(w') = U_r(w')$, then $aU_r + (1-a)U_{r*} = af(U_r(w)) + (1-a)f(U_{r*}(w) = f(aU_r(w) + (1-a) U_{r*}(w))$.

$aU_r + (1-a)U_{r*}$
$= af(U_r(w)) + (1-a)f(U_{r*}(w)$
$= a[P(w) U_r(w) + (1-P(w)) U_r(w')] + (1-a)[P(w) U_{r*}(w) + (1-P(w)) U_{r*}(w')]$
$= P(w) [aU_r(w) + (1-a)U_{r*}] + a(1-P(w)) U_r(w') + (1-a)(1-P(w)) U_{r*}(w')$
$= P(w) [aU_r(w) + (1-a)U_{r*}] + a(1-P(w)) U_r(w') + (1-a)(1-P(w)) U_r(w')$ (since $U_{r*}(w') = U_r(w')$)
$= P(w) [aU_r(w) + (1-a) U_{r*}] + (1-P(w)) U_r(w')$
$= f(aU_r(w) + (1-a) U_{r*}(w))$

We consider only the special case described in (33).

(33) Receiver's utility in an alarm situation w, with the assumption that exacty one alarm call is always produced, with a 50% chance that the call is from a conspecific and a 50% chance that it is from a heterospecific:
$P(<a, loud> | w) = P(<blah, loud> | w) = .5$ (and these are mutually exclusive events)

We assume that a receiver i adopts an alarm-appropriate reaction only if i takes the probability of w to be $\geq .5$. In w, the utility obtained is 1 if an alarm-appropriate reaction is adopted and 0 otherwise, hence if m is a message,
$U_{r\_i}(w \text{ and } m) = 1$ iff $P_i(w | m) \geq .5$; $= 0$ otherwise.

Resident C's (<u>c</u>all-based) interpretive strategy:
$P_C(w | <a, loud>) = 1$
$P_C(w | <blah, loud>) = P(w) < .5$
As a result,
$U_{r\_C}(w \text{ and } <a, loud>) = 1$
$U_{r\_C}(w \text{ and } <blah, loud>) = 0$

Mutant F's (<u>f</u>eature-based) interpretive strategy:
$P_F(w | <\bullet, loud>) = 1$ irrespective of what $\bullet$ is, and thus:
$P_F(w | <a, loud>) = 1$
$P_F(w | <blah, loud>) = 1$
As a result,
$U_{r\_F}(w \text{ and } <a, loud>) = 1$
$U_{r\_F}(w \text{ and } <blah, loud>) = 1$

As before, we assume that encounters are symmetric: when two individuals $\langle x, y \rangle$ interact, in half the cases x is the signaler and y the receiver, and in the other half it's the other way around. Importantly, encounters are among conspecifics (whether residents or mutants). Heterospecifics do not take part in the encounters, but they do contribute something to the situations, namely the *blah* call which residents cannot interpret but which mutants can interpret because it is produced as loud.

(34) a. Resident C_1 interacting with resident C_2

$$U_{r\_C}(w) \quad = P(\langle a, loud \rangle \mid w)\, U_{r\_C}(w \text{ and } \langle a, loud \rangle) + P(\langle blah, loud \rangle \mid w)\, U_{r\_C}(w \text{ and } \langle blah, loud \rangle)$$
$$= .5 * 1 + .5 * 0$$
$$= .5$$

hence in the end

$$\langle C_1\text{'s utility}, C_2\text{'s utility}\rangle \quad = \langle .5U_{r\_C} + .5U_{r\_C}, .5U_{r\_C} + .5U_{r\_C}\rangle$$
$$= \langle U_{r\_C}, U_{r\_C}\rangle$$
$$= \langle f(U_{r\_C}(w)), f(U_{r\_C}(w))\rangle \qquad \text{(using the notation in (31))}$$
$$= \langle f(.5), f(.5)\rangle$$

b. Mutant F interacting with resident C

–Resident C in receiver role

$$U_{r\_C}(w) \quad = .5 \qquad \text{(as in a. above, since F has the same behavior as C in terms of production)}$$

–Mutant F in receiver role

$$U_{r\_F}(w) \quad = P(\langle a, loud \rangle \mid w)\, U_{r\_F}(w \text{ and } \langle a, loud \rangle) + P(\langle blah, loud \rangle \mid w)\, U_{r\_F}(w \text{ and } \langle blah, loud \rangle)$$
$$= .5 * 1 + .5 * 1$$
$$= 1$$

hence

$$\langle F\text{'s utility}, C\text{'s utility}\rangle$$
$$= \langle .5f(U_{r\_C}(w) + .5U_{r\_F}(w)), .5f(U_{r\_C}(w) + .5U_{r\_F}(w))\rangle$$
$$= \langle f(.5U_{r\_C}(w) + .5U_{r\_F}(w)), f(.5U_{r\_C}(w) + .5U_{r\_F}(w))\rangle \qquad \text{(using the technical remark in (32))}$$
$$= \langle f(.75), f(.75)\rangle$$

It is clear that the mutant produces greater utility in alarm situations thanks to its ability to decode heterospecific calls. Since there are otherwise no differences between mutants and residents (in particular in greeting situations), this gives the mutant an edge: the mutant's strategy against the resident yields greater utility than the resident's strategy does against the resident. This means that Maynard Smith's Condition (1) for evolutionary stability is violated: the resident strategy will be invaded by the mutant, feature-based interpretive strategy.

Conversely, the feature-based strategy will be evolutionarily stable, as seen in (35).

(35) a. Resident F_1 interacting with resident F_2

$$U_{r\_F}(w) \quad = P(\langle a, loud \rangle \mid w)\, U_{r\_F}(w \text{ and } \langle a, loud \rangle) + P(\langle blah, loud \rangle \mid w)\, U_{r\_F}(w \text{ and } \langle blah, loud \rangle)$$
$$= .5 * 1 + .5 * 1$$
$$= 1$$

$$\langle F_1\text{'s utility}, F_2\text{'s utility}\rangle \quad = \langle f(U_{r\_F}(w)), f(U_{r\_F}(w))\rangle$$
$$= \langle f(1), f(1)\rangle$$

b. Mutant C interacting with resident F

This case is identical to (34)b, hence in the end

$$\langle F\text{'s utility}, C\text{'s utility}\rangle = \langle f(.75), f(.75)\rangle$$

In this highly simplified case, then, Featural Interpretation will arise. If the feature-based strategy is applied separately to several features with different semantics, then a bundle of features may be interpreted as the conjunction of its designated interpretable features—a case of featural compositionality.

Appendix II. Comparing Featural Interpretation with Similarity-based Interpretation: Ties

In Section 5.2, we argued that a flexible similarity metric may allow a similarity-based interpretation to emulate Featural Interpretation by giving outsized importance to one acoustic over others. In Section 5.3, we argued that even in this case there remains an important difference between Featural Interpretation and similarity-based interpretations: thanks to featural compositionality, the former predicts that new feature combinations yield meanings that are different from those of familiar calls. But this conclusion came with a caveat: in principle, several familiar calls C_1 and C_2 may be tied as being 'maximally similar' to a novel call C^*, and it's conceivable that in such a case the target birds take the meaning of C^* to be the conjunction of the meanings of C_1 and C_2.

A slight modification of example (27) (from the main text) gives rise to such a tie: with C_2 replaced with C_2', the novel call C^* is equally distant from the two known calls C_1 and C_2', as seen in (36). This is the case with the simple-minded similarity metric that just counts the features C^* that has in common with C_1 vs. C_2'.

(36) a. $C_1 \quad\quad = <1, 0, 0, 0>$
$\quad\quad C_2' \quad\quad = <0, 1, 0, 0>$
$\quad\quad C^* \quad\quad = <1, 1, 0, 0>$
$\quad$ b. $\text{Sim}(C_1, C^*) = \text{Sim}(C_2', C^*) = 3/4$

As noted, it will be hard to delineate Featural Interpretation from similarity-based interpretations if the target bird reacts to C^* as to the conjunction of the tied calls C_1 and C_2'. This somewhat perverse assumption is stated in (37).

(37) Treatment of ties in similarity-based interpretations
Upon hearing a novel call C^*, if Γ is the set of familiar calls that are maximally similar C^*,[20] the target bird takes the meaning of C^* to be the conjunction of the meanings of the calls in Γ.

The problem will arise in quite a few cases. Suppose that we use the similarity metric in (23) in the main text, copied as (38), and let us consider the situation in (39)

(38) For some family of parameters $(\lambda_i)_{1 \leq i \leq 4}$ such that $\Sigma_{1 \leq i \leq 4} \lambda_i = 1$,
$\text{Sim}_\lambda(C_1, C_2) = \quad \Sigma_{1 \leq i \leq 4} \lambda_i(1-|C^1_i - C^2_i|)$

(39) Novel call: $C^* = <1, 1, 1, e^0_4, \ldots, e^0_n>$
Known calls:
$\quad\quad\quad C_1 = <1, 0, 0, e^1_4, \ldots, e^1_n>$
$\quad\quad\quad C_2 = <0, 1, 0, e^2_4, \ldots, e^2_n>$
$\quad\quad\quad C_3 = <0, 0, 1, e^3_4, \ldots, e^3_n>$

The predictions of Featural Interpretation are straightforward, as stated in (40).

(40) Featural Interpretation
For $i = \{1, 2, 3\}$, the target bird interpret C_i as meaning F_i and C^* as F_1 & F_2 & F_3, or equivalently as C_1 & C_2 & C_3.

The predictions of a similarity-based interpretation depend on the value of the parameters λ_i. If these ensure that each of the three interpretable features $\{F_1, F_2, F_3\}$ has weight 1/3 while non-interpretable features get weight 0, we will replicate precisely the predictions of Featural Interpretation of the novel call C^*. An example of precisely this appears in (41).

(41) Similarity-based Interpretation
For $i = \{1, 2, 3\}$, target birds display a C_i-like behavior when hearing C_i because of the C_i's lexical specification.
Similarity metric: Using (23), for $i = \{1, 2, 3\}$, $\lambda_i = 1/3$; for $i > 3$, $\lambda_i = 0$.
$\text{Sim}_\lambda(C_1, C^*) = \quad \Sigma_{1 \leq i \leq n} \lambda_i(1-|C^1_i - C^*_i|) = 1/3\ [1-|C^1_1 - C^*_1| + 1-|C^1_2 - C^*_2| + 1-|C^1_3 - C^*_3|] = 1/3\ [1 + 0 + 0] = 1/3$
and by symmetry, $\text{Sim}_\lambda(C_2, C^*) = \text{Sim}_\lambda(C_3, C^*) = 1/3$.

---

[20] Familiar all C' is maximally similar to novel call C^* just in case for every familiar call C", $\text{Sim}(C', C^*) \geq \text{Sim}(C'', C^*)$.

Prediction: the target bird interprets C^* as C_1 & C_2 & C_3.

The problem is more general: in this type of case, a novel call that contains any combination of features from $\{F_1, F_2, F_3\}$ will be interpreted as their conjunction, irrespective of whether one adopts Featural Interpretation or a similarity-based interpretation with a choice of λ_i's as in (41).[21]

To address this problem, we may seek to create a novel call whose meaning according to Featural Interpretation couldn't be obtained by the conjunction of any familiar calls, as this is all Similarity-based Interpretation can ever deliver. One possible case involves a novel call that contains an interpretable feature that is not present in any familiar calls. Another possible case can be constructed if a familiar call contains positive specifications of several interpretable features. If so, these can sometimes be teased apart in a novel call. The simplest case is illustrated in (42), which contains just one familiar call, but with two interpretable features F_1 and F_2. A novel call that contains just one of these two features will give rise to a novel behavior.

(42)   a. Novel call: $C^* = \langle 1, 0, \ldots \rangle$ (where … are uninterpretable features)
           Familiar call: $C = \langle 1, 1, \ldots \rangle$
           b. Featural Interpretation
           C is interpreted as F_1 & F_2, and C^* should be interpreted as F_1.
           c. Similarity-based Interpretation
           Any similarity metric can only yield the conjunction of the meanings of the maximally similar familiar calls to C^*, so here just C.

---

[21] Suppose there are k interpretable features $F_1, \ldots, F_k$, and k^* of them appear in a novel call C^* (in the text, we consider the case $k = k^* = 3$). We assume that for each $i \le k$, C_i has feature $F_i = 1$, and features $F_{i'} = 0$ if $i' \ne i$ and $1 \le i' \le k$ (other features don't matter because for $i' > k$, $F_{i'}$ is uninterpretable). This is represented in (i).

$$\begin{array}{cccc} F_1 & F_i\ F_{i+1} & & F_k \\ \end{array}$$
(i) $C_i = \ \langle 0, \ldots, 0, \mathbf{1}, \mathbf{0}, \ldots \quad\quad \mathbf{0}, F_{k+1}, \ldots, F_n \rangle$

We further assume that for $k^* \le k$, the novel call C^* is specified by $F_1 = 1, \ldots F_{k^*} = 1$, $F_{k^*+1} = 0, \ldots, F_k = 0$ (values of F_i for $i > k$ don't matter); this is represented in (ii).

$$\begin{array}{cccc} F_1 & F_{k^*}\ F_{k^*+1} & & F_k \\ \end{array}$$
(ii) $C^* = \langle 1, \ldots, 1, \mathbf{1}, \ \mathbf{0}, \ldots \quad\quad \mathbf{0}, F'_{k+1}, \ldots, F'_n \rangle$

We take the similarity metric to have the weights in (iii); this metric gives equal weight to all interpretable features, and null weight to all other features.

(iii) Definition of λ: for each i such that $1 \le i \le k$, $\lambda_i = 1/k$. For each i such that $i > k$, $\lambda_i = 0$.

We note that for each $i \le k^*$, C^* has positive feature F_i in common with C_i, and also null features $F_{k^*+1}, \ldots, F_k$ (but no other interpretable features); these common features are boldfaced in (i) and (ii). In sum, C^* has a total of $(k-k^*)+1$ features among the k that 'matter', i.e. have positive weight in the similarity metric. By contrast, for each $i > k^*$, C^* only has the null features $F_{k^*+1}, \ldots, F_k$ in common with C_i among the k features that matter—hence $(k-k^*)$ interpretable features in common between C^* and C_i. These results give rise to the similarity measures in (iv):

(iv) For each $i \le k^*$, $\mathrm{Sim}_\lambda(C_i, C^*) = (k-k^*+1)/k$, and for each $i \ge k^*+1$, $\mathrm{Sim}_\lambda(C_i, C^*) = (k-k^*)/k$.

It follows that all the features $C_1, \ldots, C_{k^*}$ are tied as maximally similar to C^*, and the latter is interpreted as the conjunction of the former. This is precisely the result obtained with Featural Interpretation of C^*.

References

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A. L., & Poeppel, D.: 2015, Human screams occupy a privileged niche in the communication soundscape. *Current Biology*, 25(15), 2051-2056.

Aubin, Thierry: 1991, Why do distress calls evoke interspecific responses? An experimental study applied to some species of birds. *Behavioural Processes* 23(2), 103–111. https://doi.org/10.1016/0376-6357(91)90061-4

Aubin, Thierry and Brémond, Jean-Claude: 1989, Parameters used for recognition of distress calls in two species: Larus argentatus and Sturnus vulgaris. *Bioacoustics* 2, 1: 23-33.

Bergen, Leon; Levy, Roger; Goodman, Noah D.: 2016, Pragmatic reasoning through semantic inference. *Semantics & Pragmatics*, 9.

Biedenweg TA, Parsons MH, Fleming PA, Blumstein DT: 2011, Sounds scary? Lack of habitua-tion following the presentation of novel sounds. *PLoS One* 6:e14549

Blesdoe, E. K., & Blumstein, D. T.: 2014, What is the sound of fear? Behavioral responses of white-crowned sparrows *Zonotrichia leucophrys* to synthesized nonlinear acoustic phenomena. *Current Zoology*, 60(4), 534-541.

Blumstein, D. T., & Récapet, C.: 2009, The sound of arousal: The addition of novel non-linearities increases responsiveness in marmot alarm calls. *Ethology*, 115(11), 1074-1081.

Blumstein, D. T., & Chi, Y. Y.: 2012, Scared and less noisy: glucocorticoids are associated with alarm call entropy. *Biology Letters*, 8(2), 189-192.

Blumstein, Daniel T., Bryant, Gregory A. and Kaye, Peter: 2012, The sound of arousal in music is context-dependent. *Biol. Lett.* 8, 744-747

Bowling DL, Gill K, Choi J, Prinz J, Purves D: 2010, Major and minor music compared to excited and subdued speech. *Journal of the Acoustical Society of America* 127: 491–503.

Bowling DL, Sundararajan J, Han S, Purves D: 2012, Expression of Emotion in Eastern and Western Music Mirrors Vocalization. *PLoS ONE* 7(3): e31942. doi:10.1371/journal.pone.0031942

Cook, Norman D.: 2007, The Sound Symbolism of Major and Minor Harmonies, *Music Perception*, 24: 315–19.

Coye, C., Zuberbuehler, K., & Lemasson, A.: 2022, The evolution of vocal communication: Inertia and divergence in two closely related primates. *International Journal of Primatology*, 43(4), 712-732.

Cross, I. and Woodruff, G. E.: 2008, Music as a communicative medium. In Botha, R. and Knight, C. (Eds.), *The Prehistory of Language*, Vol. 1, pp. 113–144.

Culicover, Peter W.: 2021, *Language Change, Variation, and Universals*. Oxford University Press.

Dubois, A., & Martens, J.: 1984, Ein Fall möglicher stimmlicher Konvergenz zwischen Fröschen und einem Vogel an Sturzbächen des Himalaya. *Journal für Ornithologie*, 125, 455-463.

Fallow, P. M., Pitcher, B. J., & Magrath, R. D.: 2013, Alarming features: birds use specific acoustic properties to identify heterospecific alarm calls. *Proceedings of the Royal Society B: Biological Sciences*, 280(1754), 20122539.

Fitch, W. T., Neubauer, J., & Herzel, H.: 2002, Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour*, 63(3), 407-418.

Franke, Michael and Jäger, Gerhard: 2016, Probabilistic pragmatics, or why Bayes' rule is probably important for pragmatics. *Zeitschrift für Sprachwissenschaft*, 35(1), 3-44.

Goodman, Noah D. and Frank, Michael C.: 2016, Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11), 818-829.

Griffin AS: 2004, Social learning about predators: a review and prospectus. *Learn Behav* 32:131–140

Haff, T. M., & Magrath, R. D.: 2013, Eavesdropping on the neighbours: fledglings learn to respond to heterospecific alarm calls. *Animal behaviour*, 85(2), 411-418.

Hauser MD: 1988, How infant Vervet monkeys learn to recognize starling alarm calls: the role of experience. *Behaviour* 105:187–201

Hirth, D. H., & McCullough, D. R.: 1977, Evolution of alarm signals in ungulates with special reference to white-tailed deer. *The American Naturalist*, 111(977), 31-42.

Hollén, L. I., & Radford, A. N.: 2009, The development of alarm call behaviour in mammals and birds. *Animal Behaviour*, 78(4), 791-800.

Huron, David: 2015, Cues and Signals: an Ethological Approach to Music-Related Emotion. In Brandt and Carmo (eds), *Music and Meaning, Annals of Semiotics* 6/2015, Presses Universitaires de Liège.

Ilie, G. and Thompson, W. F.: 2006, A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, 23, 319–29.

Ito, R., & Mori, A.: 2010, Vigilance against predators induced by eavesdropping on heterospecific alarm calls in a non-vocal lizard Oplurus cuvieri cuvieri (Reptilia: Iguania). *Proceedings of the Royal Society B: Biological Sciences*, 277(1685), 1275-1280.

Ito, R., Ikeuchi, I., & Mori, A.: 2013, A day gecko darkens its body color in response to avian alarm calls. *Current herpetology*, 32(1), 26-33.

Jouventin, P., & Aubin, T.: 2000, Acoustic convergence between two nocturnal burrowing seabirds: experiments with a penguin Eudyptula minor and a shearwater Puffinus tenuirostris. *Ibis*, *142*(4), 645-656.

Jurisevic, M. A.: 2003, Convergent characteristics of begging vocalisations in Australian birds. *Lundiana: International Journal of Biodiversity*, 4(1), 25-33.

Juslin P, Laukka P.: 2003, Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*; 129(5):770–814.

Karp D, Manser M, Wiley EM, Townsend SW: 2014, Nonlinearities in meerkat alarm calls prevent receivers from habituating. *Ethology* 120:189–196

Lemasson, A., & Hausberger, M.: 2011, Acoustic variability and social significance of calls in female Campbell's monkeys (Cercopithecus campbelli campbelli). *The Journal of the Acoustical Society of America*, 129(5), 3341-3352.

Magrath, R. D., Haff, T. M. & Igic, B.: 2020, Interspecific Communication: Gaining Information from Heterospecific Alarm Calls. In T. Aubin & N. Mathevon (eds), *Coding Strategies in Vertebrate Acoustic Communication*, Springer.

Manser MB: 2001, The acoustic structure of suricates'alarm calls varies with predator type and thelevel of response urgency. Proc R Soc B 268:2315–2324

Marler, P. : 1955, Characteristics of some animal calls. *Nature*, *176*, 6-8.

Marler, P. : 1957, Specific distinctiveness in the communication signals of birds. *Behaviour*, 11(1), 13-38.

Maynard Smith, J. and Price, G. R.: 1973, The Logic of Animal Conflict. *Science*.

Smith, J. M., & Harper, D.: 2003, *Animal signals*. Oxford University Press.

Migotti, Léo: 2019, *Towards a theory of music semantics*. MA thesis, Cogmaster, Paris.

Migotti, Léo and Zaradzki, Léo: 2019, Walk-denoting music: refining music semantics. *Proceedings of the Amsterdam Colloquium 2019*.

Morton, E. S.: 1977, On the occurrence and significance of motivation-structure rules in some bird and mammal sounds. *Am. Nat.* 111, 855—869.

Ouattara, Karim, Alban Lemasson and Klaus Zuberbühler: 2009, Campbell's monkeys concatenate vocalizations into context-specific call sequences. *PNAS* 106(51): 22026-22031.

Perelman, Polina, Warren E. Johnson, Christian Roos, Hector N. Seuánez, Julie E. Horvath, Miguel A. M. Moreira, Bailey Kessing, Joan Ponitus, Melody Roelke, Yves Rumpler, Maria Paula C. Schneider, Artur Silva, Stephen J. O'Brien and Jill Pecon-Slattery: 2011. A molecular phylogeny of living primates. *PLoS genetics* 7(3): e1001342.

Rainey, Hugo J., Klaus Zuberbühler and Peter J. B. Slater: 2004a, Hornbills can distinguish between primate alarm calls. *Proceedings of the Royal Society B: Biological Sciences* 271 : 755-759.

Rainey, Hugo J., Klaus Zuberbühler and Peter J. B. Slater: 2004b, The responses of black-casqued hornbills to predator vocalisations and primate alarm calls. *Behaviour* 141: 1263-1277.

Ramakrishnan, U., & Coss, R. G.: 2000, Recognition of heterospecific alarm vocalization by Bonnet Macaques (Macaca radiata). *Journal of Comparative Psychology*, *114*(1), 3.

Rendall, D., Owren, M. J., & Ryan, M. J.: 2009, What do animal signals mean?. *Animal Behaviour*, *78*(2), 233-240.

Rizzi, Luigi: 2016, Monkey morpho-syntax and merge-based systems. Commentary, *Theoretical Linguistics* 42(1-2). 139–145.

Rodriguez, Hugo: 2021, *Sémantique et pragmatique de la musique. Une approche cognitive basée sur le travail de Philippe Schlenker et sur les œuvres de Franz Liszt*. Doctoral dissertation, Université Libre de Bruxelles.

Sauerland, Uli. 2016, On the definition of sentence. Commentary, *Theoretical Linguistics* 42(1-2). 147–153.

Schlenker, Philippe: 2017, Outline of Music Semantics. *Music Perception: An Interdisciplinary Journal* 35, 1: 3-37 DOI: 10.1525/mp.2017.35.1.

Schlenker, Philippe: 2019, Prolegomena to Music Semantics. *Review of Philosophy & Psychology* 10 (1): 35-111. https://doi.org/10.1007/s13164-018-0384-5

Schlenker, Philippe: 2022, Musical Meaning within Super Semantics. *Linguistics and Philosophy* 45:795–872.

Schlenker, Philippe, Emmanuel Chemla, Kate Arnold, Alban Lemasson, Karim Ouattara, Sumir Keenan, Claudia Stephan, Robin Ryder and Klaus Zuberbühler: 2014, Monkey semantics: two 'dialects 'of Campbell's monkey alarm calls. *Linguistics and Philosophy* 37(6): 439-501.

Schlenker, Philippe; Chemla, Emmanuel; Arnold, Kate; Zuberbühler, Klaus: 2016a, Pyow-Hack Revisited: Two Analyses of Putty-nosed Monkey Alarm Calls. *Lingua* 171: 1--23

Schlenker, Philippe; Chemla, Emmanuel; Schel, Anne; Fuller, James; Gautier, Jean-Pierre; Kuhn, Jeremy; Veselinovic, Dunja; Arnold, Kate; Cäsar, Cristiane; Keenan, Sumir; Lemasson, Alban; Ouattara, Karim; Ryder, Robin; Zuberbühler, Klaus: 2016b, Formal Monkey Linguistics: the Debate. (Replies to commentaries). *Theoretical Linguistics* 42, 1-2:173–201, DOI: 10.1515/tl-2016-0010

Schlenker, Philippe; Chemla, Emmanuel; Schel, Anne; Fuller, James; Gautier, Jean-Pierre; Kuhn, Jeremy; Veselinovic, Dunja; Arnold, Kate; Cäsar, Cristiane; Keenan, Sumir; Lemasson, Alban; Ouattara, Karim; Ryder, Robin; Zuberbühler, Klaus: 2016c, Formal Monkey Linguistics: the Debate. (Replies to commentaries). *Theoretical Linguistics* 42, 1-2:173–201, DOI: 10.1515/tl-2016-0010

Schlenker, Philippe; Coye, Camille; Leroux, Maël; Chemla, Emmanuel: to appear, The ABC-D of Animal Linguistics. *Biological Reviews*.

Seiler, M., Schwitzer, C., Gamba, M., & Holderied, M. W.: 2013, Interspecific semantic alarm call recognition in the solitary Sahamalaza sportive lemur, Lepilemur sahamalazensis. *PloS One*, *8*(6), e67397.

Shriner, W. M. (1999). Antipredator responses to a previously neutral sound by free-living adult golden-mantled ground squirrels, Spermophilus lateralis (Sciuridae). *Ethology*, *105*(9), 747-757.

Skyrms, Brian: 1996, *Evolution and the Social contract*. Cambridge: Cambridge University Press.

Slaughter EI, Berlin ER, Bower JT, Blumstein DT: 2013, A test of the nonlinearity hypothesis in great-tailed grackles (Quiscalus mexicanus). *Ethology* 119:309–315

Townsend, S. W., & Manser, M. B.: 2011, The function of nonlinear phenomena in meerkat alarm calls. *Biology Letters*, *7*(1), 47-49.

Vencl, F.: 1977, A case of convergence in vocal signals between marmosets and birds. *The American Naturalist*, 111(980), 777-782.

Vitousek, M. N., Adelman, J. S., Gregory, N. C., & Clair, J. J. S.: 2007, Heterospecific alarm call recognition in a non-vocal reptile. *Biology Letters*, 3(6), 632-634.

Waser, Peter M., and Mary S. Waser: 1977, Experimental Studies of Primate Vocalization: Specializations for Long-distance Propagation. *Zeitschrift für Tierpsychologie* 43(3): 239-263.

Wheatcroft, D., & Price, T. D.: 2013, Learning and signal copying facilitate communication among bird species. *Proceedings of the Royal Society B: Biological Sciences*, 280(1757), 20123070.

Wheatcroft, D., & Price, T. D.: 2015, Rates of signal evolution are associated with the nature of interspecific communication. *Behavioral ecology*, 26(1), 83-90.

Zaradzki, Léo: 2021, *Les évènements en sémantique linguistique et musicale*. Doctoral dissertation, University of Paris-Diderot.

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

Zuberbühler, Klaus: 2002,A syntactic rule in forest monkey communication. *Animal Behaviour* 63(2): 293-299.

Zuberbühler, Klaus: 2009, Survivor signals: the biology and psychology of animal alarm calling. *Advances in the Study of Behavior* 40: 277-322.

Zuberbühler, Klaus: 2020, Syntax and compositionality in animal communication. *Philosophical Transactions of the Royal Society B*, *375*(1789), 20190062.

Supplementary Materials: Literature Review

Goals: We sought to find examples in the literature illustrating the three mechanisms discussed in the text: Understand Thy Neighbor, Call Convergence, and Featural Interpretation

Methods: The literature review was conducted by CC and LR, who first looked for relevant publications using Google scholar. For each table, CC and LR used various combinations of the keywords listed (in alphabetical order) below. Second, they selected relevant articles cited in the publications that they found thanks to Google scholar. No article was excluded on the basis of its publication year (oldest reference included was published in 1984).

Mechanism	Keywords
Understand Thy Neighbor (= Table 1)	Mobbing alarm call, heterospecific, eavesdropping, birds, mammals
Call Convergence (= Table 2)	Convergence, Convergent structure, Morton, birds, alarm call, birds, mammals
Featural Interpretation (= Table 3)	Acoustic feature, alarm call, animal, bird, decoding, experiment, heterospecific, recognition, synthetic sound (or signal), vocal signal, call recognition, emotion, valence, arousal

Results: The results appear in a separate Excel file. Summaries appear in the main text.

Note: In Table 3, the same shades of grey are used for similar acoustic parameters.

Excel file:

https://www.dropbox.com/s/qmdwz1gyvrezzu3/Anti-Babel%20Tables-23.03.06-3-PS.xlsx?dl=0

References of the Literature Review

Table 1: Understand Thy Neighbor
Dutour, M., Léna, J. P., Dumet, A., Gardette, V., Mondy, N., & Lengagne, T. (2019). The role of associative learning process on the response of fledgling great tits (Parus major) to mobbing calls. *Animal Cognition*, *22*, 1095-1103.
Haff, T. M., & Magrath, R. D. (2013). Eavesdropping on the neighbours: fledglings learn to respond to heterospecific alarm calls. *Animal behaviour*, *85*(2), 411-418.
Hauser, M.D., 1988. How infant vervet monkeys learn to recognize starling alarm calls: the role of experience. *Behaviour*, *105*(3-4), pp.187-201.
Ito, R., & Mori, A. (2010). Vigilance against predators induced by eavesdropping on heterospecific alarm calls in a non-vocal lizard Oplurus cuvieri cuvieri (Reptilia: Iguania). *Proceedings of the Royal Society B: Biological Sciences*, *277*(1685), 1275-1280.
Ito, R., Ikeuchi, I., & Mori, A. (2013). A day gecko darkens its body color in response to avian alarm calls. *Current herpetology*, *32*(1), 26-33.
Keen, S. C., Cole, E. F., Sheehan, M. J., & Sheldon, B. C. (2020). Social learning of acoustic anti-predator cues occurs between wild bird species. *Proceedings of the Royal Society B*, *287*(1920), 20192513.
Magrath, R. D., & Bennett, T. H. (2012). A micro-geography of fear: learning to eavesdrop on alarm calls of neighbouring heterospecifics. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1730), 902-909.
Magrath, R. D., Haff, T. M., McLachlan, J. R., & Igic, B. (2015). Wild birds learn to eavesdrop on heterospecific alarm calls. *Current Biology*, *25*(15), 2047-2050.
Liu, J. and Liang, W., 2022. Free-range domestic chickens can distinguish between different alarm calls of Japanese tits. Animal Cognition, pp.1-6.

Pollock, H. S., Martínez, A. E., Kelley, J. P., Touchton, J. M., & Tarwater, C. E. (2017). Heterospecific eavesdropping in ant-following birds of the Neotropics is a learned behaviour. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1865), 20171785.

Potvin, D. A., Ratnayake, C. P., Radford, A. N., & Magrath, R. D. (2018). Birds learn socially to recognize heterospecific alarm calls by acoustic association. *Current Biology*, *28*(16), 2632-2637.

Ramakrishnan, U., & Coss, R. G. (2000). Recognition of heterospecific alarm vocalization by Bonnet Macaques (Macaca radiata). *Journal of Comparative Psychology*, *114*(1), 3.

Seiler, M., Schwitzer, C., Gamba, M., & Holderied, M. W. (2013). Interspecific semantic alarm call recognition in the solitary Sahamalaza sportive lemur, Lepilemur sahamalazensis. *PloS One*, *8*(6), e67397.

Shriner, W. M. (1999). Antipredator responses to a previously neutral sound by free-living adult golden-mantled ground squirrels, Spermophilus lateralis (Sciuridae). *Ethology*, *105*(9), 747-757.

Szymkowiak, J. (2021). Wood warblers learn to recognize mobbing calls of an unfamiliar species from heterospecific tutors. *Animal Behaviour*, *171*, 1-11.

Vitousek, M. N., Adelman, J. S., Gregory, N. C., & Clair, J. J. S. (2007). Heterospecific alarm call recognition in a non-vocal reptile. *Biology Letters*, *3*(6), 632-634.

Wheatcroft, D., & Price, T. D. (2013). Learning and signal copying facilitate communication among bird species. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1757), 20123070.

Zuberbühler, K., 2000. Causal knowledge of predators' behaviour in wild Diana monkeys. *Animal Behaviour*, *59*(1), pp.209-220.

Table 2: Call Convergence

Dubois, A., & Martens, J. (1984). Ein Fall möglicher stimmlicher Konvergenz zwischen Fröschen und einem Vogel an Sturzbächen des Himalaya. *Journal für Ornithologie*, *125*, 455-463.

Jouventin, P., & Aubin, T. (2000). Acoustic convergence between two nocturnal burrowing seabirds: experiments with a penguin Eudyptula minor and a shearwater Puffinus tenuirostris. *Ibis*, *142*(4), 645-656.

Jurisevic, M. A. (2003). Convergent characteristics of begging vocalisations in Australian birds. *Lundiana: International Journal of Biodiversity*, *4*(1), 25-33.

Jurisevic, M. A., & Sanderson, K. J. (1994). Alarm vocalisations in Australian birds: convergent characteristics and phylogenetic differences. *Emu-Austral Ornithology*, *94*(2), 69-77.

Russ, J. M., Jones, G., Mackie, I. J., & Racey, P. A. (2004). Interspecific responses to distress calls in bats (Chiroptera: Vespertilionidae): a function for convergence in call design?. *Animal Behaviour*, *67*(6), 1005-1014.

Vencl, F. (1977). A case of convergence in vocal signals between marmosets and birds. *The American Naturalist*, *111*(980), 777-782.

Table 3: Featural Interpretation

Bastian, A., & Schmidt, S. (2008). Affect cues in vocalizations of the bat, Megaderma lyra, during agonistic interactions. *The Journal of the Acoustical Society of America*, *124*(1), 598–608.

Briefer, E. F. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology*, *288*(1), 1–20.

Collins, K. T., McGreevy, P. D., Wheatley, K. E., & Harcourt, R. G. (2011). The influence of behavioural context on Weddell seal (Leptonychotes weddellii) airborne mother-pup vocalisation. *Behavioural Processes*, *87*(3), 286–290.

Esch, H. C., Sayigh, L. S., Blum, J. E., & Wells, R. S. (2009). Whistles as potential indicators of stress in bottlenose dolphins (Tursiops truncatus). *Journal of Mammalogy*, *90*(3), 638–650.

Fallow, P. M., & Magrath, R. D. (2010). Eavesdropping on other species: Mutual interspecific understanding of urgency information in avian alarm calls. *Animal Behaviour*, *79*(2), 411–417.

Friel, M., Kunc, H. P., Griffin, K., Asher, L., & Collins, L. M. (2019). Positive and negative contexts predict duration of pig vocalisations. *Scientific Reports*, *9*(1), 1–7.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770.

Lemasson, A., Remeuf, K., Rossard, A., & Zimmermann, E. (2012). *Cross-taxa similarities in affect-induced changes of vocal behavior and voice in arboreal monkeys*.

Manser, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1483), 2315–2324.

Randler, C., & Förschler, M. I. (2011). Heterospecifics do not respond to subtle differences in chaffinch mobbing calls: Message is encoded in number of elements. *Animal Behaviour*, *82*(4), 725–730.

Schneider, J. N., & Anderson, R. E. (2011). Tonal vocalizations in the red wolf (Canis rufus): Potential functions of nonlinear sound production. *The Journal of the Acoustical Society of America*, *130*(4), 2275–2284.

Stoeger, A. S., Charlton, B. D., Kratochvil, H., & Fitch, W. T. (2011). Vocal cues indicate level of arousal in infant African elephant roars. *The Journal of the Acoustical Society of America*, *130*(3), 1700–1710.