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!*