

Minimal Compositionality versus Bird Implicatures: Two Theories of ABC-D Sequences in Japanese Tits*

Philippe Schlenker^{ab}, Ambre Salis^{a*}, Maël Leroux^a, Camille Coye^a, Luigi Rizzi^{cd}, Shane Steinert-Threlkeld^e, Emmanuel Chemla^f

^a *Institut Jean-Nicod (ENS - EHESS - CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, 29, rue d'Ulm, 75005 Paris, France; ; PSL University*

^b *Department of Linguistics, New York University, 10 Washington Place, New York, NY 10003, USA*

^c *Collège de France, 11 Place Marcelin Berthelot, 75005 Paris, France*

^d *CISCL, University of Siena, Via Roma 56, 53100 Siena, Italy*

^e *Department of Linguistics, University of Washington, Guggenheim Hall 4th Floor Box 352425, Seattle, WA, USA.*

^f *LSCP (ENS - EHESS - CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France; PSL University*

* *Corresponding author (E-mail: salis.ambre87@gmail.com; Tel.: 0033 602 293 513)*

Abstract. It was argued in a series of experimental studies that Japanese tits (*Parus minor*) have an ABC call that has an alert function, a D call that has a recruitment function, and an ABC-D call that is compositionally derived from ABC and D, and has a mobbing function. A key conclusion was that ABC-D differs from the combination of separate utterances of ABC and of D (e.g. as played by distinct but close loudspeakers). While the logic of the argument is arguably sound, no explicit rule has been proposed to derive the meaning of ABC-D from that of its parts. We compare two analyses. One posits a limited instance of semantic compositionality ('Minimal Compositionality'); the other does without compositionality, but with a more sophisticated pragmatics ('Bird Implicatures'). Minimal Compositionality takes the composition of ABC and D to deviate only minimally from what would be found with two independent utterances: ABC means that 'there is something that licenses an alert', D means that 'there is something that licenses recruitment', and ABC-D means that 'there is something that licenses both an alert and recruitment'. By contrast, ABC and D as independent utterances yield something weaker, namely: 'there is something that licenses an alert, and there is something that licenses recruitment', without any 'binding' across the two utterances. The second theory, Bird Implicatures, only requires that ABC-D should be more informative than ABC, and/or than D. It builds on the idea, proposed for several monkey species, that a less informative call competes with a more informative one ('Informativity Principle'): when produced alone, ABC and D trigger an inference that ABC-D is false. We explain how both Minimal Compositionality and Bird Implicatures could have evolved, and we compare the predictions of the two theories. Finally, we extend the discussion to some chimpanzee and meerkat sequences that might raise related theoretical problems.

Keywords: animal linguistics, animal semantics, compositionality, minimal compositionality, implicatures, Informativity Principle, bird calls, meerkat calls

* Grant acknowledgments

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

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

CONTENTS

- I. Introduction
- II. Arguments for Non-trivial Compositionality in Japanese Tits
- III. Minimal Compositionality
 - III.1. Analysis with Minimal Compositionality
 - III.2. How Minimal Compositionality could have evolved
 - III.3. Summary, predictions and outlook
- IV. Bird Implicatures
 - IV.1. From Minimal Compositionality to Bird Implicatures
 - IV.2. Analysis with Bird Implicatures
 - IV.3. How the Informativity Principle could have evolved
 - IV.4. Summary, predictions and outlook
- V. Extensions to Chimpanzees and Meerkats
 - V.1. Chimpanzees
 - a. Experimental results of Leroux et al. 2023
 - b. Analysis
 - V.2. Meerkats
- VI. Results and Perspectives
 - VI.1. Comparing Minimal Compositionality and Bird Implicatures
 - VI.2. The typology of Merge
- VII. Conclusion
- VIII. References
 - Appendix I. No Conjunction Without Juxtaposition
 - Appendix II. Formal Analyses
 - A. Minimal Compositionality
 - B. Evolutionary stability of Minimal Compositionality
 - C. Informativity Principle
 - D. Evolutionary stability of the Informativity Principle

I. Introduction

In a striking series of articles, Suzuki and colleagues have argued that Japanese tit calls display an instance of compositionality, the property by which the meaning of a complex expression is derived from the meaning of its parts and the way they are put together (Suzuki, Wheatcroft & Griesser, 2016, 2017, 2018; Suzuki & Matsumoto 2022; for a textbook introduction in the human case, see Heim & Kratzer 1998). Specifically, the claim is that Japanese tits (*Parus minor*) have an ABC call that has an alert function, a D call that has a recruitment function, and an ABC-D call that has a mobbing function, and whose meaning is compositionally derived from those of ABC and D. The logic of the argument for the existence of a compositional rule is one of the strongest and most complete in the literature on animal communication. Our goal is to take the next step, and propose an explicit rule by which the meaning of ABC-D is derived from the meanings of ABC and D. We make a semantic proposal ('Minimal Compositionality'), and compare it with an alternative in which the action does not lie in the semantics but in the pragmatics ('Bird Implicatures').¹

Minimal Compositionality takes the composition of ABC and D to deviate only minimally from what would be found with two independent utterances. In a nutshell, as is stated in (1), ABC means that 'there is something that licenses an alert', D means that 'there is something that licenses recruitment', and ABC-D means that 'there is something that licenses both alert and recruitment'. By contrast, ABC and D co-occurring as independent utterances yield something weaker, namely: 'there is something that licenses an alert, and there is something that licenses recruitment', without any 'binding' across the two utterances. Technically, we will say that ABC and D contribute two existential statements (each of the form: "there exists something that ..."), while ABC-D contributes a single existential statement (rather than the conjunction of two existential statements).

(1) Minimal Compositionality

a. ABC-D:

For some x, x licenses an alert and x licenses recruitment.

b. ABC and D co-occurring as separate utterances:

For some x, x licenses an alert and for some y, y licenses recruitment.

The competing theory, Bird Implicatures, just starts from the assumption that ABC-D is more informative than ABC, and than D: no binding is required across ABC and D (as we will see later, it is enough that ABC-D is more informative than ABC or than D). It then builds on the idea, proposed for several monkey species, that a less informative call competes with a more informative one ('Informativity Principle'): if ABC or D are produced alone, they trigger an inference that ABC-D was false (a bird implicature); no such inference is of course produced by a combined instance of ABC-D. While this idea is compatible with different semantic implementations, in (2) we assume for concreteness (against Minimal Compositionality) that ABC-D involves two existential statements (we will explain in due course how the 'mobbing' reaction can be explained).

¹ To our knowledge, this is one of the first attempts to propose an explicit rule to derive in a non-trivial way the meaning of a combination from the meaning of its parts. An earlier compositional rule was proposed for the suffix -oo in Campbell's monkeys (Schlenker et al. 2014). And an earlier pragmatic treatment (based on the 'Urgency Principle', discussed below) derived the meaning of Putty-nosed monkey pyow-hack sequences from the meaning of their parts (Schlenker et al. 2016a).

- (2) Bird Implicatures
- a. ABC-D:
For some x, x licenses an alert and for some y, y licenses recruitment.
 - b. ABC and D used as separate utterances:
ABC:
Literal meaning: For some x, x licenses an alert
Implicature: No x licenses recruitment
D:
Literal meaning: For some y, y licenses recruitment
Implicature: No y licenses an alert.

The rest of this article is organised as follows. In Section II, we summarise the logic of the remarkable experiments of Suzuki and colleagues. In Section III, we define our semantic rule of composition, Minimal Compositionality, and we explain how it might have evolved. In Section IV, we lay out our alternative pragmatic analysis based on the Informativity Principle, and explain how the latter might have evolved. In Section V, we sketch potential extensions of the analysis to two further cases, involving chimpanzee (*Pan troglodytes*) and meerkat (*Suricata suricatta*) calls, and we draw some conclusions in Section VI.

II. Arguments for Non-trivial Compositionality in Japanese Tits

Suzuki and colleagues propose that ABC-D mobbing sequences involve a syntactic rule of combination ('Merge') and a compositional rule of semantic interpretation; let us call this theory "Non-trivial compositionality". To fully argue for it, two deflationary alternatives must be refuted (Schlenker et al., 2023): according to one ("Only one expression"), ABC-D is made of a single call; according to the other ("Separate utterances"), ABC-D is made of two calls, but these are separate utterances and thus no non-trivial syntactic or semantic rule is needed to combine them. The latter analysis is sometimes called 'trivial compositionality' in the literature (e.g., Schlenker et al. 2016d, Steinert-Threlkeld 2020, Zuberbühler 2020) because the meaning of ABC-D is derived from the meaning of ABC and of D, but simply because the informational contents of separate utterances add up in a conjunctive fashion.

The basic finding is that pure ABC sequences serve as an alarm call, pure D sequences are used as a recruitment call, and ABC-D sequences are used as a mobbing call. In the words of Suzuki et al. 2016, "ABC calls serve as warning calls that elicit predator-scanning behaviour, whereas D calls serve as recruitment calls that attract". ABC calls trigger scanning behaviour while D calls trigger an approach towards the sound source. In addition, D calls are produced more often when the caller is alone, and they tend to elicit mate visits. In experiments with a predator model near a loudspeaker, ABC-D triggers approach towards the predator (Suzuki & Matsumoto 2022); and in other playback experiments, ABC-D triggers both scanning and approach towards the speaker, a combination that is arguably appropriate to mob a predator.

We summarise in Fig. 1 the main empirical findings by Suzuki and colleagues, and explain how they could be handled by the target theory, Non-trivial compositionality, and by the two deflationary theories, Only one expression and Separate utterances. We highlight in grey the results that are most problematic for each theory.

Figure 1

Findings	Only one expression	Separate utterances	Non-trivial compositionality
1. Non-sequential response ABC-D triggers the simultaneous rather	Unsurprising: the meaning of ABC-D may be unrelated to that of ABC and of D.	Unsurprising on a variety of theories: produced in close succession, the two	Unsurprising: the meaning of ABC-D is derived from those of ABC and of D, but

than sequential production of scanning and approach behaviours. (Suzuki et al. 2016, 2018)		utterances provide information about the same moment and there is no reason for the two reactions to be sequential.	distinct from them.
2. Ordering constraint ABC-D triggers the behaviour, D-ABC doesn't (Suzuki et al. 2016 [behaviour = scanning and speaker approach], Suzuki & Matsumoto 2022 [behaviour = approach towards a model predator, wing flicking]).	Unsurprising: ABC-D is a call, D-ABC isn't.	Surprising, unless one adopts an independently motivated principle: Urgency Principle = calls that provide information about the nature/location of a threat should come before those that don't (Schlenker et al. 2016a, d, Narbona Sabate et al. 2022)	Unsurprising: compositional interpretation depends on syntax, and D-ABC may be syntactically ill-formed.
3. Productivity Hybrid ABC-D* sequences trigger the behaviour [= combination of scanning and approach] even though they are never heard in nature, as D* is a call heard in willow tits (<i>Poecile montanus</i>), a sympatric species; D*-ABC doesn't trigger the behaviour. (Suzuki et al. 2017)	Surprising: neither ABC-D* nor D*-ABC is a call and both should behave in the same way.	Assuming the Urgency Principle in row 2, this is partly unsurprising, and partly surprising. (i) Since D* produces the same kind of information as D, it's unsurprising that its production can trigger the same behaviour as D, including in the context of ABC. (ii) But it is surprising that the Urgency Principle is applied to the combination of utterances that come from different species.	Partly unsurprising, partly surprising: (i) A compositional rule is only sensitive to syntax and meaning, hence it can treat ABC-D and ABC-D* in the same way because D and D* have the same meaning. (ii) But it's unclear why ABC-D* isn't recognised as syntactically ill-formed.
4. Mobbing response ABC-D triggers mobbing of a model predator [model approach and wing flicking], unlike separate productions of ABC and D, or productions of D-ABC (Suzuki & Matsumoto 2022).	Unsurprising, as the meaning of ABC-D needn't be related to that of ABC and D.	Possibly unsurprising: if both an alert and recruitment are called for in a given situation, this is probably because a predator needs to be mobbed.	Unsurprising: the meaning of ABC-D is derived but distinct from that of ABC and D.
5. 2-speaker	Unsurprising: a single	Surprising: as	Unsurprising: a

<p>experiment ABC-D triggers the behaviour [= approach towards the model predator] when produced from 1 but not from 2 speakers (Suzuki & Matsumoto 2022).</p>	<p>expression should be produced from a single source.</p>	<p>designed, the experiment allows the two utterances to provide information about the same threat (a predator model), and their informational content should be combined.</p>	<p>compositional rule should apply within a single utterance but not across two utterances.</p>
---	--	--	---

Let us briefly go through the main findings and arguments, using the same numbering as in the table, and without repeating the main references.

1. Non-sequential response: ABC-D triggers scanning and approach, but not in a sequential fashion: the D-appropriate behaviour, approach, need not follow the ABC-appropriate behaviour, scanning. This is unsurprising for Only one expression because the meaning of ABC-D needn't be related to that of ABC and D, and for Non-trivial compositionality, for which the meaning of ABC-D may be distinct from the sequential interpretation of ABC and D. While the finding might appear surprising for Separate utterances, this is arguably not so: when uttered in close succession, ABC and D provide information about the same moment and reaction should aggregate the information from the two calls. This is clear if the two calls have a declarative semantics, e.g. *There is an alarm. Help is needed here*. But this is also the case if the calls have an imperative semantics, e.g. *Watch out for a threat! Come here!*. In both cases, the target birds should update their cognitive state and act accordingly, and there is no reason they should react first to the first call and then to the second.

2. Ordering constraint: ABC-D triggers the combination of scanning and speaker approach, and also (in another experiment) approach towards a model predator; D-ABC does not trigger the target behaviour. This can be seen as evidence for the existence of a syntactic rule, and thus indirectly for Non-trivial compositionality. But this finding is also unsurprising for Only one expression, as ABC-D is a call but D-ABC isn't. The finding is surprising for Separate utterances without additional principles, as both ABC-D and D-ABC should be interpretable and should have the same meaning. But it has been argued that there is independent evidence in other species (Putty-nosed monkeys, *Cercopithecus nictitans*, Titi monkeys, *Callicebus nigrifrons*) for an 'Urgency Principle' according to which calls that provide information about the nature or location of a threat should come before calls that don't (Schlenker et al. 2016a, d, Narbona Sabate et al. 2022). D-ABC could fail to target the behaviour either because it violates the principle, or because it suggests that the ABC call doesn't provide information about the nature/location of a threat (or else it should have come before the D call).

3. Productivity: Remarkably, Japanese tits respond with a mobbing-appropriate behaviour to sequences they have never heard before, namely ones made from an ABC call followed by a D* call, the recruitment call of a sympatric species, the Willow tit; by contrast, D*-ABC fails to be effective. This result is surprising for Only one expression, as ABC-D* and D*-ABC are both non-existent calls. This result has been claimed to be compatible with Non-trivial compositionality, as the meaning of ABC-D is derived from the meaning of ABC and the meaning of D, and the way they are put together. In ABC-D*, D* has the same meaning as D and is in the same position as D in ABC-D, so the compositional rule can treat ABC-D and ABC-D* in the same way. What is surprising, however, is that ABC-D* is not detected as being syntactically ill-formed, as it combines calls from different species and thus shouldn't count as one utterance to begin with.² The theory based on Separate utterances is in a somewhat similar situation. It has no trouble explaining why ABC-D* is understood, as it is made of two utterances with essentially the same semantic content as ABC-D. But it is surprising that the

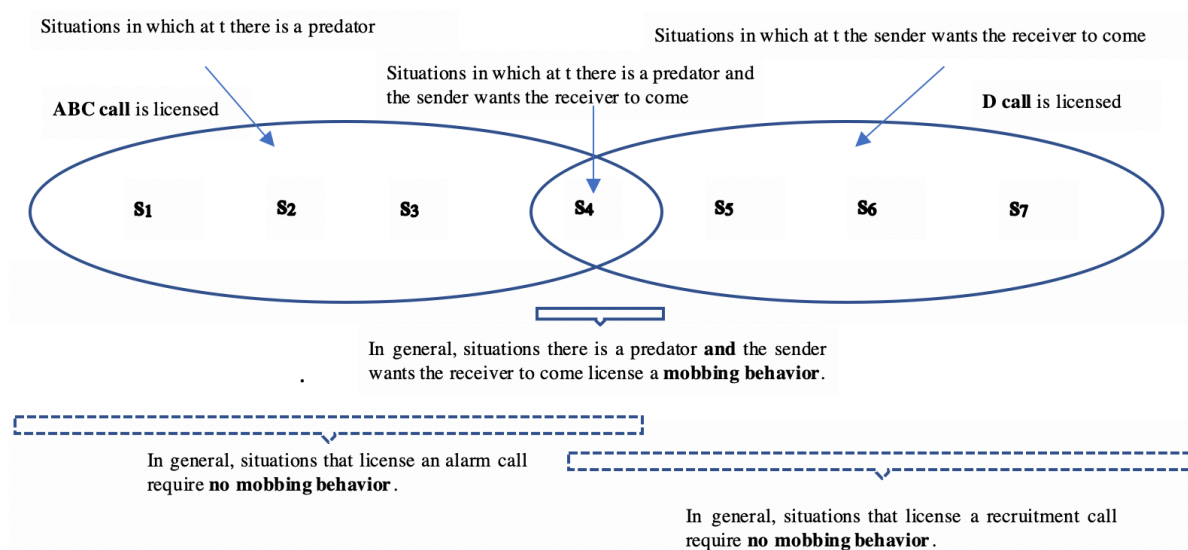
² For comparison, in playback experiments reported in Zuberbühler 2002, Diana monkeys (*Cercopithecus diana*) disregarded the heterospecific component of a hybrid sequence (the sequence was made of Diana monkey alarm calls preceded by a Campbell's monkey (*Cercopithecus campbelli*) non-predation call; the latter was disregarded).

Urgency Principle is applied to a sequence made of calls from different species; it is thus unexpected that D*-ABC fails to trigger the behaviour.

4. Mobbing response: Importantly, one must explain why ABC-D doesn't just trigger scanning and speaker approach, the behaviours triggered by ABC and D respectively, but also mobbing of a predator. The difference in reaction is unsurprising for Only one expression and Non-trivial compositionality: in the former case, because the meaning of ABC-D needn't be related to that of ABC and D; in the latter case, because the meaning of ABC-D is derived but distinct from those of ABC and D. One might think that the mobbing effect of ABC-D is surprising for Separate utterances, but this needn't be so. ABC-D produced contiguously but analysed as separate utterances provide information about a single situation, and should trigger the reaction that is most likely to be appropriate for it.

As an example (minimally modified from Schlenker et al. 2022), suppose a receiver adopts the reaction that is appropriate for most situations that license the signal, and that the possible situations are as depicted in Fig. 2. Most situations that license the alarm call ABC involve a predator but not one that needs to be mobbed. Most situations that license the recruitment call D also fail to involve a predator that needs to be mobbed. But most situations that license both ABC and D involve a predator that needs to be mobbed. So in the end Separate utterances can in principle account for the mobbing response. (This analysis might not be open if calls are imperatives in a narrow sense, i.e. imperatives that fully determine the action to be taken irrespective of the state of the world; see Schlenker et al. 2022, Appendix S1, A15.)

Figure 2



5. 2-speaker experiment: Finally, a comparison between playbacks of ABC-D from one and from two speakers has provided remarkable evidence against Separate utterances. Suzuki & Matusmoto (2022) reasoned that if ABC and D are separate utterances, they should give rise to a mobbing reaction even when they are played back from different speakers. In a 1-speaker baseline, an ABC-D sequence triggers mobbing of a predator model (a bull-headed shrike, *Lanius bucephalus*, a passerine bird), as illustrated in Fig. 3a. Crucially, ABC-D *fails* to be effective when ABC and D are played from different loudspeakers, as illustrated in Fig. 3b (D-ABC sequences are ineffective as well under such conditions).³ While unexpected for Separate utterances, this result is unsurprising for Only one expression, since in

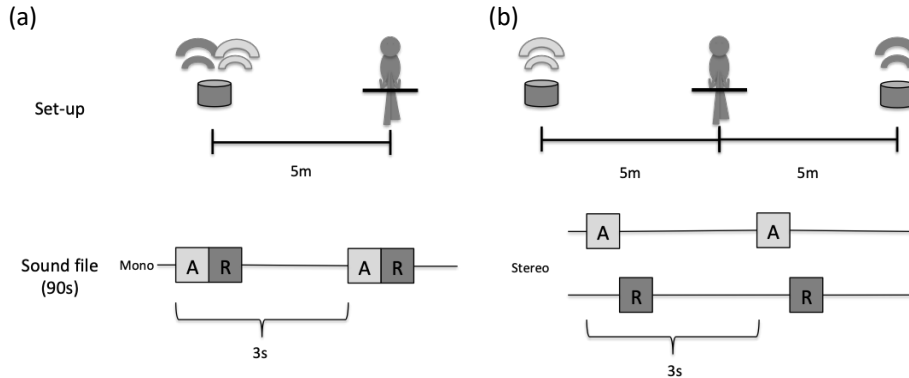
³ One might object that in the 2-speaker condition, the speakers are too far from each other to provide information about one and the same external event. But the baseline (1-speaker) experiment shows that a speaker that is positioned 5m away from the shrike model is close enough to provide information about it, since it triggers mobbing. In the 2-speaker experiment, each speaker is also 5m away from the shrike model, so the information should be understood to refer to the model just as well as in the 1-speaker experiment.

the 2-speaker condition the single expression is split among two sources. The result is also unsurprising for Non-trivial compositionality, as a compositional rule presumably shouldn't apply across separate utterances, let alone separate callers.

Figure 3

Crucial part of the experimental design of Suzuki and Matsumoto 2022 (redrawn from Suzuki & Matsumoto 2022)

Alarm-recruitment sequences are produced from a single loudspeaker (a) or from two loudspeakers (b). Only the single loudspeaker condition triggers a mobbing behaviour.



In sum, the productivity argument (from hybrid ABC-D* sequences) makes Only one expression very unlikely. The 2-speaker experiment makes Separate utterances implausible. This leaves Non-trivial compositionality, which accounts for the data, although it is faced with a puzzle: why is it that Japanese tits treat hybrid ABC-D* as a single sequence even though they are made of calls from two species, while they refuse to treat ABC-D as a single sequence when its two components are played from two different speakers?

With this proviso, the next question is to determine the precise nature of the compositional rule. In Section III, we will make a concrete semantic proposal, Minimal Compositionality. In Section IV, we will develop an alternative analysis in which the non-trivial part of the action lies in the pragmatics (pertaining to competition among calls) rather than in the semantics. (We consider in Appendix I a far more deflationary analysis, according to which birds just fail to aggregate information from different sources; as we explain, this seems implausible in view of what is currently known of animal communication and concepts.)

III. Minimal Compositionality

According to our first theory, Minimal Compositionality, the meaning of ABC-D is derived from the meaning of ABC and D in a non-trivial fashion: ABC alone means that *something licenses an alert*, D alone means that *something licenses recruitment*, and ABC-D means that *something licenses an alert and recruitment*. By contrast, combined separate utterances of ABC and D mean that *something licenses an alert and something licenses recruitment*, without an implication that it's one and the same thing that licenses both.

III. 1. Analysis with Minimal Compositionality

The four relevant meanings can be represented in logical notation using the quantifier $\exists x$, for 'there exists an x such that ...'. ABC means $\exists x x \text{ licenses an alert}$, D means $\exists x x \text{ licenses recruitment}$, and ABC-D means $\exists x [x \text{ licenses an alert and } x \text{ licenses recruitment}]$. Here the two occurrences of the variable x depend on the same quantifier $\exists x$. By contrast, contiguous separate utterances of ABC and D yield the result: $[\exists x x \text{ licenses an alert}] \text{ and } [\exists y y \text{ licenses recruitment}]$. Here the two variables x and y depend on different existential quantifiers, and as a result the combined statement is strictly weaker than ABC-D (it is entailed by but does not entail ABC-D, as the combined statement does not guarantee

that one and same thing licenses an alert and recruitment). We have again highlighted in grey assumptions that we will later question. (As far as we can tell, nothing would change in the analysis below if we replaced our declarative meanings with the following imperative meanings: ABC = *Pay attention to something that licenses an alert!* D: *Pay attention to something that licenses recruitment!* ABC-D = *Pay attention to something that licenses an alert and recruitment!*)

Figure 4. Minimal Compositionality

Meaning of ABC	Meaning of D	Meaning of ABC-D	Meaning of a ABC and D combined as separate utterances
$\exists x$ x licenses an alert	$\exists x$ x licenses recruitment	$\exists x$ [x licenses an alert and x licenses recruitment]	[$\exists x$ x licenses an alert] and [$\exists y$ y licenses recruitment]

How do the details work? The key idea is that ABC and D start out with a predicative meaning: like the adjective *dangerous*, they are true of certain things and not others. So ABC means in essence *licenses an alert*, D means in essence *licenses recruitment*. ABC-D involves a mere operation of semantic conjunction, but at the level of predicative expressions. If S and S' are two predicative signals, S-S' just means: *is an x that satisfies S and S'*. In this way, ABC-D starts out with the meaning: *licenses an alert and licenses recruitment*.

These meanings do not convey information about the world because they cannot be true or false on their own (technically, they are predicative, not propositional). Similarly, in human language, *is red* or *is dangerous* uttered on their own do not provide information about the world; one needs to specify a subject. In the bird case too, one needs to further specify *what* licenses an alert, recruitment, or both. We take this additional information to be provided (in these cases, and perhaps more generally) at the level of entire utterances, which are defined by the presence of a pause before and after one call or two calls. Writing *_S_* for a signal preceded and followed by a pause, a single rule applies to all three cases (namely ABC alone, D alone, ABC-D):

- (3) If a signal S is preceded and followed by a pause, *_S_* is true just in case there is some object x that satisfies X.

So all our calls and call combinations have an existential reading: *_ABC_* means that *some object x in the environment licenses an alert*, *_D_* means that *some object x in the environment licenses recruitment*, and *_ABC-D_* means that *some object y in the environment licenses an alert and recruitment*. A more formal definition can be found in Appendix II-A.

How does this analysis derive the Suzuki results? A summary is given in the table below, where we have again highlighted unexpected findings, as well an auxiliary assumption needed to make the theory work.

Figure 5. Summary of the analysis of Minimal Compositionality

Findings	Explanation with Minimal Compositionality
<p>1. Non-sequential response ABC-D triggers the simultaneous rather than sequential production of scanning and approach behaviour. (Suzuki et al. 2016, 2018)</p>	<p>The meaning of ABC-D, $\exists x$ [x licenses an alert and x licenses recruitment], makes reference to one event that licenses both components, and the response should take both into account.</p>
<p>2. Ordering constraint ABC-D triggers the behaviour, D-ABC doesn't (Suzuki et al. 2016 [scanning and speaker</p>	<p>There are two possibilities to explain why D-ABC fails to trigger the behaviour. Possibility 1 - Syntax: D-ABC is syntactically ill-formed.</p>

<p>approach], Suzuki & Matsumoto 2022 [approach towards a model predator, wing flicking]).</p>	<p>Possibility 2 - Pragmatics: the Urgency Principle applies within utterances.</p>
<p>3. Productivity Hybrid ABC-D* sequences trigger the behaviour [= combination of scanning and approach] even though they are never heard in nature, as D* is a call heard in willow tits, a sympatric species; D*-ABC doesn't trigger the behaviour. (Suzuki et al. 2017)</p>	<p>(i) Since D* has the same meaning as D, the semantic rule can treat ABC-D* as ABC-D. (ii) But it's unclear why ABC-D* isn't recognised as syntactically ill-formed.</p>
<p>4. Mobbing response ABC-D triggers mobbing of a model predator [model approach and wing flicking], unlike separate productions of ABC and D, or productions of D-ABC (Suzuki & Matsumoto 2022).</p>	<p>(i) By a version of the reasoning illustrated in Fig. 2, most situations in which there is an x that licenses both an alarm and recruitment are situations in which x is a predator that needs to be mobbed. (ii) But by the same reasoning, it could be that most situations in which there is an x that licenses alarm and there is a y that licenses recruitment are situations in which x = y and x/y requires mobbing. An auxiliary assumption is needed:</p> <p>Auxiliary assumption: because it involves two existential statements, [$\exists x$ x licenses an alert] and [$\exists y$ y licenses recruitment] is harder to relate to a single predator than $\exists x$ [x licenses an alert and x licenses recruitment].</p>
<p>5. 2-speaker experiment ABC-D triggers the behaviour [= approach towards the model predator] when produced from 1 but not from 2 speakers (Suzuki & Matsumoto 2022).</p>	<p>(i) The compositional rule should apply within a single utterance but not across two utterances. (ii) But it could be that separate utterances of ABC and D (in any order) could trigger the behaviour if they could be more easily related to the predator.</p>

1. Non-sequential response: It is unsurprising that ABC-D does not trigger a sequential response, as it provides information about (i) a single situation and (ii) a single thing (because in $\exists x$ [x licenses an alert and x licenses recruitment], each occurrence of x depends on the same quantifier). We note for future reference that the result would be derived with (i) even in the absence of (ii): the fact that the two calls provide information about the same situation suffices to explain why the information they provide gets aggregated.

2. Ordering constraint: As in all theories that posit a compositional rule and thus a syntax, one may posit that ABC-D triggers the behaviour while D-ABC does not because the latter is syntactically ill-formed. Alternatively, one may posit that the Urgency Principle applies within utterances, as the predicative meaning of ABC is *licenses an alert* while that of D is *licenses recruitment*.

3. Productivity: As in all theories that posit a compositional rule, the semantic system can treat D* in the same way as D because they have the same meaning. But it is surprising that the syntactic rule does not treat ABC-D* as ill-formed.

4. Mobbing response: Minimal compositionality can adopt a version of the reasoning illustrated in Fig. 2, to the effect that most situations in which there is an x that licenses both an alarm and recruitment are situations in which x is a predator that needs to be mobbed. But this raises a question: by the same

reasoning, it might be (although it does not logically follow) that most situations in which there is an x that licenses alarm and there is a y that licenses recruitment are situations in which $x = y$ and x/y requires mobbing.

Let us make the problem more concrete. Suzuki & Matsumoto's (2022) experiment shows that with ABC-D in the 1-speaker case, Japanese tits manage to relate the existential information *there is an x that satisfies ABC and D* to a predator model located 5 meters away from the sound source. In effect, they complement the call-based information they receive with a cognitive search that yields: ... *and this x is the bull-headed shrike that's 5 meters away from the source*. But this raises the following possibility about the 2-speaker case: (i) the target birds hear ABC, derive from it the information that *there is an x that licenses an alert*, and obtain with a cognitive step: ... *and this x is the shrike*; similarly, (ii) they hear D, hence *there is a y that licenses recruitment*, and obtain with a cognitive step: ... *and this y is the shrike*. Putting (i) and (ii) together, they get the information that the shrike licenses both an alert and recruitment, the very same result as in the 1-speaker case.

To avoid this unwelcome result, Minimal Compositionality needs an Auxiliary Assumption, to the effect that connecting the two existential statements ABC and D to the shrike is harder than doing so with the single existential statement ABC-D. This is a natural assumption because ABC-D provides more information than the combination of ABC and D.⁴ The caller has information that the recipient doesn't have, and the recipient's task is to identify the threat highlighted by the caller. For instance, there might be a hard to distinguish bird in the distance, a potential food source, and a shrike nearby. ABC produced alone could be made true by the bird in the distance or by the shrike nearby, hence an uncertainty on the recipient's part. D produced alone could be made true by the food source or by the shrike, hence a similar uncertainty. The uncertainty remains with ABC and D produced as separate utterances, since nothing requires that they should be made true by one and the same thing. But in this situation, ABC-D is made true by the shrike alone, hence no uncertainty (or in the general case a lesser uncertainty because the call is more specific).⁵

5. 2-speaker experiment: As in all compositional accounts, it is unsurprising that the semantic and syntactic rules do not apply across utterances. But one must still account, as in 4. above, for the fact that separate utterances of ABC and D do not trigger the target behaviour. This analysis leaves the possibility open that *if* ABC and D became easier to connect to the model predator to be mobbed (for instance by being positioned closer to it), a mobbing behaviour would arise.

It should be noted that Minimal Compositionality requires that both calls (ABC and D) should have an existential meaning. This is not at all obvious: one might well have thought that *_D_* alone just means something like *come here!*. The problem is that on this view, we lose any meaning difference between ABC-D as a single utterance, and ABC D as two separate but contiguous utterances. ABC-D means in essence: *there is an x such that [x licenses an alert and you should come here]*; and this is the very same meaning we get for the conjunction of the separate utterances ABC and D. (In logical terms, the problem is that the existential construction in the meaning of ABC-D cannot 'bind' anything in *you should come here*, as this is already a complete proposition.⁶) On the proposed analysis, then, one might expect the D call on its own to trigger a search for whatever it is that triggers the recruitment, for instance a food source; this is a bit broader than a 'come here!' meaning.⁷ We come back to this point below.

⁴ In logical terms, $\exists x [ABCx \text{ and } Dx]$ entails $\exists x ABCx$ and $\exists x Dx$, but the converse isn't true.

⁵ In the general case, we assume that (i) the sender uses a call (or call combination) C to provide information about the object most worthy of attention, and (ii) from the recipient's perspective, there might be several objects that make C true. As a result, the more informative C is, the lesser uncertainty there will be about the object that the recipient should turn its attention to.

⁶ More precisely still, the problem is that in logic $\exists x [Px \text{ and } q]$ has the same meaning as $[[\exists x Px] \text{ and } q]$ when q is a propositional expression that does not contain the variable x .

⁷ Still, the very meaning of the D call, pertaining to recruitment, suggests that whatever licenses the call should be near the caller, and thus the two candidate meanings (*there is something that licenses recruitment* vs. *come here*) are rather close.

III.2. How Minimal Compositionality could have evolved

Minimal Compositionality might seem particularly natural in view of the following observation: When two existential statements about noteworthy things are uttered in close succession (e.g., *for some x, Px* and *for some x, Qx*), they are likely to be about the same thing, at least if their meanings are compatible. When this is the case, “misinterpreting” the two existential statements as a single existential statement instead (*for some x, Px and Qx*) will arguably be more efficient. We provided an example of this above: ABC and D understood as separate existential statements could in principle refer respectively to a bird in the distance or a shrike (for ABC), and to a food source or a shrike (for D); ABC-D understood as a single existential statement could only refer to the shrike. The point is more general: when two existential statements in fact refer to the same thing, understanding them as a single existential statement makes it possible to react more appropriately and/or more quickly.

But there is a downside: sometimes, the two existential statements are in fact made true by different things—e.g., something satisfies ABC (e.g. a flying raptor), and something else satisfies D (e.g. a food source). Under such circumstances, interpreting the combination of the two calls as a single existential statement will lead to a cost, as the recipient will for instance prepare for a mobbing behaviour that’s inappropriate in the situation at hand. But this case is unlikely to arise when the two calls are produced in close succession. If so, this rare cost might be outweighed by the frequent benefit of processing a single existential statement, and ABC-D will come to be interpreted as a single existential statement in comprehension. These ideas are made precise in Appendix II, using notions of evolutionary stability proposed by Maynard Smith and Price (1973).

III.3. Summary, predictions and outlook

In sum, Minimal Compositionality offers a possible explanation of the results of Suzuki and colleagues, and vindicates their main insight: ABC-D is not the mere juxtaposition of ABC and D, but involves a non-trivial semantic composition. Semantically, ABC-D means in essence *something is of type ABC and of type D*, with a single existential quantification over noteworthy events, whereas ABC and D produced as separate utterances involve two (*something is of type ABC and something is of type D*). Furthermore, there is a plausible evolutionary scenario that might explain how ABC and D produced in close succession could have led to the semantic rule we posited.

Several empirical questions need to be investigated. First, we had to posit above that in Suzuki & Matsumoto’s (2022) experiment, it is harder to relate the calls’ content to the model predator in the 2-speaker condition than in the 1-speaker condition (because the information is less precise in the former than in the latter case). But there is nothing in the analysis that fully *prevents* the connection to be made. In particular, it might be expected that when the two calls are produced from distinct speakers, but closer to the predator model, mobbing behaviour does arise in the end.

Second, it is essential for the analysis that there should be a difference between *something is of type ABC and of type D* (= the meaning of ABC-D) and *something of type ABC and something is of type D* (= the meaning of separate utterances of ABC and D). If the D call just means *come here*, the distinction evaporates. So the D call alone should have an existential meaning, akin to *something (nearby) licenses recruitment*. One might thus expect that the D call licenses a search. There is no currently evidence for this. Suzuki et al. 2016 (Figure 3) show that the D-call triggers much speaker approach but little scanning. On the present view, both behaviours might be expected.

Finally, a problem must be noted for any compositional analysis of ABC-D. Suzuki et al. 2017 show that Japanese tits interpret hybrid ABC-D* sequences as mobbing calls. But since ABC and D* are produced by different species, they couldn't be part of the same utterance. Since syntactic/semantic composition is usually thought to be limited to the confines of a single utterance, this is a somewhat odd state of affairs.

IV. Bird Implicatures

IV.1. From Minimal Compositionality to Bird Implicatures

As we saw, Minimal Compositionality must posit that (i) both ABC and D have an existential semantics; in addition, (ii) we need to explain why the two existential statements ABC and D are harder to relate to a shrike than the single existential statement ABC-D (this was crucial in the interpretation of Suzuki & Matsumoto's 2-speaker experiment). What if one were to deny (i) or (ii) or both? There would still be a way out... but it would lead to a different theory.

Recent studies of animal semantics have proposed an Informativity Principle, according to which monkeys use the most informative call applicable (e.g. Schlenker et al. 2014, 2016c). Take Titi monkeys, for instance (Schlenker et al. 2016b). They have a B call that occurs in predation and non-predation situations alike, and thus seems to have an unspecific meaning of alert. By contrast, the Titi A call seems to be restricted to serious aerial alerts. In a flying raptor situation, the Titis could in principle start their calling sequence with a series of B calls, but they don't, presumably because they have a more specific A call at their disposal. A consequence of the Informativity Principle is that the B call triggers an inference that *there isn't a serious aerial alert*, for otherwise the A call would have been used. Such inferences derived by the Informativity Principle are called 'implicatures' in human and increasingly in animal linguistics. In the end, then, the B call comes to have an enriched meaning of corresponding to 'there is an alert, but not a serious aerial one'.⁸

Following the same logic, ABC could be taken to compete with ABC-D. The latter is clearly more informative: if *there is something that licenses an alert and recruitment*, then *a fortiori there is something that licenses an alert*, but not conversely. As a result, ABC uttered as a separate utterance should in essence have the enriched meaning: *There is something that licenses an alert, but there is nothing that licenses recruitment* (... for otherwise, ABC-D would have been produced!). And similarly, D uttered as a separate utterance should have an enriched meaning, namely: *There is something that licenses recruitment, but there is nothing that licenses an alert* (... for otherwise, ABC-D would have been produced!). This will derive the result we want: ABC-D is predicted to trigger mobbing, and ABC and D produced as separate utterances are predicted *not* to trigger this behaviour, as their enriched meaning includes a non-mobbing component. (We note for future reference that one implicature would be enough; as soon as ABC triggers a not-ABC-D inference, *or* D triggers a not-ABC-D inference, we get the result we want, namely that mobbing is blocked. This point will matter in Section V.2).

This logic is sound, but from the perspective of Minimal Compositionality, it is self-defeating: once the Informativity Principle is adopted, the desired results are derived irrespective of whether Minimal Compositionality holds. All we need is that (i) ABC-D is a single utterance, (ii) ABC and D uttered as separate utterances compete with ABC-D, and (iii) ABC-D is more informative than ABC, and than D. Condition (iii) can be satisfied in numerous ways. Minimal Compositionality is one. But it could also be that ABC-D is just a discourse made of two separate utterances, just like when Ann says successively: *It's hot. It's humid*—an instance of trivial compositionality. It could even be that ABC-D is composed of two existential statements, which happen to be made true by a single noteworthy event.⁹ The meaning would be that *there is something that licenses an alert and there is something that licenses recruitment*. Depending on one's assumptions, it could still be that most situations in which there is something that licenses an alert and there is something that licenses recruitment are situations that require mobbing.¹⁰ On this view, in Suzuki & Matsumoto's (2022) 2-speaker experiment, the target birds can in principle combine information from different sources to infer that the shrike must be mobbed. What defeats this inference is the fact that ABC uttered alone yields a not-ABC-D inference, and similarly for D uttered alone.

⁸ This is arguably confirmed by looking pattern: as shown by Cäsar et al. 2012 (Fig. 3), Titis look up as they hear an A call, but do not do so when they hear a B call.

⁹ As we noted in Section III.3, the analysis based on Minimal Compositionality did not claim that ABC and D uttered as separate utterances could not be related to a model predator to be mobbed, just that this was harder than a single utterance of ABC-D.

¹⁰ Nearly all such situations would also involve a single threat that licenses an alert and recruitment.

So while Bird Implicatures obviate the need for the auxiliary assumptions that must be added to Minimal Compositionality, they make Minimal Compositionality itself dispensable in the end. We will now proceed to formulate in greater detail this alternative analysis based on Bird Implicatures. As we will see, it has a conceivable empirical advantage over Minimal Compositionality, because unlike the latter it does not have to posit that the D call uttered alone has an existential meaning.

IV.2. Analysis with Bird Implicatures

To develop the analysis based on Bird Implicatures, we first define the Informativity Principle in greater generality, and then show how it can be applied to ABC-D sequences to derive the desired results.

The Informativity Principle has been used in several studies of monkey calls, not just in Titi monkeys (Schlenker et al. 2014, 2016b, c). In general, the principle states when a call C competes with a more informative call C', C' must be used rather than C if both are licensed by the caller's information. As a result, if C is used, one can infer that C' was false (in our Titi example C is the B call, C' is the more informative A call). One crucial extension is needed to apply the Informativity Principle to Japanese tits: we must take the single call ABC to compete with the 2-call sequence ABC-D, and similarly D must compete with ABC-D. In this more general form, the Informativity Principle can be stated as in (4). We state two versions of it depending on whether it applies whenever it does not make the caller's message contradictory (Version 1), or whenever it does not make the receiver's information state contradictory (Version 2; in this version the receiver applies the Informativity Principle to the message unless the result contradicts the receiver's beliefs).¹¹

(4) **Informativity Principle**

Suppose that an expression C has an expression C' as an alternative. If C is uttered and C' could have been uttered instead, infer that C' is false.

Version 1-Caller-centred: Apply the Informativity Principle unless this makes the caller's message contradictory.

Version 2-Receiver-centred: Apply the Informativity Principle unless this makes the receiver's information state contradictory, i.e. unless this contradicts the receiver's beliefs.

On the assumption that ABC-D is more informative than ABC, and than D, an utterance of ABC alone or of D alone will trigger the inference that ABC-D is not applicable, hence no mobbing is called for. This immediately explains why ABC and D played back from different loudspeakers do not trigger mobbing, whereas ABC-D uttered from a single speaker does. Importantly, on this analysis ABC-D might be the mere juxtaposition of two independent calls (hence an instance of trivial compositionality). The crucial assumption is just that ABC and D uttered from different sources give rise to a 'not ABC-D' (hence 'no mobbing') inference.

To make the analysis concrete, we must decide (i) what the meanings of ABC, of D and of ABC-D are, and (ii) what the alternatives are. On the meaning side, the details do not matter as long as ABC-D is more informative than ABC and than D. But for the sake of comparison with Minimal Compositionality, we discuss both the general case and a special case in which each call has an existential meaning, and ABC-D is interpreted through trivial compositionality: ABC means that *something licenses an alert*, D means that *something licenses recruitment*, and ABC-D means that *something licenses an alert and something licenses recruitment*. We do this for clarity only, since we highlighted in the preceding section that the existential analysis of D is probably a liability, not a virtue.

Regarding alternatives, we will assume for simplicity that the three expressions ABC, D, ABC-D are all alternatives to each other. But the Informativity Principle only has some bite if an expression has an alternative that is more informative, and this only arises when ABC competes with ABC-D and when D competes with ABC-D.

¹¹ In original applications of the Informativity Principle to Campbell's monkeys (Schlenker et al. 2014), the Informativity Principle was taken to be optional, but something a bit stronger is needed in the present case to derive the 'no mobbing' inferences.

(5) Alternatives

The three expressions ABC, D, ABC-D are all alternatives to each other.

The following table provides a summary of the literal meanings and implicatures obtained, both in the general case, and in the case in which ABC and D each have an existential meaning.

Figure 6. Bird Implicatures

		Meaning of ABC alone	Meaning of D alone	Meaning of ABC-D	Meaning of a ABC and D occurring as separate and non and non-juxtaposed utterances
General case	Literal meaning	Any meaning that licenses an alert	Any meaning that licenses recruitment (including: Come here!)	Any meaning that licenses mobbing, possibly just: ABC and D, as long as it's more informative than ABC and than D.	ABC and D
	Implicature	not ABC-D, hence: no mobbing is called for.	not ABC-D, hence: no mobbing is called for.	None (because ABC-D does not have a more informative alternative).	not ABC-D, hence: no mobbing is called for (because in each case, ABC-D could have been produced but wasn't).
Special case: Two existential meanings	Literal meaning	$\exists x$ x licenses an alert	$\exists x$ x licenses recruitment	$[\exists x$ x licenses an alert] and $[\exists y$ y licenses recruitment] (using trivial compositionality)	$[\exists x$ x licenses an alert] and $[\exists y$ y licenses recruitment]
	Implicature	not: $[\exists x$ x licenses an alert] and $[\exists y$ y licenses recruitment]	not: $[\exists x$ x licenses an alert] and $[\exists y$ y licenses recruitment]	None (because ABC-D does not have a more informative alternative).	not: $[\exists x$ x licenses an alert] and $[\exists y$ y licenses recruitment] (because in each case, ABC-D could have been

					produced but wasn't).
--	--	--	--	--	-----------------------

In the general case, the key elements are that ABC should trigger an alert, D should trigger recruitment, and ABC-D should trigger mobbing and be more informative than ABC and than D. ABC produced alone yields the implicature that ABC-D could have been produced but wasn't because it is false. This implicature is derived when ABC is not immediately followed by D, or is followed by D as produced by another bird; what is common to the two cases is that the caller could have produced ABC-D but didn't. Similarly, D produced alone yields the implicature that ABC-D could have been produced but wasn't because it is false. In both cases, it follows from the implicature ('not ABC-D') that no mobbing is called for. ABC-D obviously does not yield this implicature.

As a special case, we can take ABC to mean that *something licenses an alert*, D to mean that *something licenses recruitment*, and ABC-D to mean that *something licenses an alert and something licenses recruitment*, using trivial compositionality. ABC-D could come to trigger mobbing through the same reasoning as was sketched in Fig. 2, and in Fig. 5, row 4., (ii): *most situations in which there is an x that licenses alarm and there is a y that licenses recruitment are situations in which x = y and x/y requires mobbing*. As in the general case, when ABC is produced alone it will trigger a not ABC-D implicature, hence 'no mobbing is called for', and similarly for D produced alone. ABC-D will produce no such implicature, but separate utterances of ABC and D (produced with a pause or by different birds) will, as for each component, ABC-D could have been produced instead.

We turn to a derivation of the main results using Bird Implicatures. A summary appears in Fig. 7.

Figure 7. Summary of the analysis of Bird Implicatures

Findings	Explanation with Bird Implicatures
<p>1. Non-sequential response ABC-D triggers the simultaneous rather than sequential production of scanning and approach behaviour. (Suzuki et al. 2016, 2018)</p>	<p>Both components of ABC-D are uttered at almost the same time and thus provide information about the same situation, hence the response should take both into account.</p>
<p>2. Ordering constraint ABC-D triggers the behaviour, D-ABC doesn't (Suzuki et al. 2016 [scanning and speaker approach], Suzuki & Matsumoto 2022 [approach towards a model predator, wing flicking]).</p>	<p>There are two possibilities to explain why D-ABC fails to trigger the behaviour. Possibility 1 - Syntax: D-ABC is syntactically ill-formed. Possibility 2 - Pragmatics: the Urgency Principle applies within utterances. Possibility 2 does not require any syntactic integration between ABC and D.</p>
<p>3. Productivity Hybrid ABC-D* sequences trigger the behaviour [= combination of scanning and approach] even though they are never heard in nature, as D* is a call heard in willow tits, a sympatric species; D*-ABC doesn't trigger the behaviour. (Suzuki et al. 2017)</p>	<p>(i) Since D* has the same meaning as D, whatever semantic rule is posited can treat ABC-D* as ABC-D. (ii) But it's unclear why ABC does not give rise to a 'not ABC-D' implicature, since ABC is produced by a Japanese tit that could have uttered ABC-D but didn't.</p>
<p>4. Mobbing response ABC-D triggers mobbing of a model predator [model approach and wing</p>	<p>By a version of the reasoning illustrated in Fig. 2, most situations that license an alarm and recruitment might be situations that license</p>

flicking], unlike separate productions of ABC and D, or productions of D-ABC (Suzuki & Matsumoto 2022).	mobbing. Special case of the Two existential meanings theory: most situations in which there is an x that licenses alarm and there is a y that licenses recruitment are situations in which $x = y$ and x/y requires mobbing.
5. 2-speaker experiment ABC-D triggers the behaviour [= approach towards the model predator] when produced from 1 but not from 2 speakers (Suzuki & Matsumoto 2022).	Each separate utterance of ABC and D (played back from different speakers) gives rise to a ‘not ABC-D’ implicature, hence: no mobbing is called for. ABC-D played back from a single speaker does not trigger this implicature.

1. Non-sequential response: As in all theories, it is unsurprising that ABC-D does not trigger a sequential response, as it provides information about a single situation (because ABC and D are produced at almost the same time). Correspondingly, the information contributed by the two components should be combined before a response is determined.

2. Ordering constraint: Here the details depend on the theory under consideration. If ABC-D is taken to be syntactically combined rather than to be the juxtaposition of two utterances, the ordering may come from the syntax. If the component of ABC-D are just juxtaposed utterances, as in the Two existential meanings theory, the Urgency Principle may be responsible for the fact that ABC-D triggers the behaviour but D-ABC doesn’t.

3. Productivity: As in all theories that posit a semantic rule (whether through trivial or non-trivial compositionality), the semantic system can treat D^* in the same way as D because they have the same meaning. Depending on which version of the Informativity Principle is adopted, it may be surprising that ABC does not give rise to a ‘not ABC-D’ implicature. This result is surprising on the caller-centred version of the Informativity Principle (= Version 1 in (4)). The reason is this: ABC and D^* are produced by different birds and thus by different callers, namely a Japanese tit for ABC and a Willow tit for D^* . The ‘not ABC-D’ implicature doesn’t make the Japanese tit’s message contradictory and so it should be derived, predicting a ‘no mobbing’ behaviour. By contrast, on the receiver-centred version of the Informativity Principle (= Version 2 in (4)), the implicature shouldn’t be derived because it contradicts the informational contribution of D^* . In other words, on the caller-centred version of the Informativity Principle, a ‘not ABC-D’ implicature is triggered by ABC because the ABC and D^* are produced by different callers. On the receiver-version of the principle, ABC does not trigger this implicature because ABC and D^* are perceived by the same receiver. This gives the receiver version (= Version 2) an edge over the caller version (= Version 1).

4. Mobbing response: Bird Implicatures can adopt a version of the reasoning illustrated in Fig. 2, to the effect that most situations that license an alarm and recruitment are situations that license mobbing. In the special case of the Two existential meanings theory, this becomes: most situations in which there is an x that licenses alarm and there is a y that licenses recruitment are situations in which $x = y$ and x/y requires mobbing.

5. 2-speaker experiment: According to the caller-centred version of the Informativity Principle (= Version 1 in (4)), in the 2-speaker experiment, ABC should trigger a ‘not ABC-D’ implicature, since ABC-D could have been uttered but wasn’t; and similarly D should trigger a ‘not ABC-D’ implicature. No implicature should arise in the 1-speaker experiment, since ABC-D was in fact produced. Importantly, the prediction is that no matter how easy it is to relate ABC and D to the predator model, no mobbing should be triggered in the 2-speaker condition (as ‘not ABC-D’ implicatures will be triggered).

Things are more complex in the receiver-centred version of the Informativity Principle (= Version 2 in (4)), as the derivation of the ‘not ABC-D’ implicature might give rise to a contradiction for the receiver when information from the two speakers is aggregated. But since they are played from different locations, they might pertain to slightly different situations, which might remove the contradiction.

Importantly, the theory based on Bird Implicatures is compatible with an analysis of ABC-D as separate utterances. In the discussion of Separate Utterances in Section II, the 2-speaker experiment raised the following problem: in view of the 1-speaker experiment, the target birds ought to be able to relate the production of ABC to the model shrike, and similarly for the production of D (from a separate speaker). This, in turn, should give rise to the same information as ABC-D, and mobbing should occur. The reason Bird Implicatures avoid this consequence is that the ‘not ABC-D’ implicature triggered by each call blocks this inference.

IV.3. How the Informativity Principle could have evolved

We noted in Section III.2 that Minimal Compositionality lends itself to a natural evolutionary scenario. But this doesn't give Minimal Compositionality an edge over an analysis based on the Informativity Principle because the latter also makes excellent evolutionary sense.

The Informativity Principle presumably derives from an optimization process. In humans, a similar principle was taken by Grice 1989 and a long line of researchers to arise from cooperative rational behaviour. If I tell you that *I'll invite Ann or Bill* while knowing full well that I'll invite both, I am being less than cooperative. In animals, the principle might be an automatic process (e.g., Schlenker et al. 2014), but it presumably arose from an evolutionary process that yielded a preference for more informative calls.

For simplicity, we will think about it in two evolutionary steps (the reasoning can extend to other applications of the Informativity Principle, but we restrict attention to the case at hand). First, we consider a resident population that does not go by the Informativity Principle in production. In other words, sometimes a resident produces ABC or D alone even though ABC-D is licensed. The details of the meanings of ABC, D and ABC-D don't matter as long as ABC-D is more informative than ABC, and/or than D. In particular, we could take ABC-D to just be the conjunction of ABC and D (an instance of trivial compositionality). By being under-informative, a resident will on average produce less shared utility than a mutant that followed the Informativity Principle. For instance, in some cases in which mobbing is called for, the resident only produces ABC, whereas the mutant invariably produces ABC-D. In this way, the resident strategy will be invaded in the end, and the Informativity Principle will come to dominate in production.

In a second evolutionary step, the Informativity Principle will emerge in comprehension. On the assumption that production strategies are now fixed, we start from a resident population that applies the Informativity Principle in production but fails to apply it in comprehension. This means that a resident that hears ABC considers all possible cases of alert and corresponding reactions, including ones that involve mobbing. A mutant that follows the Informativity Principle in comprehension can tell that the appropriate reaction doesn't involve mobbing. This yields a more effective reaction and greater shared utility. In the end, the resident strategy will be invaded by the mutant strategy and the Informativity Principle will come to be applied not just in production but also in comprehension. A more detailed discussion can be found in Appendix II.

IV.4. Summary, predictions and outlook

In sum, an analysis based on Bird Implicatures (and thus on the Informativity Principle) offers an alternative account of ABC-D sequences in Japanese tits. Importantly, the analysis does not depend on the details of the semantics of ABC-D, as long as two general conditions are met: first, ABC and/or D each evoke ABC-D as an alternative; second, ABC-D is more informative than ABC, and/or than D. In particular, the implicature-based analysis is compatible with the view that ABC-D is combined by way ‘trivial compositionality’, in the sense that ABC and D are just two juxtaposed utterances.¹²

¹² Two potential differences must be highlighted relative to the case of human language. First, scalar implicatures, the best studied case of human implicatures, are usually thought to be restricted to the confines of a single sentence. For instance, in contemporary theories based on a covert exhaustivity operator, this operator is seen in essence as an unpronounced version of the word *only* (e.g. Chierchia, Fox & Spector, 2021). The word *only* can take scope over a clause but not over an entire discourse, and one expects the same constraint to apply to the covert exhaustivity operator. If ABC-D is made of two separate utterances, but ABC-D still counts as an alternative to ABC, and to D, this is a potential difference with the human case. Given that human language and

There are two respects in which the present analysis makes different predictions from Minimal Compositionality. First, in the latter analysis, the production of separate utterances of ABC and D didn't fully rule out a mobbing message, but it made it harder to infer than the production of ABC-D. By contrast, for Bird Implicatures, ABC and D on their own should trigger a 'not ABC-D' inference and thus a non-mobbing inference. Second, it was important for Minimal Compositionality that D on its own should have an existential meaning (along the lines of: *something (near me) licenses recruitment*). This was because the analysis was based on the distinction between *something licenses ABC and D* (for ABC-D) and: *something licenses ABC, something licenses D* (for separate utterances). As we saw, there is currently no evidence for this existential component, at least in the data of Suzuki et al. 2016. The theory based on Bird Implicatures needs no such assumption: as long as ABC-D is more informative than ABC, and/or than D, it can get off the ground (this remark applies, among others, to analyses that take ABC and D to have imperative meanings).

Finally, we note that, in one version of the Informativity Principle (= the caller-centred version), the analysis of ABC-D calls based on Bird Implicatures encounters a version of the issue that afflicted Minimal Compositionality with hybrid sequences. On the assumption that Japanese tits recognise that ABC-D* is made of calls from two different species, ABC should trigger a 'not ABC-D' inference because it doesn't make the caller's message contradictory. This problem does not arise in the receiver-centred version of the Informativity Principle (because a 'not ABC-D' inference contradicts the contribution of D*).

V. Extensions to Chimpanzees and Meerkats

V.1. Chimpanzees

a. Experimental results of Leroux et al. 2023

Leroux et al. 2023 argue that something rather similar to ABC-D sequences can be found in chimpanzees. In brief, chimpanzees produce alarm hoos "when they are frightened or surprised (e.g. earth tremors, snakes, dead monkeys, researcher's waterproof cloaks)"; they produce waa barks "in a range of social and ecological contexts such as hunting, predator encounters, intercommunity encounters and aggression", and this call "has been argued to play a role in recruiting individuals to the caller". But they produce hoo-waa combinations in a rather different kind of situation, namely "when encountering a snake, specifically when isolated from other individuals but still within earshot". The authors hypothesize that the meaning of hoo-waa might be compositionally derived from the meaning of hoo and of waa, possibly yielding a meaning of 'recruitment to a threat'.

Leroux et al. 2023 obtained several key experimental results.

1. They presented snake models to chimpanzees and observed that (i) in 43% of cases, they produced hoo-waa combinations, while in the remaining cases they produced hoos alone; and that (ii) hoo-waa combination triggered approach by conspecifics, but hoos alone didn't.
2. In a playback experiment, hoo-waa combinations triggered longer looking towards the loudspeaker than hoos alone or waas alone. Furthermore, the effect wasn't additive, in the sense that looking time triggered by hoo-waa was far greater than the sum of the looking times triggered by hoo and by waa.

bird calls are entirely different systems, it's unclear what this shows. Second, algorithms of alternative generation in human language take a complex expression to automatically evoke its subcomponents, at least within the confines of a single sentence (Katzir 2007, Katzir & Fox 2011). For instance, *I didn't drink a lot* automatically evokes *I didn't drink*, and since the latter is more informative, an implicature is triggered to the effect that this alternative is false, and thus that *I drank*. By contrast, in the absence of a specific context, *I drank* doesn't evoke *I drank a lot* (if it did, an implicature could conceivably be triggered to the effect that *I didn't drink a lot*, contrary to fact). In our analysis of ABC-D, by contrast, we take ABC and/or D to automatically evoke the more complex combination ABC-D. This is another point of departure from the human case, and this is so whether ABC-D is considered as a complex sentence or as a combination of utterances. Again, it is unclear how this difference from the human case should be interpreted.

Importantly, was alone triggered looking behaviour, including quicker looks towards the loudspeaker compared to hoo alone, unlike what has been shown in Japanese tits.

The authors also ask why chimpanzees recruit conspecifics in some snake encounters. They consider two (non-mutually exclusive) hypotheses. One is that this might "help chimpanzees encountering a snake ensure group members in the surroundings are aware of both the presence of the threat and its precise location". As they further note, "a recent study has shown that chimpanzees not only extract information from alarm call production on the presence of a snake, but also infer the specific location of the snake from the physical position of the signaller who orientates their body to "mark" the snake to naive individuals [Crockford et al. 2017]. This strategy seems to be very effective given chimpanzees that join an individual that has discovered a snake are not startled [Girard-Buttoz et al. 2020]." An alternative explanation is that "chimpanzees recruit individuals to the threat to help drive it away, akin to mobbing behaviour"; this would of course make chimpanzee hoo-waa sequences particularly similar to Japanese tit ABC-D sequences.

b. Analysis

As before, we can first ask whether these results can be explained if (i) hoo-waa is analysed as a single expression ('Only one expression' theory), and if (ii) hoo-waa is analysed as two separate utterances ('Separate utterances' theory).

The 'Only one expression' cannot be fully refuted, but it is a bit implausible because it makes it an accident that hoo and waa *as they appear in other contexts* can be combined to yield a hoo-waa sequence. In particular, playbacks using artificial concatenation of hoo and waa produced in isolation trigger the characteristic behaviour of hoo-waa sequences. In addition, the same waa can be combined with other calls, for instance screams (Leroux et al. 2022).

When it comes to the 'Separate utterances' theory, Leroux et al. 2023 explicitly state that they cannot reject such a conjunctive analysis. The situation is somewhat similar to that of Japanese tit ABC-D sequences before Suzuki and Matsumoto 2022 published their 2-speaker experiment. In particular, just as is the case of D-ABC, the reverse order waa-hoo is never produced in the relevant context.¹³ One natural thought would be to perform a version of Suzuki and Matsumoto's 2-speaker experiment to see if hoo-waa needs to be produced by a single source to be effective.

On the assumption that analyses based on Only one expression and on Separate utterances can be rejected, the same question will arise as in ABC-D sequences: what is the nature of the rule that makes hoo-waa effective in triggering a mobbing behaviour (or whatever the precise effect of hoo-waa turns out to be) while the conjunction of its component parts as separate utterances doesn't have this effect? As far as we can tell, both theories entertained in the present piece, Minimal Compositionality or Implicatures, could make good initial sense.

V.2. Meerkats

In the words of an important review article (Manser et al. 2014), "meerkats show one of the most sophisticated alarm call systems described for any animal species with predator-type specific and more general alarm calls". Alarm calls are discreetly distinct for different threats, including aerial predators, terrestrial predators, snakes, moving animals, and panic calls (e.g. Manser, Bell & Fletcher, 2001). Aerial and terrestrial calls are graded by levels of urgency (possibly with the same acoustic modifications affecting different calls, Manser 2001). In addition, meerkats have social calls. Altogether, meerkats are taken by Manser et al. 2014 to have over 30 call types.

Alarm calls are described more precisely by Manser, Seyfarth & Cheney, 2002, who discuss (i) a call given to mammalian predators, primarily jackals (*Canis mesomelas*), which attack from the ground; (ii) a distinct call to avian predators, primarily eagles (martial eagles, *Polemaetus bellicosus*

¹³ Leroux et al. 2023 further note that hoo-waa yields non-additive looking time reactions relative to its component parts. But this does not speak against a conjunctive analysis (Schlenker et al. 2016d, 2023)—nor do Leroux et al. claim that it does. "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." (Schlenker et al. 2016d, p. 185-186).

and tawny eagles, *Aquila rapax*) and pale chanting goshawks (*Melierax canorus*), which attack from the air; and (iii) a third call type given to snakes, which also induces other animals to approach the caller to inspect or mob the predator. Meerkats also have (iv) an “animal moving call”; it is used “when either an animal on the ground passes by or a bird flies close to the ground, but also in response to stationary animals that subtly move body parts”.

What is of immediate relevance for us is that meerkats can produce combinations of a moving animal call and a terrestrial predator call (see also Collier, Townsend & Manser, 2017):

In certain contexts, specifically when terrestrial predators are moving, meerkats combine predator-specific terrestrial alarm calls with animal moving calls into longer call sequences. Receivers seem to respond to these naturally occurring sequences by increasing their antipredator behavior in comparison to the two call types produced by themselves (Townsend & Manser, in preparation). When playing back the two different call types by themselves in comparison to their combination, meerkats took longer to relax when exposed to the potentially more threatening terrestrial+animal moving call combination than when exposed to either of the meaningful calls by themselves or any of the other artificially created combinations or their individual call components (Townsend & Manser, in preparation). This suggests that meerkats can extract meaningful compositional information from combinations of referential calls. (Manser et al. 2014)

Strikingly, these facts raise the same theoretical issues about semantic composition as alarm-recruitment sequences, but with very different call functions. As in the case of chimpanzee hoo-waa sequences, several experimental steps are needed before one can make a pronouncement about the existence of a non-trivial rule of semantic composition. Specifically, one will have to refute the usual deflationary theories by answering the following questions:

1. Can the moving predator-terrestrial predator call sequence be analysed as a single expression? ('Only one expression' theory)
2. Can this call sequence be analysed as the combination of two separate utterances? ('Separate utterances' theory). One natural thought is that one should follow Suzuki and Matsumoto's logic and determine if the behavioural effect is different in 2-speaker than in 1-speaker conditions.

If both Only one expression and Separate utterances can be rejected, there will be an argument that some non-trivial combination rule is needed. Our earlier theories based on Minimal Compositionality and on Implicatures will both have to be considered, but in modified form.

Minimal Compositionality can be adapted as follows. The moving animal call means *there is something that is a moving animal*, the terrestrial predator call means *there is something that is a terrestrial predator*, and their combination in a single utterance means *there is something that is a moving animal and a terrestrial predator*, in other words: *there is a moving terrestrial predator*. By contrast, separate utterances of the two calls would yield the global information that *there is something that is a moving animal and there is something that is a terrestrial predator*, without 'binding' between the two component parts.

The implicature-based theory needs a tiny twist. On the assumption that a 2-speaker version of moving animal-terrestrial predator sequences fails to trigger the same behaviour as a 1-speaker version (something that hasn't been shown!), one might want to posit that the animal moving call uttered alone triggers an implicature that the animal moving-terrestrial predator combination isn't applicable. But this isn't very plausible, for by the same logic it should presumably trigger an implicature that an animal moving-aerial predator combination shouldn't be applicable. The consequence is that the moving animal call should almost never be used alone.¹⁴ This makes it somewhat unlikely that the moving animal call triggers the implicature that the moving animal-terrestrial predator call sequence wasn't applicable.

¹⁴ This cancellation of implicatures through the symmetry of alternatives is well-known in human language. *p or q* has as alternatives *p* as well as *q*. This can for instance be seen in sufficiently complex examples. For instance, *Everyone read Chomsky or Montague* gives rise to the inference that *it's not the case that everyone read Chomsky*, and also *it's not the case that everyone read Montague* (in other words, there should be some diversity in what different people read). By the same logic, one would expect *Ann read Chomsky or Montague* should have as alternatives that *Ann read Chomsky*, *Ann read Montague*. So one might expect an inference that *Ann didn't read*

Still, there is a way out: it could be that the terrestrial predator call produced alone gives rise to an implicature that *it isn't the case that there is a moving animal*. In view of what we know of the meerkat repertoire, this does not appear to give rise to the same problem. (For comparison, in the case of ABC and D uttered as separate utterances, it was plausible that each triggered a 'not ABC-D' inference, but as we noted, this was overkill: just one such inference would have been sufficient. In the case of moving animal-terrestrial predator call sequences, one implicature is unlikely, but the second one suffices to do the job.)

VI. Results and Perspectives

VI.1. Comparing Minimal Compositionality and Bird Implicatures

While Suzuki and colleagues have convincingly argued that ABC and D are combined in a non-trivial fashion in Japanese tits, the precise nature of the semantic rule involved has remained elusive. The main challenge is to explain the difference between ABC-D and separate utterances of ABC and D. We have considered two simple hypotheses.

Minimal Compositionality takes ABC-D to mean that *something licenses both an alert and recruitment*, with a single existential statement; separate utterances of ABC and D yield two existential statements instead: *something licenses an alert*, *something licenses recruitment*. Minimal Compositionality makes two salient predictions. First, a mobbing inference is harder to obtain for separate utterances of ABC and D than for ABC-D, but it is not impossible. Second, ABC and D alone should have an existential component, something that is unclear for the recruitment call D.

The competing analysis, Bird Implicatures, is compatible with diverse meanings as long as ABC-D triggers mobbing and is more informative than ABC alone, and/or than D alone. In particular, it is compatible with the view that ABC-D is combined by 'trivial compositionality', in the sense that the two components are separate utterances. What is essential is that the individual calls produced alone (in particular, from separate sources) automatically evoke the combination ABC-D, and yield the inference that ABC-D is not applicable, hence a 'no mobbing' inference. Unlike Minimal Compositionality, this analysis predicts that a mobbing inference should always be blocked when ABC and D are understood to be produced from different sources. And unlike Minimal Compositionality, this analysis is not at all committed to the view that D alone has an existential component.

A summary of the debate between the two theories appears in Fig. 8.

Figure 8. Summary comparison of Minimal Compositionality and Bird Implicatures

	Minimal Compositionality	Bird Implicatures
Main mechanism	<p>ABC means <i>something licenses an alert</i></p> <p>D means <i>something licenses recruitment</i></p> <p>ABC-D means <i>something licenses an alert and recruitment</i></p>	<p>ABC-D is more informative than ABC and/or than D.</p> <p>ABC and/or D alone trigger a 'not ABC-D' inference thanks to the Informativity Principle.</p>
Auxiliary assumptions	Both ABC and D alone have an existential semantics.	ABC and/or D alone evoke ABC-D as an alternative.

Chomsky. But by the same logic, one would infer that *Ann didn't read Montague*, which would defeat the original utterance. See for instance Fox 2007 for an account of implicatures that addresses this issue.

Does the theory require non-trivial compositionality?	Yes: the meaning of ABC-D is different from the conjunction of the meanings of ABC and D.	No: as long as ABC-D is evoked as an alternative by ABC and D, and is more informative than them, an implicature can be triggered.
Salient predictions	<ul style="list-style-type: none"> • D should make reference to the existence of a noteworthy event (rather than just inducing the receiver to move towards the caller). • ABC and D produced as separate utterances should be compatible with a mobbing behaviour (but should have greater difficulty triggering it than ABC-D). 	<ul style="list-style-type: none"> • D need not make reference to the existence of a noteworthy event (it could be a pure recruitment call). • ABC and/or D produced as separate utterances should be incompatible with a mobbing behaviour.
Possible advantages of the theory	<ul style="list-style-type: none"> • There is a natural evolutionary scenario about the emergence of the rule. 	<ul style="list-style-type: none"> • There is a natural evolutionary scenario about the emergence of the rule.
Possible drawbacks of the theory	<ul style="list-style-type: none"> • The prediction about the existential nature of D is currently unsupported. 	<ul style="list-style-type: none"> • There is no independent evidence for the assumption that a simple call may evoke a call combination as an alternative.

VI.2. The typology of Merge

Rizzi 2016 proposes a typology of Merge (the operation by which two expressions are syntactically combined, e.g. Chomsky 1995) for animal linguistics. How does the debate between Minimal Compositionality and Bird Implicatures bear on Rizzi's typology?

In Rizzi's typology, the base level is '0-merge', where no two expressions are ever syntactically combined (juxtaposition of utterances, and correspondingly trivial compositionality, is all there is). By contrast, at the level of '1-merge', Merge "can apply, forming two-word expressions, but then the system stops, i.e., it lacks recursive procedures"; a schematic example appears in (6a).¹⁵

(6) Merging two versus three elements

a. Merging two elements: [C C']

Example: [ABC D] according to Minimal Compositionality, with ABC counting as a single call

b. Merging three elements (imaginary example): [C C' C'']

Minimal Compositionality comes very close to 1-merge, but with a twist. To get the meaning of ABC-D, it is essential that ABC and D should not be combined as separate utterances but as predicates, corresponding to: *licenses an alert and licenses recruitment*. Existential quantification over this conjunction yields *something licenses an alert and licenses recruitment*, as is desired within that analysis. But there is nothing in the analysis that would prevent an extension of the reasoning to series of three calls, of C C' C'', as illustrated in (6b).¹⁶ One would expect that this should yield a meaning akin to: *something licenses a C-type content and a C'-type content and a C''-type content*.

¹⁵ See Miyagawa & Clarke 2019 for a distinct but related idea, to the effect that some animal systems work by way of a non-recursive syntactic 'template'.

¹⁶ As far as we can tell, this would also be permitted by Miyagawa and Clarke's (2019) analysis.

Thus, if Minimal Compositionality were to *replace* Rizzi's 1-merge, it would give rise to a slightly more liberal system, with non-recursive combinations of multiple calls, not necessarily limited to two. On the other hand, if Minimal Compositionality is *added* to Rizzi's 1-merge (as a semantic system can be added to a syntax), the result will be non-recursive combinations of two calls only (note that there might be independent reasons to block the processing of three calls, as argued by Watson et al. 2023 for the Chestnut-crowned babbler, *Pomatostomus ruficeps*). A clear conclusion of this discussion is that the existence of 3-call combinations should be investigated.

What about the theory based on Bird Implicatures? It only relies on a non-syntactic notion of combination, to the effect that ABC-D should be sufficiently unified to count as an alternative to ABC, and to D. As we emphasized, Bird Implicatures is compatible with diverse views of the meaning of ABC, D and ABC-D as long as ABC-D is an alternative to ABC and to D, and is more informative than them. This does not involve syntax (let alone compositionality) in the usual sense.

VII. Conclusions

1. The logic of recent experiments by Suzuki and colleagues convincingly argues that in Japanese tit ABC-D sequences, the meaning of ABC-D is derived in a non-trivial way from the meanings of ABC and of D. This calls for explicit hypotheses about the nature of the rule involved.
2. A theory based on Minimal Compositionality posits a non-trivial compositional rule in the semantics. It has the effect that ABC-D means *something licenses an alert and recruitment*, whereas ABC and D mean on their own respectively: *something licenses an alert*, *something licenses recruitment*.
3. An alternative theory based on Bird Implicatures posits instead that the meaning of ABC-D can be derived through trivial compositionality, with ABC and D treated as separate but juxtaposed utterances. But it posits some complexity in the pragmatics, with ABC and D triggering a 'not ABC-D' implicature when uttered alone (e.g. from difference sources).
4. Both theories are associated with natural evolutionary scenarios. But they make different predictions about the meaning of D, and about the possibility of triggering mobbing behaviour when ABC and D are played from different sources that are close enough to a predator model.
5. Depending on the results of future experiments, the same theoretical debate might arise in chimpanzee hoo-waa sequences and in meerkat terrestrial-'animal moving' sequences.
6. Debates on animal communication can benefit from the statement of precise competing hypotheses using recent methods from animal linguistics.

VIII. References

- Cäsar, C., Byrne, R., Hoppitt, W., Young, R. J. & Zuberbühler, K. (2012). Evidence for semantic communication in Titi monkey alarm calls. *Animal Behavior* 84: 405–411. doi:10.1016/j.anbehav.2012.05.010.
- Chierchia, G., Fox, D. & Spector, B. (2012). Scalar implicatures as a grammatical phenomenon. In Maienborn, C., von Stechow, K. & Portner, P. (eds.), *Semantics: An international handbook of natural language meaning* (Vol. 3, pp. 2297–2331). Berlin: Mouton de Gruyter, Genthiner Straße 13, Berlin, Berlin 10785, DE
- Chomsky, N. (1995) *The Minimalist Program*. MIT Press.
- 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.
- Collier, K., Townsend, S.W. & Manser, M.B. (2017). Call concatenation in wild meerkats. *Animal Behaviour*, 134, pp.257–269.
- Crockford, C., Wittig, R. M. & Zuberbühler, K. (2017) Vocalizing in chimpanzees is influenced by social-cognitive processes. *Science Advances* 3, e170174.
- Fox, D. (2007). Free choice and the theory of scalar implicatures. In *Presupposition and implicature in compositional semantics* (pp. 71–120). London: Palgrave Macmillan UK.
- Girard-Buttoz, C., Surbeck, M., Samuni, L., Tkaczynski, P., Boesch, C., Fruth, B., Wittig, R.M., Hohmann, G. & Crockford, C. (2020). Information transfer efficiency differs in wild chimpanzees and bonobos, but not social cognition. *Proceedings of the Royal Society B*, 287(1929), p.20200523.
- Grice, H. P. (1989). *Studies in the way of words*. Cambridge: Harvard University Press. Heim, I. & Heim, I. and Kratzer, A. (1998). *Semantics in generative grammar*. Blackwell Publishers Inc./Blackwell Publishers Ltd., Maiden, MA, USA/Oxford, UK (199
- Katzir, R. (2007). Structurally-defined alternatives. *Linguistics and Philosophy*, 30(6), 669–690.
- Katzir, R., & Fox, D. (2011). On the characterization of alternative. *Natural Language Semantics*, 19(1), 87–107.
- Leroux, M., Chandia, B., Bosshard, A. B., Zuberbühler, K. & Townsend, S. W. (2022) Call combinations in chimpanzees: a social tool? *Behav. Ecol.* 33, 1036–1043.
- Leroux, M., Schel, A.M., Wilke, C., Chandia, B., Zuberbühler, K., Slocombe, K.E. & Townsend, S.W. (2023). Call combinations and compositional processing in wild chimpanzees. *Nature Communications*, 14(1), p.2225.
- Manser, M. B. (2001) The Acoustic Structure of Suricates' Alarm Calls Varies with Predator Type and the Level of Response Urgency. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268: 2315–2324.
- Manser, M. B, Bell, M. B. & Fletcher, L. B. (2001) The information that receivers extract from alarm calls in suricates. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, Vol. 268, No. 1484: 2485–91.

- Manser, M. B., Jansen, D. A. W. A. M., Graw, B., Hollen, L. I., Bousquet, C. A. H., Furrer, R. D. & Leroux, A. (2014). Vocal complexity in meerkats and other mongoose species. In Naguib, M., Barrett, L., Brockmann, H. J., Healy, S., Mitani, J. C., Roper, T. J. et al. (Eds.), *Advances in the study of behavior* (Vol. 46, pp. 281e310). London, U.K.: Elsevier. 125 London Wall, Barbican, London EC2Y 5AS.
- Manser, M. B., Seyfarth, R. M. & Cheney, D. L. (2002) Suricate alarm calls signal predator class and urgency. *Trends in Cognitive Sciences*. 6(2):55–7. doi: 10.1016/s1364-6613(00)01840-4.
- Maynard Smith, J. & Price, G. R. (1973), The Logic of Animal Conflict. *Nature*, 246(5427), pp.15–18.
- 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., Dezechache, 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.
- 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>
- 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. 1 & 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., Veselinović, 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., Veselinović, 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
- Schlenker, P. Coye, C., Steinert-Threlkeld, S., Klinedinst, N. & Chemla, E. (2022). Beyond Anthropocentrism in Comparative Cognition: Recentering Animal Linguistics. *Cognitive Science* 46:e13220. <https://doi.org/10.1111/cogs.13220>
- Schlenker, P., Coye, C., Leroux, M. & Chemla, E. (2023). The ABC-D of animal linguistics: are syntax and compositionality for real? *Biological Reviews*.

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.

Skyrms, B. (1996). *Evolution of the social contract*. Cambridge, MA: Cambridge University Press.

Skyrms, B. (2010). *Signals: Evolution, learning, and information*. Oxford: Oxford University Press.

Steinert-Threlkeld, S. (2020). Toward the Emergence of Nontrivial Compositionality. *Philosophy of Science*, **87**(5), 897–909. doi:10.1086/710628

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. *Nature Communications*, **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>

Watson, S.K., Mine, J.G., O’Neill, L.G., Mueller, J.L., Russell, A.F. & Townsend, S.W. (2023). Cognitive constraints on vocal combinatoriality in a social bird. *Isience*, **26**(7).

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

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

Zuberbühler, K., Cheney, D.L. & Seyfarth, R.M. (1999). Conceptual semantics in a nonhuman primate. *Journal of Comparative Psychology*, **113**(1), p.33.

Appendix I. No Conjunction Without Juxtaposition

We consider a deflationary analysis in which Japanese tits fail to aggregate information from two calls unless they are next to each other. In other words, when they hear ABC-D from a single source, they can interpret it as a conjunction of ABC and D. But when they hear ABC and D from different sources, they just fail to aggregate the informational content of the two calls.

We take this view to be implausible (but not fully impossible) in view of data from other species. The short of it is that animals often tend to integrate the meaning of calls to what they know about the world. But if so, a bird in information state S that hears ABC produced from one source and then D produced from another source should first integrate ABC to S, yielding S+ABC; and then it should integrate D to this information state, yielding [S+ABC]+D. But this means that the informational content of the two calls should be aggregate in the end (through successive modifications of the bird's information state).

Several arguments are given in the literature for the cognitive integration of the meaning of animal calls to the receiver's information state. Seyfarth, Cheney & Marler, 1980 show that Vervet monkeys (*Chlorocebus pygerythrus*) react differently to one and the same call depending on their own position, suggesting that they combine the call meaning with their own information state before selecting a response. For instance, a Vervet monkey that hears an eagle alarm call sometimes looks down if the monkey is in a tree; but if it is on the ground, it does not do so as often. Using a different paradigm, Zuberbühler, Cheney & Seyfarth, 1999 show that there is a kind of 'conceptual equivalence' in Diana monkeys between an eagle shriek and an eagle-related conspecific call. The authors start from the observation that Diana monkeys dishabituate (i.e. react less strongly) to an eagle shriek if it is preceded by an earlier eagle shriek—which is unsurprising since the second occurrence provides information that the monkeys already have. Crucially, the same dishabituation occurs if the initial shriek is replaced with an eagle-related Diana call (but not a control leopard-related Diana call). The eagle-related Diana calls seems to be cognitively integrated in the same way as an eagle shriek.

While it could be that birds fail to integrate the meaning of calls to their cognitive state, this is not a particularly plausible hypothesis in our view.

Appendix II. Formal Analyses

A. Minimal Compositionality

The key to our formal analysis is that elementary calls and their juxtaposition are predicative: they are true of certain objects and not others (technically, they are of type $\langle e, t \rangle$, where e is the type of entities and t is the type of truth values). It is only when these calls or combinations of calls are separated by pauses that they become 'sentences' with truth values (technically, they are of type t). Since any information solely pertains to the caller's context, all semantic definitions are relativized to a context c .

Syntax

ABC, D are well-formed signals
 ABC-D is a well-formed signal
 If S is a well-formed signal $_S_$ is a well-formed utterance

Semantics

• ABC and D

When uttered in a context c , a signal S <within $\{ABC, D\}$ > is true of a set of objects; thus S^c is of type $\langle e, t \rangle$. Specifically:

$ABC^c = \{x: x \text{ is an object in the environment of } c \text{ and } x \text{ licenses an alert relative to } c\}$

$D^c = \{x: x \text{ is an object in the environment of } c \text{ and } x \text{ licenses recruitment relative to } c\}$

• Combinations

When uttered in a context c , for any signals S and S'

$[S-S']^c = \{x: x \text{ is an object in the environment of } c \text{ and } x \text{ is in } S^c \text{ and } x \text{ is in } S'^c\}$

• Utterances

When uttered in a context c , for any (possibly complex) signal S , $[_S_]^c = \text{true}$ if and only if for some object x , x is in S^c .

B. Evolutionary stability of Minimal Compositionality

We assume that comprehension and production rules can be distinct, and we focus first on the evolution of interpretive strategies. We will show that under certain conditions, the single-event interpretation is beneficial even if the production rule is compatible with there being two events.

On the production side, ABC means that *something licenses an alert* and D means that *something licenses recruitment*. ABC-D just involves these two existential statements produced at the same time.

• Interpretive strategies

We consider two interpretive strategies for such ABC-D calls combined without a pause: one takes them to be two existential statements (as is the case on the production side), the other takes them to form a single complex existential statement. When ABC and D in fact pertain to a single noteworthy event, the second strategy has an edge; when there are two noteworthy events, the former strategy does, but if this is a sufficiently rare occurrence, this still makes it beneficial to adopt the 'one existential statement' strategy.

The interpretive strategies, Sep for Separate utterances, Sin for Single utterance, are defined as follows:

Sep: Separate utterances—two existential statements

Interpret ABC-D as *there is something that licenses an alert and there is something that licenses recruitment*.

Sin: Single utterance—one existential statement

Interpret ABC-D as *there is something that licenses an alert and recruitment*.

We study the relative utility of these two interpretive strategies in case ABC-D has been produced.

Situations

Call situations are of two types:

(i) with probability e , the two parts of ABC-D are triggered by two separate events;

(ii) with probability $(1-e)$, the two parts of ABC-D are triggered by a single event, e.g., a predator that needs to be mobbed.

$$P(2\text{-events}) = e$$

$$P(1\text{-event}) = 1-e$$

Since the two parts of ABC-D are produced at the same time, the ‘two separate events’ situations are very uncommon and we can take e to be small.

Utilities

We assume that the Separate utterances interpretive strategy Sep gives rise to the same utility, 1, in all cases:

$$U_{\text{Sep}}(2\text{-threats}) = U_{\text{Sep}}(1\text{-threat}) = 1$$

As a result, irrespective of the probability e with which the two parts of ABC-D are triggered by a single event, the average utility obtained by Sep is 1 as well:

$$U_{\text{Sep}} = P(1\text{-threat}) U_{\text{Sep}}(1\text{-threat}) + P(2\text{-threats}) U_{\text{Sep}}(2\text{-threats}) = 1$$

By contrast, we take the Single utterance interpretive strategy Sin to give rise to different utilities depending on the situation. When the two parts of ABC-D are triggered by a single event, Sin yields more specific information than Sep does, and this information is correct, which yields greater utility.¹⁷ We’ll assume that ‘one event’ case, this yields a utility of $1+b$ rather than 1 (where b is the benefit of the additional specificity relative to the Separate utterances interpretive strategy). When the two parts of ABC-D are triggered by two events, Sin yields somewhat incorrect information, which yields a utility of $1-c$ rather than 1 (where c is the cost of the additional and incorrect specification? in this case).

$$U_{\text{Sin}}(1\text{-event}) = 1+b$$

$$U_{\text{Sin}}(2\text{-events}) = 1-c$$

Overall, the utility produced by Sin is a weighted sum of these two utilities:

$$U_{\text{Sin}} = P(1\text{-event}) U_{\text{Sin}}(1\text{-event}) + P(2\text{-events}) U_{\text{Sin}}(2\text{-events}) = (1-e)(1+b) + e(1-c)$$

We can thus ask under what conditions Sin produces greater utility than Sep:

¹⁷ As mentioned in the main text, there might be a hard to distinguish bird in the distance, a potential food source, and a shrike nearby. ABC-D interpreted as two separate existential statements doesn’t make it possible to decide. ABC-D interpreted as a single existential statement makes it possible to adopt the reaction appropriate to a shrike, notably mobbing.

$$\begin{aligned}
U_{\text{Sin}} > U_{\text{Sep}} & \text{ iff } (1-e)(1+b) + e(1-c) > 1 \\
& \text{ iff } 1 - e + b - eb + e - ec > 1 \\
& \text{ iff } b(1-e) > ec
\end{aligned}$$

This can be interpreted as follows: relative to Sep, Sin yields an additional utility of b in proportion $(1-e)$ of situations, and it yields a utility of c in proportion e of situations. Sin has an overall advantage just in case the expected benefit outweighs the cost.

The condition can also be stated as:

$$U_{\text{Sin}} > U_{\text{Sep}} \quad \text{iff } b/c > e/(1-e)$$

Depending on the value of the parameters, it will be easier to satisfy to the extent that the benefit b of Sep outweighs its cost, and to the extent that e is small, i.e. that the odds of there being ‘two events’ are small.

Evolutionary stability

Let us assume that the Single utterance interpretive strategy yields greater utility than the Separate utterances interpretive strategy: $U_{\text{Sin}} > U_{\text{Sep}}$. It is intuitively clear that Sep will be invaded by Sin, while Sin will be evolutionarily stable.

To make this intuition clear, we use the notion of evolutionary stability proposed by Maynard Smith and colleagues. Strategy I is evolutionarily stable just in case for all alternative strategies J, either (1) the payoff of I against I is greater than the payoff of J against I, or (2) I and J have equal payoffs when played against I, but I has a greater payoff than J when played against J (e.g. Maynard Smith & Price 1973, Skyrms 1996, 2010).

We assume that encounters are symmetric: when two individuals $\langle x, y \rangle$ interact, in half the cases x is the signaller and y the receiver, and in the other half it's the other way around. We further assume that whatever utility is obtained when a message is transmitted is shared; this makes good sense for mobbing alarms, as it is in everyone's interest to deter a predator. Thus when a bird b_1 using Sep interacts with a bird b_2 using Sin, in half the cases b_1 is the sender and b_2 is the receiver; this gives rise to a utility u_2 associated with the interpretive strategy Sin, and u_2 is shared among b_1 and b_2 : each gets $u_2/2$. In the other half of cases, b_2 is the sender and b_1 is the receiver, which gives rise to a utility u_1 associated with the Sep interpretive strategy, so each bird gets $u_1/2$. Overall, in such interactions, each bird gets $(u_1+u_2)/2$.

This leads us to the following payoffs for different types of encounters.

(i) Payoffs of Sin against Sin

$$\langle \text{Payoff of Sin, Payoff of Sin} \rangle = \langle U_{\text{Sin}}, U_{\text{Sin}} \rangle$$

(ii) Payoffs of Sin against Sep (and conversely)

$$\langle \text{Payoff of Sin, Payoff of Sep} \rangle = \langle \text{Utility of Sep, Utility of Sin} \rangle = \langle .5U_{\text{Sin}} + .5U_{\text{Sep}}, .5U_{\text{Sin}} + .5U_{\text{Sep}} \rangle$$

(iii) Payoffs of Sep against Sep

$$\langle \text{Payoff of Sep, Payoff of Sep} \rangle = \langle U_{\text{Sep}}, U_{\text{Sep}} \rangle$$

On the assumption that $U_{\text{Sin}} > U_{\text{Sep}}$, we also have: $U_{\text{Sin}} > .5U_{\text{Sin}} + .5U_{\text{Sep}} > U_{\text{Sep}}$.

When the resident population uses Sin, (i) and (ii) imply that Sin is evolutionarily stable relative to Sep, since $U_{\text{Sin}} > .5U_{\text{Sin}} + .5U_{\text{Sep}}$, satisfying Condition (1) of evolutionary stability.

When the resident population uses Sep, (ii) and (iii) imply that Sep is not evolutionarily stable relative to Sin, since $.5U_{\text{Sin}} + .5U_{\text{Sep}} > U_{\text{Sep}}$, violating Condition (1) of evolutionary stability.

• Production strategies

Once the resident population has an interpretive strategy that relies on S_{in} , it is plausible that the production strategies will adapt as a result. Let us assume that the resident population uses Sep^* in production, invariably producing ABC-D whether there are two separate events that license the two parts of ABC and D at the same time, or just one that licenses both components. We argue that a mutant strategy S_{in}^* that only uses ABC-D when a single event licenses both components will have an edge, at least in certain cases.

We adopt the simplifying assumption that events occur at discrete times, say $t = 0, 1, 2, \dots$, with enough time in the interval to distinguish between ABC-D without pause and ABC D with a brief pause; we'll assume that calls are produced, if at all, at times $t = .5, 1.5, 2.5$, etc (we do allow separate events to occur at the same discrete times, however).

The two production strategies are defined as follows.

Sep^* : At $t+.5$, produce ABC-D if at t *something licenses an alert and something licenses recruitment* (irrespective of whether these are separate events). (This strategy never produces ABC D with a brief pause.)

S_{in}^* : At $t+.5$, (i) produce ABC-D (without a pause) if at t *something licenses an alert and licenses recruitment*; (ii) produce ABC D (with a brief pause) if at t *something licenses an alert and something licenses recruitment <and it is not the case that it's the same thing that licenses an alert and recruitment>*.

The bracketed part is optional, and makes the production rule deterministic.¹⁸

It is clear that S_{in}^* will produce greater utility than Sep^* , as it makes it possible for recipients to distinguish between the case in which two noteworthy events justify the calls ABC and D, and the case in which one event satisfies both parts. By the same reasoning as was conducted above regarding the evolutionary stability of interpretive strategies, S_{in}^* will come to dominate in production.

C. Informativity Principle

The analysis based on Bird Implicatures posits the very same syntax as Minimal Compositionality. With respect to the semantics, it is non-committal, as it just requires the result that ABC-D is more informative than ABC, and/or more informative than D (as discussed in Section V.2, one condition is enough, hence our use of 'and/or').

Syntax (as above)

ABC, D are well-formed signals
 ABC-D is a well-formed signal
 If S is a well-formed signal $_S_$ is a well-formed utterance

Semantics

ABC-D is more informative than ABC, and/or ABC-D is more informative than D.

¹⁸ This rule incorporates something akin to the Informativity Principle, but as part of the meaning of a call. The trade-off between the Informativity Principle and the lexical specialization of calls is discussed in Steinert-Threlkeld, Schlenker & Chemla, 2021. (Note that with the bracketed part, S_{in}^* isn't exactly the production counterpart of the interpretive strategy S_{in} , since the latter does not interpret ABC D as excluding the possibility that one and the same event licenses both calls.)

The key is the Informativity Principle, which has two sides: a production side and a comprehension side.¹⁹

Informativity Principle

a. Production: If (i) a sequence S competes with a sequence S', (ii) S' is more informative than S, and (iii) both S and S' are true in the situation, S' rather than S must be produced in that situation, unless this yields a contradiction.

b. Comprehension: If (i) a sequence S competes with a sequence S', (ii) S' is more informative than S, and (iii) S is produced in that situation, if possible infer that S is true and S' is false, unless this yields a contradiction.

D. Evolutionary stability of the Informativity Principle

We study the evolutionary stability of the Informativity Principle in production strategies, assuming that comprehension strategies are unaffected. We study two strategies: no-IP (for 'no Informativity Principle') sometimes produces an under-informative call; IP (for 'Informativity Principle') always produces the maximally informative call. As before, we assume that when two individuals interact in this communicative fashion, utilities are shared.

On the production side, as far as we can tell, there is little or no trade-off involved (and the discussion is correspondingly simpler than in our discussion in B. above): by sometimes producing calls that fail to be maximally informative, no-IP should produce less shared utility than IP: $U_{IP} > U_{no-IP}$.

Using the same reasoning as in our discussion of Sep and Sin above, we obtain the following distribution of payoffs:

(i) Payoffs of IP against IP

$$\langle \text{Payoff of IP, Payoff of IP} \rangle = \langle U_{IP}, U_{IP} \rangle$$

(ii) Payoffs of IP against no-IP (and conversely)

$$\langle \text{Payoff of IP, Payoff of no-IP} \rangle = \langle \text{Payoff of no-IP, Payoff of IP} \rangle = \langle .5U_{IP} + .5U_{no-IP}, .5U_{IP} + .5U_{no-IP} \rangle$$

(iii) Payoffs of no-IP against no-IP

$$\langle \text{Payoff of no-IP, Payoff of no-IP} \rangle = \langle U_{no-IP}, U_{no-IP} \rangle$$

On the assumption that $U_{IP} > U_{no-IP}$, we also have: $U_{IP} > .5 U_{IP} + .5 U_{no-IP} > U_{no-IP}$

This is exactly the same situation we discussed with respect to Sin and Sep above, and the same result can be derived: IP is evolutionarily stable, no-IP isn't.

On the assumption that the Informativity Principle has emerged in production, it is clear that in a second step it should emerge in comprehension as well. Re-using our notations, but now on the comprehension side, the key assumption is that a bird that uses IP will obtain greater (shared) utility than one that uses no-IP. This makes sense because the no-IP bird will fail to get some of the information conveyed by the calls. For instance, for a no-IP receiver, the ABC alert call will be compatible with the presence of a predator that requires mobbing. For the IP receiver, by contrast, the ABC call will have a narrower meaning, excluding mobbing situations. This should make it possible to react more quickly or at less cost, yielding greater utility.

¹⁹ In the main text, we defined a 'caller-centred' and a 'receiver-centred' version of the Informativity Principle. They differed with respect to what counts as a "contradiction" in a. and in b. For simplicity, we disregard here the distinction between these two versions of the Informativity Principle. (For part a., pertaining to production, only the caller-centred version makes sense. For part b., pertaining to comprehension, either version could make sense: a contradiction might be what makes the *message* contradictory, or what makes the *receiver's beliefs* contradictory.)

The structure of the payoffs is thus identical to the one we discussed on the production side, and it will yield the same result: IP is evolutionarily stable, no-IP is not.