The ABC-D of Animal Linguistics: Are Syntax and Compositionality for Real?*

Philippe Schlenker[a], Camille Coye[b*], Maël Leroux[c], Emmanuel Chemla[d]

January 23, 2023. Accepted with minor revisions in *Biological Reviews*/

Abstract. In several animal species, an alarm call (e.g. *ABC* notes in the Japanese great tit) can be immediately followed by a recruitment call (e.g. *D* notes) to yield a complex call that triggers a third behavior, namely mobbing. This has been taken to be an argument for animal syntax and compositionality (= the property by which the meaning of a complex expression depends on the meaning of its parts and the way they are put together). Several additional discoveries were made across species. First, in some cases, animals respond with mobbing to the order *alarm–recruitment* but not to the order *recruitment–alarm*. Second, animals sometimes respond similarly to functionally analogous heterospecific calls they have never heard before, and/or to artificial hybrid sequences made of conspecific and heterospecific calls in the same order, thus adding an argument for the productivity of the relevant rules. We consider the details of these arguments for animal syntax and compositionality and argue that, with one important exception (Japanese tit ABC-D sequences), they currently remain ambiguous: there are reasonable alternatives on which each call is a separate utterance and is interpreted as such ('trivial compositionality'). More generally, we propose that future studies should argue for animal syntax and compositionality by explicitly pitting the target theory against two deflationary analyses: the 'only one expression' hypothesis posits that there is no combination in the first place, e.g. just a simplex *ABCD* call; while the 'separate utterances' hypothesis posits that there are separate expressions (e.g. *ABC* and *D),* but that they form separate utterances and are neither syntactically nor semantically combined.

Keywords: animal linguistics, animal semantics, compositionality, animal syntax, combinatoriality, interspecies comprehension

* **Acknowledgments:** We are very grateful to Michael Griesser, Nathan Klinedinst, Pritty Patel-Grosz and Ambre Salis for helpful discussions, and to Lucie Ravaux for help with the figures.

Funding:

Coye, Schlenker: This research received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 788077, Orisem, PI: Schlenker).

Coye, Chemla, Schlenker: Research was conducted at DEC, Ecole Normale Supérieure - PSL Research University. DEC is supported by grant FrontCog ANR-17-EURE-0017. .

[a] Institut Jean-Nicod (ENS - EHESS - CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France; PSL Research University; New York University, New York.

[b] Institut Jean-Nicod (ENS - EHESS - CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France; PSL Research University.

[c] Department of Comparative Language Science, University of Zürich, Zürich, Switzerland; Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zürich, Zürich, Switzerland

[d] LSCP (ENS - EHESS - CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France; PSL Research University.

* Corresponding author: Email: camillecoye@gmail.com

CONTENTS

I. Introduction

I.1 Animal linguistics

Four main questions have been raised in recent studies of animal calls (similar questions extend to gestures, but we will restrict attention to calls):
1. What is the inventory and meaning of individual calls?
2. Are there rules of syntactic combination besides the juxtaposition of independent calls?
3. Are there rules of semantic composition besides the juxtaposition of independent calls? In particular, are there animal instances of 'compositionality', the principle by which the meaning of a complex expression is derived from the meaning of its parts and the way they are put together?
4. Are there pragmatic rules of competition among calls? In particular, do some animals go by the 'Informativity Principle', according to which a more informative call should be preferred to a less informative one when both can be used truly?

To avoid ambiguity, it might be useful to define at the outset some of the key linguistic terms, namely *syntax*, *semantics* and *pragmatics*. Replacing 'word' with 'call', the definitions are applicable to animal linguistics (see Berthet et al. 2022 for a further introduction for biologists).
Syntax is the set of rules that determine which sequences of words are well-formed and which are not. For instance, *Robin loves Casey* is well-formed, *Loves Robin Casey* is not.
Semantics is the set of rules that determine the meaning of individual words, and of sequences of words combined by a syntactic rule. Meaning is usually analyzed in terms of the situations a word or sequence of words is true of.
Pragmatics is the set of rules, usually derived from optimal language use, that help choose among words or sentences when several are applicable (true). Thus, *I'll invite Ann or Bill* is semantically true but pragmatically deviant when I know that *I'll invite Ann and Bill*. The reason is that the latter sentence is more informative and should thus be preferred ('Informativity Principle').

Importantly, the division between syntax, semantics and pragmatics is in part a theoretical affair. In Campbell's monkeys, the non-predation call *boom* is usually restricted to sequence-initial positions. This restriction can be viewed as a syntactic rule (Zuberbühler 2002), but the argument in favor of a syntactic rule disappears if there are other constraints that impose this ordering. Due to its production, *boom* requires that an air sac be filled, which might take time and energy; doing so might conceivably

be impossible within a sequence, in which case the sentence-initial restriction has a non-syntactic source (here: a production-related contraint; see Schlenker et al. *201*6c [fn 5] for relevant speculations). Similarly, *pyow-hack* sequences in Putty-nosed monkeys come with a specific order (a small number of *pyows* followed by a small number of *hacks*), and the sequence is associated with a particular function, involving group movement. The ordering restriction might result from a syntactic rule (Arnold & Zuberbühler 2012, Miyagawa & Clarke 2019). But an alternative is that the ordering is due to a pragmatic principle, called the 'Urgency Principle' in Schlenker et al. 2016a (in brief, it mandates that calls that provide information about the nature/location of a threat should come before calls that don't; see Appendix A1). Several studies coming out of collaborations between formal linguists and primatologists have focused on the meaning of individual calls and rules of pragmatic competition among calls, i.e. points 1. and 4. above (e.g. Schlenker et al. 2016*c,d*; see Appendix A2). They have mostly given a deflationary answer to questions of syntactic combination and of semantic composition: in the relevant case studies, no non-trivial syntactic or semantic rules were posited; one notable exception, to which we return below, pertains to the suffix *-oo* in Campbell's monkeys, which was taken to be both morphosyntactically and semantically constrained.

The message that came out of these studies was that animal linguistics has little in common with human language, and should be studied in its own right rather than by comparison with human language. But if this is so, why should specialists of animal communication expect anything of collaborations with linguists? For two reasons, highlighted in Schlenker et al. 2016c. First, using general tools of formal language theory (rather than specific tools of human linguistics) makes it possible to specify explicit and predictive models of animal communication. Second, a fundamental but very subtle issue arises in animal and human communication alike; it pertains to the division of labor between the literal meaning of words/calls, additional information due to knowledge of the environment, and rules of competition among words/calls. Linguistic expertise has a clear role to play in delineating these modules.

The deflationary view of primate syntax and compositionality might have come as a disappointment if animal linguistics is to offer an explanation of how human language came to emerge, as syntax and compositionality are two of its most remarkable hallmarks. But birds seem to have provided a new argument for animal syntax and compositionality (points 2. and 3. above). In several species, researchers have noticed that an alarm call (e.g. *ABC* notes in the Japanese tit) can be followed by a recruitment call (e.g. *D* notes) to yield a sequence with an apparently new meaning, one that triggers mobbing (Engesser, Ridley & Townsend, 2016, Suzuki, Wheatcroft & Griesser, 2016, Dutour et al., 2019*a*, Salis et al., 2021*a*,). This has been taken as an argument for bird syntax and compositionality.

Several additional discoveries were made. First, in some cases, animals respond with mobbing to the order *alarm–recruitment* but not to the order *recruitment–alarm* (Suzuki, Wheatcroft & Griesser, 2017, Salis, Léna & Lengagne, 2021*a*, Dutour, Lengagne & Léna, 2019*a*, Dutour Suzuki & Wheatcroft, 2020; but see Engesser et al., 2020). Second, animals sometimes respond to functionally and structurally analogous heterospecific calls they have never heard before (Dutour, Léna & Lengagne, 2017), and/or to artificial hybrids of conspecific and heterospecific calls in the same order (Dutour et al., 2020), thus adding an argument for the productivity of the relevant compositional rules.

To assess these arguments, we offer a systematic method to argue for or against animal syntax and compositionality (a summary of key arguments in each case study is offered at the end of this piece, in (27)). Applied to the cases above, the conclusion will be that, with one important exception (the ABC-D sequences of Japanese tits), there are reasonable deflationary alternatives on which each call is a separate utterance, and sequences of calls are interpreted one at a time ('trivial compositionality').

Concretely, we propose to systematically pit each claim about syntax and compositionality against two alternative theories (here too, summary recommendations will be offered at the end of this piece, in (28)). Consider a claim that two calls C_1–C_2 are combined by a syntactic rule and interpreted by a compositional rule, a rich theory we summarize in (1) below (see Appendix A3).

(1) Rich theory
a. Syntactic claim: two calls C_1–C_2 are combined by a syntactic rule.
b. Semantic claim: two calls C_1–C_2 are interpreted by a compositional rule, i.e. one on which the meaning of the whole is determined by the meaning of its parts and the way they are put together.

The first deflationary theory, which we call 'Only one expression' in (2) below, denies that there are two calls in the first place, and thus no syntactic combination is needed (= (2)a), and no compositional rule either (= (2)b), as the single call C_1C_2 can gets its meaning independently of the meanings of C_1 and C_2. In other words, although C_1C_2 has an acoustic structure that is similar to the concatenation of C_1 and C_2, it is not made of two calls and its meaning is not compositional (i.e. the animals need to learn the meaning of C_1, C_2 and C_1C_2 separately).

(2) **Deflationary theory 1:** 'Only one expression'
a. Syntactic status: There is no need for a syntactic rule because there is a single call C_1C_2, which is only acoustically complex.
b. Semantic status: There is no need for a semantic rule because there is a single call C_1C_2, which is only acoustically complex.

The second deflationary theory, called 'Separate utterances' in (3), grants that there are two calls but claims that they are separate utterances emitted in close succession (without forming a complex call). As a result, no syntactic rule is needed to combine them (= (3)a), as they are produced independently. And no semantic rule is needed either (= (3)b): it is a given that if an animal hears and understands a call C_1 and then a call C_2 at a later time (even as much as 5 minutes later), memory permitting, the animal's final information state will result from the (conjunctive) sum of the information of C_1 and that of C_2. If this is all that is going on, C_1 and C_2 function as separate utterances, and no compositional rule is needed to combine their meanings. On the semantic side, 'Separate utterances' has been called 'Trivial compositionality' in the literature (in this piece, we will use 'compositionality' to mean 'non-trivial compositionality', in line with the literature; see Appendix A4).

(3) **Deflationary theory 2:** 'Separate utterances' (also called 'trivial compositionality' in the literature)
a. Syntactic status: There is no need for a syntactic rule because although there are two calls, they are separate utterances.
b. Semantic status: There is no need for a semantic rule because although there are two calls, they are separate utterances.

It might be worth explaining further why 'separate utterances' (i.e. 'trivial compositionality') requires no syntax and no compositional semantics. Suppose Ann and Bill just entered my apartment and Ann tells me *It's hot outside* while Bill tells me *It's humid outside*. These are clearly separate utterances since they are produced by different individuals, and they are neither syntactically nor semantically combined. But I have no trouble aggregating the information they provide: my final information state will be one according to which it's both hot and humid outside. The situation does not change if Ann is the only speaker and first says *It's hot (outside)* and then *It's humid (outside)*: these are separate utterances that do not require a syntactic or a semantic rule to be combined (although each involves sentence-internal rules of English syntax and semantics). Of course, the meaning of Ann's two-sentence discourse is derived from the meaning of its sentential parts, and hence it is 'compositional', but in a trivial fashion as these parts can be treated as separate utterances: no syntactic or semantic rules are needed to account for the combination (see Appendix A5).

I.2 Bolhuis et al.'s critique

Bolhuis et al. 2018*a,b* expressed towards animal syntax and compositionality a skepticism that is close to ours, but with different arguments and motivations.

First, Bolhuis et al. wished to highlight differences between *ABC-D*-type data and syntax and compositionality *in human language*. We adopt instead the perspective of Schlenker et al. 2016*c,d*, and ask whether, irrespective of any differences with human language, *ABC-D*-type data display an argument for *some* kind of syntax and compositionality. The difference can be illustrated with a very simple mathematical example. One can define the set of natural numbers as the expressions {0, s0, ss0, sss0, …}, where *0* represents the number zero, and *sx* is intended to represent x+1 (also called the 'successor' of the number x, hence the choice of symbols). This set of expressions clearly has a syntax; for instance, *s0* is well-formed but *0s* isn't. This syntax can be defined by way of just two rules, stated in (4). In this way, we define as well-formed *0*, then *s0*, then *ss0*, etc. It is clear that this syntax has little to do with the syntax of human language, but it is a syntax nonetheless.

(4) Syntax[7]
(i) *0* is a well-formed expression
(ii) if *E* is a well-formed expression, *sE* is a well-formed expression

This syntax can be associated with an equally simple compositional semantics, made again of just two rules (corresponding to the two syntactic rules):

(5) Compositional semantics
(i) The expression *0* denotes the number zero.
(ii) For any well-formed expression *E*, *sE* denotes what *E* denotes plus one.

This semantics too has little to do with human language, but it is a compositional semantics nonetheless. In other words, questions of syntax and compositionality can be asked irrespective of the vast differences there are between animal and human languages—and these are interesting questions.

The second difference between Bolhuis et al.'s contribution and ours is that we seek to propose a simple way of arguing for syntax and compositionality: not just by listing criteria (often an arduous task), but by pitting the 'syntax' and 'compositionality' claims against explicit deflationary alternatives, 'only one expression' and 'separate utterances'; our hope is that this will help clarify future debates (see Appendix A6).

I.3 Structure

The rest of this piece is organized as follows. We illustrate the main notions with some of the simplest possible examples of (morpho-)syntactic combination and semantic composition in humans and in monkeys (Section II). We then discuss in turn three cases of *alarm–recruitment* combinations, in the Japanese tit (Section III), in the Southern pied-babbler (Section IV), and in the Great tit (Section V), pitting claims of syntax and compositionality against our two deflationary hypotheses. While very recent results on the Japanese tit make a strong case for morphosyntax and compositionality, the other findings are fascinating but remain ambiguous in demonstrating syntax and compositionality, in large part because the 'separate utterances' theory has not been refuted.

II. Morphosyntax and compositionality in two suffixes: English *-ish* and Campbell's *-oo*

In human language, syntax broadly construed pertains to the rules by which expressions are put together to form new expressions. But it comes in two varieties. Syntax proper could also be called 'word-external syntax', and it pertains to the ways words can be put together to form phrases and sentences. But words are themselves composed by way of rules, which are the domain of 'morphology'. The reason for the distinction is that the rules look different when it comes to building words and phrases (but see for instance Embick and Noyer 2012 for attempts to unify morphological and syntactic rules). From the verb *inform* one can construct the adjective *informative*, and from that the noun *informativity*, a term that was used above ('Informativity Principle'); the rules of combination are, at least superficially, different from those used in word-external syntax, for instance to form *Robin loves Casey* or *Casey loves Robin* from the words {*Robin*, *loves*, *Casey*}. We may thus call morphology 'word-internal syntax'. We will follow linguistic usage in talking of 'morphosyntax' when we wish to encompass both morphological and syntactic rules.

While word-external syntax superficially has a counterpart in the sophisticated songs of birds (e.g. Berwick et al. 2011 for a review), the analogy is only partial: unlike sentences, bird songs do not currently appear to have a meaning derived from that of their parts, and unlike alarm calls, they might not have any meaning at all besides their ability to mark territory and/or advertise the singer's quality (Brémond, 1968; Marler, 1998, Collier et al., 2014). By contrast, animal calls, be it in birds or in mammals, have a meaning and possibly a pragmatics (Schlenker et al. 2016*c*), but usually no syntax to speak of. There is one case of call-internal syntax which is arguably different, however: the suffix *-oo* in Campbell's monkeys. Before we lay out the logic of the argument (as well as its weaknesses), we'll start by illustrating the main notions on the example of the suffix *-ish* in English.

II.1 Arguments for morphosyntax and compositionality in English: -ish

A distant analogy to Campbell's *-oo* is the English suffix *-ish* as it applies to *blue* and *green* to form *blueish* and *greenish*; further examples could easily be summoned, for instance *-like* as in *green-like*. *-ish* has two properties that will prove crucial in our discussion: it is governed by rules of word-internal syntax, i.e. morphology; and it modifies meaning by a compositional rule (see Appendix A7). These points are summarized in (6).

(6) a. Morphosyntax of *-ish*
For any adjective *A*, *A-ish* is an adjective (by contrast, *ish-A* isn't).
b. Semantics of *-ish*
For any adjective *A*, *A-ish* holds true of things that are kind of *A*.

The argument for the existence of a morphosyntactic rule is summarized in (7). As stated above, there are two deflationary theories to consider. The first ('only one expression') is that *blueish* is only accidentally acoustically made of *blue+ish*, just like the adjective *irate* is accidentally made of *I+rate*: *irate* has nothing to do with the first person, nor with any kind of rating. On this theory, one might for instance posit that it is for historical reasons that *blueish* contains the word *blue*, but that cognitively *blueish* is not decomposable into further meaningful units and is thus a single expression. The second deflationary theory ('separate utterances') is that no combinatorial rule is needed because each component, *blue* and *ish*, forms a separate utterance. This is implausible for *blueish* (and even more so for *blue-like*; see again Appendix A7), but in the animal discussions below it will be a live contender. To put things differently, the first deflationary theory denies that there are two separate expressions to combine in the first place, while the second deflationary theory grants that there are two expressions, but denies that rules are needed to combine them because they form separate utterances.

(7) *-ish* is added to words by a (word-internal) syntactic rule:
for any adjective *A*, *A-ish* is an adjective

Deflationary Theory 1—'Only one expression': *blueish* is a word that is only accidentally pronounced with *blue* in it.
Counterargument 1: Pattern: *greenish* patterns like *blueish*, and so does *yellowish*.
Counterargument 2: Productivity: *-ish* can be added to a new adjective that one has never heard before, e.g. *alt-right,* or the invented adjective *wof*.

Deflationary Theory 2—'Separate utterances': *blue* is a separate utterance, *ish* is a separate utterance. (Implausible in English, as *blue* and *-ish* cannot be used as separate utterances.)

One weak argument ('Pattern') against the 'only one expression' view is that the same pattern is found in further adjectives, such as *greenish* and *yellowish*. In other words, if the presence of *blue* in *blueish* is an accident, the same accident must be responsible for *greenish* and *yellowish*. This is a weak argument because the same accident could in principle arise several times (in addition, what is an accident in the speaker's or caller's mind could be explained by an historical or evolutionary process which, *in view of general laws,* led two separate words or calls to become merged at some point). The strong argument ('Productivity') is that speakers of English can add *-ish* to words they are hearing for the first time—for instance someone who hears that *Ann is alt-right* will have no trouble inferring that one can also say that *Bill is alt-rightish*, and similarly for *wof* and *wofish* (where *wof* is an invented adjective).

The existence of a (word-internal) syntactic rule pertains to form alone, but it has a counterpart on the meaning side. Specifically, the meaning of *blueish* is computed from the meaning of its parts in accordance with the rule in (8), an example of compositionality. Here too, there are two deflationary semantic theories to consider: one denies that there are two meanings in the first place; the other grants that there are two meanings, but denies that they are integrated by a semantic rule as they correspond to separate utterances.

(8) *-ish* modifies the meaning of adjectives by a (compositional) semantic rule:
for any adjective *A*, *A-ish* holds true of things that are kind of *A*.

Deflationary Theory 1—'Only one expression': the meaning of *blueish* is memorised.

Counterargument 1: Pattern: the meaning of *greenish* relates to the meaning of *green* as the meaning of *blueish* relates to the meaning of *blue*.
Counterargument 2: Productivity: one may understand the meaning of a new expression *Xish* as soon as one understands *X*, e.g. *alt-rightish*.

Deflationary Theory 2—'Separate utterances': *blue* is a separate utterance, *ish* is a separate utterance, and the meaning of *blueish* is not the result of a compositional rule.
(Implausible in English, as *blue* and *ish* cannot be used as separate utterances.)

The semantic argument against the 'only one expression' view is that the way the meaning of *blueish* is derived from the meaning of *blue* and the meaning of *-ish* corresponds to a productive pattern; in particular, anyone who understands the meaning of *alt-right* understands the meaning of *alt-rightish*.

What about the 'separate utterances' view? To clarify the conceptual issue, it's worth giving it a chance, and imagine (fancifully) that we've only heard people exclaiming *Blue!* when watching blue things, and exclaiming *Blueish!* when watching things that are kind of blue. We could still argue against the 'separate utterances' view in two ways. First, although now *Blue!* is a separate utterance, *Ish!* isn't. Second, even if the latter were (as in some non-standard uses described in McCulloch 2014), a situation that satisfies both *Blue* and *Ish* should in particular satisfy *Blue*. But this predicts that when *Blueish!* is true, so is *Blue*; and this is not the case. These arguments are summarized in (9).

(9) Arguments against Deflationary Theory 2—'Separate utterances'
in the imaginary situation in which one has heard people exclaiming *Blue!* when watching blue things, and exclaiming *Blueish!* when watching things that are kind of blue
Argument 1: *Ish* is not heard on its own.
Argument 2: If *Blue* and *Ish* were separate utterances, a situation that satisfies both should in particular satisfy *Blue*. But this is not so: *blueish* things need not be blue.

II.2 Arguments for animal morphosyntax and compositionality: -oo

It was argued in Schlenker et al. 2014, 2016*c* that Campbell's monkeys have a suffix *-oo* which (i) is morphosyntactically attached to two calls, *krak* and *hok*, and (ii) modifies their meaning by way of a compositional rule. Schlenker et al. 2014 consider two theories of the meaning of *-oo*, and for simplicity we will only discuss the second and final one, disregarding various details.

Considering data from two sites, the Tai forest in Ivory Coast and Tiwai island in Sierra Leone, the authors propose the stylized generalizations in (10). They explain them by way of the morphosyntactic and semantic rules in (11).

(10) a. Calls found: *boom, krak*, *hok*, *krak-oo*, *hok-oo*.
b. Call meaning (according to the final analysis only):
boom: there is a disturbance but not of a predator
krak: there is a disturbance
hok: there is a non-terrestrial disturbance
krak-oo: there is a weak disturbance
hok-oo: there is a weak non-terrestrial disturbance

(11) a. Calls and call meanings
boom: there is a disturbance but not of a predator
krak: there is a disturbance
hok: there is a non-terrestrial disturbance

b. Morphosyntactic rule
If *C* is the call *krak* or *hok*, *C-oo* is a call.

c. Semantic rule
If *C* is the call *krak* or *hok*, *C-oo* is true just in case there is a disturbance that licences *C* and is weak among disturbances that licence *C*.

What are the arguments for positing these rules? As in the case of *-ish* above, we need to consider two deflationary theories on the morphosyntactic and on the semantic side alike. One is that *krak-oo* and *hok-oo* are each made of a single expression to begin with. The second is that they are made of two expressions, but that these are separate utterances.

The argument against the 'only one expression' view is entirely based on patterns, both on the morphosyntactic and on the semantic side. First, *-oo* can be added to two calls, namely *krak* and *hok*, to form *krak-oo* and *hok-oo* respectively. Second, on one plausible analysis, *-oo* modifies the meaning of *krak* and *hok* in the same way, as is stated by the rule in (11)c (but see Appendix A8).This is an animal counterpart of the weaker of the two arguments we mentioned in relation with *-ish*. It would be far better to have an argument based on productivity, to the effect that animals understand that *-oo* can be added to new calls and can modify their meaning in a regular way. This argument does not currently exist. On the other hand, the main argument against the 'separate utterances' view is simple: *-oo* is never found on its own, making it implausible that it can form a separate utterance (but see Appendix A9).

These arguments are summarized in (12).

(12) Summary of arguments and objections: morphosyntax and compositionality in Campbell's monkeys' *-oo* suffix

Nature of the rule	**Main arguments**	**Alternative 1: 'Only one expression'**	**Alternative 2: 'Separate utterances'**
Morphosyntactic	Pattern: *-oo* is never found on its own but can be added to *krak* and *hok*.	• A pattern with 2 instances (*krak-oo*, *hok-oo*) could be an accident. • No argument for productivity exists.	Implausible as: *-oo* never appears on its own (but see Sauerland 2016)
Semantic	Pattern: *-oo* plausibly modifies the meaning of *krak* in the same as it does the meaning of *hok*.	• A pattern with 2 instances (*krak-oo*, *hok-oo*) could be an accident. • Auxiliary assumptions are needed to get the compositional analysis to work.	• Implausible as: *-oo* never appears on its own (but see Sauerland 2016) • On some analyses of the meaning, *hok-oo* does not entail the purported meaning of *-oo* (≈ non-serious alarm)

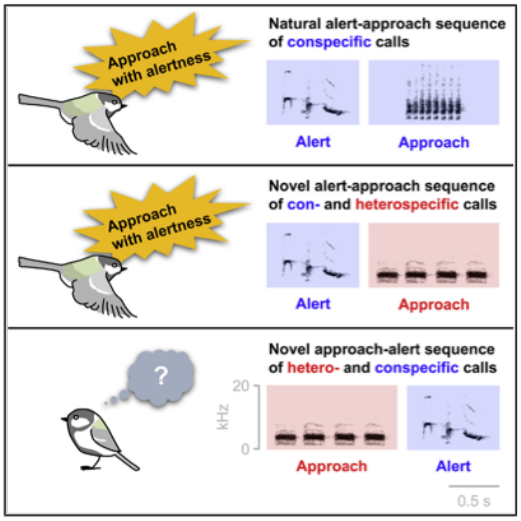
III. Syntax and compositionality in ABC-D sequences of the Japanese tit

We turn to ABC-D-type sequences in birds and argue that the argument for a combination of parts is very strong: 'only one expression' is implausible. In initial experiments, the 'separate utterances' view remained a live contender: the combination might have been of the 'trivial' kind. But recent and very important results present new challenges for the 'separate utterances' view, making ABC-D sequences the best argued case for syntax and compositionality in the animal world. We will explain in turn why the initial results left open a 'separate utterances' analysis, and why new results challenge it.

Suzuki et al. 2016 show that the Japanese tit (*Parus minor*) reacts with increased vigilance to an ABC sequence, an alarm call, and they tend to approach the speaker when hearing a D sequence, a recruitment call. When hearing an ABC-D sequence, they react with increased vigilance and approach; but when the order is reversed, with a D-ABC sequence, they do not react much.

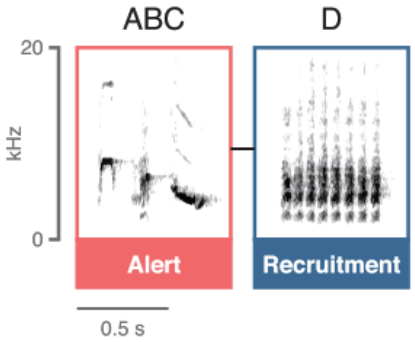
Remarkably, Suzuki et al. 2017 show that these results are replicated when researchers play back synthetic sequences made of an *ABC*-sequence immediately followed by a *D**-sequence which is used by the Willow tit, a sympatric species (one that lives in the same area as the Japanese tits under study). Importantly, the Willow tit *D**-sequence sounds rather different from the Japanese tit *D*-sequence (here and throughout, we will use *D** for an alternative to *D* that is *not* acoustically similar to *D*; we will later use *ABC*' and *D*' for alternatives to *ABC* and to *D* that *are* acoustically similar to them). Specifically, when hearing *ABC-D** sequences that are never found in nature, the Japanese tits display increased vigilance and approach the speaker, but when hearing *D***-ABC* sequences that are equally non-existent in nature, they display neither behavior. The main experiment is summarized in (13), and the comparison between the acoustic form of Japanese and Willow tit calls is made in (14), where D*, the Willow tit counterpart of D, is called *tää*.

(13) Logic of the experiments of Suzuki et al. 2017

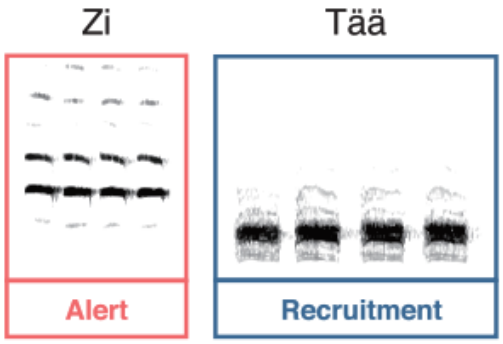


(14) Japanese tit vs. Willow tit calls (Suzuki et al. 2017)

a. Japanese tit calls



b. Willow tit calls



A key assumption is that despite the presence of repetitions in both species, the Willow tit D* *(= tää)* sounds very different from the Japanese tit D. Suzuki et al. 2017 provide direct evidence that acoustic similarity between *tää* and D isn't at stake. Since *tää* calls are longer than D calls, *tää* can be artificially shortened so as to resemble D more than the unshortened *tää* does. If acoustic similarity to D were driving the response, Japanese tits should react more strongly to the shortened *tää* call than to the real *tää*. But the opposite is the case, suggesting that acoustic similarity to D isn't at stake. It should be added that Suzuki et al. 2017 include an additional control to show that not just anything of the form ABC-blah triggers a reaction. Specifically, a hybrid sequence of the form ABC-zi, where zi is the Willow tit counterpart of ABC, fails to trigger a reaction comparable to ABC-tää.

The authors conclude that this is a case of syntax associated with a compositional semantics. We will discuss separately the syntactic and the semantic argument (see also Appendix A10).

On the syntactic side, there are, as before, two deflationary views to be refuted, as is stated in (15).

(15) Japanese tit *ABC-D* sequences involve a syntactic rule

Deflationary Theory 1—'Only one expression'
Counterargument: Productivity
Japanese tits treat artificial ABC-D* sequences in the same way as ABC-D sequences, showing that they have a productive rule.

Deflationary Theory 2—'Separate utterances'
Counterargument: Japanese tits react to ABC-D/D* sequences but not to D/D*-ABC sequences.
Counter-counterargument: the distinction might be due to a non-syntactic rule (e.g. acoustic or pragmatic)

The first deflationary view is that ABC-D is only one expression. This view would immediately explain why Japanese tits react to ABC-D but not to D-ABC: on the proposed view, ABC-D is a call but D-ABC isn't. Here Suzuki et al. 2017 provide the strongest kind of argument (unlike Schlenker et al. 2014 for *-oo*): Japanese tits apply a rule to sequences they have never heard before, namely the hybrid ABC-D* and D*-ABC sequences. Both are unattested in nature, and yet the Japanese tits react to ABC-D* but not to D*-ABC. We thus agree with Suzuki et al. that Deflationary Theory 1 is extremely implausible.

But what about the second deflationary theory, 'separate utterances'? Here the argument against treating *ABC-D* as two utterances is that *D-ABC* does not give rise to a reaction. It does seem plausible that this contrast involves a rule, but does it have to be a *syntactic* rule? There are at least three alternatives to consider. One is that lack of reaction is due to lack of familiarity. The second is that the rule is driven by acoustic constraints. The third one is that it is driven by pragmatic considerations.

The least exciting hypothesis is that lack of reaction to *D-ABC* sequences is due to lack of familiarity. But since Japanese tits display a differential behavior to two equally unfamiliar sequences, namely the hybrids created by the researchers, this does not account for the data.

A second hypothesis is acoustic in nature. Dutour et al. 2020 mention for another species, the Great tit, that ABC-type notes might be hard to perceive after D-type notes (see Appendix A11). If so, the effect found by Suzuki et al. might result from limitations of perception, and might not speak against a treatment of *ABC* and *D* as separate utterances. Importantly, for this hypothesis to work, it should be the case that *both* the Japanese tit D and the Willow tit D* have a masking effect. We doubt that this is the case, for three reasons. First, the articles cited in Dutour et al. 2020 mention masking mechanisms pertaining to call overlap (Grafe, 1996 and Klump, 1992) and to the specific acoustic structure of trills and whistles, two tonal units (Brown and Handford, 1996). But these mechanisms do not apply to *D* and *D**, which are repetitive broadband structures uttered without overlap. Second, given the metabolic and eavesdropping costs associated with long, conspicuous sequences of easily localized calls (Klump & Shalter, 1984; Randler, 2012; Jones & Hill, 2001), it seems unlikely that such structures would be maintained if they were not functionally beneficial (still, a possible counterargument is that these long sequences could be a mere by-product of high arousal in the emitter). Finally, *D* and *D** share a repetitive, broadband structure and are emitted at similar intensity (~ 75dB, Suzuki 2017) and with a short inter-note interval. However, when we explored the acoustic properties of *D* and *D** notes, none were likely to trigger a masking effect (see also Appendix A12).

A third possible hypothesis is pragmatic in nature. We briefly mentioned that Schlenker et al. 2014, 2016*c* posit rules such as the Informativity Principle, according to which when several calls are appropriate, one should choose the most informative. But in their study of *pyow-hack* sequences in Putty-nosed monkeys, Schlenker et al. 2016*a,c* also proposed another principle, the Urgency Principle, "which mandates that calls that provide information about the nature/location of a threat must come before calls that don't" (see Appendix A13). These two pragmatic principles were in the background of theories that treated *pyow-hack* sequences as having no semantic structure. The Urgency Principle has also found support in the call sequences of Titi monkeys, as argued by Narbona Sabaté et al. 2022.

Schlenker et al. 2016*d* briefly mentioned that the Urgency Principle might account for Suzuki et al.'s findings: *ABC* is an alarm call and might thus provide some information about the nature/location of a threat, while the *D* part is a recruitment call, and thus the *ABC* part should come before the *D* part. Furthermore, the principle should apply productively to any sequence that is understood, hence also

*ABC-D** by contrast with *D*-ABC*. Japanese tits might fail to react to *D/D*-ABC* sequences not because they violate a principle of bird syntax, but because they violate the Urgency Principle. In fact, this can be interpreted in two ways. One is that the sequence is (pragmatically) deviant because it violates the Urgency Principle. Another is that in the *D/D*-ABC* order, the *ABC*-part is interpreted as not providing information about the nature/location of a threat (see Appendix A14). Either way, lack of reaction might be explained.

Turning to the semantics, the 'only one call' theory is again very implausible given that the birds generalize the rule from their system to new, artificially-constructed structures. But in view of the data discussed up to this point, the 'separate utterances view' is far more plausible. It takes *ABC* and *D/D** to each form a separate utterance. The calls are produced in close succession: ABC and D calls are naturally produced with a 0.5- to 0.15-s interval (Suzuki et al. 2018), and Suzuki et al. 2017 used a 0.1s interval to create their stimuli. As a result, ABC and D provide information about essentially the same time of utterance. In essence, *ABC* produced at time t conveys the information that there is an alarm at t, and *D/D** produced at t conveys information that the caller needs help at t. On the simplest theory, the two uttered in close proximity are thus expected to trigger an alarm-appropriate behavior, namely scanning, and a recruitment-appropriate behavior, namely approaching. Suzuki, Wheatcroft & Griesser, 2018 clearly state that in response to *ABC-D* sequences, the Japanese tits "produce both behaviors". They add:

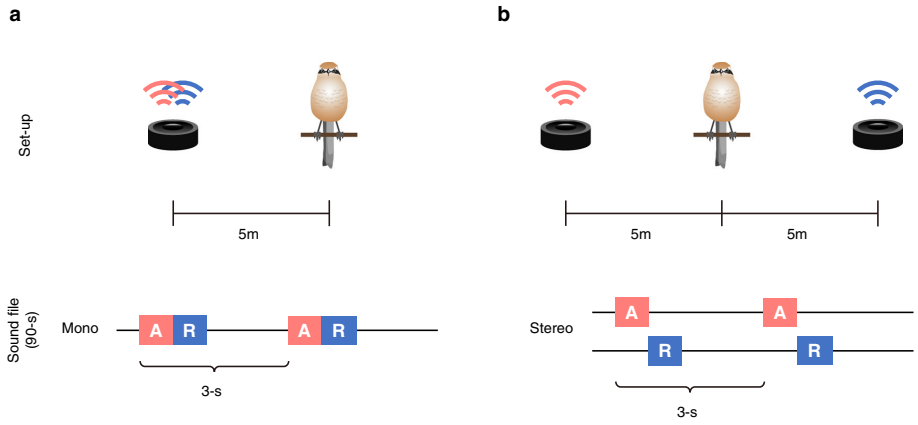
Importantly, tits do not first scan and then approach, as would be predicted if they perceived ABC-D sequences as linear, ordered strings. Instead, they progressively approach the sound source while continuously scanning.

However, one could expect a sequential response only if there were a long pause between *ABC*, produced at time t, and *D*, produced at t+d. But such isn't the case, as the calls are produced in close proximity. As for the fact that *D/D*-ABC* fails to trigger a reaction might, as mentioned, be due to the Urgency Principle (although we would need to know more to exclude purely acoustic reasons).

In addition, we should note that, on the assumption that the Japanese tits recognize ABC and D* as coming from two different sources (namely conspecifics and heterospecifics), the fact that they respond similarly to ABC-D and ABC-D* suggests that they aggregate the two sets of information independently from the source of the calls: whether ABC and D are from the same caller or from distinct species does not seem to alter the meaning extracted by the birds. This might appear to lend support to the 'separate utterances' theory. The latter is still faced with an important puzzle: why is ABC-D* effective while D*-ABC isn't if ABC and D* are perceived as separate utterances produced by different species? In particular, it seems unlikely that the Urgency Principle could constrain the ordering of two utterances made by different individuals, not to mention different species.

Still, at this point the 'separate utterances' theory is a live contender. But Suzuki and Mastumoto 2022 (published after the initial version of this piece was written) develop a new and remarkable argument to refute it. As a baseline, they set up an experiment in which an ABC-D sequence triggers mobbing of a predator model (a bullheaded shrike, a passerine bird), as illustrated in (16)a. Remarkably, they show that ABC-D sequences *fail* to be effective when ABC and D are played from different loudspeakers, as illustrated in (16)b (D-ABC sequences are ineffective as well under such conditions). This is precisely the opposite of what we saw in the human language case. If Ann says *It's hot* while Bill says *It's humid*, we naturally aggregate the information from the two sources. And it makes no difference whether the information is coming from one source (e.g., Ann) or two sources (Ann and Bill), as long as they are produced from locations that are close to each other (if Ann and Bill are on Zoom and are talking from different locations, aggregation of the information fails).

(16) Crucial part of the experimental design of Suzuki and Matsumoto 2022
Alarm-recruitment sequences are produced from a single loudspeaker (a) or from two loudspeakers (b).
Only (a) triggers mobbing behavior.



Suzuki and Matsumoto conclude that there is a crucial difference between separate utterances of ABC and then D, and a single utterance of ABC-D. This is possibly the most important argument to have been published as part of this debate. The main objection that comes to mind is that the two speakers might be too distant (10 meters) to allow for the contents they broadcast to be aggregated. In the human case, if Ann says inside the house that *it's hot* while Bill says outside the house that *it's humid*, one cannot infer that it's both hot and humid in one and the same place. But within Suzuki and Matsumoto's experimental set-up, both (16)a (= the 1-speaker condition) and (16)b (= the 2-speaker condition) involve speakers that are 5 meters away from a predator model that could license mobbing behavior (the authors also note that 10m "is a natural distance between two individuals within a flock"). This distance doesn't prevent the target birds from relating ABC-D to the predator in the 1-speaker condition, and thus it's unclear why it should have a different effect in the 2-speaker condition. The logic of Suzuki and Matsumoto's argument thus appears to be strong.

The dialectical situation is summarized in (17), where the counterarguments to the 'separate utterances' theory are divided into A (corresponding to Suzuki et al. 2016, 2017, 2018) and B (corresponding to Suzuki and Matsumoto 2022—the most decisive finding in our view, and thus the strongest argument against the 'separate utterances' view).

(17) Japanese tit *ABC-D* sequences involve a compositional rule

Deflationary Theory 1—'Only one expression'
This is implausible in view of (15).

Deflationary Theory 2—'Separate utterances': *ABC* is a separate utterance, *D* is a separate utterance, with the following meanings:
ABC produced at time t means: There is an alarm at t.
D produced at time t means: The receiver's presence is needed at t.

Counterargument A (Suzuki et al. 2018): "Tits do not first scan and then approach, as would be predicted if they perceived ABC-D sequences as linear, ordered strings."
Counter-counterargument 1: The sequential behavior could be predicted only if ABC and D were separated by a long pause, which isn't the case.
Counter-counterargument 2: Subjects react similarly to ABC-D and ABC-D*, although the latter is composed of utterances from two emitters, suggesting that they aggregate the information independently from the source of the call.

Counteragument B (Suzuki and Matsumoto 2022): When ABC and D are played from different loudspeakers, ABC-D fails to trigger the target behavior.
Counter-counterargument: Birds might fail to aggregate the information because the two speakers are too far apart (10m), thus blocking the inference that alert and recruitment are needed *in the same place*. But this worry might be far-fetched because the speakers are relatively close (5m) to the predator model that should license the mobbing behavior.

We summarize our discussion of Japanese tit ABC-D sequences in (18).

(18) Summary of arguments and objections: morphosyntax and compositionality in Japanese tit *ABC-D* mobbing calls

Nature of the rule	**Main arguments**	**Alternative 1: 'Only one expression'**	**Alternative 2: 'Separate utterances'**
Morphosyntactic	• Ordering: Japanese tits react to *ABC-D* but not to *D-ABC*. • Productivity: They extends this to hybrid sequences *ABC-D** (versus *D*-ABC*) that do not resemble their own.	Implausible, as this wouldn't account for the productivity of the rule—unless initial *D* and *D** acoustically mask *ABC*.	Ordering restrictions might come from non-syntactic principles: • acoustic if *D* and *D** acoustically mask *ABC;* • pragmatic (Urgency Principle)
Semantic	A. *ABC-D*/*D** gives rise to the simultaneous production of scanning and approaching. B. ABC-D triggers the target behavior when ABC and D are produced from the same source but not when they are produced from distinct but spatially close sources.	Implausible as: • this wouldn't account for the productivity of the semantic effect; • the effect of the complex call is directly related to the effect of its parts.	A. Whether the semantics is imperative or declarative, *ABC-D*/*D** produced in quick succession provide an order/a statement about a single moment t, hence simultaneity of the response is expected. B. However, the distinction between ABC-D played back from one source (effective) and from two spatially close sources (ineffective) makes an analysis based on separate utterances implausible.

Several important points should be kept in mind in future research. First, the 'separate utterances' view is not without a potential reply. It could invoke the Informativity Principle and competition among calls to explain Suzuki and Matsumoto's new finding. An utterance of *ABC* alone, or of *D* alone, might conceivably compete with the utterance of *ABC-D*, which is more informative than either of its component parts (each viewed as a separate utterance). If so, each separate utterance would yield a non-*ABC-D* inference, and this would explain the absence of a mobbing reaction. But more analytical work is needed to make this view precise.

Second, a full analysis would need to explain the details of the target birds' reactions in Suzuki and Matsumoto's experiment. In the 2-speaker condition, one might expect that approach is triggered in the direction of the speaker that plays D, since D is a recruitment call (and it was indeed shown in Suzuki et al. 2017 [Figure 3] that D played alone triggers speaker approach). As we understand it, in their experiment, Suzuki and Matsumoto assessed whether subject birds approached the *predator,* but not whether the approached the *speaker*, something that could be assessed in future experiments.

Third, in none of the studies reviewed here do Suzuki and colleagues state the specific compositional rule that they take to be involved in the interpretation of *ABC-D* calls. A precise statement of (i) the meaning of ABC and D, and (ii) the compositional rule would be helpful to assess whether the detailed findings are explained (e.g. why D fails to trigger approach in Suzuki and Matsumoto's experiment), and to delineate the compositional theory from the 'separate utterances' view. By comparison, in the analysis of Campbells' *-oo*, the argument against 'separate utterances' was in part based on a precise proposal about the way the suffix affects the meaning of the calls it attaches to (see Appendix A9).

Fourth, putting all the results together, we have an intriguing conceptual situation. On the one hand, Japanese tits have no trouble integrating into a single utterance calls from two different species, since they treat a hybrid sequence ABC-D* in the same way as normal ABC-D sequences. On the other hand, they refuse to integrate into a single utterance ABC and D calls coming from two nearby locations.

Why they are so sensitive to caller location but so insensitive to caller species is a mystery and ought to be explored in future research.

Finally, we adopted in our discussion the view of call meaning espoused by a long line of research on monkey calls, from the field experiments Seyfarth, Cheney & Marler, 1980 to the formal analyses of Schlenker et al. 2016*c*. According to these diverse analyses, calls have declarative meanings and thus provide information about the world. But an alternative possibility is that calls are in essence imperatives, telling the receivers what to do irrespective of the environment (here we have in mind imperatives in a very narrow sense: an imperative fully determines the action to be taken irrespective of the state of the world; see Appendix A15). In fact, data collection and meaning attribution in the field of animal behavior traditionally assigned imperative meanings to animal signals: the meaning of a signal is mostly determined by its context of production and the behavioral response of the receiver (Fröhlich al. 2019; Hobaiter, Graham & Byrne, 2022; Jäger 2016). Seyfarth et al. 1980 discussed data that made this possibility unlikely for Vervet monkeys: a Vervet that heard an eagle alarm call had differential reactions depending on its own position. For instance, if it was in a tree, it sometimes looked down; but if it was on the ground, it didn't do so as often. The situation is different with Suzuki et al.'s Japanese tit data: in view of the observed reactions, the 'separate utterances' view could posit that *ABC* produced at t is an imperative meaning in essence: *Scan!* (i.e. *now, at t)* while *D* produced at t means: *Approach!* (i.e. *now, i.e. at t).* As in the declarative analysis, the fact that both imperatives are produced at the same time suffices to explain why the two behaviors co-occur, rather than appearing in a sequence. Similarly, nothing precludes an imperative analysis in Suzuki and colleagues' target theory either, as long as there is a semantic composition of some sort (to explain the results of Suzuki and Mastumoto 2022). By contrast, data available about the Southern pied-babbler make an analysis based on imperatives (in a narrow sense) less plausible, as we will now see (see also Appendix A16).

IV. Syntax and compositionality in the mobbing sequences of the Southern Pied babbler

Engesser et al. (2016, 2020) display related patterns in the Southern pied babbler (*Turdoides bicolor*), by no means a close cousin of the Japanese tit (the two species diverged approximately 30 million years ago [Selvatti, Gonazaga & Moraes Russo, 2015]). The striking point of convergence between the two species is that they form mobbing sequences with a combination *alarm–recruitment*. Unlike Suzuki et al., Engesser et al. find the same types of responses to *alarm–recruitment* sequences and to *recruitment–alarm* sequences. But they find something that Suzuki and colleagues didn't, namely that *alarm–recruitment* sequences give rise to very different behavioral responses from their component parts when the parts are presented alone. They reason that this shows that a compositional rule is involved:

> To investigate whether babblers process the sequence in a compositional way, we conducted systematic experiments, playing back the individual calls in isolation as well as naturally occurring and artificial sequences. Babblers reacted most strongly to mobbing sequence playbacks, showing a greater attentiveness and a quicker approach to the loudspeaker, compared with individual calls or control sequences. We conclude that the sequence constitutes a compositional structure, communicating information on both the context and the requested action. (Engesser et al. 2016)

The question is whether this conclusion (namely that Southern pied babbler calls display compositionality) is justified.

Engesser et al.'s main finding is that the target birds "responded most strongly to playbacks of mobbing sequences, revealing the highest attentiveness and fastest approach toward the sound source" compared to alarm calls alone or recruitment calls alone. According to the authors, these results support their "hypothesis that the call sequence tested conforms to the definition of basic compositional syntax, with the high vigilance response to mobbing sequences and the fast approach to the loudspeaker being directly related to the contextual information and function of both individual calls".

To assess Engesser et al.'s findings, we pit their claims about syntax and compositionality against our two usual deflationary alternatives—'only one expression' and 'separate utterances'. While Engesser et al. do not propose a specific compositional rule, we can nonetheless assess the strength of their argument.

On the syntactic side, there is an argument against the 'only one expression' analysis, but no argument against the 'separate utterances' view, as is summarized in (19). First, to refute the claim that mobbing sequences are made of a single expression, the authors confirm that the calls forming the combination are acoustically identical to the calls occurring in isolation, assuming such accidental resemblance is highly unlikely. In addition, they conduct playbacks of artificially created *recruitment-alarm* sequences and obtain similar reactions as to sequences that appear in the natural order. Second, however, there is no argument against treating mobbing sequences as made of two separate utterances. Not only is there no argument from ordering restrictions, unlike in Suzuki et al. 2016; there is an explicit argument to the opposite conclusion, since adult Southern pied-babblers react both to *alarm–recruitment* and to *recruitment–alarm* sequences. The authors take this to show that the compositional semantics can be "open", but a simpler explanation is that the calls constitute separate utterances and do not involve (non-trivial) compositionality to begin with.

(19) Southern pied babbler *alarm–recruitment* sequences involve a syntactic rule

Deflationary Theory 1—'Only one expression'
Counterargument: Southern pied babblers respond both to naturalistic alarm-recruitment sequences and to artificial recruitment-alarm sequences, suggesting that alarm-recruitment is made of two parts.

Deflationary Theory 2—'Separate utterances'
No counterargument—in fact, the Southern pied babbler reacts to recruitment–alarm sequences (unlike the Japanese tit).

Still, there are fascinating results on the semantic side. Unlike what we saw in Suzuki et al.'s data, the effect of an *alarm–recruitment* sequence is very different from the effect of its component parts, both in terms of vigilance and movement in the direction of the speaker. If the mobbing sequence is in fact made of a single expression, there is of course nothing particularly surprising to explain, since we are just talking about three calls, *alarm*, *recruitment,* and *mobbing*, which accidentally happen to have some acoustic parts in common. By contrast, if mobbing sequences are made of two calls, there is something to be explained. The authors consider the hypothesis that the component parts are interpreted as separate utterances, but they rule it out:

We are confident that we can rule out alternative explanations related to a sequential or additive processing of calls, because responses to played back mobbing sequences exceeded those elicited by the independent calls or their sum.

The dialectical situation is summarized in (20).

(20) Southern pied babbler *alarm–recruitment* sequences involve a compositional rule

Deflationary Theory 1—'Only one expression'
Counterargument (identical to the argument against Deflationary Theory 1 in the syntactic case): Southern pied babblers respond both to naturalistic alarm-recruitment sequences and to artificial *recruitment-alarm* sequences, suggesting that alarm-recruitment is made of two parts.

Deflationary Theory 2—'Separate utterances':
alarm produced at time t means: There is an alarm at t.
recruitment produced at time t means: The receiver's presence is needed at t.
Counterargument: Target birds react far more strongly to *alarm–recruitment* than to *alarm* or to *recruitment*.
Counter-counterargument: The counterargument against 'separate utterances' is valid if calls have an imperative meaning. But it is *not* valid if calls have a declarative meaning: the most appropriate reaction to the information that *there is an alarm and receiver presence is needed at t* may be very different from the most appropriate reaction to the individual parts.

Here it is important to be more specific about two possible views of the meanings of calls, as stated in (21).

(21) Two views of the meaning of calls
a. Narrow imperative analysis: a call is an instruction to adopt a certain behavior, irrespective of the state

of the environment.
b. Declarative analysis (Seyfarth et al. 1980): a call provides information about the world

In our discussion of Suzuki et al.'s Japanese tit data, we argued that a declarative and an imperative analysis alike could account for the data on the 'separate utterances' view. The situation is different for Engesser et al.'s findings: if *alarm* and *recruitment* are separate utterances, an analysis based on imperatives in a narrow sense is unlikely to account for the data (as before, by 'imperatives in a narrow sense', we mean expressions that trigger the same behavioral response irrespective of the environment). On this view, *alarm–recruitment* produced at t is a double instruction to do what *alarm* mandates at t and what *recruitment* mandates at t. It is natural to think that the two calls uttered in close succession should display roughly the sum of the effects of the independent calls, but not much larger effects, as is in fact found. It is important to note the authors seem to adopt a narrow imperative-based analysis in this case (Engesser et al., 2016, Townsend et al., 2018).

But things are different on the declarative view, on which a call typically provides propositional information about the world. It is not hard to find cases in which separate utterances processed in a close succession are far more alarming than their member parts. This point was made in Schlenker et al. 2016*d*:

To take a human analogy: *Little Johnny is on the pedestrian crossing* might not trigger a human alarm; nor need the sentence *There is a car coming* be alarming when uttered on its own. But the conjunction *Little Johnny is on the pedestrian crossing and there is a car coming* might require immediate action: the effect of the conjunction is not additive in terms of the effects of the conjuncts.

To put it in more bird-compatible terms, consider the utterance, *You should come*. My interlocutor might comply if they are polite, but my utterance might not be enough to move them. Similarly, *There's someone I don't know here* might raise my interlocutor's level of alarm, but this might not be enough to move them either. But things might be different if I utter these sentences in close succession: *There's someone I don't know here*. *You should come*. A reasonable inference would be that I need your support in case the stranger turns out to be aggressive (see also Appendix A17).

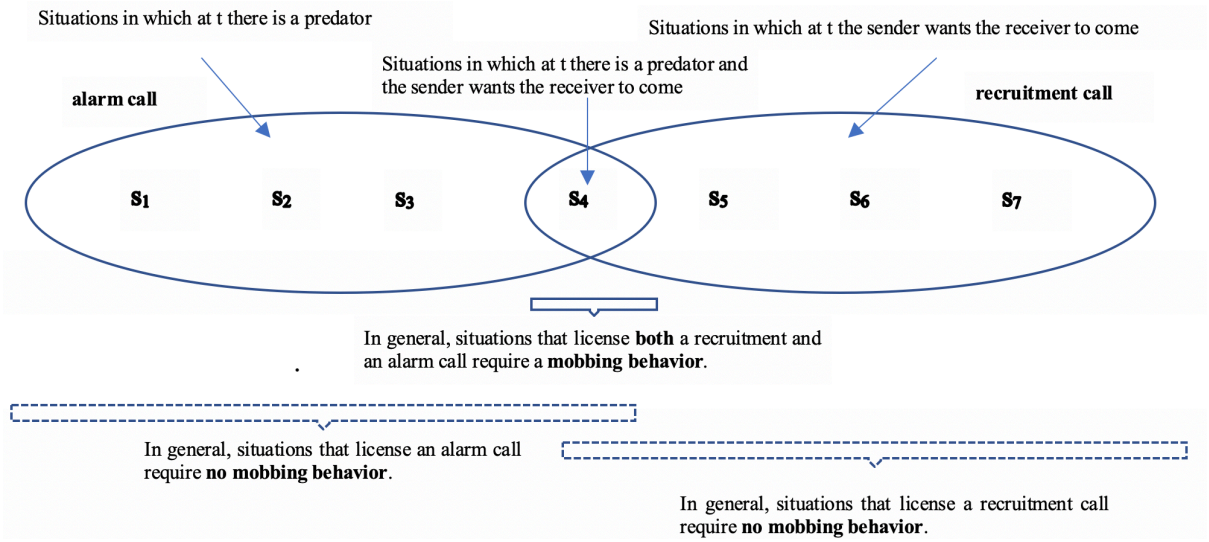
A salient possibility is that this is in essence what happens in mobbing sequences: the alarm call combined with the recruitment call mostly singles out situations in which the source of the alarm is a predator that should be mobbed.

The logical point can be viewed as follows. A call produced at a time t, just like a sentence uttered at a time t, is true in some situations but not others. Taken together, two calls uttered in close succession around the same time t will be true of the intersection of the situations that make each call true. For simplicity, we'll assume that the alarm call is true in situations in which there is a predator, while the recruitment call is true in situations in which the sender wants the receiver to come, as is depicted in **Error! Reference source not found.**. On the standard declarative view, the calls modify the information state of the receiver, who presumably adopts the behavior most likely to be appropriate in view of the information it has. The optimization problem might be simple or complex (e.g. the receiver might adopt the behavior that's optimal in *most* of the relevant situations, or the behavior that will maximize its expected utility, or something else). The important point is that the behavior depends on the information state, or in other words on the entire set of situations compatible with the receiver's knowledge. Now it is likely that most situations that license an alarm call do not require mobbing, as it is only certain predators in certain situations that are best dealt with through mobbing. Similarly, most situations that license a recruitment call do not require mobbing (for instance, they may involve mating or foraging). On the other hand, most of the situations that license both an alarm call and a recruitment call require a mobbing behavior, as is also illustrated in.

To be very concrete, suppose there are just 7 possible situations, labelled s_1, …, s_7 in (22), and assume they are equiprobable. In s_1, s_2, s_3, there is a disturbance that does not require mobbing (e.g. a non-predator such as another passerine, a predator that is too far to be worth mobbing, an unknown loud sound); in s_5, s_6, s_7, there is a reason to recruit conspecifics, for instance a food source, but no predator to be mobbed. Only in s_4 is there a predator that needs to be mobbed, which licenses both the alarm call and the recruitment call. If all 7 situations are equally likely, upon hearing an alarm call alone, the receiver can only infer that there is a 25% chance that mobbing is called for, as this is the case in situation s_4 but not in the equally likely situations s_1, s_2, s_3. This low probability might justify *not*

adopting the behavior. Similarly, upon hearing a recruitment call alone, there is again just a 25% chance that mobbing is called for, as this is the case in situation s_4 but not in the equally likely situations s_5, s_6, s_7. But upon hearing an alarm call and a recruitment call in close succession, there is a 100% chance that mobbing is called for, as mobbing is definitely called for in situation s_4.

(22) Alarm and recruitment calls uttered separately versus in close succession at time t



We conclude that even if one grants that mobbing calls are made of two parts, it does not follow that a syntactic or semantic rule is involved. These two parts might be separate utterances with a declarative semantics, and this might be enough to explain why *alarm* and *recruitment* produced in close succession have a radically different effect than either call produced alone. On the other hand, on a narrow imperative analysis (which Townsend et al. 2018 seem to assume), it is true that the two calls are unlikely to form separate utterances. One could thus seek to refute the 'separate utterances' theory by showing on independent grounds that the calls have a meaning that falls under the narrow imperative analysis. For this line of research to be explicit, one would also need to propose a specific compositional rule that the birds use to combine the two imperatives (alarm and mobbing).

Here too, we summarize the dialectical situation in a table, given in (23).

(23) Summary of arguments and objections: morphosyntax and compositionality in Southern pied-babbler *alarm–recruitment* mobbing calls

Nature of the rule	**Main arguments**	**Alternative 1: 'Only one expression'**	**Alternative 2: 'Separate utterances'**
Morphosyntactic and semantic	The behavior effect of *alarm–recruitment* is not additive relative to the behavior effects of *alarm* and *recruitment*.	Implausible, as the birds react in the same way to alarm-recruitment sequences and to artificial *recruitment-alarm* sequences, suggesting that alarm-recruitment is made of two parts.	• Non-additivity might be unexpected on an imperative semantics. • Non-additivity might be expected on a declarative semantics: both calls are produced at the same moment, and situations in which there is both an alarm and a need for receiver presence might generally require mobbing.

V. Syntax and compositionality in the mobbing sequences of the Great tit

Taken together, the Great tit studies of Dutour et al. 2020 and Salis et al. 2021*a* combine (parts of) the arguments discussed above for the Japanese tit and the Southern pied-babbler (we disregard some subtleties, especially the effect of seasonality; see for instance Salis et al. 2021b and Dutour et al. 2019b, 2022). The Great tit (*Parus major*) is closely related to the Japanese tit (estimated divergence date between the Parus major and Parus minor groups = 3Myr, Kvist et al., 2003, 1999). One crucial

innovation is that both studies investigate Great tit reactions to Great tit calls but also to the calls of Black-capped chickadees, an allopatric species that lives in North America. The target Great tits, from France, could not have been in contact with Black-capped chickadees. Still, to interpret the data, it is important to keep in mind that the calls of the two species might be sufficiently close that acoustic similarity might account for some of the findings (the two species have a divergence date of approximately 14 million years ago, Päckert et al. 2007; see also Appendix A18).

On the syntactic side, Dutour et al. 2020 show that playback of a Great tit *alarm–recruitment* sequence triggers a mobbing behavior in the Great tits, but that Chickadee *alarm'–recruitment'* sequence does as well. By contrast, an inverse Chickadee or Great tit *recruitment'–alarm'* does not produce as much reaction (Salis et al., 2021*b*). On the semantic side, Salis et al. 2021*a* show that Great tits display far more mobbing reactions to Great tit *alarm–recruitment* sequences than to their component parts alone. And this finding extends to Great tit reactions to Chickadee *alarm'–recruitment'* sequences compared to their parts alone.

To present things in greater detail, let us start with Great tit syntax. The results obtained by Dutour and colleagues (Dutour et al., 2020, Salis et al., 2021*b*) show that Great tit reactions to *alarm–recruitment* sequences trigger more vigilance, more signs of excitement such as wing-flicking and tail wagging, and (for Salis et al., and only in winter) more approach than *recruitment–alarm* sequences, irrespective of whether these sequences are made of conspecific Great tit calls or of allopatric Chickadee calls.

Importantly, the logic of the experiment is completely different from Suzuki et al.'s use of hybrid sequences. The latter, made of Japanese tit *ABC* alarm calls combined with a neighbor's *D** recruitment call, did not acoustically resemble the Japanese tit's *ABC-D* mobbing call (as *D** was very different from *D*). But no such argument is offered in Dutour et al.'s Great tit experiment. What is shown is that Great tits discriminate between *alarm'–recruitment'* and *recruitment'–alarm'*. But since *alarm'* is acoustically similar to *alarm* and *recruitment'* is acoustically similar to *recruitment*, everything might be driven by the Great tits' knowledge of their own calls, combined with a similarity measure that ensures that they assimilate Chickadee calls to their own calls (see also Appendix A19).

As a result, contrary to the Japanese tit argument in (15) above, there is no argument against the 'only one expression' theory, as summarized in (24). On the other hand, as in the original arguments made for the Japanese tit, the case against the 'separate utterances' theory is that Great tits discriminate between the order *alarm'–recruitment'* and the order *recruitment'–alarm'*. But as discussed for the Japanese tit, this rule need not be syntactic. Importantly, as yet there is no Great tit version of the argument from playbacks from spatially distinct sources, which yielded Suzuki and Matsumoto's (2022) crucial argument against the 'separate utterances' theory.

(24) Great tit *alarm–recruitment* sequences involve a syntactic rule

Deflationary Theory 1—'Only one expression'
No counterargument. In particular, the fact that Great tits react to allopatric *alarm'–recruitment'* but not to *recruitment'–alarm'* might be entirely driven by acoustic similarity to conspecific calls and is not an argument for the application of a productive rule.

Deflationary Theory 2—'Separate utterances'
Counterargument: Great tits react to *alarm–recruitment* and *alarm'–recruitment'* sequences but not as strongly to *recruitment'–alarm'* sequences.

Counter-counterargument: the distinction might be due to a non-syntactic rule, as the lack of reaction to *recruitment–alarm* and *recruitment'–alarm'* might in principle be due to (i) lack of familiarity, (ii) acoustic masking by the large frequency bandwidth of *alarm* and *alarm'*, (iii) a pragmatic principle, e.g. the Urgency Principle.

In our discussion of the Japanese tit data, we argued that lack of familiarity can't explain the contrast between the reactions to hybrid *ABC-D** sequences and hybrid *D*-ABC* sequences: all are unfamiliar but Japanese tits react to the former much more than to the latter. In the case of the Great tit data, things are different, since *alarm–recruitment* is definitely familiar while *recruitment–alarm* isn't (see also Bolhuis et al. 2018*a* for a similar argument). Furthermore, since Chickadee *alarm'* and *recruitment'* acoustically resemble Great tit *alarm* and *recruitment* respectively, to a Great tit ear,

natural Chickadee *alarm*'–*recruitment*' sequences presumably resemble a familiar (conspecific) call while reversed Chickadee *recruitment*'–*alarm*' sequences don't.

In addition, the non-syntactic rules discussed in relation to the Japanese tit data are contenders in the present case as well. Dutour et al. 2020 explicitly discuss the possibility that acoustic masking underlies the phenomenon (see also Appendix A20):

> Another explanation for our results could be a perception bias (i.e., the first D notes of the call mask the notes that follow them, preventing the receiver from perceiving the second part of the call; Grafe 1996; Klump and Gerhardt 1992). Indeed, D notes, which have large frequency bandwidths and are produced in long, repetitive sequences, may mask the FME notes (Marler 1955; Brown and Handford 1996) given the relative short delay between both sequences. As a result, tits may no longer perceive FME notes when they are artificially placed after D notes.

Great tits' D sequences are relatively loud repeated and broadband structures which resemble those of Japanese tits. Following the same rationale as above, we have no reason to believe that a masking effect is the most likely explanation for these results (see Appendix A11). And as mentioned in relation to the Japanese tit, yet another possibility is that the lack of reaction to reversed sequences is due to the fact that these violate a pragmatic principle, such as the Urgency Principle.

Turning to the semantics, Salis et al. 2021*a* replicate the argument of Engesser et al. 2016 regarding the non-additivity of mobbing reactions to Great tit *alarm–recruitment* sequences relative to its individual parts. They further show that this finding extends to allopatric *alarm*'–*recruitment*' sequences of the Chickadee, but as before the acoustic similarity among conspecific and allopatric calls makes it hard to argue that anything but acoustic similarity is at stake.

In sum, neither the 'only one expression' theory nor the 'separate utterances' theory can be ruled out, as is summarized in (25). It is true that *if* calls have an imperative semantics, *alarm–recruitment* is unlikely to be made of separate utterances. But this conclusion does not follow if calls have a declarative semantics. Furthermore, as in our discussion of Great tit syntax, the fact that the target birds react to allopatric calls does not show that they apply a compositional rule, as reactions might be driven by the acoustic similarity between allopatric and conspecific calls.

(25) The Great tit *alarm–recruitment* sequences involves a compositional rule

Deflationary Theory 1—'Only one expression'
No counterargument. In particular, the fact that Great tits react to allopatric *alarm*'–*recruitment*' is not an argument for productivity as reactions might be driven by acoustic similarity to conspecific calls.

Deflationary Theory 2—'Separate utterances': *alarm* is an utterance, *recruitment* is a separate utterance,
alarm produced at time t means: There is an alarm at t.
recruitment produced at time t means: The receiver's presence is needed at t.

Counterargument: target birds react far more strongly to *alarm–recruitment* than to *alarm* or to *recruitment*, and to *alarm*'–*recruitment*' than to *recruitment*'–*alarm*'.

Counter-counterarguments [as in the discussion of the Southern pied-babbler]:
–The argument against 'separate utterances' is valid if calls have an imperative meaning. But it is *not* valid if calls have a declarative meaning: the most appropriate reaction to the information that *there is an alarm and receiver presence is needed at t* may be very different from the most appropriate reaction to the individual parts.
–The fact that the Great tit reacts to allopatric calls does not change the argument because this might be driven by acoustic similarity to conspecific calls.

Salis et al. 2021*a* take their results to support the idea of animal compositionality, which may be correct if the calls are semantically interpreted as imperatives (in a narrow sense), but it seems hard to rule out that the calls are interpreted as separate utterances if their meaning is declarative. We summarize the dialectical situation in (26).

(26) Summary of arguments and objections: morphosyntax and compositionality in Great tit *alarm–recruitment* mobbing calls

Nature of the rule	**Main arguments**	**Alternative 1: 'Only one expression'**	**Alternative 2: 'Separate utterances'**
Morphosyntactic	• Ordering: Great tits react to *alarm–recruitment* but not to *recruitment–alarm*. • 'Productivity': They extends this to allopatric *alarm'–recruitment'* (versus *recruitment'–alarm'*)	Possible, as the apparent 'productivity' might entirely be driven by acoustic similarity between *alarm–recruitment* and *alarm'–recruitment'*.	Ordering restrictions might come from non-syntactic principles: • acoustic if *recruitment* and *recruitment'* acoustically mask *alarm* and *alarm'*. • pragmatic (Urgency Principle)
Semantic	• The behavior effect of *alarm–recruitment* is not additive relative to the behavior effects of *alarm* and *recruitment*. • 'Productivity': This extends to allopatric *alarm'–recruitment'*	Possible, as the apparent 'productivity' might entirely be driven by acoustic similarity between *alarm–recruitment* and *alarm'–recruitment'*.	• Non-additivity might be unexpected on an imperative semantics. • Non-additivity might be expected on a declarative semantics: both calls are produced at the same moment, and situations in which there is both an alarm and a need for receiver presence might generally require mobbing.

VI. Recommended steps and future prospects

Setting aside the new findings of Suzuki and Matsumoto 2022, the results reviewed here on *alarm–recruitment* mobbing sequences remain ambiguous when arguing for animal syntax or compositionality. On the syntactic side, mobbing sequences can be analyzed as being made of separate utterances whose order is constrained by non-syntactic principles (e.g. acoustic or pragmatic ones; as noted by a reviewer, the order could also be fixed by a 'template', as in Miyagawa & Clarke 2019—but since a template just stipulates the order of two call types, its explanatory status is unclear). On the semantic side, the non-additivity of the behavioral effect of *alarm–recruitment* relative to its component parts is expected if these parts are separate utterances that provide information about the same moment (due to their temporal proximity) and have a declarative semantics, providing information about the world. On the other hand, if *alarm* and *recruitment* have an imperative semantics (in a narrow sense), the non-additivity is unexpected and might require a compositional rule. The importance of providing a declarative rather than imperative semantics for bird calls dovetails with conclusions reached by Seyfarth et al. 1980 about the calls of Vervet monkeys. Contrary to what a narrow imperative-based semantics would lead one to expect, the monkeys displayed different reactions depending on the context: a Vervet that heard a leopard alarm call while on the ground typically ran for cover or looked up, and not down (unlike for a snake alarm call); but it could look down if it heard the leopard call while in a tree.

Crucially, Suzuki and Matsumoto's recent results change the dialectical situation. Suzuki and colleagues had earlier shown (by way of hybrid sequences) that Japanese tits use a productive rule to understand ABC-D sequences, but they had not convincingly ruled out an analysis based on 'separate utterances'. Their new findings offer a powerful argument against this possibility, as Japanese tits fail to react to ABC followed by D, played from two different (but spatially close) locations. The final result is puzzling, however: if genuine syntactic integration is at stake in this case, Japanese tits can treat calls coming from different species as a single utterance, but they cannot do the same with calls coming from a single species but from slightly different locations. More work will be needed to fully understand the situation.

On a methodological level, our main point is that arguments should be given by explicitly pitting the main claim—existence of a syntactic rule, existence of a compositional rule of interpretation—against two deflationary alternatives, the 'only one expression' theory and the 'separate utterances' theory. General criteria of syntax and compositionality are of course helpful, but in the end the test of arguments in favor of animal syntax or compositionality lies in their ability to rule out both of these deflationary alternatives. In addition, if one wishes to argue for a compositional rule, stating the rule explicitly might help distinguish it from the 'separate utterances' view, i.e. from trivial compositionality.

A summary of the arguments discussed in this piece for diverse animal sequences appears in (27); the table collates summaries that appear at the end of each section, and it adds a similar summary for Campbell's monkeys' *boom* (mentioned in passing at the beginning of this piece).

(27) **Summary of the main arguments**

Case study	**Nature of the rule**	**Main arguments**	**Alternative 1: 'Only one expression'**	**Alternative 2: 'Separate utterances'**
Campbell's monkey *-oo* suffix	Morphosyntactic	Pattern: *-oo* is never found on its own but can be added to *krak* and *hok*.	• A pattern with 2 instances (*krak-oo*, *hok-oo*) could be an accident. • No argument for productivity exists.	Implausible as: *-oo* never appears on its own (but see Sauerland 2016)
	Semantic	Pattern: *-oo* plausibly modifies the meaning of *krak* in the same as it does the meaning of *hok*.	• A pattern with 2 instances (*krak-oo*, *hok-oo*) could be an accident. • Auxiliary assumptions are needed to get the compositional analysis to work.	• Implausible as: *-oo* never appears on its own (but see Sauerland 2016) • On some analyses of the meaning, *hok-oo* does not entail the purported meaning of *-oo* (≈ non-serious alarm)
Campbell's monkey *boom*	Morphosyntactic	*Boom boom* appears at the beginning of sequences.	Probably implausible (sequences can be very long, and there is a long interval between two booms (see also Appendix A21).	The ordering restriction might have an articulatory source, as time and energy are needed to fill an air sac.
	Semantic	n/a [no argument has been given that a compositional rule is needed]		
Japanese tit *ABC-D* mobbing calls	Morphosyntactic	• Ordering: Japanese tits react to *ABC-D* but not to *D-ABC*. • Productivity: They extends this to hybrid sequences *ABC-D** (versus *D*-ABC*) that do not resemble their own.	Implausible, as this wouldn't account for the productivity of the rule—unless initial *D* and *D** acoustically mask *ABC*.	Ordering restrictions might come from non-syntactic principles: • acoustic if *D* and *D** acoustically mask *ABC;* • pragmatic (Urgency Principle)
	Semantic	A. *ABC-D*/*D** gives rise to the simultaneous production of scanning and approaching. B. ABC-D triggers the target behavior when ABC and D are produced from the same source but not when they are produced from distinct but spatially close sources.	Implausible as: • this wouldn't account for the productivity of the semantic effect; • the effect of the complex call is directly related to the effect of its parts.	A. Whether the semantics is imperative or declarative, *ABC-D*/*D** produced in quick succession provide an order/a statement about a single moment t, hence simultaneity of the response is expected. B. However, the distinction between ABC-D played back from one source (effective) and from two spatially close sources (ineffective) makes an analysis based on separate utterances implausible.

Southern pied-babbler *alarm–recruitment* mobbing calls	Morphosyntactic and semantic	The behavior effect of *alarm–recruitment* is not additive relative to the behavior effects of *alarm* and *recruitment*.	Implausible, as the birds react in the same way to alarm-recruitment sequences and to artificial *recruitment-alarm* sequences, suggesting that alarm-recruitment is made of two parts.	• Non-additivity might be unexpected on an imperative semantics. • Non-additivity might be expected on a declarative semantics: both calls are produced at the same moment, and situations in which there is both an alarm and a need for receiver presence might generally require mobbing.
Great tit *alarm–recruitment* mobbing calls	Morphosyntactic	• Ordering: Great tits react to *alarm–recruitment* but not to *recruitment–alarm*. • 'Productivity': They extends this to allopatric *alarm'–recruitment'* (versus *recruitment'–alarm'*)	Possible, as the apparent 'productivity' might entirely be driven by acoustic similarity between *alarm–recruitment* and *alarm'–recruitment'*.	Ordering restrictions might come from non-syntactic principles: • acoustic if *recruitment* and *recruitment'* acoustically mask *alarm* and *alarm'*. • pragmatic (Urgency Principle)
	Semantic	• The behavior effect of *alarm–recruitment* is not additive relative to the behavior effects of *alarm* and *recruitment*. • 'Productivity': This extends to allopatric *alarm'–recruitment'*	Possible, as the apparent 'productivity' might entirely be driven by acoustic similarity between *alarm–recruitment* and *alarm'–recruitment'*.	• Non-additivity might be unexpected on an imperative semantics. • Non-additivity might be expected on a declarative semantics: both calls are produced at the same moment, and situations in which there is both an alarm and a need for receiver presence might generally require mobbing.

Our practical recommendations are summarized in (28).

(28) **How to argue for animal syntax and compositionality**

Step 1: State 3 competing theories

Rich theory: There is syntax/compositionality.
Deflationary Theory 1—'Only one expression'
Deflationary Theory 2—'Separate utterances' (= trivial compositionality)

If arguing for compositionality, *state the compositional rule in detail*, thus helping to distinguish it from trivial compositionality.

Step 2: Compare the predictions of the three theories, and the plausibility of any auxiliary hypotheses they might need.
–To refute Deflationary theory 1, one can for instance construct an argument based on pattern or productivity by showing that the same component calls appear in other naturalistic or artificial constructions and give rise to the meaning predicted by the Rich theory.
–To refute Deflationary theory 2, one can for instance show that some of the component parts cannot occur on their own (as for the Campbell's *-oo* suffix), or that the meaning obtained cannot be analyzed as the conjunction of the component parts. Another argument against this theory is to show that subjects react to a combined call from one source (using stimuli rebuilt artificially to control for any effect of manipulation) but not to a call created by concatenating two units from distinct sources (this is the argument in Suzuki and Matsumoto 2022).

In the end, one needs to pit fully specified theories against each other, and weigh the plausibility of any auxiliary assumptions one might need to make them work.

Two further directions could be further developed in future research. First, we highlighted that some instances of Deflationary Theory 2 ('separate utterances') crucially rely on a declarative analysis of call meanings. One way to argue for a (non-trivial) compositional analysis would be to show that some or all of the calls involved a semantics that falls under what we called the 'narrow imperative analysis'. This requires developing clear criteria for (different types of) imperative vs. declarative meanings in animals, an important but non-trivial task for the future (see for instance Steinert-Threlkeld, Schlenker & Chemla 2021 for discussion, and Appendix A22).

Second, as highlighted by Suzuki and Matsumoto 2022, declarative versions of 'separate utterances' rely on the fact that the informational content of two calls C_1 and C_2 can be aggregated without being thereby combined by a non-trivial compositional rule. As Suzuki and Matsumoto argue, this leads one to expect that even when C_1 and C_2 are emitted by different sources (e.g. different birds; see Appendix A23), their behavioral effect on the receiver should remain the same as long as the sources are collocated (we discussed a simple human example in which Ann says *It's hot* while Bill says about the same location *It's humid*: we naturally aggregate the two pieces of information). If C_1 and C_2 are *not* independent utterances, it is less clear that their informational content could be integrated across separate utterances; in fact, in Zuberbühler's (2002) experiment on Diana monkeys' understanding of hybrid Campbell's and Diana sequences, the receivers *failed* to perform the aggregation (but see Appendix A24). Suzuki and Matsumoto 2022 used such a failure of integration to argue against 'separate utterances' in Japanese tit ABC-D sequences, but the same experimental paradigm could profitably be used in further species.

Stepping back, the general debate can definitely benefit from greater interaction between ethologists and linguists, but the role of the latter is in some ways paradoxical. We adopted the view that animal languages should be studied with the primary goal of understanding their specific properties, and thus we were entirely open to the possibility that the kinds of syntax and semantics they display are very different from human language. In this respect, we resisted the urge to focus on similarities and differences with human language. At the same time, however, arguments for or against syntax and compositionality are very subtle irrespective of whether they apply to animal or to human languages, and for this reason there might be genuine added value in linguists' expertise.

VII. Conclusion

1. With the exception of Japanese tits, *alarm–recruitment* mobbing sequences remain ambiguous when arguing for animal syntax or compositionality.
2. Japanese tit ABC-D sequences display productivity (ruling out an analysis based on 'only one expression') while requiring a single source to be effective (ruling out an analysis based on 'separate utterances'). They are a good candidate for a case of syntax and compositionality, but the semantic rule involved has yet to be specified. In addition, it is puzzling that Japanese tits integrate information from two different species but not from two different locations.
3. Here, we propose two deflationary hypotheses to analyze animal combinatorial systems: the 'only one expression' theory and the 'separate utterances' theory.
4. We suggest that future work pits their findings against the two deflationary hypotheses introduced in this article.

VIII. References

ARNOLD, K. & ZUBERBÜHLER, K. (2012). Call combinations in monkeys: Compositional or idiomatic expressions? *Brain and Language* **120**(3), 303–309.

BACH, K. & HARNISH, R. M. (1979). *Linguistic Communication and Speech Acts*. MIT Press.

BERTHET, M., COYE, C., DEZECACHE, G. & KUHN, J. (2022). Animal Linguistics: a Primer. Accepted for publication in *Biological reviews*.

BERWICK, R. C., OKANOYA, K., BECKERS, G. J. L. & BOLHUIS, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends in Cognitive Sciences* **15**(3), 113–121.

BOLHUIS, J. J., BECKERS, G. J. L., HUYBREGTS, M. A. C., BERWICK, R. C. & EVERAERT, M. B. H. (2018a). Meaningful syntactic structure in songbird vocalizations? *PLOS Biology* **16**(6), e2005157. https://doi. org/10.1371/journal.pbio.2005157

BOLHUIS, J. J., BECKERS, G. J. L., HUYBREGTS, M. A. C., BERWICK, R. C., & EVERAERT, M. B. H. (2018b). The slings and arrows of comparative linguistics. *PLOS Biology*, **16**(9), e3000019. https://doi.org/10.1371/ journal.pbio.3000019

BREMOND, J. (1968). Recherches sur la semantique et les elements vecteurs d'information dans les signaux acoustiques du rouge- gorge (Erithacus rubecula L.). *La Terre et la vie* **2**, 109–220.

BROWN, T. J. & HANDFORD, P. (1996) Acoustic signal amplitude patterns: a computer simulation investigation of the acoustic adaptation hypothesis. *The Condor* ***98***(3), 608–623. https://doi.org/10.2307/1369573

CHARLOW, N. (2014) The Meaning of Imperatives. *Philosophy Compass* 9/8:540–555, 10.1111/phc3.12151

COLLIER, K., BICKEL, B., VAN SCHAIK, C. P., MANSER, M. B. & TOWNSEND, S. W. (2014). Language evolution: syntax before phonology? *Proceedings of the Royal Society B: Biological Sciences* **281**(1788), 20140263.

DUTOUR, M., LÉNA, J. P. & LENGAGNE, T. (2017). Mobbing calls: A signal transcending species boundaries. *Animal Behaviour* **131**, 3–11. https://doi.org/10.1016/j.anbehav.2017.07.004

DUTOUR, M., LENGAGNE, T. & LENA, J. P. (2019a). Effect of syntax manipulation on response to mobbing calls in passerine birds. *Ethology* **125**(9), 635–44. https://doi.org/10.1111/eth.12915

DUTOUR, M., CORDONNIER, M., LÉNA J. P. & LENGAGNE, T. (2019b). Seasonal variation in mobbing behaviour of passerine birds. *Journal of Ornithology* **160**(2), 509–514

DUTOUR, M., SUZUKI, T. N. & WHEATCROFT, D. (2020). Great tit responses to the calls of an unfamiliar species suggest conserved perception of call ordering. *Behavioral Ecology and Sociobiology*, **74**(3), 1–9.

DUTOUR, M. (2022). Season does not influence the response of great tits (*Parus major*) to allopatric mobbing calls. *Journal of Ethology,* 1–4. https://doi.org/10.1007/s10164-022-00752-3

EMBICK, D. & NOYER, R. (2012). Distributed Morphology and the Syntax—Morphology Interface. In *The Oxford Handbook of Linguistic Interfaces* (eds: G. RAMCHAND & C. REISS), 289324. Oxford University Press, UK. DOI: 10.1093/oxfordhb/9780199247455.013.0010

ENGESSER, S., RIDLEY, A. R. & TOWNSEND, S. W. (2016). Meaningful call combinations and compositional processing in the southern pied babbler. *Proceedings of the National Academy of Sciences* **113**(21): 5976–5981.

ENGESSER, S., RIDLEY, A. R., WATSON, S. K., KITA, S. & TOWNSEND, S. W. (2020). Open compositionality in pied babbler call combinations. In *The Evolution of Language: Proceedings of the 13th International Conference* (eds RAVIGNANI, A. et al.), 81–83. Evolang 13 Scientific Committee, Brussels.

FRÖHLICH, M., SIEVERS, C., TOWNSEND, S. W., GRUBER, T. & VAN SCHAIK, C. P. (2019). Multimodal communication and language origins: integrating gestures and vocalizations. *Biological Reviews* **94**(5), 1809–1829.

GRAFE, T. U. (1996). The function of call alternation in the African reed frog (*Hyperolius marmoratus*): precise call timing prevents auditory masking. *Behavioral Ecology and Sociobiology*, **38**(3), 149–158.

HOBAITER, C., GRAHAM, K. E. & BYRNE, R. W. (2022). Are ape gestures like words? Outstanding issues in detecting similarities and differences between human language and ape gesture. *Philosophical Transactions of the Royal Society B*: *Biological Sciences*. **377**(1860), 20210301.

JÄGER, G. (2016). Grice, Occam, Darwin. *Theoretical Linguistics* **42**(1–2), 111–115. https://doi.org/10.1515/tl-2016-0004

JONES, K. J. & HILL, W. L. (2001). Auditory perception of hawks and owls for passerine alarm calls. *Ethology* **107**(8), 717–726.

KLUMP, G.M. & GERHARDT, H. C. (1992). Mechanisms and function of call-timing in male-male interactions in frogs. In *Playback and studies of animal communication* (ed P. K. MCGREGOR) 153–174. Springer US, New York.

KLUMP, G.M. & SHALTER, M.D. (1984). Acoustic behaviour of birds and mammals in the predator context; I. Factors affecting the structure of alarm signals. II. The functional significance and evolution of alarm signals. *Zeitschrift für Tierpsychologie* **66**(3), 189–226.

KUHN, J., KEENAN, S., ARNOLD, K. & LEMASSON, A. (2018). On the-oo suffix of Campbell's monkeys. *Linguistic Inquiry* **49**(1), 169–181.

KVIST, L., RUOKONEN, M., LUMME, J. & ORELL, M. (1999). The colonisation history and present day colonisation structure of the European great tit (Parus major major). *Heredity* **82**(5), 495–502.

KVIST, L., MARTENS, J., HIGUCHI, H., NAZARENKO, A. A., VALCHUK, O. P. & ORELL, M. (2003). Evolution and genetic structure of the great tit (Parus major) complex. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**(1523), 1447–1454.

LUCAS, J. R., FREEBERG, T. M., LONG, G. R. & KRISHNAN, A. (2007). Seasonal variation in avian auditory evoked responses to tones: a comparative analysis of Carolina chickadees, tufted titmice, and white-breasted nuthatches. *Journal of Comparative Physiology A* **193**(2), 201–215.

MAGRATH, R. D., HAFF, T. M. & IGIC, B. (2020). Interspecific communication: gaining information from heterospecific alarm calls. In *Coding strategies in vertebrate acoustic communication* (eds N. MATHEVON, & T. AUBIN), 287–314. Springer, Cham.

MARLER, P. (1955). Characteristics of some animal calls. *Nature* 176(4470), 6–8.

MARLER, P. (1998). Animal communication and human language. In *The origin and diversification of language* (eds N. G. JABLONSKI & L. AIELLO), 1–20. San Francisco, CA: California Academy of Sciences.

MCCULLOCH, G. (2014). Ish: How A Suffix Became A Word. Slate.com, June 9, 2014. https://slate.com/human-interest/2014/06/ish-how-a-suffix-became-an-independent-word-even-though-it-s-not-in-all-the-dictionaries-yet.html

MIYAGAWA, S. & CLARKE, E. (2019). Systems Underlying Human and Old World Monkey Communication: One, Two, or Infinite. Front Psychol. 2019 Sep 3;10:1911. doi: 10.3389/fpsyg.2019.01911.

NARBONA SABATÉ, L., MESBAHI, G., DEZECACHE, G., CÄSAR, C., ZUBERBÜHLER, K., & BERTHET, M. (2022). Animal linguistics in the making: the Urgency Principle and titi monkeys ' alarm system. *Ethology Ecology and Evolution* **34**(3), 378–394. ⟨10.1080/03949370.2021.2015452⟩

OUATTARA, K., LEMASSON, A. & ZUBERBÜHLER, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, **106**(51), 22026–22031.

PÄCKERT, M., MARTENS, J., TIETZE, D. T., DIETZEN, C., WINK, M. & KVIST, L. (2007). Calibration of a molecular clock in tits (Paridae)—Do nucleotide substitution rates of mitochondrial genes deviate from the 2% rule? *Molecular phylogenetics and evolution*, **44**(1), 1–14.

RANDLER, C. (2012). A possible phylogenetically conserved urgency response of great tits (Parus major) towards allopatric mobbing calls. *Behavioral ecology and sociobiology* **66**(5), 675–681.

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

SALIS, A., LENA, J. P. & LENGAGNE, T. (2021a). Great tits (Parus major) adequately respond to both allopatric combinatorial mobbing calls and their isolated parts. *Ethology* **127**(3), 213–222. https://doi.org/ 10.1111/eth.13111

SALIS, A., LENGAGNE, T., LENA, J. P. & DUTOUR, M. (2021b). Biological conclusions about importance of order in mobbing calls vary with the reproductive context in Great Tits (Parus major). *Ibis* **163**(3), 834–844.

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

SCHLENKER, P., CHEMLA, E., ARNOLD, K., LEMASSON, A., OUATTARA, K., KEENAN, S., STEPHAN, C., RYDER, R. & ZUBERBÜHLER, K. (2014). Monkey semantics: two 'dialects 'of Campbell's monkey alarm calls. *Linguistics and Philosophy* **37**(6), 439–501.

SCHLENKER, P., CHEMLA, E., ARNOLD, K., ZUBERBÜHLER, K. (2016a). Pyow-Hack Revisited: Two Analyses of Putty-nosed Monkey Alarm Calls. *Lingua* **171**, 1–23.

SCHLENKER, P., CHEMLA, E., CÄSAR, C., RYDER, R. & ZUBERBÜHLER, K. (2016b). Titi Semantics: Context and Meaning in Titi Monkey Call Sequences. *Natural Language & Linguistic Theory* **35**(1), 271–298. doi:10.1007/s11049-016-9337-9

SCHLENKER, P., CHEMLA, E., SCHEL, A., FULLER, J., GAUTIER, J. P., KUHN, J., VESELINOVIC, D., ARNOLD, K., CÄSAR, C., KEENAN, S., LEMASSON, A., OUATTARA, K., RYDER, R. & ZUBERBÜHLER, K. (2016c). Formal Monkey Linguistics. *Theoretical Linguistics* **42**(1–2), 1–90. DOI: 10.1515/tl-2016-0001

SCHLENKER, P., CHEMLA, E., SCHEL, A., FULLER, J., GAUTIER, J. P., KUHN, J., VESELINOVIC, D., ARNOLD, K., CÄSAR, C., KEENAN, S., LEMASSON, A., OUATTARA, K., RYDER, R. & ZUBERBÜHLER, K. (2016d). Formal Monkey Linguistics: the Debate. (Replies to commentaries). *Theoretical Linguistics* **42**(1–2), 173–201. DOI: 10.1515/tl-2016-0010

SELVATTI, A. P., GONZAGA, L. P. & DE MORAES RUSSO, C. A. (2015). A Paleogene origin for crown passerines and the diversification of the Oscines in the New World. *Molecular phylogenetics and evolution*, **88**, 1–15.

SEYFARTH, R. M., CHENEY D. L. & MARLER, P. (1980). Monkey responses to three different alarm calls: evidence for predator classification and semantic communication. *Science*. **210**(4471), 801–803.

SOARD, C. M. & RITCHISON, G. (2009). 'Chick-a-dee' calls of Carolina chickadees convey information about degree of threat posed by avian predators. *Animal Behaviour* **78**(6), 1447–1453.

STEINERT-THRELKELD, S., SCHLENKER, P. & CHEMLA, E. (2021). Referential and General Calls in Primate Semantics. *Linguistics & Philosophy* **44**(6), 1317–1342.

SUZUKI, T. N. & MATSUMOTO, Y. K. (2022). Experimental evidence for core-Merge in the vocal communication system of a wild passerine. *Nat Commun* 13, 5605 https://doi.org/10.1038/s41467-022-33360-3

SUZUKI, T. N., WHEATCROFT, D. & GRIESSER, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nature Communications*, **7**(1), 1–7. https://doi.org/10.1038/ncomms10986

SUZUKI, T. N., WHEATCROFT, D. & GRIESSER, M. (2017). Wild birds use an ordering rule to decode novel call sequences. *Current Biology*, **27**(15), 2331–2336. e3. https://doi.org/10.1016/j.cub.2017.06.031

SUZUKI, T. N., WHEATCROFT, D. & GRIESSER, M. (2018). Call combinations in birds and the evolution of compositional syntax. *PLOS Biology*, **16**(8), e2006532. https://doi.org/10.1371/journal.pbio.2006532.

TEMPLETON, C. N., GREENE, E. & DAVIS, K. (2005). Allometry of alarm calls: black-capped chickadees encode information about predator size. *Science* **308**(5730), 1934–1937.

TOWNSEND, S. W., ENGESSER, S., STOLL, S., ZUBERBÜHLER, K. & BICKEL, B. (2018). Compositionality in animals and humans. *PLoS Biology* **16**(8), e2006425. https://doi.org/10.1371/journal.pbio.2006425 PMID: 30110319

ZUBERBÜHLER, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour* **63**(2), 293–299.

ZUBERBÜHLER, K. (2009). Survivor signals: the biology and psychology of animal alarm calling. *Advances in the Study of Behavior* **40**, 277-322.

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

The ABC-D of Animal Linguistics:
Are Syntax and Compositionality for Real?

Appendix

A1. In essence, the pragmatic view is that *pyow-hack* sequences are semantically true whenever there is an important non-ground movement, which could be raptor movement or movement of the monkey group. But in the former case, the non-ground call *hack* would provide information about the nature/location of a threat, and thus it should come first in virtue of the Urgency Principle (see also A13 below).

A2. In a separate collaboration between linguists and primatologists, Miyagawa & Clarke 2019 focused on syntax rather than semantics. They proposed that 'animal syntax' allows for limited combinations by way of two ordered templates. Thus in Putty-nosed monkey *pyow-hack* sequences, a single 'pyow' compartment is followed by a single 'hack' compartment, and each compartment allows for repetitions. (See also Rizzi (2016) for a different mechanism, with a limited application of Merge in some animal systems, without recursion [= '1-merge'].)

A3. In principle, our rich theory could come in several varieties. The key claim is that there are cognitively real rules that determine the presence of new forms and/or new meanings based on old ones. This claim could take different forms. Thus instead of taking two expressions C_1 and C_2 to be concatenated, one could think of C_1C_2 as an elementary expression, *but connected to C_1 and C_2* by cognitively real rules. On this view, the lexicon of the language contains $\{C_1, C_2, C_1C_2\}$. But lexical rules specify (i) that if a C_1-type call and a C_2-type call are parts of the lexicon, a C_1C_2-type call must be as well; and/or (ii) that if C_1, C_2 and C_1C_2 are part of the lexicon, the meaning of C_1C_2 is derived from the meanings of C_1 and C_2 by a certain semantic rule. If (i) and (ii) are adopted, we have a near-notational variant of a morphosyntactic analysis based on complex calls. If we have (ii) but not (i), we have a morphological rule on the meaning side but not on the form side (this is conceptually non-standard because the semantics must make reference to component parts which, for the morphosyntax, are not cognitive real). Having (i) but not (ii) would most naturally be treated as a case of phonological rather than morphological complexity, as the meanings of the component parts do not make their effects felt (however, see Arnold and Zuberbühler 2012 for a related case that they characterize as being syntactically combinatorial but not semantically compositional). For reasons of clarity and simplicity, we leave aside these variants of the rich theory in what follows.

A4. See for instance Schlenker et al. 2016*d* and Zuberbühler 2020. Schlenker et al. 2016*b* write the following:
"Although monkey sequences can be quite long, we take the “null hypothesis” to be that each call contributes its informational content independently from the others, by way of a propositional meaning. (…) this leads one to expect that the semantic content of a sequence should be the conjunction of the meanings of its component parts, evaluated at their respective times of utterance. This is the most trivial notion of “compositionality” that one can imagine, which is not indicative of the existence of genuine rules of combination (since each call can be interpreted independently)."

A5. In human and animal linguistics alike, a further property of separate utterances is that they provide information about different moments of utterance. For instance, Ann can't say *It's raining and not raining* without contradicting herself because this involves a single utterance, and it thus talks about a single moment. But no contradiction arises if Ann first utters *It's raining*, and then later looks out the window and says *It's not raining* (or more naturally: *Now it's not raining*). This property played a key role in the analysis of Titi calls in Schlenker et al. 2016*b,c*: each call was taken to provide information about the very moment at which it was uttered. On this view, a flying raptor gives rise to a shorter

sequence of A-calls than a perched raptor because in the former situation the threat disappears more quickly.

A6. Let us give an example of the benefits of pitting a target theory against deflationary theories, rather than just relying on a list of criteria. Salis et al. 2021*a* cite the following criteria for semantic compositionality:
(a) "a different order should trigger a different response";
(b) "the whole sequence should not only be the sum of its different parts, but have a new emergent meaning";
(c) "the two parts, when isolated, should still be meaning-bearing units".
These criteria are indeed in line with the requirement that the meaning of a complex expression is *derived* from the *meaning of its parts* and *the way they are put together*. (a) pertains to the fact that syntax matters, or in other words: the way the parts are put to together matters. (b) pertains to the fact that the meaning of the whole should be derived in a non-trivial way from the meaning of its parts: it shouldn't just be their (conjunctive) addition. (c) pertains to the fact that the derivation is based on more elementary meanings. The criteria primarily help exclude deflationary theories based on 'only one expression' (especially (c)) and 'separate utterances' ((a) and (b)). Still, these criteria are not perfect. *It rained yesterday* has the same meaning as *Yesterday, it rained*, against (a), but the combination is still compositional. The key is that neither 'only one expression' nor 'separate utterances' has any plausibility in this case. *Everyone eats and drinks* involves a phrase, *eats and drinks*, whose meaning is the (conjunctive) sum of its parts, against (b), but it is still compositional. Here too, neither 'only one expression' nor 'separate utterances' has any plausibility (in the latter case, because *eats* on its own is not a possible utterance, nor is *drinks*). And as we discuss below, *blueish* and *blue-like* are compositional, but proving that *-ish* or *-like* are meaning-bearing requires an analysis, since the suffixes cannot usually appear on their own, making (c) delicate. Here 'separate utterances' has no plausibility, and an analysis of the distribution and productivity of *-ish* or *-like* shows that 'only one expression' is incorrect.

A7. Two remarks should be added. First, the similarity between *-oo* and *-ish* (or *-like*, for that matter) is particularly striking in the first theory of *-oo* entertained by Schlenker et al. 2014. In that theory, *-oo* broadens the meaning of the call it applies to. So, if *hok* indicates that one is in a situation in which there is an aerial predator, *hok-oo* indicates that one is in a *hok*-ish situation, in the sense that the situation licenses the same attentional state as if there were an aerial predator—e.g. one should look up. We caution that in the present piece we follow Schlenker et al. 2016*c* in discussing the *second* (and preferred) theory of Schlenker et al. 2014, in which the similarity between *-oo* and *-ish* is more remote.

Second, *-ish* has a broader and more interesting distribution than is discussed here; for instance, (i) it can turn nouns into adjectives, and (ii) in uses described by McCulloch 2014, it can even appear on its own. (i) also applies to *-like*, but we are not aware that (ii) does. In any event, since our object is animal rather than human linguistics, we allow ourselves some simplifications.

A8. This is a simplification, in two respects. First, as Kuhn et al. 2018 note, *-oo* is separated by *krak* and *hok* by a very tiny pause, making the 'complex call' analysis plausible. Second, as Schlenker et al. 2014 discuss at length, auxiliary hypotheses are needed, in particular the view that there is a pragmatic rule of competition among calls, the Informativity Principle.

A9. Sauerland 2016 proposes that *-oo* forms a separate utterance and means: *there is a weak disturbance*. As a result, the two utterances *hok* and *oo* could only be satisfied by a situation in which there is a non-ground disturbance and there is a disturbance (presumably the same one) which is weak. As Schlenker et al. 2016*d* note, this does not make exactly the same predictions as the analysis in **Error! Reference source not found.**c. Even on the assumption that only one disturbance is at stake, the analysis in **Error! Reference source not found.**c allows for *hok-oo* to be true in case there is a threat that counts as weak *among non-ground threats*. But on the assumption that non-ground threats are raptor-related and thus very serious in general, a weak raptor threat might still count as serious relative to the entire set of threats (by the same logic, a *cheap diamond* might not count as a *cheap object*: it is cheap relative to the set of diamonds, but not relative to the set of all objects). This prediction is hard to test directly. But Schlenker et al. 2016*d* argue that, when combined with the Informativity Principle, Sauerland's theory makes the

wrong predictions (in a nutshell, it predicts that *krak-oo* should compete with *hok-oo* and should only be true of ground-related disturbances, contrary to fact).

A10. Three general remarks should be added about Suzuki et al.'s analyses. First, we follow Suzuki et al. 2016 in taking ABC to form an unanalyzed morphosyntactic and semantic unit, but this hypothesis might have to be revised in future research. As the authors write: "A, B and C notes are typically produced in combination with other note types, resulting in AC, BC or ABC calls (…). In contrast, D notes are produced as a string of seven to ten notes (…)."

Second, a full argument would also need to show that Japanese tit *ABC* does not sound like Willow tit *ABC**, which Suzuki et al. 2017 call *zi*. The reason is this: Japanese tits are familiar with Willow tit *ABC***-D** (i.e. *zi-tää*) sequences. If *ABC* sounds like *ABC**, they might interpret hybrid *ABC-D** sequences as a variant of Willow tit ABC*-D* sequences.

Third, Suzuki et al.'s argument predicts that hybrid sequences *ABC-shortened D** (i.e. *ABDC-shortened tää*) should differ from *ABC-D** (= *ABC-tää*) in *not* triggering the target behavior (namely increased vigilance and approach). To our knowledge, this experiment has not been performed or reported.

A11. As Dutour et al. 2020 write, "the first D notes of the call mask the notes that follow them, preventing the receiver from perceiving the second part of the call" because "D notes, which have large frequency bandwidths and are produced in long, repetitive sequences, may mask "the alarm call part" given the relative short delay between both sequences".

A12. Two remarks should be added. First, we assume for the sake of simplicity that any masking effect is caused by a shared property of *D* and *D**. It could in principle be that it is for separate reasons that masking arises in the two cases, which would require a longer discussion.

Second, to establish our conclusion that masking effects are unlikely to be at stake, we explored four key parameters: repetitive structure, broadband spectrum, intensity and short inter-element interval. We failed to find any convincing case for a potential masking effect.
(i) Repetitive structure and broadband spectrum: The length of D notes sequences encodes urgency and/or influences the receiver's reaction in some bird species (Templeton, Green & Davis, 2005; Soard & Ritchison, 2009), which strongly suggests that the number of iterations is perceived by the receivers. For instance, the distance at which great tits approach a playback speaker decreases linearly as the number of D notes in the mobbing sequence of black-capped chickadees increases (Randler, 2012).
(ii) Sound intensity and inter-note interval: all notes (D/D* and ABC) were broadcast with the same intensity, and the inter-note interval used in the stimuli was the same in ABC-D/D* and D/D*-ABC stimuli (0.1s). A masking effect due to these parameters alone is thus unlikely.

A13. Schlenker et al. 2016*c* summarize the main analysis as follows: "Semantically, *pyow-hack* sequences are compatible with any kind of situation involving (moving) aerial predators or (arboreal) movement of the monkeys themselves. But in the former situation, *hacks* provide information about the location of a threat, and hence should appear at the beginning of sequences. As a result, *pyow-hack* sequences can only be used for non-risk-related situations involving movement, hence a possible *inference* that they (often) involve group movement. While it is too early to adjudicate this debate, we will argue that a formal analysis of the competing theories should help produce new predictions to be tested in future field studies."

A14. A version of the second alternative was pursued in Schlenker et al.'s (2016*a,c*) analysis of *pyow-hack* sequences in Putty-nosed monkeys. The idea was that because of the Urgency Principle *hack* is not predator-related in this context. Rather, it pertains to an important non-ground movement which isn't that of a raptor, but that of the group of monkeys. This, in turn, explains why *pyow-hack* sequences announce group movement.

A15. We gloss over complex questions. First, some imperatives in a broader sense do not determine an action irrespective of the state of the world—e.g. *Stay safe!* might require different actions in different environments. Second, in human language some sentences that do not involve the imperative mood do

have something close to an imperative meaning, e.g. *You should climb up* is close to *Climb up!* While we leave these issues for future research, we briefly revisit them in the conclusion. For old and new views on the distinction between imperatives and declaratives, see for instance Bach and Harnish 1979 and Charlow 2014.

A16. Two further remarks are in order. First, on the imperative analysis, just as in the declarative theory, something must be said about the failure of the reverse order *D/D*-ABC*, and here too various hypotheses can be entertained—including an imperative version of the Urgency Principle. For instance, one could posit that information that pertains to the receiver's survival should come first.

Second, Suzuki et al.'s remarkable result about hybrid sequences should be contrasted with an experiment with the same logic but opposite results in Diana monkeys. In a nutshell, Zuberbühler 2002 showed that Diana monkeys understand the calls of male Campbell's monkeys, whose acoustic properties are very different from those of Diana monkeys. In fact, they understand them down to the details: in the Tai forest, *krak* signals the presence of a ground threat, and Diana monkeys react appropriately. But they also know that the Campbell's call *boom* (which comes in pairs at the beginning of sentences) is only used in situations of non-predation, and thus the Dianas fail to react with alarm when they hear a series of *kraks* preceded by *boom boom*. But when *boom boom* precedes Diana alarm calls, they just ignore the *boom boom* part in this hybrid sequence. One open question is why the Japanese tits studied by Suzuki et al. do not do the same thing, just ignoring the heterospecific calls interspersed in the conspecific sequence. (One possible explanation for the Japanese tit–Diana difference is that Dianas don't have any equivalent of *booms*, and they have no simple way of interpreting the hybrid sequence. A second possible explanation is that the Dianas have reasons to trust conspecifics more than heterospecifics. Additional explanations should be explored.)

A17. Importantly, the same situation could be replicated with imperatives that give rise to differential responses depending on the context (and thus do not fall under the 'narrow imperative analysis' in our terminology). For instance, a mother talking to her child may say *Watch out!* and elicit mild reactions (for instance if the child is spilling food); and similarly, if she says *Come here!* (for instance, if the child is being asked to help with house chores). But *Watch out! Come here!* might elicit a much stronger reaction because the two imperatives in combination suggest that there is serious danger for the child. See also A15 above for complexities arising from the discussion between imperatives and declaratives (a point we briefly revisit in the conclusion of the main text).

A18. As surveyed in Magrath et al. 2020, there are multiple cases in which birds appear to interpret a designated acoustic feature, and thus general measures of similarity among calls might not be optimally relevant. Rather, one might want to determine whether a certain designated acoustic feature is shared.

A19. On this analysis, one would definitely expect that Great tits fail to react to Great tit inverse sequences, i.e. to *recruitment–alarm*. But it's unclear which theory would *not* predict this in view of the Great tits' reaction to Chickadee sequences.

A20. In the following discussion, Dutour et al. 2020 add that seasonality might play a role as well, but we don't see how this particular point speaks speak against the masking hypothesis.
"However, perception bias is unlikely to fully explain our results because mobbing call responsiveness also depends on the social context and the season (Lucas et al. 2007; Dutour et al. 2019*b*). Japanese tits are more likely to approach loudspeakers playing back FME-D calls than the D-FME calls during the non-breeding season (Suzuki et al. 2016, 2017). In the present study, tests were conducted during the breeding season, which may offer a partial explanation for why great tits approached playbacks of D-FME calls, without needing to invoke perception bias" (Dutour et al. 2020).

A21. The average time interval between two booms is about 7 seconds, and other call types generally follow booms within 25 seconds (Zuberbühler, 2002). Vocal sequences, especially when signaling a predator, can count up to 40 calls (Ouattara et al., 2009).

A22. Needless to say, it is not enough to observe that, say, researchers have 'traditionally' thought that recruitment calls have an imperative semantics, along the lines of 'Come here!'. The problem is that in simple cases this imperative meaning makes essentially the same predictions as a declarative meaning such as 'Help is needed here'. Real criteria and predictions are thus needed.

A23. Some calls may not contain identity cues, or some species may not be able to identify callers based on their voice alone. In this case, an alternative approach would be to broadcast the distinct parts of the "composed" stimuli from two distinct locations, equidistant from the caller (to control for intensity, degradation due to propagation etc) but far enough from each other to ensure that a single individual could not travel from one to the other in the short time laps between the two sounds broadcast.

A24. Some caution is needed because in some non-standard cases one can take a speaker to continue another speaker's utterance. For instance, little Ann and her mother could have the following dialogue: Ann: *You will buy me an ice-cream!* Mother: … if *you do your homework!*