

# Ancestral Meanings: A Prelude to Evolutionary Animal Linguistics\*

Philippe Schlenker<sup>ab</sup>, Christina Pawlowitsch<sup>c</sup>, Luc H. Arnal<sup>d</sup>, Keny Chatain<sup>a</sup>, Lucie Ravaux<sup>a</sup>, Robin Ryder<sup>e</sup>, Ambre Salis<sup>a</sup>, Shane Steinert-Threlkeld<sup>f</sup>, Léo Wang<sup>a</sup>, Emmanuel Chemla<sup>g</sup>

**Abstract.** We ask how the very first meaning types could have arisen in animals, and argue that answers interact in interesting ways with existing data on current or ancestral animal communication systems. Using standard ideas of evolutionary stability in biology, we develop a simple framework (restricted to cooperative situations) to analyze the emergence of three meaning components: individual signals, non-trivial combinations, and pragmatic principles of competition among signals. We show that for elementary signals to arise, roughly, one of two things should be true: they had null cost, or they were understood from the start. While this largely dovetails with the traditional idea that signals often originate in cues, i.e. informative by-products of non-communicative processes, the two scenarios can be distinguished in case studies in which ancestral meanings were partly reconstructed. For non-trivial combinations, we start from two recent cases in which a sequence CC' was been argued not to be reducible to the separate utterance of C and then C'. One case (pyow-hack sequences in putty-nosed monkeys) was claimed to involve idioms, while the other (ABC-D sequences in Japanese tits) was argued to involve non-trivial compositionality. We show that the emergence of such non-trivial combinations is heavily constrained by the fact that they will initially give rise to miscommunication due to the possibility of treating C and C' as separate utterances; we take this to make certain idiom-based analyses implausible. Finally, for pragmatic principles, we study two principles that were proposed in recent animal linguistics, namely the Informativity Principle and the Urgency Principle. We argue that for both there is a clear evolutionary path, especially if they start appearing in production, and then in comprehension. Overall, we suggest that recent work in animal linguistics can be fruitfully combined with simple principles of evolutionary stability and, increasingly, with methods of ancestral signal reconstruction, to address in a precise fashion questions about the very first meaning operations that arose in nature.

---

\* **Acknowledgments:** We are very grateful to Jean-Baptiste André, Kirsty Graham, Catherine Hobaiter, Nicolas Mathevon, Luigi Rizzi, and Klaus Zuberbühler for very helpful discussions.

**Funding:**

Arnal: LHA is supported by the Fondation pour l'Audition (RD-2020-10) and Agence Nationale de la Recherche under the France 2030 program (ANR-23-IAHU-0003).

Schlenker, Chatain, Ravaux, 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).

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

<sup>a</sup> Institut Jean-Nicod (ENS - EHESS - CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France; PSL University.

<sup>b</sup> Department of Linguistics, New York University.

<sup>c</sup> Université Panthéon-Assas, Paris II, LEMMA-Laboratoire d'Economie Mathématique et de Microéconomie Appliquée.

<sup>d</sup> Université Paris Cité, Institut Pasteur, AP-HP, Inserm, Fondation Pour l'Audition, Institut de l'Audition, IHU reConnect, F-75012 Paris, France.

<sup>e</sup> Centre de Recherche en Mathématiques de la Décision (CEREMADE), CNRS, UMR 7534, Université Paris-Dauphine, PSL University, Paris, France.

<sup>f</sup> Department of Linguistics, University of Washington, Seattle, WA, USA.

<sup>g</sup> LSCP (ENS - EHESS - CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France; PSL University.

1	Three Problems about Meaning Emergence .....	3
1.1	Goals .....	3
1.2	Models and Conditions for Stability or Invasion .....	4
1.3	Structure .....	5
2	How Elementary Signals can Arise .....	5
2.1	Assumptions and notations .....	5
2.2	Payoffs .....	6
2.3	Evolutionary stability of Silent .....	6
2.4	Total invasion of Silent by Caller .....	7
2.5	Signals, cues and sensory manipulation .....	8
3	Emergence of Three Types of Elementary Signals .....	9
3.1	Whole signal meanings .....	10
3.2	Iconic meanings .....	13
3.3	Featural meanings .....	16
3.4	Stepping back .....	18
4	How Non-trivial Combinations can Arise .....	18
4.1	The existence of non-trivial combinations .....	19
4.2	Conditions for emergence of non-trivial combinations .....	19
5	Emergence of two combinations: Japanese tits and Putty-nosed monkeys .....	24
5.1	ABC-D sequences in the tit family .....	24
5.2	Pyow-hack sequences in Putty-nosed monkeys .....	26
6	How Competition Principles can Arise I: the Informativity Principle .....	27
6.1	Production first vs. Comprehension first .....	27
6.2	Emergence of the Informativity Principle in production .....	28
6.3	Emergence of the Informativity Principle in comprehension .....	30
6.4	Stepping back .....	31
6.5	A cautionary note .....	31
7	How Competition Principles can Arise II: the Urgency Principle .....	32
7.1	Emergence of the Urgency Principle in production .....	32
7.2	Emergence of the Urgency Principle in comprehension .....	34
7.3	Stepping back: evolution of pragmatic principles .....	36
8	Conclusion .....	37
	Appendix I. Evolutionary Stability and Invasion .....	38
	Appendix II. Comparison with Scott-Phillips et al. 2012 .....	39
	Appendix III. Other Evolutionary Paths for the Informativity Principle .....	41
	References .....	46

# 1 Three Problems about Meaning Emergence

## 1.1 Goals

How did the first meaningful signals arise? While this might seem like a hopelessly speculative question, we will argue that there are precise analytical and empirical tools that can make the problem tractable within evolutionary animal linguistics. We will discuss three main problems. The first pertains to the emergence of individual signals, the second to the emergence of non-trivial combinations of signals, and the third to that of pragmatic principles of competition among signals. Throughout, we restrict attention to cases in which communication is entirely cooperative (the sender's and the receiver's interests are aligned).<sup>1</sup>

In a nutshell, the problem of individual signals is this. Suppose there is a resident population that doesn't produce any signal. Suppose further that an extremely tiny population of mutants appears who do produce and understand a signal—say a general alarm call. Producing this signal has a tiny cost, for instance because it requires exertion, or makes the caller more conspicuous to predators. If a mutant produces the signal when encountering a mutant, both share the benefit. Alas, this case arises extremely rarely, and in most cases the mutant encounters a resident, who just fails to understand the signal. This means that the signal yields a negative overall utility. This is just a way of saying that, on this simple-minded story, the mutants will never invade the residents, and the signal will not stick. But there are two alternatives in which the signal does propagate. One is that the signal is in fact understood from the start. Another is that the signal has no cost. We will derive this result more formally, and explore its empirical consequences for the reconstruction of ancestral animal signals. While this initial study largely dovetails with traditional ideas, it serves to establish a simple framework in which our further questions can be addressed.

The problem of non-trivial signal combinations is slightly different. In a series of influential articles, it was argued that Japanese tits combine alert calls (ABC) with recruitment calls (D) to obtain an ABC-D call with a different meaning, one that triggers mobbing. It was further argued on the basis of sophisticated experiments that the ABC-D mobbing call is compositionally derived from {ABC, D}, and cannot just be treated as the conjunction of two separate utterances, ABC on the one hand, D on the other. In a different case, Arnold and Zuberbühler (2012) argued that Putty-nosed monkeys concatenate a small number of general calls (*pyow*) with a small number of eagle alert calls (*hack*) to obtain a call that triggers group movement. They took these "pyow-hack" sequences to be combinatorial but not compositional, and they analogized them to idioms because the meaning of the whole seems entirely different from the meaning of the parts. In both cases, the evolutionary question is how such non-trivial combinations (whether compositional or idiomatic) could emerge in the first place. The problem once again involves a resident population that does not have non-trivial combinations, and thus interprets the calls as separate utterances. A mutant may well produce such combinations with a new meaning, and another mutant will interpret them correctly. But in most cases, a mutant encounters a resident, and the resident will systematically misunderstand the combinations. We will argue that this puts severe constraints on the emergence of non-trivial combinations, and makes implausible certain (though not all) idiom-based analyses.

The problem of the emergence of pragmatic principles is probably easier. Recent literature on animal linguistics has argued that two competition principles could be analytically useful. One, the Informativity Principle, posits (in production and in comprehension) that when a call  $A^+$  is more informative than a call  $A^-$ ,  $A^+$  is used if both are true. The Urgency Principle posits that in a sequence involving calls B and  $B^+$ , if  $B^+$  provides information about the nature or location of a threat but B doesn't, then the order  $B^+B$  should be preferred to the order  $BB^+$ . We will argue that there is a clear evolutionary path for both principles, particularly if they arise in production first, and in comprehension second.

While the theoretical questions are rather open, we will argue that in several cases they can be made concrete when combined with the reconstruction of ancestral signals or signal combinations. In brief, animal signals are often preserved over millions of years, with clearly homologous forms in closely or less closely related species—to the point that, when DNA methods were less readily available than they are today, call similarity was sometimes used as a proxy for species relatedness (e.g., Gautier

---

<sup>1</sup> For a textbook survey of the non-cooperative case, see for instance Searcy and Nowicki 2007.

1988). The situation is very different today: DNA methods have established relatively clear phylogenies, and we can use them to reconstruct the evolutionary history of animal signals, sometimes to the point where we can ask where these signals originated from.

Thus we propose to extend recent work on animal linguistics with an evolutionary component combining explicit models of signal evolution with an empirical component based on current and ancestral animal communication systems.

## 1.2 Models and Conditions for Stability or Invasion

Our models involve populations with two types of individuals. With some regularity, pairs of individuals meet. These encounters are asymmetrical: one of the individuals, the signaler, notices an event, say a threat. This individual may produce a call (or more generally some signal, e.g. a gesture) that the other individual in the encounter, the receiver, may observe, and may use to react as well.

*Notation:* We write the utility of an individual using strategy I interacting with an individual using strategy J as  $U_{I,J}$ . Assuming that the asymmetric roles (observer of the event or not) are sampled homogeneously in encounters,  $U_{I,J}$  will be the mean of the utility obtained by I in encounters in which I is the signaler and J the receiver, and in encounters in which J is the signaler and I the receiver. Importantly, because signalers may incur a cost for producing signals, and because receivers may obtain a benefit from receiving a call that may not be shared with the signaler, it may be that  $U_{I,J} \neq U_{J,I}$ . But these represent benefits from the same encounters (both of the encounters (I, J) and (J, I)), just for different individuals (the I-individual or the J-individual, whichever appears first in the subscript).

Throughout, we use the notion of evolutionary stability proposed by Maynard Smith and colleagues (see for instance Maynard Smith and Price 1973, Taylor and Jonker (1978), Hofbauer, Sigmund, and Schuster (1979), Hofbauer and Sigmund (1988, 1998) and Weibull (1995), and Skyrms 1996, 2010). This notion is meant to capture the idea that an established population purely made of individuals with a strategy I will remain established even if a few individuals using an alternative strategy J arise. Formal conditions have been calculated for this to hold, as given in (1).

### (1) Evolutionary stability conditions

Strategy I is evolutionarily stable just in case for all alternative strategies J, one of the following conditions hold:

(Condition 1)  $U_{I,I} > U_{J,I}$ : the payoff of I against I is greater than the payoff of J against I.

(Condition 2)  $U_{I,I} = U_{J,I}$  and  $U_{I,J} > U_{J,J}$ : I and J have equal payoffs when played against I, but I has a greater payoff than J when played against J.

The intuition is that if a strategy I happens to be established in a population, then most encounters are with residents following strategy I, which is why the first condition is what primarily matters for fitness. Only in case the first condition does not give a fitness advantage to residents or mutants does the second condition kick in, as it involves these rare occasions on which encounters are with mutants.

In the cases we study below, we ask under what condition a resident population using a possibly archaic strategy I (e.g., one that uses no signals at all) is *not* stable relative to a mutant population with some innovative strategy J (e.g., one that produces and understands a certain signal). We will thus calculate conditions under which some mutants J could penetrate the resident population I. But we will also want to go one step further and ask under what conditions I is not just unstable (in the sense that it could allow for *some* mutants), but totally invaded (meaning that it gets entirely replaced by the mutant population). A necessary and sufficient condition is stated in (2) (see Appendix I).

### (2) Condition for total invasion<sup>2</sup>

Strategy I is totally invaded by strategy J if and only if the following three conditions hold:

<sup>2</sup> In terms of evolutionary game theory, Condition (2) is equivalent to the following other two conditions:

(i) strategy J is evolutionarily stable while I is not (in other words, in a world with these two strategies only, J is the unique evolutionarily stable strategy);

- (Condition 1)  $U_{J-I} \geq U_{I-I}$ : payoff of J against I  $\geq$  payoff of I against I  
 (Condition 2)  $U_{J-J} \geq U_{I-J}$ : payoff of J against J  $\geq$  payoff of I against J  
 (Condition 3) at least one the inequalities above is strict.

The intuition can be described as follows. Suppose that the population includes a proportion  $\pi$  of Mutants. For a large population, both Mutants and Residents interact with Mutants in a proportion  $\pi$  of encounters, and with Residents in a proportion  $(1-\pi)$  of encounters. The condition in (2) guarantees that, over both types of encounters, mutants will have a systematic fitness advantage over residents, which means that no matter what  $\pi$  is, their proportion will grow. So a non-null proportion  $\pi(0)$  at time 0, however small, will be followed by a higher proportion  $\pi(1)$  at time 1, and so on and so forth until the entire population of residents is invaded by the mutants. (See Appendix I to see why  $\pi$  will not only grow, but converge all the way up to 1).

### 1.3 Structure

The rest of this piece is organized as follows. We first discuss elementary signal emergence from a theoretical and then from an empirical perspective (Sections 2-3). We then turn to the emergence of non-trivial combinations, again from a theoretical and then from an empirical perspective (Section 4-5). At several junctures, pragmatic principles of competition (specifically, the Informativity Principle and the Urgency Principle) offer an alternative to semantics-heavy analyses, and we thus ask how these principles themselves could have arisen in the first place (Sections 6-7), before drawing some conclusions (Section 8).

## 2 How Elementary Signals can Arise

We start by addressing the problem of the emergence of elementary signals. Signals (may) have a cost, and we argue that whatever this cost is, it must be outweighed by whatever understanding they give rise to in a population that hasn't yet evolved to understand them. This gives rise to two salient possibilities: new signals, if they are to stay, are to some extent understood from the start, or they have non-positive cost.

### 2.1 Assumptions and notations

We assume a cooperative communicative situations in which the signaler and the receiver share the benefits of signal transmission, though not the cost of signal production (thus when a type is silent and the other isn't, payoffs won't be symmetric). We assume that a resident population has no signals, but that some mutants start producing and understanding a signal. We study the conditions under which the resident strategy, which is to not produce or understand any signal, is evolutionarily stable; and then we ask under what conditions the signal-using mutants can fully invade the resident population. In both cases, we rely on the definitions and results of (1)-(2) above.

Importantly, both the form and the meaning of the new signals is entirely specified in the mutants. Our framework is thus very different from standard signaling games à la Lewis (1969) (as well as Hurford (1989), and Nowak and Krakauer (1999), where the issue is to determine how signals can be paired with meanings/actions. To get something closer to our framework in a Lewisian signaling game, we would have to restrict the choices available to the signaler and receiver, as is done in Barrett and Skyrms 2017 with explicit reference to cues and sensory manipulation (two mechanisms we come back to in Section 2.5).

We will consider the emergence of a new signal A, and we will write as **T** the set of situations in which A is true, and as **F** the set of situations in which A is false. Correspondingly, the probability weights of these sets of situations will be written as  $p_T$  and  $p_F$  respectively.

---

(ii) strategy J at least weakly dominates strategy I (which implies in particular that the strategy profile (J, J) is a Nash equilibrium).

## 2.2 Payoffs

In our first model, we compare archaic Silent individuals, who cannot produce any call, and Caller individuals, who not only produce a call A in certain situations, but also understand that call A. We assume that throughout encounters, involving a Signaler and a Receiver, for any of the types Silent or Caller:

- A signaler producing the call A will incur a cost  $c$ .
- A receiver receiving the call A in an A situation will obtain a benefit, specifically: a benefit  $b_A$  if it perfectly understands that call (that is, if it is of the Caller type), and a benefit  $b_A \leq b_A$  if it partially understands that call (that is, if it is of the Silent type). This lower benefit may, among others, arise because the receiver only derives a rough approximation of the correct meaning. For instance, a receiver that hasn't evolved to understand a general alarm call may still find it salient enough to start looking around its environment.
- The signaler will obtain a portion  $\alpha$  of the benefit obtained by the receiver. For  $\alpha = 0$ , the signaler gets no benefit. For  $\alpha = 1$ , all the benefit is fully shared (an illustrative situation for this is when the call indicates the presence of a predator that could be mobbed if the signaler and the receiver coordinate for their joint benefit).

In the table in (3), we summarize, for each encounter, the payoff obtained by each of the players in such a game.

(3) **Summary of the payoffs for each of the different encounters**

When the payoffs are not the same for signaler and receiver, we first provide the payoff of the signaler, and then the payoff of the receiver after a comma.

Signaler	Silent	Caller	Silent	Caller
Receiver	Silent	Caller	Caller	Silent
<b>Call A</b>				
<b>F</b> false	0	0	0	0
<b>T</b> true	0	$\alpha.b_A - c, b_A$	0	$\alpha.b_A - c, b_A$
<b>Expected payoff</b>	$U_{\text{Silent-Silent}} = 0$	$U_{\text{Caller-Caller}} = p_T[(1+\alpha)b_A] - c/2$	$U_{\text{Caller-Silent}} = p_T[\alpha.b_A - c] / 2$ $U_{\text{Silent-Caller}} = p_T b_A / 2$	

Let us explain how this table is obtained.

— When the call A is false (top row with numbers), a situation we note as **F**, no call is produced, no cost is incurred and no benefit ensues in any of the interactions, for either player.

— When the call A is true (next row with numbers), a situation we note as **T**, things are more interesting. If the Signaler is Silent, then no call will be produced and, again, there will be no cost or benefit for any player. If the Signaler is Caller, then the Signaler will produce a call and immediately incur a cost  $c$ . If the Receiver is Caller too, then the receiver will obtain a benefit  $b_A$  for full understanding of that call, and the Signaller will obtain a proportion  $\alpha.b_A$  of it; if the receiver is Silent, it will obtain a benefit  $b_A$  for partially understanding the call A, and the signaler will obtain a portion  $\alpha.b_A$  of this benefit.

We can then sum up the expected benefit obtained overall from such interactions, by each individual; this appears in the row Expected payoff. In Caller-Caller interactions, an individual will have the role of Signaler half the time, and the role of Receiver half the time as well. And the expected benefit obtained is weighted by the probability of each situation occurring. In this way, a final benefit for the Caller, in a Caller-Caller encounter, is  $p_T[\alpha.b_A - c]/2 + p_T[b_A]/2 = p_T[(1+\alpha)b_A - c]/2$ . In mixed encounters between a Caller and a Silent individual, half the time the Caller will be the signaler, and since this is the only source of benefit for the Caller, it gets a weighted benefit of  $p_T[\alpha.b_A - c]/2$ . Silent obtains an expected benefit of  $p_T[b_A]/2$  from such mixed encounters.

## 2.3 Evolutionary stability of Silent

We can ask under what conditions Silent is stable. Using (1) we obtain the condition:

- (4) Silent is stable against Caller iff  $c > \alpha \cdot b_A$ .

If we write  $b^- = \alpha \cdot b_A$  for the payoff obtained by a caller interacting with a receiver that hasn't evolved to understand the message, the condition is just:  $c > b^-$ .

*Argument:* Here we assume that  $p_T$  is not null. The conditions of stability are, in general:

(Condition 1)  $U_{\text{Silent-Silent}} > U_{\text{Caller-Silent}}$ , or

(Condition 2)  $U_{\text{Silent-Silent}} = U_{\text{Caller-Silent}}$  and  $U_{\text{Silent-Caller}} > U_{\text{Caller-Caller}}$ .

According to the table in (3), the conditions are equivalent to (Condition 1)  $0 > [\alpha \cdot b_A - c] / 2$ , i.e.  $c > \alpha \cdot b_A$ , or (Condition 2)  $\boxed{\alpha \cdot b_A = c}$ , and  $b_A / 2 > [(1+\alpha)b_A - c] / 2$ . But condition 2 can never hold, since the second (boldfaced) part reduces to the following being negative (adding underlined terms that sum to 0):  $(1+\alpha)b_A - c - b_A = (1+\alpha)b_A - b_A - \underline{\alpha b_A} + \underline{\alpha b_A} - c = (1+\alpha)(b_A - b_A) + (\alpha b_A - c)$ . Because of the first (boxed) part of Condition 2, this is equal to  $(1+\alpha)(b_A - b_A)$ , which can at worst be null, but never negative.

Going back to the Stability Condition  $c > \alpha \cdot b_A$ , the key message is that Silent will be unstable just in case the signaler's benefit  $\alpha \cdot b_A$  (obtained as a share of the partial understanding of its call by Silent) is superior or equal to the cost  $c$  of sending a signal ( $\alpha \cdot b_A \geq c$ ). In the next section, we will in fact show that these conditions are not only those that make Silent unstable: they are also the conditions that make Silent totally invaded.

#### 2.4 Total invasion of Silent by Caller

Taking one further step, we can ask under what conditions Silent is totally invaded. Using (2), we obtain the conditions in (5).

- (5) Silent is totally invaded by Caller  
iff  
 $\alpha \cdot b_A \geq c$ , or with the notation  $b^- = \alpha \cdot b_A$ ,  $b^- \geq c$ .

*Argument (similar to the argument of Section 2.3 regarding the stability Silent):* The conditions for total invasion are in general:

(Condition (i))  $U_{\text{Caller-Silent}} \geq U_{\text{Silent-Silent}}$

(Condition (ii))  $U_{\text{Caller-Caller}} \geq U_{\text{Silent-Caller}}$

(Condition (iii)) one of these inequalities is strict

According to the table in (3) (disregarding the probability  $p_T$ , which is assumed to be positive), the conditions are equivalent to the following:

(Condition (i))  $[\alpha \cdot b_A - c] / 2 \geq 0$ , i.e.  $\alpha \cdot b_A - c \geq 0$

(Condition (ii))  $[(1+\alpha)b_A - c] / 2 \geq b_A / 2$ .

So, Condition (ii) is equivalent to the following being positive:  $(1+\alpha)b_A - c - b_A = (b_A - b_A) + (\alpha \cdot b_A - c)$ .

The first term is strictly positive by hypothesis ( $b_A > b_A$ ) and the second term is strictly larger than:  $\alpha \cdot b_A - c$ . As a result, when Condition (i) holds, Condition (ii) also holds and, in fact, the inequality is strict, so

Condition (iii) also holds. In sum, the three conditions together are equivalent to Condition (i):  $\alpha \cdot b_A \geq c$ .

We thus observe that the conditions for instability of Silent (see Section 2.3) are the same as the conditions for its total invasion by Caller: Silent will be totally invaded just in case the signaler's benefit  $\alpha \cdot b_A$  (obtained as a share of the partial understanding of its call by Silent) is superior or equal to the cost  $c$  of sending a signal ( $\alpha \cdot b_A \geq c$ ).

We highlight some salient subcases.

- If the signal has a negative cost (i.e. the signal/behavior is beneficial to the caller independently of communication), then Silent will be invaded.
- If the benefit of communication for Silent is null ( $b_A = 0$ ), or not passed to the Signaler ( $\alpha = 0$ ), then Silent will be invaded just in case the signal has a non-positive cost ( $c \leq 0$ ).

## 2.5 Signals, cues and sensory manipulation

The question of the very first emergence of signals is an old one. In the words of Maynard Smith and Harper (2003), "a major problem in the evolution of signals is that, even if a model predicts a stable signalling equilibrium, a 'no-signalling' equilibrium is usually a stable alternative: that is, if no one signals, do not evolve the capacity to respond, and if no one responds, do not bother to signal". This is the problem that our condition  $b^- \geq c$  is meant to address (as it yields both lack of stability of Silent and total invasion of Silent by Caller). But a traditional response is that signals arise from cues, with a transition often called 'ritualization', assuming the definitions in (6), from Maynard Smith and Harper (2003, p. 15).

- (6) a. Signal: an act or structure that alters the behaviour of another organism, which evolved because of that effect, and which is effective because the receiver's response has also evolved.
- b. Cue (Hasson 1994): a feature of the world, animate or inanimate, that can be used by an animal as a guide to future action.
- c. Ritualization (Tinbergen 1952): evolutionary process whereby a cue may be converted into a signal.

Interestingly, the idea that signals originate in cues is related to, but not identical with, the condition  $b^- \geq c$  (the signaler's benefit when interacting with a receiver that hasn't evolved to understand the signal is at least as great as the production cost  $c$ ). This could be either because (i)  $b^-$  is comparatively large, or (ii)  $c$  is comparatively small. Let us consider these two cases in turn, assuming for greater simplicity that in  $b^- = \alpha \cdot b_A$ ,  $\alpha = 1$  (the signaler gets the same benefit as the receiver in such cases).

(i): First, having a signal originate in a cue might be a way to guarantee that a receiver that hasn't evolved to understand the signal still reacts to it appropriately. In essence, understanding of how the world works could ensure that  $b^-$  is 'high enough' even if there isn't full comprehension yet. So it is safe to assume that if a signal originates as a cue,  $b^-$  will be non-null. But the converse need not be true: there could be ways in which a signal has an effect on the receiver without necessarily being a cue of anything; (this happens with sensory manipulation, a case we return to below).

(ii) Second, having a signal originate in a cue might sometimes guarantee that the primitive form of the signal arose for non-communicative reasons, which in turn might make the cost  $c$  of signal production null or negative. A simple example is given by Maynard Smith and Harper (2003, p. 69): "The increased activity likely in many social interactions may prompt an increased respiration rate. Rapid breathing might be loud enough to provide a cue that can be ritualized into a call." Still, the relation between cues and signals with small or null cost is not at all direct. In numerous cases, such as woodpecker drumming and ritualized ape gestures, discussed below in Section 3.1, a signal developing through ritualization may lose any physical efficacy (thus woodpecker drumming does not serve to forage, contrary to pecking, N. Mathevon, p.c.); and ritualized gestures are not physically efficacious, contrary to manual actions). So there is no implication from 'signal  $x$  originates in a cue' to 'signal  $x$  currently has a small or null cost' (on the other hand, it could be in early stages of evolution the signal-to-be was efficacious, and thus had a null net cost).<sup>3</sup>

In sum, the condition for total invasion  $b^- \geq c$  is related to, but distinct from, the idea that signals originate in cues. Finally, we note that when  $c > 0$ , the communicative behavior emerges *despite* its cost, and thus it qualifies as a signal in the technical sense of (6)a.

Several authors have emphasized that signals do not always arise from cues, but sometimes from sensory manipulation (e.g. Krebs and Dawkins 1984, Scott-Phillips et al. 2012, Barrett and Skyrms 2017). In a cue, a state of the world triggers an action by the signaler. In sensory manipulation, an action by the signaler triggers a reaction on the receiver's part. Scott-Phillips et al. (2012) provide the examples of mating displays that build on pre-existing preferences for a certain color in receivers. We will discuss in Section 3 the example of nonlinearities, which increase the effectiveness of some alert calls because they make them harder to habituate to. From the perspective of our model, the key in such

---

<sup>3</sup> Still, one might ask whether 'signal  $x$  has null cost' implies that 'signal  $x$  originates in a cue'. The reasoning might be that if signal  $x$  has null cost, its appearance must be motivated by something other than communication, and it might thus serve as a cue. But this point would require further research.



cases is that the manipulatory nature of the signal guarantees that the pre-communication benefit  $b^-$  is significant, and may thus outweigh any non-null cost  $c$  associated with signal production (with the result that  $b^- \geq c$ , guaranteeing total invasion by mutants that produce and fully understand the signal).

### 3 Emergence of Three Types of Elementary Signals

To illustrate our results, we apply them to three types of signals that have been described in animal communication. One includes elementary signals (notably, calls and gestures) whose atomic meaning involves the whole signal, and cannot be derived from deeper considerations (in linguistic parlance, they have 'lexical' meanings). A second type involves signals that have an iconic meaning, in the sense that there is a structure-preserving map between the form of the signal and its meaning. A third type involves signals whose meaning is derivable from the presence of certain distinguished acoustic features, such as roughness or nonlinearities; in such cases, the semantic atoms are the acoustic features, not the whole call they are a component of. Although these examples serve to illustrate our findings, the latter pertain to a narrower case, involving the very first signals that appeared; but establishing such an anteriority over all other signals is beyond current knowledge, and we allow ourselves a departure from our narrow hypotheses.

Throughout, we focus on the conditions for total invasion of Silent by Caller in (5). We will consider three case studies of whole meaning emergence, two of iconic meaning emergence, and two of featural meaning emergence, as summarized in (7). When available, we provide estimates of the appearance of the relevant signals to highlight the extraordinary time depth afforded by recent studies. In all but one case, we will speculate that the condition  $b^- \geq c$  (the benefit of communication for the mutant Caller is greater than or equal to the cost of signal production) is satisfied because, despite a non-null cost, the benefit of pre-comprehension signaling is significant. But there is a possible exception: as we will explain, several researchers have argued that the waggle dance of bees arose for non-communicative reasons; if so, it could be that  $c$  is null or negative in this case (which doesn't preclude the possibility that  $b^-$  is significant as well).

(7) **Sample examples of elementary signal emergence**

(mya = million years ago;  $b^-$  is as defined in (5) (= part of the benefit that goes to the signaler, when the receiver partially understands the call))

Meaning type	Case study	Time of appearance	How is $b^- \geq c$ plausibly satisfied?	Why?
<b>Whole call meanings</b>	Woodpecker drumming	22.5 mya	$b^-$ is high	Similarity to foraging makes it easy for drumming to advertise species identity and mate quality.
	Ritualized ape gestures: present body part (grooming) present (climb on)	$\pm 17$ mya	$b^-$ is high	Similarity to actions makes it easy to understand the signal.
	Cercopithecine booms	$\geq 2.5$ -5 (depending on the subfamily)	$b^-$ is high?	Unclear. Possibly, the use of air sacs makes it easy to recognize male presence.
<b>Iconic meanings</b>	Female bonobo foot pointing and pantomime	Unclear	$b^-$ is high	Salience and similarity might make the signal easy to understand.
	Waggle dance of bees	$\geq 20$ mya	<b><math>c</math> is null or negative</b>	The waggle dance plausibly arose independently of communication.
<b>Featural meanings</b>	Non-linearities	Unclear, probably very ancient due to presence	$b^-$ is high	Non-linearities are attention-grabbing because they block habituation.

		in phylogenetically diverse species		
	Roughness	Unclear, probably very ancient due to presence in phylogenetically diverse species	$b^-$ is high	Roughness is salient and aversive because of high speech repetitions below the threshold for a continuous percept.

### 3.1 Whole signal meanings

#### 3.1.1 Woodpecker drumming

In the words of Garcia et al. 2020, "drumming is an exaptation, which derived through ritualization (exaggerated amplitude, rhythm stereotypy) from pecking on tree trunks, a foraging behaviour typical of woodpeckers (...). A phylogenetic reconstruction of drumming suggests that this was the ancestral behaviour in this family (95% probability of being present in the common ancestor, 22.5 million years ago (...)." Garcia et al. take its function to be "to communicate species identity in territorial and mating contexts" (see also Wilkins and Ritchison 1999).

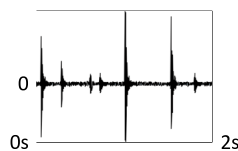
Starting from an even older ancestral state without drumming, the question is how this signaling strategy could have come to totally invade one without it, and in particular how the benefit  $b^-$  obtained without communication could be greater than or equal to the cost  $c$  to the signaler.

The first idea that comes to mind is that the cost to the signaler is null because drumming is a version of pecking. But drumming is not used to forage, which means that its only value lies in communication (N. Mathevon, p.c.).

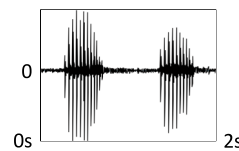
It is thus essential that drumming gave rise to a significant communicative benefit even when the receiver had not yet evolved to understand the signal. The fact that drumming is a ritualized version of pecking probably guarantees this: the link to woodpecker foraging would ensure that species identity is readily understood; similarly, mate attraction might stem from the fact that a woodpecker that can peck with energy is a good potential mate. This is all in line with the view that 'signals originate in cues'.

#### (8) Comparison between pecking (top) and drumming (bottom)

a. Pecking (foraging)



b. Drumming (communication)



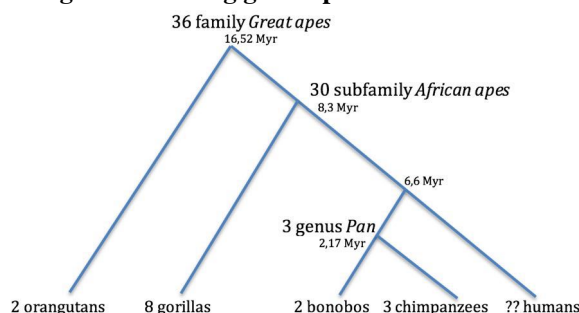
"Pecking, a foraging behaviour used in woodpeckers to excavate prey from tree trunk, has acquired a stereotypical and rhythmic structure, thus exapting into drumming, which holds a communicative function during mating and territorial defence." (Garcia et al. 2020, Fig. 1a)

Our discussion is probably overly simplified, however. There could well have been multiple evolutionary steps in which pecking became ritualized while preserving a foraging benefit. This would have initially kept the production cost  $c$  very low, making  $b^- \geq c$  easier to achieve. Receivers might have adapted to this ritualized pecking, which in turn might have made communication more effective. We leave such refinements for future research.

#### 3.1.2 Ritualized ape gestures

Ape researchers have uncovered a striking repertoire of communicative gestures that are largely shared among great ape species, and are thus likely to be very ancient—in some cases, 16 million years old, as illustrated in (9) (Byrne et al. 2017).

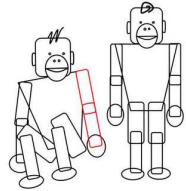
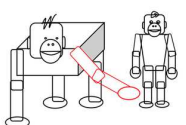
(9) **Shared gestures among great apes**



The distribution of gestures across living great ape species and genera, based on current knowledge: numbers of gestures specific to each species are shown, revealing extensive overlap. Notably, 36 gestures are shared among all apes, 30 additional ones are shared by African apes, and 3 additional ones are shared by bonobos and chimpanzees. (Byrne et al. 2017; figure redrawn by L. Ravaux)

The functions of such gestures in chimpanzees and bonobos were compared in great detail in Graham et al. 2018. Some but not all can be explained as instances of ritualization, whereby in ontogenetic or phylogenetic development, an action becomes truncated so that it stops being physically efficacious but can still be used as a cue about the producer's intention. Two striking examples appear in (10); one is an invitation to initiate grooming, while the other invites the receiver to climb on the signaler.<sup>4</sup> Both gestures are analyzed as varieties of a more general gesture, "present body part", in Byrne et al. 2017 ("body part is moved to deliberately expose an area to recipient's attention"); and this more general gesture is found in orangutans in Knox et al. 2019, suggesting that it is ape-specific and approximately 17 million years old.<sup>5</sup> (In line with the authors' methodology, the functions of the gestures are determined by 'apparently satisfactory outcomes', reactions that make the signaler stop gesturing.)

(10) **Potential examples of ritualization (in ontogeny or phylogeny)**

Gesture Type	Bonobo ASOs	Chimpanzee ASOs
<b>Present (grooming)</b> 	<b>Initiate grooming</b> 100%  <i>Tight</i>  [47(1058): $f=7835$ , $df=12,552$ $p<0.0001$ ]	<b>Initiate grooming</b> 100% <sup>1</sup>  <i>Tight</i>  [f=2384.95, $df=14,364$ $p<0.001$ ]
<b>Present (climb on)</b> 	<b>Climb on me</b> 100%  <i>Tight</i>  [7(34): $f=4720$ , $df=12,72$ $p<0.0001$ ]	<b>Climb on me</b> 100%  <i>Tight</i>  [f=1820.37, $df=14,28$ $p<0.001$ ]

Figures from Graham et al. 2018. For video examples, see: <https://greatapedictionary.ac.uk/gesture-videos2/>

<sup>4</sup> In Hobaiter and Byrne 2011 (pertaining to the chimpanzee repertoire), these gesture forms are defined as follows:

*Present (grooming)*: "Body is moved to deliberately expose an area to the recipients attention which is immediately followed by grooming of the area."

*Present (climb on)*: "Arm or leg is extended to young recipient in order to facilitate them climbing onto the signallers body (normally mother to infant)."

<sup>5</sup> Special thanks to C. Hobaiter for help with the above discussion.

Interestingly, there is disagreement among researchers about the status of ritualized gestures. For Tomasello and colleagues, ritualization involves learning at the individual level, along the following lines:

An individual actually performs some social act toward a recipient, and over repeated instances the recipient starts anticipating what the actor intends (or will do) based on some initial part of the act; the actor notes that the recipient anticipates his intention (or what he will do) on the basis of this initial sequence, and infers a causal link in the sense that he understands that it is this initial part of the act that instigates the reaction. The initial part of the act becomes “emancipated” from the physically efficacious dimensions of the original social act; it becomes ritualized. (Tomasello and Call 2019)

Other researchers, notably Hobaiter and Byrne 2011, take instead these gestures to be innately specified for an entire species and, sometimes, to be very ancient, as we saw above.

If individual learning is all there is, an evolutionary story is not needed. But from the perspective of Hobaiter and Byrne 2011, we must ask how such innate signals might have come to evolve in the first place. And here Tomasello's ideas might prove essential. What matters for our purposes is that, by definition, communicative gestures are not physically efficacious.<sup>6</sup> It is also clear that they come with a cost, although it might be a small one; so in this case,  $c > 0$ . For these gestures to emerge, they should come with a benefit even when the receiver has not evolved to understand the signal yet. The type of reasoning outlined above by Tomasello and Call 2019 might be just enough to guarantee that the signal has an effect on a receiver that does not yet fully understand it. Thus even if one sides with Hobaiter and Byrne and takes ape gestures (including ritualized ones) to be innate, one must explain how these signals could have evolved in the first place, and here Tomasello and Call's idea offer a plausible analysis. It is clear that this case too is in line with the view that 'signals originate in cues'.

As in the case of woodpecker drumming, future research should investigate a multi-step evolutionary process in which physically efficacious actions are ritualized before becoming pure communicative signals.

### 3.1.3 *Cercopithecine booms*

A strikingly costly signal is the boom call produced by numerous primate species using laryngeal air sacs. Among cercopithecines, booms are usually thought to be non-predation calls, and possible functions involve group cohesion; intergroup spacing; and mate attraction (see for instance Fuller 2013 and Gautier-Hion et al. 1999). Relatedly, in a particularly detailed study of booms in Blue monkeys, Fuller and Cords 2019 conclude that their primary function is to facilitate interactions between adult males and group members.

On the production side, booms require highly developed air sacs. Booms are produced at a very low pitch (120–140 Hz) and have low attenuation rates, as emphasized by Waser and Waser (1977). It is clear that the effort and time it takes to fill the air sacs make boom very costly (see Gautier 1971 and Gautier-Hion et al. 1999 for further production details).

Both air sacs and booms are rather ancient. Hewitt et al. argue that the most recent common ancestor of the entire family of cercopithecines had air sacs, maybe around 7.5-10 million years ago. Schlenker et al. 2016b note that it is unclear whether the common ancestor of all cercopithecines had booms, but that in two cases, an entire subfamily of cercopithecines has booms, which suggests that their most recent common ancestor did too. Depending on the subfamily, this makes cercopithecine booms at least 2.5-5 million years old.

There are two natural evolutionary questions to ask. First, why did air sacs evolve? Second, how did booms evolve? We have nothing to contribute to the first question, and we can only cite one commonly accepted view. Hewitt et al. 2002 mention that "the best supported suggestion is that the sacs of Old World monkeys help to amplify vocalisations, possibly by acting as resonance chambers (...). For example, the intensity of the boom vocalisation of *Cercopithecus neglectus* was reduced when

---

<sup>6</sup> A standard definition in ape research is given in Hobaiter and Byrne 2011: "Gestures were defined as discrete, mechanically ineffective physical movements of the body observed during periods of intentional communication."

the air sac was punctured experimentally and male cercopithecines, which give loud calls, have larger sacs than the quieter females (Gautier 1971)".<sup>7</sup> For his part, Gautier 1971 noted (p. 262) that by their mere presence, air sacs "modify the laryngeal tract and therefore the nature of the vocalizations of the entire repertoire".

Assuming that air sacs were present first, the next question is how booms evolved. This is a very clear case in which the cost  $c$  of signal production is strictly positive, and a necessary condition for the signal to evolve is that the pre-comprehension benefit  $b$  is greater than or equal to  $c$ . On the assumption that air sacs already affected the quality of other vocalizations, it might be that booms could initially function as a cue to the presence of a male. There might also be an effect of a well-known biological code by which lower-frequency sounds are associated with larger bodies (for related discussions, see Fitch and Hauser 1995 and Hewitt et al. 2002). Be that as it may, the conditions of emergence of boom should be investigated in greater detail.

### 3.2 *Iconic meanings*

By definition, an expression is iconic if it conveys information about a situation by resembling certain aspects of that situation, or more abstractly: if there is a structuring-preserving map from the form of the expression to situations it is true of. Iconicity has some 'bite' only to the extent that the form can be modified (often gradiently) and yield corresponding changes to the meaning. Proving this is a tall order.

Pantomime and pointing have been argued to exist in female bonobos; these seem to raise the same kind of issues as ritualized gestures, and to our knowledge strict iconicity (with gradient changes in form entailing gradient changes in meaning) has not been proven. More famously perhaps, bees use a waggle dance that iconically represents the direction and distance of the food source, and here the arguments for gradient dependency of meaning on form are overwhelming. As we will see, the literature suggests that this might conceivably be a case in which the cost of signal production was initially null or negative—a different road to emergence than the other ones we have seen so far.

#### 3.2.1 *Pointing and pantomime in female bonobos?*

As Douglas and Moscovice 2015 argue, bonobo females that wish to initiate mutual genital rubbing with another female may use two signals that are potentially iconic. One involves foot pointing (with the heel or toe) towards the initiator's sexual swelling. The other involves a pantomime: by way of hip movement, the initiator will represent the action that takes place during genital rubbing. In order to show that these signals are genuinely iconic, one would need to show that their form can be gradiently changed to modify their meaning: some modulations of the pantomime or of the foot pointing should come with small meaning differences. Showing this might be difficult. Thus we cannot exclude the possibility that these are whole signal meanings, which would then raise exactly the same issue as ritualized gestures discussed in Section 3.1.2: their resemblance to the desired action might be enough to explain why something is communicated even to a receiver that doesn't fully understand the signal. If they are not part of a whole signal repertoire, the very first step, namely ensuring a benefit of pre-comprehension understanding, might be the same, but one would still need to explain how a fully iconic system ended emerging—something that lies beyond the present study.

#### 3.2.2 *The waggle dance of bees*

The most striking example of iconic communication in the animal world involves the waggle dance of bees.<sup>8</sup> Following a long line of research (including von Frisch's Nobel Prize-winning work, see von Frisch 1967), it was found that "the dance, performed by a honey bee upon returning to the colony

---

<sup>7</sup> Hewitt et al. 2002 develop an alternative hypothesis: "air sacs may have evolved in some primate lineages to enable fast call sequences in larger-bodied species".

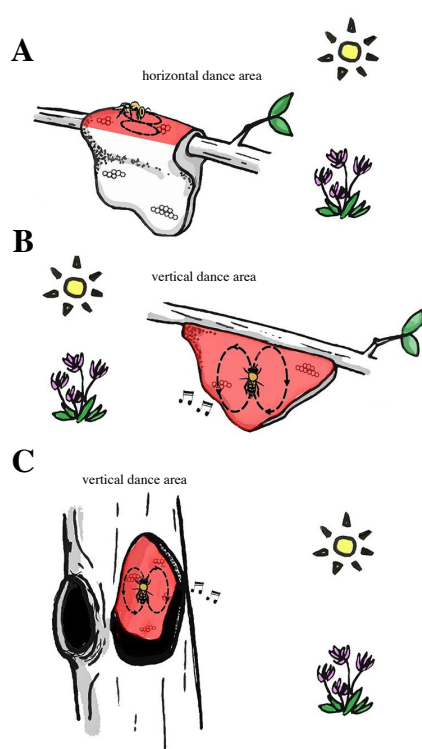
<sup>8</sup> Much research (e.g. von Frisch 1967) distinguished between the round dance, which indicates that a food source has been found near the hive, and the waggle dance, used for more distant food sources. But as Gardner et al. 2008 note, "the round and waggle dances are merely two ends of a continuum; 'adjustable waggle dance' is sufficient. Distance and direction are encoded using the same mechanisms in both dances, and the predictability of dance behaviour gradually increases as the food-source distance increases."

having successfully located a food source, offers information on the presence, odor, quality, direction, and distance of said food source, enabling nest-mates to exploit it (...). The direction information is conveyed through the orientation of the waggle run, whereas the distance information is expressed through the duration of the waggle run" (I'Anson Price and Grüter 2015).

The waggle dance is probably at least 20 million years old (I'Anson Price and Grüter 2015). Depending on the bee species, the dance orientation that determines the direction of the food source is interpreted in one of two ways. In one subfamily of bees, the dance is performed horizontally and points towards the food source. In another subfamily, the dance is performed vertically, and the angle of the dance relative to 'upwards' is interpreted as the angle of the food direction relative to the sun. Thus if the dance is upwards and vertical, the food is in the direction of the sun.

To illustrate, consider in (11) three dance types summarized by I'Anson Price and Grüter 2015. In (11)A, the dance is horizontal, and is oriented towards the food source (the flowers). In (11)B, the dance is vertical and in the open, in an upwards direction. This indicates that the food source is towards the sun from the bee's position. The dance in (11)C is similar to that in (11)B, with the difference the dance is performed by cavity-nesting bees, and thus in darkness. Here too, the dance is oriented upwards, indicating that the food source is oriented toward the sun.

(11) **Three types of waggle dance** (modified from I'Anson Price and Grüter 2015)



(A) Dwarf honey bees perform dances on horizontal surface in the direction of the food source. (B) Giant honey bees perform dances on vertical surface and orient dances with gravity and celestial cues. (C) Cavity-nesting honey bees perform dances in darkness on a vertical surface. Dances are oriented with gravity and celestial cues. All species produce auditory cue when dancing. Bees in (B,C) are both dancing directly upwards, indicating that the food source is in the direction of the sun. Areas of the comb on which bees dance are indicated by an arrow (and appear in red in the online version) (after I'Anson Price and Grüter 2015)

We may once again ask how the waggle dance originated. Since it is a complex signal, we may in principle ask the question about the directional component alone or about the distance component alone, unless one has reason to assume a specific evolutionary order among these two properties. Either way, the key question will be to explain how the benefit of communication when the receiver hasn't evolved to understand the dance outweighs the cost of producing the signal.

But here a remarkable suggestion was made by several researchers, in particular von Frisch 1967 (e.g. pp. 327-328), Lindauer 1971 (p. 85), and Menzel 2019: a dance behavior may exist in species without the corresponding signal. In Menzel's words (p. 2), "the most basic component is thought to be food-induced arousal, seen, for example, in the circling "dances" of the fly *Phormia regina* (...). Hungry flies run around a droplet of sugar water and continue performing such "dances" even when the sugar droplet is removed. When the surface is tilted from horizontal to vertical, they transfer the elongated runs they do, relative to light on the horizontal surface, to the upward direction on the vertical surface. When other flies are close, the "dancer" may regurgitate some of the sugar syrup, leading to similar circling movements in the other flies." To put it differently, independently of any signaling behavior, there seems to be a common transfer from 'orientation towards the sun' to 'upwards orientation towards the vertical'.

Menzel 2019 goes one step further and suggests that the waggle dance could be viewed as an "intended flight": "the waggle run mirrors its next outbound flight, a performance that will follow when she leaves the hive".

Taken together, these observations suggest that a behavior akin to the waggle dance might have existed before any signaling behavior (why it existed is an important but separate question). Furthermore, the transposition from 'towards the light' to 'upwards' might also have existed before signaling. If so, when primitive versions of the waggle dance arose, signal production might have had a null or negative cost, in the sense that it was for non-communicative reasons—maybe the simulation of a bee's next flight, as suggested by Menzel 2019—that the dance arose. This, in turn, guarantees that our condition  $b \geq c$  for emergence of signaling in (5) would have been automatically satisfied.<sup>9</sup> This means that mutant bees that understood the waggle dance would have invaded a non-communicative population. Of course this begs the question of precisely how the iconic interpretive rule could have developed in comprehension. Here too, a full account of the emergence of iconic systems goes beyond the present study.

One last point should be mentioned. Several researchers, including Lindauer 1956, von Frisch 1967 (e.g. p. 304) and Menzel 2019, have expressed a strong suspicion that the horizontal dance, which points directly towards the food source, is more primitive than the vertical dance, which requires a transposition of the angle of the food source relative to the sun to the angle of the dance relative to 'upwards'. This makes much intuitive sense, but bee phylogenies don't clearly bear this out. Two appear in (12). The phylogeny from Engel 2006 is somewhat neutral, as it exhibits two independent monophyletic clades as structural sisters, one for the 'vertical' family and one for the 'horizontal' family.<sup>10</sup> The phylogeny from Hedtke et al. 2013 suggests that the vertical dance is more primitive, because the monophyletic clade with a horizontal dance is embedded in the clade with a vertical dance. Refining the the phylogenetic analyses might help deliver clearer conclusions.

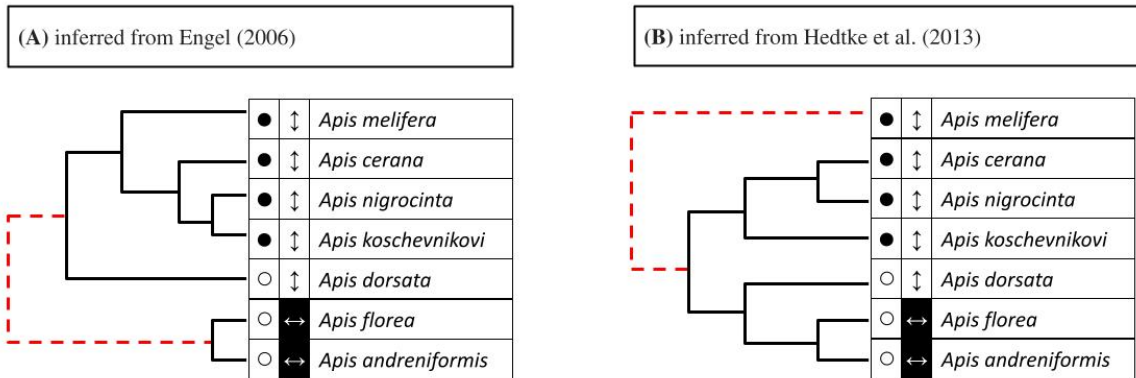
---

<sup>9</sup> In fact, the inequality might well be strict. Menzel's (2019) simulation view once again offers an elegant analysis. For him, "the recruit would learn a simulated outbound flight by repeating the movements of the dancer, store it as an intended flight, and then apply it". For a recruit that hasn't mastered the signal, the information might be very imperfect, but any non-null benefit will guarantee that  $b > c = 0$ .

<sup>10</sup> If anything, the phylogeny from Engel 2006 might weakly suggest that the vertical dance is more primitive. This is because in this phylogeny the vertical dance must be ancient, whereas the horizontal dance might be ancient or more recent (as it could have emerged anywhere on the bottom-most horizontal line).

(12) **Phylogenetic relations among modern bee species, from different sources**

Nesting structure is indicated as ● close vs ○ open, and the direction of the dance as ↓vertical vs ↔



### 3.3 Featural meanings

We turn to a third type of semantic system, whose elementary elements are acoustic features rather than whole calls. This mechanism was sometimes postulated to explain cases, in particular in birds, in which a species can understand the calls of other species, including ones that are neither geographically nor phylogenetically close to it.

To give an example of featural interpretation (often called feature decoding),<sup>11</sup> Aubin and Brémond (1989) focused on some bird distress calls, which evoke interspecific reactions. They proposed that this is because different species use the same basic feature: "a simple slope applied to a carrier frequency that corresponds to the acoustic shape of a distress call is sufficient to confer a distress meaning to the signal". Building on these ideas, Aubin (1991) synthesized a general distress call that elicits responses in five species of birds (*Larus argentatus*, *L. ridibundus*, *Vanellus vanellus*, *Corvus frugilegus* and *Sturnus vulgaris*). Aubin's synthetic call crucially involved an ascending slope, which led him to conclude that feature decoding is at stake.

Magrath et al. 2020 mention several cases in which featural interpretation might be involved.

"The alarm calls of many species contain acoustic features that are inherently attention-grabbing or frightening, and this may also prompt immediate response to unfamiliar calls (...). Many mobbing, distress, "panic," or general alarm calls, for example, contain elements that are harsh, abrupt, broadband, or nonlinear. These acoustic features tend to be honest signals of caller distress, increase listener attention and response, and prevent habituation (...)

A key question is whether these cases just involve evolutionary convergence among the calls of different species, or also featural interpretation per se. These two possibilities can in principle be disentangled by creating novel calls that are dissimilar to familiar calls on any reasonable measure, but contain the crucial feature (Schlenker et al., to appear)<sup>12</sup>. Doing so in a systematic fashion is non-trivial, however (for mixed results, see Salis et al., to appear).

Without taking a stand on the general issue, we will consider two cases in which there are strong arguments for featural interpretation, and we will argue that call/feature emergence can be explained because pre-comprehension of the feature would have come with a communicative benefit. We will distinguish two cases: emergence of a whole call with the designated feature, with the associated cost of any whole call (and of the feature itself); and emergence of the feature in an existing call, where all that matters is the net cost of *adding* this feature to the call.

<sup>11</sup> Here we follow the discussion of Schlenker et al., to appear.

<sup>12</sup> Concretely: if we start from a familiar call C defined by 4 binary features as in (i), and the assumption that the first feature gets interpreted, we predict that C\* will trigger the same behavior as C while C\*\* will not, despite the fact that in terms of overall similarity, C\*\* is more similar to C than C\* is.

(i) C = <1, 1, 1, 1>  
 C\* = <1, 0, 0, 0>  
 C\*\* = <0, 1, 1, 1>



### 3.3.1 Non-linearities

In line with Magrath et al.'s (2020) remark that some features "increase listener attention and response, and prevent habituation", Blumstein and Récapet 2009 devised an elegant experiment to show that non-linearities per se increase the effect of marmot alarm calls. The authors started from a natural marmot alarm call, and then created two synthetic versions of it: a target call with white noise (non-linearities) inserted in the middle of the call, and a control call with silence inserted in the middle. The target call with white noise had a *greater* alarm effect than the natural call (in terms of reduction of time devoted to foraging), which suggests that the acoustic feature added was effective on its own. By contrast, the control call with silence inserted in the middle had a comparable alarm effect to the natural call. Importantly, the effect of non-linearities is by no means restricted to marmots. Thus Blumstein et al. 2012 show that adding non-linearities to human music increases arousal and decreases valence.

Turning to the issue of emergence, the first question is how a new alarm call with non-linearities could come to emerge. If the call is new, it will of course have a cost, and thus our condition  $b^- \geq c$  in (5) will have some 'bite': there should be a pre-comprehension benefit to communication. But here Blumstein and Récapet's driving hypothesis (following the literature, see e.g. Fitch et al. 2002) might offer a clue. For them, "vocalizations containing non-linearities are more variable from one rendition to the next, and thus are harder to habituate to", a property reflected in Magrath et al.'s (2020) description. In the terms of Section 2.5, nonlinearities are an instance of sensory manipulation. If correct, this suggests that upon first exposure, a new alarm call with non-linearities might be particularly salient to a receiver that hasn't evolved to understand the signal, simply because habituation is more difficult. Whether this effect could outweigh the cost of producing a call remains to be seen, of course.

But since we allow for the possibility that designated individual features rather than entire calls are interpreted, we should ask how a new feature involving non-linearities could be added to an *existing* call. To apply our framework to this case, we must take the new signal to be the feature itself, and thus we must assess the communicative benefit of adding the feature to the call, and the cost of doing the same thing.<sup>13</sup> The same condition  $b^- \geq c$  requires that the additional pre-comprehension benefit of the feature should outweigh the cost of producing the feature. As before, the benefit of adding the feature to an existing call might be to magnify its salience. But what about the cost of producing non-linearities? Here Fitch et al. 2002 make an important remark: "much of the complexity evident in animal vocal repertoires results from nonlinearities in the production system, meaning that rather simple neural commands to the vocal tract can lead to highly complex and individually variable acoustic output". In other words, adding non-linearities to an existing signal might only add a modest cost to it, in which case the condition  $b^- \geq c$  might be 'easy' to satisfy. This is of course speculative at this point (see Fitch et al. 2002 for further possible benefits of non-linearities in animal communication).<sup>14</sup>

### 3.3.2 Roughness

Arnal et al. 2019 focus on a special case of non-linearities, roughness, "an acoustic texture that arises from fast repetitive acoustic transients" (i.e. from fast modulations of the intensity of the sound) in a restricted frequency [30-150 Hz] range.<sup>15</sup> When discrete modulations are repeated faster, they are perceived as continuous. When repetition speed is below this sampling limit (sensory fusion), the modulations are perceived as discrete and their high rate makes them salient and, according to Arnal et al. 2019, aversive.<sup>16</sup> Here too, we are dealing with a case of sensory manipulation (in the terminology

<sup>13</sup> Note that this requires a departure from the details of our formal framework; we leave a more detailed formal analysis of feature emergence for future research.

<sup>14</sup> A further issue is whether nonlinearities could be used as a cue, for instance about the sender's emotional state. We leave this question for future research.

<sup>15</sup> Arnal et al. 2019 add: "Although the delimitation of the roughness range—whether psychoacoustic or perceptual—may slightly vary depending on experimental settings (...), empirical observations consistently suggest that sensory systems and perception are exceedingly well tuned to recurring temporal features in the 30–150 Hz range (...)."

<sup>16</sup> As Arnal et al. 2019 write, "although salience may not systematically result in aversive percept, we argue that in this specific context, temporal salience—owing to the imperative effect of exogenously saturating perceptual systems in time—constitutes a valid proxy of aversion".

of Section 2.5). While Arnal et al. 2015, 2019 focus on the effects of roughness on humans, including at the neural level, fast amplitude-modulated (AM) sounds have been found in numerous animals. Postal et al. 2020 note that "in other mammals (otters, primates or bats), studies indicate that fast AM sounds, such as a baby's cry, are also produced in alarm calls, and/or antagonistic interactions (Leinonen et al., 2003; Mumm and Knörnschild, 2017; Hechavarría et al., 2019)" Testing how this feature impacts mice behavior, Postal et al. show that rough modulations increase mice attraction to the source of the sound.

Strikingly, Hechavarría et al. 2020 observe that "roughness-like structure is also present in vocalizations emitted by bats (species *Carollia perspicillata*) in distressful contexts". They further show that in bats, "rough-like vocalizations entrain brain potentials and are more effective in accelerating the bats' heart rate than slow amplitude modulated sounds". The prevalence of roughness in diverse animals suggests that this feature has been preserved over very long periods, and/or gave rise to strong convergent evolution.

The evolutionary question is similar to that of non-linearities in the preceding section. We can first consider a new call with roughness, and ask whether the pre-comprehension benefit of communication could outweigh the cost of producing the call. The perceptual and neural mechanism evidenced by Arnal et al. 2015, 2019 suggests that the new call will have an aversive effect on receivers that have not yet evolved to fully understand the signal, which means that the pre-comprehension benefit  $b^-$  of communication in such cases might conceivably outweigh the cost  $c$ , although this is of course speculative.

Second, we should consider the addition of roughness to an existing call. If the call is to have an aversive/alarm function, adding roughness to it should certainly strengthen its effect, including for receivers that do not fully decode the feature. Since the call already exists, we must ask about the specific cost of adding roughness to it. Here Hechavarría et al. 2020 suggest that the production of roughness might be 'cheap', in the sense that it might be a by-product of other physiological processes.<sup>17</sup> If this remark is on the right track, the cost  $c$  of adding roughness to an existing call might be small, which in turn might make the condition  $b^- \geq c$  relatively easy to satisfy. But this issue ought to be investigated further, as humans (or rather adult ones) typically seem reluctant to produce sustained rough sounds in a normal social context.

### 3.4 Stepping back

In sum, the framework developed in Section 2 offers a very simple way to analyze the emergence of diverse semantic systems. The condition  $b^- \geq c$  can be achieved in different ways. When  $c > 0$ , there must be a pre-comprehension benefit of the signal, which in turn can arise in two ways: because signals are to some extent understood because they are cues (cases discussed above involve ritualization), or because they are based on sensory manipulation (nonlinearities and roughness might be cases in point). There are also cases in which arguably  $c \leq 0$ , with a possible example in the waggle dance of bees, where the behavior might have arisen for non-communicative reasons, making the condition  $b^- \geq c$  automatically satisfied.<sup>18</sup>

## 4 How Non-trivial Combinations can Arise

We turn to the question of the emergence of non-trivial combinations. But since their very existence is under debate, we first mention two cases that seem convincing: Putty-nosed monkey *pyow-hack*

<sup>17</sup> Specifically, Hechavarría et al. 2020 write that "in humans, growl-like voices (related to anger) could be a byproduct of abdominal muscle contraction which changes resonances in the vocal tract. Abdominal muscle contraction is a mechanism for enhancing spine stability, which in turn is fundamental for achieving advantageous postures to produce and/or withstand physical attacks in distressful contexts. The mechanism that accounts for growl-like voices in humans is largely reactive and could account as well for roughness in human screams and even for the fast amplitude modulations reported in the present study in bat distress calls."

<sup>18</sup> Importantly, our framework should be extended in the future to take into account the case—highly relevant for applications—in which a new signal arises in the presence of pre-existing repertoire, with non-trivial issues pertaining to the added benefit of the new signal relative to the old signals.

sequences, which were proposed to be idioms (Arnold and Zuberbühler 2012); and Japanese tit ABC-D sequences, which were argued to involve non-trivial compositionality (Suzuki et al. 2016, 2017, 2018, Suzuki and Matsumoto 2022). We develop an evolutionary analysis, and then apply it to these two cases, with non-trivial results: we argue that one recent compositional analysis of ABC-D sequences lends itself to a plausible evolutionary scenario, but that *pyow-hack* sequences as originally analyzed would have trouble getting off the ground from an evolutionary perspective.

#### 4.1 *The existence of non-trivial combinations*

A key question in contemporary studies of animal communication pertains to the existence of non-trivial combinations of signals. Here 'non-trivial' means that the concatenation of two signals C and C' cannot be analyzed as the consecutive but separate utterance of C and C'. Three salient cases have been discussed in the literature.

In Putty-nosed monkeys, *pyow* serves as a general call (or possibly as an attention-getter), *hack* serves as a raptor alarm call (or possibly as a high arousal call), but a small number of *pyows* followed by a small number of *hacks* trigger group movement. Arnold and Zuberbühler 2012 proposed that such series are combinatorially complex (as shown by the fact that the number of *pyows* and *hacks* is variable, making it unlikely that these just form a very long call), but not compositional: they proposed instead that they should be analogized to idioms (such as *kick the bucket*) in human language.

In Campbell's monkeys, it was argued that a suffix *-oo* can modify the meaning of two roots, *krak* and *hok*, in a possibly regular fashion (Ouattara et al. 2009, Schlenker et al. 2014). Importantly, it was difficult to exclude the possibility that *krak-oo* and *hok-oo* are just elementary calls without a root-suffix structure. More essentially for present purposes, *-oo* never occurs on its own, with the result that *krak-oo* and *hok-oo* fall outside the scope of this section.

Finally, a series of articles by Suzuki and colleagues on Japanese tits made the best case to date for the existence of compositionality in animal calls (Suzuki et al. 2016, 2017, 2018, Suzuki and Matsumoto 2022). Japanese tits have an ABC call that raises an alert, and a D call that triggers recruitment. But ABC-D serves to trigger mobbing of a predator, an apparently very different function. Importantly, Suzuki et al. 2017 gave an argument for the productivity of ABC-D by constructing novel sequences of the form ABC-D\*, where D\* is the recruitment call of a sympatric species, one which is understood by Japanese tits. These took the novel sequence ABC-D\* to call for recruitment as well. This strongly suggested that the meaning of ABC-D is computed from the meaning of ABC and D, for otherwise the extension to D\* would have been surprising.

In a further development, Suzuki and Matsumoto 2022 excluded the possibility that ABC-D is understood as the mere concatenation of two separate utterances ('trivial compositionality'). In a nutshell, they showed that ABC followed by D is effective in triggering predator mobbing when played back from a single loudspeaker 5m away from a predator model, but not when played back from two speakers at the same (5m) distance of the model (for a positive review of the arguments, see Schlenker et al. 2023, and Beckers et al. 2024 for a more critical view). In other words, ABC-D was effective as a single utterance, but not as two separate utterances.

In the case of *pyow-hack* sequences and ABC-D sequences alike, a key evolutionary question is: How did a combination of two calls C C' start being produced or understood differently from two separate utterances of C and C'? As we will see, the puzzle is that this would have led to systematic errors. The problem does not arise with the *krak-oo* and *hok-oo* in Campbell's monkeys because the suffix *-oo* is never produced on its own.

#### 4.2 *Conditions for emergence of non-trivial combinations*

In the case of individual signals, the basic question was how they could ever arise without being understood first. In the case of non-trivial combinations, the question is a bit different: how could they ever arise, without causing systematic errors? This is because a population that initially does not have non-trivial combinations will produce and understand sequences as separate utterances, and thus

differently from any mutants that treated some of them as non-trivial combinations.<sup>19</sup> We will argue that, if all interpretive errors are detrimental to the communicative exchange, non-trivial combinations will never arise. We will conclude that some interpretive errors should increase shared utility, potentially more than offsetting cases in which they decrease it.

#### 4.2.1 Assumptions and notations

We consider one population (the residents), *Trivial*, which has calls A and B, as well as trivial combinations in which they are treated as separate utterances; we write these trivial combinations as A&B, using the symbol for conjunction because separate utterances are conjunctively combined. Another population (the mutants), *Combinatorial*, can produce and understand A and B, A&B, and another combination of A and B with a meaning different from A&B. We write this as A\*B, where \* represents whatever non-conjunctive operation yields the meaning of the combination A\*B. Our notations are summarized in (13). It is important to remember that in all cases only the concatenations A B are produced, with different pause lengths between A and B. In addition, with an eye to maximum simplicity, we solely consider the case in which A&B and A\*B are produced (we thus leave for future research the case in which signalers can for instance produce a sequence A A\*B, made of A together with the non-trivial combination A\*B).

(13) **Trivial and non-trivial combinations**

A&B: combination of A and B interpreted conjunctively

A\*B: combination of A and B interpreted non-conjunctively by Combinatorial.

We assume that Combinatorial distinguishes in production and comprehension alike between A&B and A\*B. It may insert a longer pause in A&B, thus making it distinct from its production of A\*B (thus we assume that Combinatorial always interpreted correctly A&B and A\*B as produced by Combinatorial). In situations in which both are true, Combinatorial has a choice: it will produce A&B at rate  $r_{\&}$  (and thus A\*B at rate  $(1-r_{\&})$ ).<sup>20</sup> By contrast, Trivial collapses A&B and A\*B into A&B, in the following sense: in perception, it understands both as meaning A&B, and in production, it occasionally produces A&B with a (short) pause that corresponds to what Combinatorial understands as A\*B, at rate  $r_*$ . These assumptions are summarized in (14).

(14) **Cases of optionality**

- a. When A&B and A\*B are both true: Combinatorial produces A&B at rate  $r_{\&}$ , A\*B at rate  $(1-r_{\&})$ .
- b. Trivial produces A&B with a timing corresponding to Combinatorial's A\*B at rate  $r_*$ , and it produces A&B with a timing corresponding to Combinatorial's A&B at rate  $(1-r_*)$ . As a result, Combinatorial perceives A&B produced by Trivial as A\*B at rate  $r_*$ , and as A&B at rate  $(1-r_*)$ .

(For mnemonic purposes, it may help to think of  $r_{\&}$  and  $r_*$  as the rates of suboptimal behavior:  $r_{\&}$  corresponds to the rate at which Combinatorial could make use of its innovation A\*B, but chooses not to, and to produce A&B instead.  $r_*$  corresponds to the case in which Combinatorial misinterprets a production of A&B by Trivial, understanding it, wrongly, as A\*B.)

Some of our discussions will be simplified by assuming that  $r_*$  is smaller than  $2(1-r_{\&})$ . In words, this corresponds to the following: the rate of misunderstanding of A&B by Combinatorial is smaller than twice the rate at which Combinatorial produces A\*B when it could just as well produce A&B. It is reasonable to assume that  $r_*$  is small, because this will correspond to cases in which Trivial

<sup>19</sup> Conceptually, this is part of a broader pattern in which mutants repurpose an old signal to have a new meaning. For simplicity, we only consider the case of combinations, leaving the more general cases for future research.

<sup>20</sup> It is for the sake of generality that we do not take a position on the value of  $r_{\&}$ , the rate at which Combinatorial produces A&B even though it could produce A\*B. Assuming the Informativity Principle, whose emergence is discussed in Section 6, when A\*B is more informative than A&B and both are true, Combinatorial should always produce A\*B. More generally, on the assumption that  $b_* > b_{\&}$ , producing A&B in this case is a suboptimal behavior. We show below that assuming that  $r_{\&} = 0$  (and thus that Combinatorial is optimal in this respect) does not change the conclusions of this section.

accidentally hits the 'wrong' timing (from Combinatorial's viewpoint) for the A&B combination. It is also plausible to assume that  $r_{\&}$  is relatively small (and thus  $1-r_{\&}$  relatively large), as it corresponds to cases in which Combinatorial could use its innovation A\*B but decides to use A&B instead (this is especially the case under the assumption that A\*B yields stronger benefits). These make the assumption in (15) plausible, and it will simplify some of the technical discussions (for our purposes, assuming the stronger condition  $r_* < 1-r_{\&}$  would work just as well, it just wouldn't be minimal):

(15) **Technical assumption**

$$r_* < 2(1-r_{\&})$$

As a final set of notations, we will label four types of situations, FF, TT, TF and FT, with the first letter indicating the truth value in that situation of A&B (separate utterances), and the second letter indicating the truth value of A\*B (non-trivial combination). For each situation type, we write its probability weight as  $p_i$ , hence as  $p_{FF}$ ,  $p_{TT}$ , etc.

#### 4.2.2 Payoffs

We assume that issues of production cost won't matter. Furthermore, we will now assume for simplicity that the benefits of communication are fully shared. Thus there will be no cost asymmetry between sender and receiver (unlike in Section 2), and no difference in payoffs either. We take the benefit of conveying a true message to be  $b_{\&}$  for A&B and  $b_*$  for A\*B. When a false message is conveyed, a negative benefit ensues,  $-e$  (this can be thought of as the cost of the error).<sup>21</sup> These assumptions are summarized in (16).

(16) **Benefits of communication in different situations**

Information inferred	Actual truth value	Benefit
<b>A&amp;B</b>	True (successful communication)	$b_{\&}$
	False (misunderstanding)	$-e$
<b>A*B</b>	True (successful communication)	$b_*$
	False (misunderstanding)	$-e$

or

Information inferred	A&B		A*B	
Actual truth value	true (successful)	false (error)	true (successful)	false (error)
Benefit of communication	$b_{\&}$	$-e$	$b_*$	$-e$

We will now motivate the distribution of payoffs summarized in (17).

(17) **Summary of the payoffs**

Sender	Trivial		Combinatorial	
Receiver	Trivial		Combinatorial	
<b>A&amp;B</b> <b>A*B</b>				
<b>FF</b> false   false	0	0	0	0
<b>TT</b> true   true	$b_{\&}$	$r_{\&}b_{\&} + (1-r_{\&})b_*$	$r_*b_* + (1-r_*)b_{\&}$	$b_{\&}$
<b>FT</b> false   true	0	$b_*$	0	$-e$
<b>TF</b> true   false	$b_{\&}$	$b_{\&}$	$r_*(-e) + (1-r_*)b_{\&}$	$b_{\&}$

<sup>21</sup> For simplicity, we only consider cases in which  $e > 0$  and thus conveying false information has a strict cost ( $-e < 0$ ). We leave for future research the case in which some false information can still provide a benefit.

<b>Weighted shared payoff</b>	$(p_{TT}+p_{TF})b_{\&}$	$p_{TT}[r_{\&b_{\&}} + (1-r_{\&})b_{*}]$ $+ p_{FT} b_{*}$ $+ p_{TF} b_{\&}$	$p_{TT} [(2-r_{*})b_{\&} + r_{*}b_{*}]/2$ $- p_{FT} e/2$ $+ p_{TF} [(2-r_{*})b_{\&} - r_{*}e]/2$
-------------------------------	-------------------------	---	--

The weighted payoffs for homogeneous encounters of Trivial vs. Trivial and Combinatorial vs. Combinatorial are just obtained by weighting the benefits of the four rows FF, TT, FT and TF by the probabilities of the situations. For Combinatorial vs. Trivial, in half the cases the sender is Trivial, and in the other half the sender is Combinatorial. We must thus take the average of these two cases, and weigh the payoffs by row, as for the other columns.

The remaining task is to justify the content of the four rows FF, TT, FT and TF.

**FF:** A&B: false, A\*B: false

In all cases, neither A&B nor A\*B can be produced. And all parties get a payoff of 0.

**FT:** A&B: false, A\*B: true

- When Trivial is the sender, no call is produced, and all payoffs are then 0.
- When Combinatorial is the sender, it produces A\*B.
- If the receiver is Combinatorial, A\*B will be interpreted as such and yield a payoff  $b_{*}$ .
- If the receiver is Trivial, A\*B will be misinterpreted as A&B, and the payoff will be  $-e$ .

**TF:** A&B: true, A\*B: false

- When Trivial is the sender, it always produces A&B.
- If the receiver is also Trivial, the payoff is  $b_{\&}$ .
- If the receiver is Combinatorial, it interprets the message as A\*B with probability  $r_{*}$ , which is true, hence a payoff of  $b_{*}$ . With probability  $(1-r_{*})$ , Combinatorial interprets the message as A&B, which is false, hence a payoff of  $-e$ . Overall, the payoff is  $r_{*}b_{*} + (1-r_{*})(-e)$ .
- When Combinatorial is the sender, it produces A&B. Whether the receiver is Trivial or Combinatorial, the message will be interpreted as A&B, which is true, and the payoff will be  $b_{\&}$ .

**TT:** A&B: true, A\*B: true

- When Trivial is the sender, it always produces A&B.
- If the receiver is also Trivial**, the payoff is  $b_{\&}$ .
- If the receiver is Combinatorial**, it interprets the message as A\*B with probability  $r_{*}$ , which is true, hence a payoff of  $b_{*}$ . With probability  $(1-r_{*})$ , Combinatorial interprets the message as A&B, which is true as well, hence a payoff of  $b_{\&}$ . Overall, the payoff is  $r_{*}b_{*} + (1-r_{*})b_{\&}$ .
- When Combinatorial is the sender, it produces A&B with probability  $r_{\&}$  and A\*B with probability  $(1-r_{\&})$ .
- If the receiver is Combinatorial**, it will interpret both messages appropriately in all cases, and the payoff will be  $r_{\&} b_{\&} + (1-r_{\&}) b_{*}$ .
- If the receiver is Trivial**, it will interpret both messages as A&B in all cases, which is true, and therefore get a reward of  $b_{\&}$ .

#### 4.2.3 Evolutionary stability of Trivial

As before, Trivial is evolutionarily stable relative to Combinatorial just in case:

Condition (1): payoff of Trivial vs Trivial > payoff of Combinatorial vs Trivial, or

Condition (2): these payoffs are equal, but payoff of Trivial against Combinatorial > payoff of Combinatorial against Combinatorial.

**Condition (1):** It is clear by inspecting (17) that, as soon as  $e > 0$  and  $r_{*} > 0$ , the weighted shared payoffs of Combinatorial vs Trivial are strictly lower than those of Trivial vs Trivial in rows FT and TF, and identical (and null) in row FF. The only possibility for Trivial to be invaded is for row TT to give an advantage to Combinatorial, and furthermore one that offsets the disadvantage of Combinatorial in the other cases.

Condition (1) for the stability of Trivial is given in (18).

- (18) Trivial gets a higher payoff against Trivial than Combinatorial does against Trivial  
 $(p_{TT} + p_{TF})b_{\&} > p_{TT}[(2-r_*)b_{\&} + r_*b_*/2 - p_{FT}e/2 + p_{TF}[(2-r_*)b_{\&} - r_*e)]/2$ , which is equivalent to  
 $2(p_{TT} + p_{TF})b_{\&} > p_{TT}[(2-r_*)b_{\&} + r_*b_*] - p_{FT}e + p_{TF}[(2-r_*)b_{\&} - r_*e]$   
 $p_{TT}[2b_{\&} - 2b_{\&} + r_*b_{\&} - r_*b_*] + p_{FT}e + p_{TF}(2b_{\&} - 2b_{\&} + r_*b_{\&} + r_*e) > 0$   
 $p_{FT}e + p_{TF}r_*(b_{\&} + e) > p_{TT}r_*(b_* - b_{\&})$   
 $r_*(p_{TT}b_{\&} + p_{TF}b_{\&}) > r_*(p_{TT}b_* - p_{TF}e) - p_{FT}e$

**Condition (2):** We can once again inspect the payoffs in (17), and note that:

- the situations FF continue to be irrelevant, as every interaction yields payoff 0.
- when  $b_* > b_{\&}$ ,  $r_{\&} \neq 1$ ,  $r_* \neq 0$ , and  $r_* < 2(1-r_{\&})$  (= the technical assumption in (15)), Combinatorial has strictly higher payoffs against Combinatorial than Trivial does against Combinatorial in situations TT, FT, TF. This is immediate for FT and TF. For TT, this follows because of the computation in (19).

- (19) Line TT of (17): condition on which payoff of Combinatorial against Combinatorial > payoff of Trivial against Combinatorial  
 $r_{\&}b_{\&} + (1-r_{\&})b_* > [r_*b_* + (1-r_*)b_{\&} + b_{\&}] / 2$ , which is equivalent to  
 $2r_{\&}b_{\&} + 2(1-r_{\&})b_* > r_*b_* + (1-r_*)b_{\&} + b_{\&}$   
 $2r_{\&}b_{\&} - (2-r_*)b_{\&} + 2(1-r_{\&})b_* - r_*b_* > 0$   
 $[-2 + 2r_{\&} + r_*]b_{\&} + [2(1-r_{\&}) - r_*]b_* > 0$   
 $-[(1-r_{\&}) - r_*]b_{\&} + [2(1-r_{\&}) - r_*]b_* > 0$   
 $[2(1-r_{\&}) - r_*](b_* - b_{\&}) > 0$   
 This is satisfied on the assumption that  $b_* > b_{\&}$  and  $r_* < 2(1-r_{\&})$ , which is the technical assumption in (15).

In sum, condition (2) is always violated, and Trivial is evolutionarily stable just in case condition (1) is satisfied.<sup>22</sup> Several consequences can be drawn from this analysis.

1. We note that if the two meanings A&B and A\*B are non-overlapping (in the sense that no situation makes them both true), i.e.  $p_{TT} = 0$ , then Trivial will always be evolutionarily stable and will not be invaded. Technically, this is because only row TT can provide an advantage to Trivial vs Combinatorial over Trivial vs Trivial. In other words, a meaning that is incompatible with the conjunction of A and B cannot evolve for the combination A\*B.
2. But having  $p_{TT} > 0$  is not at all enough to alter the stability of Trivial. The advantage obtained by Combinatorial in the cases in which both A&B and A\*B are true must at least outweigh the disadvantage that Combinatorial has in the other cases. (We write 'at least outweigh' rather than 'more than outweigh' because if Trivial interacting with Trivial has the very same payoff as Combinatorial interacting with Trivial, the second condition of evolutionary stability kicks in, and the latter gives a clear advantage to Combinatorial, as noted).

#### 4.2.4 Total invasion of Trivial by Combinatorial

Using (2), we can ask under what conditions Trivial is totally invaded by Combinatorial.

- (20) Trivial is fully invaded by Combinatorial  
 iff  
 (i) payoff of Combinatorial against Trivial  $\geq$  payoff of Trivial against Trivial,  
 (ii) payoff of Combinatorial against Combinatorial  $\geq$  payoff of Trivial against Combinatorial  
 and (iii) at least one of these inequalities is strict.

Since (i) and (ii) immediately follow from conditions (1) and (2) of the previous subsection (Evolutionary Stability of Trivial), the results are immediate:

1. By our discussion of condition (1) in Section 4.2.3 (specifically, (18)), condition (i) is equivalent to:  
 $p_{FT}e + p_{TF}r_*(b_{\&} + e) \leq p_{TT}r_*(b_* - b_{\&})$ .

<sup>22</sup> Echoing fn. 20, we can check that nothing in our results is affected when  $r_{\&} = 0$ , as  $r_{\&}$  plays no role in (18), and when  $r_{\&} = 0$  the assumptions of (19) are easier to satisfy than when  $r_{\&} \neq 0$ .

2. By our discussion of condition (2) in Section 4.2.3 (especially the underlined parts), when  $b_* > b_{\&}$ ,  $t \neq 1$ ,  $r_* \neq 0$ ,  $r_* < 2(1-t_{\&})$  and  $p_{FF} < 1$ , condition (ii) is satisfied in its strict version, which satisfies condition (iii).

#### 4.2.5 Discussion

Our main general finding is that for a combination  $A*B$  to emerge, it must be true in some of the same situations as  $A\&B$ . This is a non-trivial constraint. For example, Schneider et al. 2010 discuss head shakes used by some bonobos to accompany "an active effort to terminate or prevent the recipient from engaging in a particular behavior"—let us abbreviate this as *you shouldn't do that* (relatedly, Patel-Grosz 2023 discusses chimpanzee gestures that could be taken to mean *stop that*). One might well expect the bonobo head shake to combine with another gesture to negate it.<sup>23</sup> But within our framework, a concatenation of  $A = \textit{you shouldn't do that}$  and another gesture, for instance  $B = \textit{you should climb on me}$  (as in (10) above), could *not* give rise to a new combination  $A*B$  meaning *you shouldn't climb on me*. The problem is that  $A\&B$  entails *you should climb on me*;  $A*B$  entails its negation. So it is clear that no situation makes the two meanings true, which was a crucial condition for the meaning of the combination to emerge. It should be noted in this connection that although some non-human primates, such as captive baboons, can learn the meaning of negation in artificial learning experiments (Dautriche et al. 2022), no gesture or call with the meaning of negation (i.e. as applied to another gesture or call) has been described to date in natural animal communication. We will see in Section 5.2 another example with the same logic, pertaining not to negation, but to a purported idiom in putty-nosed monkeys.

Finally, while  $A*B$  and  $A\&B$  must be true in some of the same situations, this is not enough: using  $A*B$  over  $A\&B$  must have an added benefit that outweighs the penalties that would incur from misinterpretations in other situations.

More refined models could be developed in future research, in particular if incorrect information can sometimes be useful (which would change modify the assumption that  $e > 0$  and thus  $-e < 0$ ), if more than two strategies are considered,<sup>24</sup> and if signalers can produce sequences that contain individual calls in addition to  $A\&B$  and  $A*B$ .

## 5 Emergence of two combinations: Japanese tits and Putty-nosed monkeys

We now apply the results of our analysis to two cases (briefly introduced in Section 4.1) in which non-trivial combinations have been argued to exist in animals: one type, in Japanese tits, might involve non-trivial compositionality (ABC-D sequences); the other, in putty-nosed monkeys, has been argued to involve idioms (pyow-hack sequences). As we will see, there is a plausible evolutionary scenario for the emergence of ABC-D sequences, but not necessarily for pyow-hack sequences analyzed as idioms.

### 5.1 ABC-D sequences in the tit family

In a series of articles, Suzuki and colleagues have argued that Japanese tits (*Parus minor*) have ABC calls that raise alarms, D calls that trigger recruitment, and ABC-D calls that are derived from ABC and D by way of non-trivial compositionality, and trigger mobbing of certain predators (Suzuki, Wheatcroft and Griesser, 2016, 2017, 2018; Suzuki and Matsumoto 2022). As outlined in Section 4.1, the authors gave arguments to the effect that Japanese tits make use of a productive rule to derive the meaning of

<sup>23</sup> This is not an outlandish possibility: in the human case, Benitez-Quiroz et al. 2016 argue that the head shake found in several sign languages originate in non-linguistic facial expressions of negative moral judgment.

<sup>24</sup> Note that if a third strategy, 'Mixed', behaved like Trivial in production but like Combinatorial in comprehension, to the extent that Combinatorial invades Trivial in our model, Mixed would fare even better against Trivial than Combinatorial does. Specifically, it would fare like Combinatorial in receiver role, but better than Combinatorial in sender role. This is because in (17), the Combinatorial-Trivial column (= the last column) yields strictly less utility than the Trivial-Trivial column owing to the negative payoff  $-e$  in the FT row. In sum, Mixed is also a strong candidate for invading Trivial. For a partly related point, see Schlenker et al. 2024, Appendix B (residents treat ABC-D as ABC & D in production and comprehension alike; mutants treat ABC-D in production, but as ABC \* D in comprehension). See also Sections 6-7 below, where for a different phenomenon (competition principles), invasion can occur in stages, first in production, and then in comprehension.



ABC-D from that of its parts, and furthermore that these two components do not display the behavior of separate utterances. The logic of these experiments provides the strongest argument to date for non-trivial compositionality in animals.

It remains to ask what the compositional rule in question is. Schlenker et al. 2024 propose an analysis ('Minimal Compositionality') that posits a minimal distinction between ABC and D uttered as separate utterances, and ABC-D uttered as a single utterance. ABC means *for some x, x licenses an alert*. D means *for some y, y licenses recruitment*. By trivial compositionality, ABC and D uttered as separate utterances mean in essence: *for some x, x licenses an alert, and for some y, y licenses recruitment*, without an implication that  $x = y$ , so to speak. By contrast, ABC-D uttered as a single utterance means something, namely: *for some x, x licenses an alert and x licenses recruitment*. The mobbing reaction is obtained on the assumption that most things that license both an alert and recruitment involve a predator to be mobbed.

What matters for present purposes is the possibility that ABC-D (the non-trivial combination) might evolve from ABC D (the trivial combination of ABC and D as separate utterances). Schlenker et al. 2024 consider a special case in which the production strategies are fixed, and in a first step only the interpretive strategies evolve. A resident population interprets ABC D as two separate utterances, and thus as: *for some x, x licenses an alert, and for some y, y licenses recruitment*. Mutants interpret this instead as: *for some x, x licenses an alert and x licenses recruitment*, which is stronger.

Owing to the entailment (namely:  $ABC-D \Rightarrow ABC\ D$ ), it couldn't be that ABC-D is true and ABC D is false, and thus there are just three possibilities to consider, corresponding to rows **FF**, **TT** and **TF** in (17).

- (i) ABC D: false; ABC-D: false;
- (ii) ABC D: true; ABC-D: true;
- (iii) ABC D: true; ABC-D: false.

In case (i), there is no communication at all. In case (iii), mutants misinterpret the message produced by residents: ABC D is made true in this case by virtue of two separate events x and y, and mutants incorrectly take this to involve a single event and thus mobbing. This has to lead to a lower payoff. But if this occurrence is sufficiently rare, this cost could be more than outweighed by the gain obtained in case (ii): in those cases, there is one x that licenses an alert and mobbing, presumably an x that needs to be mobbed. ABC-D presents this information directly. By contrast, ABC D is underinformative, and might make it more arduous for the receiver to zero in on the appropriate reaction. In terms of the discussion of Section 4, ABC D yields a benefit  $b_x$  and ABC-D a higher benefit  $b^*$  in this case.

The argument in Schlenker et al. 2024 was thus a special case of the one we developed above: if the frequencies and benefits are 'right', the added benefit in case C&C' and C\*C' are both true could outweigh the cost of the miscommunication that arises in some of the other cases.

We should say for completeness that Schlenker et al. 2024 consider an alternative theory of the Japanese tit data, one that does not rely on non-trivial compositionality, but rather on competition among calls and call sequences. The basic idea is based on the the Informativity Principle, which states that if an expression E competes with an expression E', if E' is more informative than E, it should be preferred if it is true. In the Japanese tit case, the idea was that ABC competes with ABC D, which is more informative on a variety of theories, including trivial compositionality. As a result, uttered on its own, ABC yields a 'not ABC D' inference, and similarly for D uttered on its own. Importantly, Schlenker et al. 2024 needed the non-standard assumption (from the perspective of human language!) that a less complex form (e.g. ABC, or D) competes with a more complex form (ABC D) in order to derive the desired result, and accordingly they did not adjudicate between Minimal Compositionality and bird implicatures in the analysis of Japanese tit calls.

Let us add that it is now possible to obtain extraordinary time depth about the emergence of ABC-D sequences. As shown in ongoing work (Salis et al., in preparation), sequences similar to ABC-D are produced by the vast majority of members of the Paridae family, to which Japanese tits belong. These cousin species associate high-frequency, often modulated pure tones (i.e. ABC-like notes) with broadband, noisy elements (D-like notes). The associated function of the combination is often related to mobbing, although several species produce it in much broader contexts (e.g., when foraging). Thus ABC-D-like combinations are probably an ancestral trait of the Paridae family, emerging about 15

million years ago.<sup>25</sup> Whether ABC-D-type calls are treated as a non-trivial combinations in other species than Japanese tits has yet to be determined, however.

## 5.2 *Pyow-hack sequences in Putty-nosed monkeys*

We turn to the case of putty-nosed monkey pyow-hack sequences, briefly introduced in Section 4.1; by contrast with the case of Japanese tit ABC-D sequences, we will argue that evolutionary scenarios might make some analyses of the meaning of pyow-hack sequences implausible.

To reiterate, pyows have a broad distribution suggestive of a general call, or potentially of attention-getters (Arnold and Zuberbühler 2012). Hacks, by contrast, are often indicative of eagles, or possibly more generally of high arousal situations (Arnold and Price 2015, Schlenker et al. 2016a). Arnold and Zuberbühler 2012 showed that putty-nosed monkeys sometimes produce distinct pyow-hack sequences made of a small number of pyows followed by a small number of hacks ( $P^+H^+$ ); and these were shown both in quantitative observational data and in field experiments to be predictive of group movement.

The key question is whether the meaning of pyow-hack sequences can be derived from the meaning of their parts. In view of the radical difference in meaning, Arnold and Zuberbühler 2012 claim that it cannot be. For them, pyow-hack sequences are syntactically combinatorial but not semantically compositional, and they should thus be likened to idioms in human language. Initially, it might seem to make equal sense to say that pyow-hack sequences are phonologically complex but lexically simple. But Arnold and Zuberbühler do not use this terminology and speak instead of *syntactically* complex sequences interpreted as idioms. This is presumably because pyow-hack sequences are relatively slow, with pauses between calls; and because they are not fully stereotyped: they involve a variable number of pyows followed by a varying number of hacks.

From the perspective of our earlier discussions, we must ask whether the idiomatic analysis of pyow-hack sequences lends itself to a plausible evolutionary scenario; for maximal similarity with the formal framework we developed, we will restrict attention to the case of a single pyow followed by a single hack (these correspond to the calls A and B in Section 4). The plausibility of an evolutionary scenario depends on the details.<sup>26</sup>

If we start from the original description of hack as an eagle-related call, it's unlikely that the evolution of the idiomatic call could get off the ground. Let us recall the framework of Section 4, in which a resident population Trivial only goes by trivial compositionality, while mutants Combinatorial go by a non-trivial combinatorial rule, which could involve anything, including idioms. The key was to find cases in which the communicative errors induced by the mutants' new rule could produce greater utility when they interact with residents than when the latter interact with each other. This only arose in one type of case, corresponding to row **TT** in (17), and thus when  $A*B$  and  $A\&B$  were both true; in the putty-nosed case: when both pyow hack as separate utterances and pyow-hack as an idiom are true. But on the analysis of hack as relating to eagle presence, pyow hack analyzed as separate utterances ought to be false in most group movement situations. This is because eagle presence inhibits any movement, as it would make the prey more easily detectable by the eagle;<sup>27</sup> thus if there is group movement, there is likely no eagle. As a result, on this analysis there just are no situations in the crucial (**TT**) row of (17). This goes to show that the evolutionary analysis we developed in Section 4 puts non-trivial constraints on the possible emergence of idiomatic meanings.

Things are different if the meaning of the individual calls is made much weaker. For instance, if pyow is a general call or an attention-getter, and hack just has a meaning of high arousal, then there

<sup>25</sup> By contrast, the sister clades of Paridae (namely Stenostiridae and Remizidae) do not produce similar combinations, which suggests that ABC-D-type combinations might not much predate the most recent common ancestor of all Paridae.

<sup>26</sup> Importantly, nothing in our account hinges on the fact that pyow-hack sequences are syntactically rather than phonologically complex. All that matters is that their meaning—not matter how it is obtained—deviates in certain ways from trivial compositionality.

<sup>27</sup> Thanks to Klaus Zuberbühler for discussion. See also Arnold and Zuberbühler 2012: "‘hacks’ tend to inhibit movement (probably because they indicate eagle presence)".

could be lots of situations in which both pyow hack construed as separate utterances and the idiom pyow - hack are simultaneously true. But on such analyses, the challenge is to provide a precise analysis that explains why hack is often related to eagle presence.

An analysis has been developed along these lines, but it entirely dispenses with an idiomatic meaning.<sup>28</sup> In a nutshell, pyows were taken to trigger unspecific alerts, and hacks to warn of (serious) non-ground, movement-related events. Semantically, then, a pyow-hack sequence warns of a non-ground, movement-related event. Importantly, this includes two very different event types: the impending movement of an attacking raptor, and also the movement of the (arboreal) monkeys themselves. At this point, a crucial hypothesis was that sequences are ordered according to an Urgency Principle: in any sequence, calls that convey information about the nature or location of a threat should come before those that don't (this principle has been argued to have relevance for other species, notably Japanese tits [Schlenker et al. 2023] and Titi monkeys [Narbona Sabaté et al. 2022]). Now if a raptor were present, hacks would convey information about the location of the threat and hence (by the Urgency Principle) it should come before pyows. In the end, this helped explain why pyow-hack sequences were indicative of group movement; no idioms were posited, but competition principles were crucial.

On this view, the problem of explaining how idioms can arise in evolution is replaced with a different one: how could the Urgency Principle arise? We turn to this question, and to the more general issue of the emergence of competition principles, in the next sections.

## 6 How Competition Principles can Arise I: the Informativity Principle

### 6.1 *Production first vs. Comprehension first*

We mentioned at two junctures competition principles among calls or call sequences. The Informativity Principle mandates that, when a call  $A^+$  is more informative than a call  $A^-$  and both are appropriate,  $A^+$  should be preferred over  $A^-$ . The Urgency Principle mandates that, in any sequence, calls that convey information about the nature or location of a threat should come before calls that don't.

We alluded to the Informativity Principle in our discussion of Japanese tits because it offers an alternative analysis of the results of Suzuki and colleagues on ABC-D sequences. But this version of the principle involved competition among call sequences (ABC competes with ABC-D, D competes with ABC-D), whereas the standard version, stated in the previous paragraph, only involves individual calls. It has been analytically useful in several cases studies, surveyed for instance in Schlenker et al. 2016b. We also alluded to the Urgency Principle because it is a key mechanism in an account of *pyow-hack* sequences that analyzes them without appeal to idioms; and here too, the principle has found a modicum of support in other species, such a Japanese tits and Titi monkeys (Schlenker et al. 2023, Narbona Sabaté et al. 2022).

In this section and the next, we ask how each principle could have evolved (a special case was considered for the Informativity Principle in Schlenker et al. 2024). We consider a particular evolutionary sequence in which pragmatic principles arise in production first, and in comprehension second. There is a reason for this. A pragmatic principle is built on top of optionality that arises from the rest of the system. For instance, the Informativity Principle states that when two calls  $A^-$  and  $A^+$  can be used truly and  $A^+$  is more informative than  $A^-$ ,  $A^+$  must be used. This means that applying the principle in production first should not lead to negative results, and could in fact be useful (in this case, by providing more information). By contrast, if the principle were applied in comprehension relative to a resident population that didn't already have it in production, it would sometimes result in miscommunication. In our example, a mutant would understand  $A^-$  to mean  $A^-$  *and not*  $A^+$ , even though the resident sometimes produces  $A^-$  when  $A^+$  could be used as well. As a result, the emergence of the Informativity Principle (and more generally of pragmatic principles) in production gives rise to a much cleaner state of affairs than was the case in earlier sections; involving comprehension in the first stage would make it necessary to weigh the costs and benefits of adopting the principle.

---

<sup>28</sup> Here we follow the discussion of Schlenker et al. 2016a.

Focusing for the moment on the Informativity Principle, there are in principle three paths of evolution depending on whether it initially develops (i) simultaneously on the production and on the comprehension side, (ii) in comprehension only, (iii) in production only. For the reasons we just stated, we restrict attention to case (iii), but discuss the other two cases in Appendix II. We will then show that once the Informativity Principle has emerged in production, it will emerge in comprehension as well.<sup>29</sup>

## 6.2 *Emergence of the Informativity Principle in production*

### 6.2.1 *Assumptions and notations*

We consider populations with two calls  $A^+$  and  $A^-$ , where  $A^+$  is strictly more informative than  $A^-$ . A population Inf ('Informativity Principle') goes by the Informativity Principle in production but not in comprehension: when  $A^+$  and  $A^-$  are both true, Inf always uses  $A^+$ . A population No-Inf ('no Informativity Principle') does not go by the Informativity Principle, be it in production or in comprehension. In production, when given a choice between  $A^+$  and  $A^-$ , No-Inf uses  $A^-$  at rate  $r$ , and thus  $A^+$  at rate  $(1-r)$ . Since neither population applies the Informativity Principle in comprehension, they have the same interpretive rule: just interpret  $A^+$  as  $A^+$  and  $A^-$  as  $A^-$ , without strengthening the latter to:  $A^-$  and not  $A^+$ . These assumptions are summarized in (21).

- (21) **Production and comprehension rules for Inf and No-Inf**
- a. Production: only produce true calls, and furthermore:  
 Inf: if  $A^+$  and  $A^-$  are both true, use  $A^+$   
 No-Inf: if  $A^+$  and  $A^-$  are both true, use  $A^-$  at rate  $r$ , and  $A^+$  at rate  $(1-r)$
  - b. Comprehension: for Inf and No-Inf alike,  
 interpret  $A^+$  as  $A^+$  and  $A^-$  as  $A^-$  (no strengthening)

As a final set of notations, we will label four types of situations, FF, TT, TF, with the first letter indicating the truth value in that situation of  $A^-$ , and the second letter indicating the truth value of  $A^+$ . Since  $A^+ \Rightarrow A^-$ , there are no situations in which  $A^-$  is false while  $A^+$  is true, which is the reason we do not need to consider situations of type FT. As in earlier sections, for each situation type, we write its probability weight as  $p_I$ , hence as  $p_{FF}$ ,  $p_{TT}$ , etc.

### 6.2.2 *Payoffs*

We disregard issues of call production cost. Furthermore, since both populations under consideration only produce true calls, and always interpret calls correctly, we needn't worry about the cost of communication errors. The only assumption that will matter is that when  $A^-$  and  $A^+$  can both be used, the more informative call  $A^+$  yields benefit  $b^+$  while the less informative call  $A^-$  yields benefit  $b^-$ , with  $b^+ > b^-$ .

- (22) **Benefits of communicating different meanings** (all are true)  
 Assumption:  $b^+ > b^-$

Information inferred	Actual truth value	Benefit
$A^-$	True (successful communication)	$b^+$
$A^+$	True (successful communication)	$b^-$

We will now motivate the distribution of payoffs summarized in (23).

- (23) **Summary of payoffs**

Sender	No-Inf	Inf	No-Inf	Inf
--------	--------	-----	--------	-----

<sup>29</sup> For a sophisticated game-theoretic analysis of the evolution of the Informativity Principle relative to lexical specifications in human language, see Brochhagen et al. 2018.

Receiver	No-Inf	Inf	Inf	No-Inf
<b>FF</b> ( $A^-$ : false, $A^+$ : false)	0	0	0	0
<b>TT</b> ( $A^-$ : true, $A^+$ : true)	$rb^- + (1-r)b^+$	$b^+$	$rb^- + (1-r)b^+$	$b^+$
<b>TF</b> ( $A^-$ : true, $A^+$ : false)	$b^-$	$b^-$	$b^-$	$b^-$
<b>Weighted shared payoff</b>	$p_{TT} [rb^- + (1-r)b^+] + p_{TF} b^-$	$p_{TT} b^+ + p_{TF} b^-$	$p_{TT} [rb^- + (2-r)b^+]/2 + p_{TF} b^-$	

**FF:**  $A^-$ : false,  $A^+$ : false

In all cases, neither call can be produced, and all parties get a payoff of 0.

**TT:**  $A^-$ : true,  $A^+$ : true

- When No-Inf is the sender, it produces  $A^-$  at rate  $r$ , yielding benefit  $b^-$ , and  $A^+$  at rate  $(1-r)$ , yielding benefit  $b^+$ , hence a payoff (for sender and receiver alike) of  $rb^- + (1-r)b^+$ .
- When Inf is the sender, it always produces  $A^+$ , yielding a payoff of  $b^+$ .

**TF:**  $A^-$ : true,  $A^+$ : false

In all cases, the sender produces  $A^-$ , yielding a payoff of  $b^-$ .

Weighted shared (symmetric) payoffs are computed as expected, namely by weighting situations FF, TT and TF by their probability weight for No-Inf vs. No-Inf and Inf vs. No-Inf. For No-Inf interacting with Inf, we must first take the average of the payoffs for No-Inf = the sender and Inf = the sender, after which weight the result by the probability of FF, TT and TF.

### 6.2.3 Evolutionary stability of No-Inf

As before, No-Inf is evolutionarily stable relative to Inf just in case:

Condition (1): payoff of No-Inf vs No-Inf > payoff of Inf vs No-Inf, or

Condition (2): these payoffs are equal, but payoff of No-Inf vs Inf > payoff of Inf vs Inf.

By inspecting rows FF and TF of (23), it is clear that in these two cases, payoff of No-Inf vs No-Inf = payoff of Inf vs No-Inf (the payoff is 0 in row FF and  $b^-$  in row TF). In row TT, if  $r \neq 0$  and  $b^+ > b^-$ , payoff of No-Inf vs No-Inf < payoff of Inf vs No-Inf. So as soon as there are TT situations, or in other words as soon as  $p_{TT} > 0$ , overall, **payoff of No-Inf vs No-Inf < payoff of Inf vs No-Inf**. Therefore No-Inf is not evolutionarily stable relative to Inf.

### 6.2.4 Total invasion of No-Inf by Inf

Using (2), we can ask under what conditions No-Inf is totally invaded by Inf.

(24) No-Inf is fully invaded by Inf

iff

(i) payoff of Inf vs No-Inf  $\geq$  payoff of No-Inf vs No-Inf,

(ii) payoff of Inf vs Inf  $\geq$  payoff of No-Inf vs Inf

and (iii) at least one of these inequalities is strict.

By the boldfaced part of the discussion of Section 6.2.3, condition (i) is satisfied in its strict version as soon as  $r \neq 0$ ,  $b^+ > b^-$  and  $p_{TT} > 0$ . Regarding condition (ii), by inspecting the payoffs in (23) again, we see that in rows FF and TF payoff of Inf vs Inf = payoff of No-Inf against Inf, and in row TT, payoff of Inf vs Inf  $\geq$  payoff of No-Inf against Inf, a condition which is strict as soon as  $b^+ > b^-$  and  $r \neq 0$ . It is clear that in all cases, condition (ii) is satisfied (and if  $p_{TT} \neq 0$ ,  $b^+ > b^-$  and  $r \neq 0$ , it is satisfied in its strict version).

In sum, if  $r \neq 0$ ,  $b^+ > b^-$  and  $p_{TT} > 0$ , No-Inf is totally invaded by Inf.

### 6.3 Emergence of the Informativity Principle in comprehension

It remains to be checked that the Informativity Principle must arise in comprehension after it has emerged in production.

#### 6.3.1 Assumptions and notations

We use the same notations, but now started to refer to:

a population No-Inf\* that goes by the Informativity Principle in production but not in comprehension;  
a population Inf\* that goes by the Informativity Principle both in production and in comprehension.

#### (25) Production and comprehension rules for Inf\* and No-Inf\*

a. Production: for Inf\* and No-Inf\* alike, only produce true calls, and furthermore:

if  $A^+$  and  $A^-$  are both true, use  $A^+$

No-Inf: if  $A^+$  and  $A^-$  are both true, use  $A^-$  at rate  $r$ , and  $A^+$  at rate  $(1-r)$

b. Comprehension:

No-Inf\*: interpret  $A^+$  as  $A^+$  and  $A^-$  as  $A^-$  (no strengthening)

Inf\*: interpret  $A^+$  as  $A^+$  and  $A^-$  as ( $A^-$  and not  $A^+$ ) (strengthening)

#### 6.3.2 Payoffs

We continue to disregard issues of call production cost. Furthermore, since both populations under consideration only produce true calls and go by the Informativity Principle in production, we needn't worry about the cost of communication errors. In Section 6.2, there were two benefits of communication to discuss: that produced by  $A^-$  (without strengthening), namely  $b^-$ , and that produced by  $A^+$ , namely  $b^+$ , with  $b^+ > b^-$ . But now we also need to treat separately the case in which  $A^-$  is (correctly) interpreted with strengthening to mean  $A^-$  and not  $A^+$ . We take the corresponding benefit to be  $b^*$ , and we assume that  $b^* > b^-$  (on the other hand, we do not need assumptions about the ordering of  $b^*$  vs.  $b^+$ ).

#### (26) Benefits of communicating different meanings (all are true)

Assumptions:

$b^+ > b^-$

$b^* > b^-$

Information inferred	Actual truth value	Benefit
$A^-$ (without strengthening)	True (successful communication)	$b^-$
$A^-$ and not $A^+$	True (successful communication)	$b^*$
$A^+$	True (successful communication)	$b^+$

We will now motivate the distribution of payoffs summarized in (27).

#### (27) Summary of payoffs

Sender	No-Inf*	Inf*	No-Inf*	Inf*
Receiver	No-Inf*	Inf*	Inf*	No-Inf*
FF ( $A^-$ : false, $A^+$ : false)	0	0	0	0
TT ( $A^-$ : true, $A^+$ : true)	$b^+$	$b^+$	$b^+$	$b^+$
TF ( $A^-$ : true, $A^+$ : false)	$b^-$	$b^*$	$b^*$	$b^-$
Weighted shared payoff	$p_{TT} b^+ + p_{TF} b^-$	$p_{TT} b^+ + p_{TF} b^*$	$p_{TT} b^+ + p_{TF} (b^- + b^*)/2$	

**FF:**  $A^-$ : false,  $A^+$ : false

In all cases, neither call can be produced, and all parties get a payoff of 0.

**TT:**  $A^-$ : true,  $A^+$ : true

In all cases, the sender goes by the Informativity Principle in production, using  $A^+$  and yielding payoff  $b^+$ .

**TF:**  $A^-$ : true,  $A^+$ : false

In all cases, the sender produces  $A^-$ .

- Whenever No-Inf\* is the receiver, it doesn't strengthen the message to  $A^-$  and not  $A^+$ , and yields a payoff of  $b^-$ .
- Whenever Inf\* is the receiver, it strengthens the message to  $A^-$  and not  $A^+$ , and yields a payoff of  $b^*$ .

As before, the weighted shared (symmetric) payoffs are computed in the expected fashion in view of the probability of the different situation types, and of the fact that when No-Inf\* interacts with Inf\*, each has the 'sender' role in 50% of cases.

### 6.3.3 Evolutionary stability of No-Inf\*

As before, No-Inf\* is evolutionarily stable relative to Inf just in case:

Condition (1): payoff of No-Inf\* vs No-Inf\* > payoff of Inf\* vs No-Inf\*, or

Condition (2): these payoffs are equal, but payoff of No-Inf\* vs Inf\* > payoff of Inf\* vs Inf\*.

By inspecting the payoffs in (27), we see that as soon as  $p_{TF} > 0$  and  $b^* > b^-$ , **payoff of Inf\* vs No-Inf\* > No-Inf\* vs No-Inf\***, violating condition (1): No-Inf\* is not evolutionarily stable relative to Inf\*.

### 6.3.4 Total invasion of No-Inf\* by Inf\*

As before, using (2), we ask under what conditions No-Inf\* is totally invaded by Inf\*.

(28) No-Inf\* is fully invaded by Inf\*

iff

(i) payoff of Inf\* vs No-Inf\*  $\geq$  payoff of No-Inf\* vs No-Inf\*,

(ii) payoff of Inf\* vs Inf\*  $\geq$  payoff of No-Inf\* vs Inf\*

and (iii) at least one of these inequalities is strict.

By the boldfaced part of the discussion of Section 6.3.3, we already know that as soon as  $p_{TF} > 0$  and  $b^* > b^-$  condition (i) is satisfied in its strict version (which therefore satisfies (iii)). By inspecting the payoffs in (27) once again, we see that under the same conditions, payoff of Inf\* vs. Inf\* > payoff of No-Inf\* vs Inf\*, satisfying condition (ii) (in its strict version). It is thus clear that No-Inf\* will be fully invaded by Inf\*.

## 6.4 Stepping back

In sum, we have seen that when the Informativity Principle arises in production first, it can only increase the utility of communication, thus guaranteeing that it should emerge. Once the principle has emerged in production, applying it in comprehension will again only increase the utility of communication, guaranteeing emergence again. The reasoning is in fact more general and holds of other pragmatic principles, as we will now see in the case of the Urgency Principle.

## 6.5 A cautionary note

An important disclaimer should be added. For the Informativity Principle to arise in our framework, there must be at least two calls  $A^-$  and  $A^+$  that stand in an asymmetric entailment relation, in the sense that  $A^+$  is strictly more informative than  $A^-$ . We saw that, on the assumption that the lexical meaning of calls  $A^-$  and  $A^+$  is fixed, the Informativity Principle will generate greater utility and will arise in production (and then in comprehension). But if lexical meanings too can evolve, this could show

something different, namely that the  $A^-$  call will gradually come to mean  $A^-$  *and not*  $A^+$ . The overall semantic effect will be the same, but with the consequence that the Informativity Principle has nothing to apply to!

This question was addressed in detail in Steinert-Threlkeld et al. 2021, who model general calls and informativity via probabilistic mappings between states and calls. They ask whether general calls (i.e. calls with a weaker meaning, such as  $A^-$ ) could be stable. With minimal assumptions, their answer is negative: "it is indeed difficult to come up with scenarios in which general calls plus optional Informativity reasoning can be evolutionarily stable". They consider relaxing some of their hypotheses to see if general calls might be stable in the end, and mention two conceivable options. One is to take into account uncertainty on the sender's side, which might make a general call useful in the end; for instance, if the sender is uncertain whether an eagle or a non-eagle was spotted, neither an eagle call nor a non-eagle call will do, but a general call will. A second option is to posit that certain lexical items are costly. Intuitively, 'there is an eagle' corresponds to a natural concept, 'there is an alert' does too, but 'there is a non-eagle-related alert' might not.

Steinert-Threlkeld et al. 2021 show that a lexicon with atomic messages may be optimized in such a way that the Informativity Principle has nothing to apply to. But one can also ask what happens if complex messages are allowed. In this case, Wang 2024 finds that the Informativity Principle may be useful again, although he does so from the perspective of information theory, rather than by considering utility as we do here.

These issues are entirely open, and thus our arguments for the emergence of the Informativity Principle should be taken with a grain of salt, as they presuppose that there can be stable situations in which a call  $A^+$  is strictly more informative than a call  $A^-$ .

## 7 How Competition Principles can Arise II: the Urgency Principle

We turn to an analysis of the emergence of Urgency Principle. Suppose for simplicity that a 2-call sequence involving  $\{A^-, A^+\}$  is produced, and  $A^+$  provides information about the nature or location of a threat but  $A^-$  doesn't. The Urgency Principle states that the order  $A^+ A^-$  should be preferred over the order  $A^- A^+$ :

### (29) Urgency Principle

In a 2-call sequence involving  $\{A^-, A^+\}$ , if  $A^+$  provides information about the nature or location of a threat but  $A^-$  doesn't, the order  $A^+ A^-$  should be used.

In our key motivating example, a putty-nosed sequences made of pyows and hacks can be true in any situation that makes both calls true. Pyow is a general alert call, and hack signals a (serious) non-ground, movement-related event. Events that license both pyow and hack might involve a threat, such as an aerial predator, or no threat, such as a movement of the group of monkeys themselves. If the order is  $\text{pyow}^+ \text{hack}^+$ , however, we cannot be in an eagle-related situation, for if so the hacks would provide information about the nature and location of a threat and should come before the pyows.

As in the case of the Informativity Principle, we argue that the Urgency Principle could evolve in production first and then in comprehension, only adding utility to the communicative exchange at each step. The key idea is that in production, when there is a threat, time is of the essence, and knowing first what or where the threat is might give make all the difference. Once the principle is established in production, it will emerge in comprehension as well.

### 7.1 Emergence of the Urgency Principle in production

#### 7.1.1 Assumptions and notations

We consider populations with two calls  $A^+$  and  $A^-$ , where  $A^-$  never provides information about the nature or location of a threat, while  $A^+$  does in some situations but not in others. We restrict attention to the 2-call sequences  $A^+ A^-$  and  $A^- A^+$ . Semantically, these two sequences provide the same information, but in situations in which  $A^+$  provides information about the nature or location of a threat,  $A^+ A^-$  obeys the Urgency Principle while  $A^- A^+$  violates it. In such situations, a population Urge goes by the Informativity



Principle in production, and only produces  $A^+A^-$ ; a population No-Urge doesn't go by the Urgency Principle in production, and produces  $A^-A^+$  at rate  $r$  and  $A^+A^-$  at rate  $(1-r)$ .

(30) **Production and comprehension rules for Urge and No-Urge**

a. Production: only produce sequences that are true, and furthermore:

Urge: if  $A^+$  provides information about the nature or location of a threat, use  $A^+A^-$ .

No-Urge: if  $A^+$  provides information about the nature or location of a threat, use  $A^-A^+$  at rate  $r$ , and  $A^+A^-$  at rate  $(1-r)$ .

*Note:* If  $A^+$  does not provide information about the nature/location of a threat, Urge and No-Urge alike might produce  $A^+A^-$  vs  $A^-A^+$  at certain rates, but this won't matter as they will yield the same payoff (as ordering only matters when  $A^+$  provides information about the nature/location of a threat).

b. Comprehension: for Urge and No-Urge alike,

interpret  $A^+A^-$  and  $A^+$  and  $A^-A^+$  in the same way (as the conjunction of  $A^+$  and  $A^-$ ).

Restricting attention to situations that make both  $A^+$  and  $A^-$  true, we write as  $p_\theta$  the probability weight of the situations in which  $A^+$  provides information about the nature or location of a threat (taking  $\theta$  as mnemonic for 'threat'), and  $(1-p_\theta)$  the probability weight of the situations in which  $A^+$  doesn't provide information about the nature or location of a threat, as is summarized in (31).

(31) **Probability weights within the set of situations that make both  $A^+$  and  $A^-$  true**

$p_\theta$  = probability weight of the situations in which  $A^+$  provides information about the nature or location of a threat

$1-p_\theta$  = probability weight of the situations in which  $A^+$  does not provide information about the nature or location of a threat

### 7.1.2 Payoffs

We continue to disregard issues of call production cost. As before, calls are only produced when they are true. Furthermore, since both populations interpret calls correctly, and without pragmatic strengthening, we needn't worry about communication errors. All payoffs are symmetric, i.e. obtained by sender and receiver alike.

We assume that there are three potential benefits of communication. If  $A^+A^-$  or  $A^-A^+$  are produced in a situation in which  $A^+$  does not provide information about the nature or location of a threat, a payoff  $a$  is obtained (order doesn't matter, hence it's the same payoff for both). If  $A^+A^-$  or  $A^-A^+$  are produced in a situation in which  $A^+$  provide information about the nature or location of a threat,  $A^-A^+$  brings payoff  $b^-$  and  $A^+A^-$  brings payoff  $b^+$ , with  $b^+ > b^-$ .

(32) **Benefits of communicating different meanings (all are true)**

Assumptions:

$b^+ > b^-$

Situation type: $\theta$	Information inferred	Benefit
<b>T: <math>A^+</math> does not provide information about the nature/location of a threat</b>	$A^-A^+$	$a$
	$A^+A^-$	$a$
<b>F: <math>A^+</math> provides information about the nature/location of a threat</b>	$A^-A^+$	$b^-$
	$A^+A^-$	$b^+$

We will now motivate the distribution of payoffs summarized in (33).

(33) **Summary of payoffs**

Sender	No-Urge	Urge	No-Urge	Urge
Receiver	No-Urge	Urge	Urge	No-Urge

<b>No <math>\theta</math>: <math>A^+</math> is not about a threat</b>	$a$	$a$	$a$	$a$
<b><math>\theta</math>: <math>A^+</math> is about a threat</b>	$r b^- + (1-r)b^+$	$b^+$	$r b^- + (1-r)b^+$	$b^+$
<b>Weighted shared payoff</b>	$(1-p_\theta)a$ $+ p_\theta(r b^- + (1-r)b^+)$	$(1-p_\theta)a$ $+ p_\theta b^+$	$(1-p_\theta)a$ $+ p_\theta[r b^- + (2-r)b^+]/2$	

**No  $\theta$ :  $A^+$  does not provide information about the nature/location of a threat:** in all cases, the sender produces  $A^+A^-$  or  $A^-A^+$ , and the choice doesn't matter because both yield the payoff  $a$ .

**$\theta$ :  $A^+$  provides information about the nature/location of a threat**

- When No-Urge is the sender, it produces  $A^-A^+$  at rate  $r$  and  $A^+A^-$  at rate  $(1-r)$ , yielding a payoff of  $r b^- + (1-r)b^+$ .
- When Urge is the sender, it always produces  $A^+A^-$ , yielding benefit  $b^+$ .

The weighted shared (symmetric) payoffs are then computed as expected in view of the probabilities of the two situation types, namely No  $\theta$  and  $\theta$ .

### 7.1.3 Evolutionary stability of No-Urge

As before, No-Urge is evolutionarily stable relative to Urge just in case:

Condition (1): payoff of No-Urge vs No-Urge  $>$  payoff of Urge vs No-Urge, or

Condition (2): these payoffs are equal, but payoff of No-Urge vs Urge  $>$  payoff of Urge vs Urge.

By inspecting the payoffs in (33), we see that in situations No  $\theta$ , payoff of No-Urge vs No-Urge = payoff of Urge vs No-Urge, while in situations  $\theta$ , if  $r > 0$ ,  $b^+ > b^-$ , payoff of No-Urge vs No-Urge  $<$  payoff of Urge vs No-Urge. In sum, as soon as  $p_\theta > 0$ ,  $r > 0$ ,  $b^+ > b^-$ , **payoff of No-Urge vs No-Urge  $<$  payoff of Urge vs No-Urge**, and thus No-Urge is not evolutionarily stable relative to Urge.

### 7.1.4 Total invasion of No-Urge by Urge

Using (2), we can ask under what conditions No-Urge is totally invaded by Urge.

- (34) No-Urge is fully invaded by Urge  
iff  
(i) payoff of Urge vs No-Urge  $\geq$  payoff of No-Urge vs No-Urge,  
(ii) payoff of Urge vs Urge  $\geq$  payoff of No-Urge vs Urge  
and (iii) at least one of these inequalities is strict.

By the boldfaced part of the discussion of Section 7.1.3, we see that as soon as  $p_\theta > 0$ ,  $r > 0$ ,  $b^+ > b^-$ , condition (i) is satisfied in its strict version, thus satisfying (iii) as well. By inspecting again the summary of payoffs in (33), we see that under the same conditions, condition (ii) is satisfied in its strict version as well. In sum, No-Urge will be totally invaded by Urge.

## 7.2 Emergence of the Urgency Principle in comprehension

It remains to be checked that the Urgency Principle must arise in comprehension after it has arisen in production.

### 7.2.1 Assumptions and notations

We use the same notations and general assumptions (disregarding costs), but now started to refer to: a population No-Urge\* that goes by the Urgency Principle in production but not comprehension; a population Urge\* that goes by the Urgency Principle both in production and in comprehension.

- (35) **Production and comprehension rules for Urge\* and No-Urge\***  
a. Production: for Urge\* and No-Urge\* alike, only produce sequences that are true, and furthermore:

–if  $A^+$  provides information about the nature or location of a threat, use  $A^+A^-$ ;  
 –if  $A^+$  doesn't provide information about the nature or location of a threat, use  $A^-A^+$  at rate  $r^+$ , and  $A^+A^-$  at rate  $(1-r^+)$  (menomonicallly, we can think of  $r^+$  as the rate at which the more useful order is picked, as  $A^-A^+$  conveys information about the fact that the situation is *not* threat-related, for otherwise  $A^+A^-$  would have been used).

*Note:* The rate  $r$  could have been included in Section 7.1 but it wouldn't have made a difference, as in 'no threat' situations  $A^-A^+$  and  $A^+A^-$  were interpreted in the same way; this is no longer the case here due to the Urgency Principle in comprehension, and thus this rate will matter.

b. Comprehension:

No-Urge\*: interpret  $A^+A^-$  and  $A^+$  and  $A^-A^+$  in the same way (as the conjunction of  $A^+$  and  $A^-$ ).

Urge\*: interpret  $A^+A^-$  as:  $A^+$  and  $A^-$ , and  $A^-A^+$  as:  $A^+$  and  $A^-$ , and the situation doesn't involve a threat.

As summarized in (31), we continue to use  $p_\theta$  for probability weight of the situations in which  $A^+$  provides information about the nature or location of a threat, and  $(1-p_\theta)$  the probability weight of the situations in which  $A^+$  doesn't provide information about the nature or location of a threat,

### 7.2.2 Payoffs

We assume that there are three potential benefits of communication (but they are not the same as in section 7.1). When  $A^+$  provides information about the nature/location of a threat ( $= \theta$  situations), all senders produce  $A^+A^-$  because they go by the Urgency Principle in production, and this yields a benefit  $b^+$ . When  $A^+$  does not provide information about the nature/location of a threat ( $= \text{No } \theta$  situations), all senders produce  $A^-A^+$  (at rate  $r^+$ ) and  $A^+A^-$  (at rate  $(1-r^+)$ ). In the first case, Urge\* but not No-Urge\* correctly strengthens the information to:  ***$A^-$  and  $A^+$  and the situation doesn't involve a threat.*** Without the boldfaced strengthening, the benefit of communication is  $a^-$ , with the boldfaced strengthening, it is  $a^+$ , which we assume to be greater.

#### (36) Benefits of communicating different meanings (all are true)

Assumptions:

$$a^+ > a^-$$

Information inferred	Situation type	Benefit
$A^-$ and $A^+$ (no strengthening)	<b>No <math>\theta</math>:</b> $A^+$ does not provide information about the nature/location of a threat	$a^-$
$A^-$ and $A^+$ and the situation doesn't involve a threat		$a^+$
$A^+A^-$	<b><math>\theta</math>:</b> $A^+$ provides information about the nature/location of a threat	$b^+$

We will now motivate the distribution of payoffs summarized in (37).

#### (37) Summary of payoffs

Sender	No-Urge*	Urge*	No-Urge*	Urge*
Receiver	No-Urge*	Urge*	Urge*	No-Urge*
<b>No <math>\theta</math>: <math>A^+</math> not about a threat</b>	$a^-$	$r^+a^+ + (1-r^+)a^-$	$r^+a^+ + (1-r^+)a^-$	$a^-$
<b><math>\theta</math>: <math>A^+</math> about a threat</b>	$b^+$	$b^+$	$b^+$	$b^+$
<b>Weighted shared payoff</b>	$(1-p_\theta)a^- + p_\theta b^+$	$(1-p_\theta)[r^+a^+ + (1-r^+)a^-] + p_\theta b^+$	$(1-p_\theta)[r^+a^+ + (2-r^+)a^-]/2 + p_\theta b^+$	

**No  $\theta$ :  $A^+$  does not provide information about the nature/location of a threat**

All senders produce  $A^-A^+$  at rate  $r^+$ , and  $A^+A^-$  at rate  $(1-r^+)$ .

–When No-Urge\* is the receiver, it infers in either case the unstrengthened information ( $A^-$  and  $A^+$ ), hence a benefit of  $a^-$ .

–When Urge\* is the receiver, it infers from  $A^-A^+$  (produced at rate  $r^+$ ) but not from  $A^+A^-$  (produced at rate  $(1-r^+)$ ) that the situation doesn't involve a threat, hence an overall benefit of  $r^+a^+ + (1-r^+)a^-$ .

### **$\theta$ : $A^+$ provides information about the nature/location of a threat**

All senders produce the sequence  $A^+A^-$ , yielding a benefit of  $b^+$ .

(Here too, the weighted shared payoffs are computed as expected.)

#### *7.2.3 Evolutionary stability of No-Urge\**

As before, No-Urge\* is evolutionarily stable relative to Urge just in case:

Condition (1): payoff of No-Urge\* vs No-Urge\* > payoff of Urge\* vs No-Urge\*, or

Condition (2): these payoffs are equal, but payoff of No-Urge\* vs Urge\* > payoff of Urge\* vs Urge\*.

By inspecting the payoffs in (33) (with special attention to the crucial row No  $\theta$ ),  $(1-p_\theta) \neq 0$ , i.e.  $p_\theta \neq 1$ , if  $r^+ \neq 0$  and  $a^+ > a^-$ , overall **payoff of No-Urge\* vs No-Urge\* < payoff of Urge\* vs No-Urge\***. Thus No-Urge\* is not evolutionarily stable relative to Urge\*.

#### *7.2.4 Total invasion of No-Urge\* by Urge\**

As before, using (2), we ask under what conditions No-Urge\* is totally invaded by Urge\*.

- (38) No-Urge\* is fully invaded by Urge\*  
iff  
(i) payoff of Urge\* vs No-Urge\*  $\geq$  payoff of No-Urge\* vs No-Urge\*,  
(ii) payoff of Urge\* vs Urge\*  $\geq$  payoff of No-Urge\* vs Urge\*  
and (iii) at least one of these inequalities is strict.

By the boldfaced part of the discussion of Section 7.2.3, we see that when  $p_\theta \neq 1$ , if  $r^+ \neq 0$  and  $a^+ > a^-$ , condition (i) is satisfied in its strict version, hence also satisfying (iii). By inspecting the summary of payoffs in (37), we see that under the same conditions, condition (ii) is satisfied (in fact, in its strict version). In sum, No-Urge\* will be totally invaded by Urge\*.

We note that the worry we raised about the stability of general (i.e. weak) calls in Section 6 does not arise in the present case because the pragmatic enrichment comes from entire sequences rather than individual calls, and hence it could not be emulated by changing the meaning of individual calls.

### **7.3 Stepping back: evolution of pragmatic principles**

While our discussion of elementary signals and non-trivial combinations highlighted the constrained nature of scenarios that could lead to their emergence, things are very different with competition principles: both in the case of the Informativity Principle and of the Urgency Principle, there is a clear evolutionary path for them if they arise in production first and in comprehension second (as discussed in Appendix II in the case of the Informativity Principle, if different evolutionary sequences are considered, quantitative conditions weighing the costs and benefits of the principles come into play). The main worry at this point is that the Informativity Principle might have nothing to operate on if general (i.e. weak) calls are not evolutionarily stable when lexical meanings are allowed to change, but in line with Steinert-Threlkeld et al. 2021, we sketched plausible constraints that might prevent the evolutionary collapse of general calls, and thus guarantee that the Informativity Principle would have something to apply to.

## 8 Conclusion

Building on standard conditions of evolutionary stability in biology (but within a purely cooperative framework), we have proposed very simple models that put constraints on the emergence of elementary signals, of non-trivial combinations and of principles of competition among signals. First, we found that for a signal to emerge in a population that doesn't initially have it, signal-producing mutants must get a benefit from communication with residents that outweighs signal production cost. Since residents haven't evolved to understand the signal, this means that the pre-comprehension benefit of the signal must be high and/or its production cost must be low. Both cases can be related to the old intuition that signals originate in cues (or in sensory manipulation), but there are still two distinct cases, which we illustrated with several examples.

Second, we found that for a non-trivial combination with meaning  $A*B$  to arise in a population that only has trivial compositionality, and hence only the meaning  $A\&B$ , a basic difficulty must be overcome: due to the homophony between  $A*B$  and  $A\&B$  (both produced as  $AB$ ), mutants will create misunderstandings and thus utility loss. This must be outweighed by cases in which  $A*B$  and  $A\&B$  are both true, and  $A*B$  produces greater utility. An evolutionary scenario along precisely these lines was recently developed to account for ABC-D sequences in Japanese tits. By contrast, some idiom-based analyses of putty-nosed monkey pyow-hack sequences seem to be ruled out by this condition because the purported idiom is never true in the same situation as the combination of pyow and hack as separate utterances. One may be tempted to think that an idiom-based theory is a good initial analysis of call combinations because it does not posit complex combinatorial abilities in animals. But our results show that the opposite is sometimes true: some idiom-based analyses have low prior plausibility for evolutionary reasons, and they must thus be supported by independent evidence to remain viable.

Finally, we saw that when it comes to the emergence of pragmatic principles, things are arguably simpler: for the Informativity Principle and the Urgency Principle alike, there is a clear path of evolution as long as the principles arise in production first and in comprehension second. The reason is that such an evolutionary sequence avoids the risk that an emerging principle would create misunderstandings.

While our framework is maximally simple, we hope that it can highlight the fruitfulness of the interplay between easily interpretable analytical conditions and rich empirical findings about current animal communication as well as reconstructed ancestral meanings. This should help extend recent animal linguistics with a precise and rich evolutionary component.

Corresponding to these results, there are two directions to extend this line of research in the future. One is to develop more fine-grained models by considering less narrow assumptions, for instance by investigating non-cooperative situations, by allowing for more than two strategies to interact simultaneously, by investigating more diverse parameters, and by taking into account more steps in evolutionary processes (something that would be necessary for a better analysis of ritualized cues, for instance). The other direction is to consider many more case studies, and in particular to build on recent efforts to reconstruct ancestral signals and meanings over millions of years to test the predictions of competing evolutionary scenarios.

### Appendix I. Evolutionary Stability and Invasion

We start from the replicator dynamic equation (Taylor and Jonker 1978, see also Hofbauer and Sigmund 1988, 1998). This equation essentially models the fact that the sub-population following strategy 1 will grow ( $dr_1/dt$ ) proportionally to the current proportion of that strategy in the population ( $r_1$ ) and to the added utility that picking this strategy provides. For an infinite population with a proportion  $r_1$  of individuals following strategy  $S_1$ , and  $r_2 = 1 - r_1$  following strategy  $S_2$ , we assume that the dynamic of the population is governed by the law:

$$dr_1/dt = r_1 [ U_1(r_1) - U_{\text{population}} ]$$

where  $U_i(r_i)$  is the expected utility for an individual  $i$  following strategy  $S_i$  in a population with  $r_i$  individuals  $i$ , and  $U_{\text{population}}$  is the expected utility for any individual in the population,  $U_{\text{population}} = r_1 U_1(r_1) + r_2 U_2(r_2)$ . A 1-individual encounters on average  $r_1$  1-individuals and  $r_2$  2-individuals, hence  $U_1(r_1) = r_1 U_{1-1} + r_2 U_{1-2}$ , with  $U_{I-J}$  the utility of  $I$  in  $I-J$  encounters, as before. Similarly,  $U_2(r_2) = r_2 U_{2-2} + r_1 U_{2-1}$ . Hence,

$$\begin{aligned} dr_1/dt &= r_1 [ U_1(r_1) - (r_1 U_1(r_1) + r_2 U_2(r_2)) ] \\ &= r_1 [ (1 - r_1) U_1(r_1) - (1 - r_1) U_2(r_2) ] \\ &= r_1 (1 - r_1) [ U_1(r_1) - U_2(r_2) ] \\ &= r_1 (1 - r_1) [ (r_1 U_{1-1} + r_2 U_{1-2}) - (r_2 U_{2-2} + r_1 U_{2-1}) ] \\ &= r_1 (1 - r_1) [ r_1 (U_{1-1} - U_{2-1}) + r_2 (U_{1-2} - U_{2-2}) ] \end{aligned}$$

We note that this derivative is null when the population is homogenous ( $r_1 = 0$  or  $r_1 = 1$ ), and otherwise of the sign of:

$$S(r_1) = r_1 (U_{1-1} - U_{2-1}) + (1 - r_1) (U_{1-2} - U_{2-2})$$

**Evolutionary stability.** The subpopulation of 1-individuals is evolutionarily stable if when it coexists with few 2-individuals, the subpopulation of 2-individuals will be eliminated. Formally, this translates into the fact that the derivative  $dr_1/dt$  should be and remain positive when  $r_1$  is close to 0 (so that the population of 1-individuals will tend to increase despite the occasional 2-individuals). The sign of the derivative  $S(r_1)$  will remain positive if and only if:

- its limit when  $r_1$  goes to 0 is positive, that is  $(U_{1-2} - U_{2-2}) > 0$
- or its limit is zero, but it is approached from above, that is  $(U_{1-2} - U_{2-2}) = 0$  and  $(U_{1-1} - U_{2-1}) > 0$ .

These are the Maynard Smith conditions of evolutionary stability, as reported in (1) in the main paper.

**Total invasion.** The 1-individuals will totally invade the 2-individuals if, for all heterogeneous populations ( $0 < r_1 < 1$ ), the dynamics pushes the population towards having only 1-individuals. We note that the sign of the derivative is the barycenter of  $(U_{1-1} - U_{2-1})$  and  $(U_{1-2} - U_{2-2})$  with weights  $r_1$  and  $(1 - r_1)$  respectively. The sign of this barycenter is governed by the signs of the terms. Most importantly, if both terms are positive or null, and one is strictly positive, then the derivative is always strictly positive and the population is pushed towards  $r_1 = 1$ . **This corresponds exactly to the necessary and sufficient conditions given in (2) for total invasion in the main text:  $(U_{1-2} - U_{2-2}) \geq 0$  and  $(U_{1-1} - U_{2-1}) \geq 0$  and one of these inequalities is strict.** There is indeed no other configuration of total invasion. For instance, if both terms of the barycenter are of different signs (or if both are null) there will be a value  $0 < r_1^* < 1$  for which the derivative is null. This is sufficient to show that there will not be a total invasion. And one can even show that then either both strategies are evolutionarily stable, or none is. (The final case where both terms are negative or null can be treated by symmetry). A more classic presentation of this result can be found, for instance, in Weibull 1995 p. 84; see Hofbauer and Sigmund 1988 for a more general overview.

## *Appendix II. Comparison with Scott-Phillips et al. 2012*

Scott-Phillips et al. 2012 offer an interesting analysis of signal emergence. They start from the observation that "a state of non-interaction is evolutionarily stable, and so communication will not necessarily emerge even when it is in both parties' interest"; in this respect, they address the same initial problem that we do in Section 2. Here we highlight important differences between the two approaches.

First, Scott-Phillips et al. posit that signal production and signal comprehension are separate evolutionary steps. By contrast, with a view towards maximal simplicity, we assumed that residents neither produce nor understand signals, while mutants do both.

Second, Scott-Phillips et al. distinguish among three varieties of information exchange, taking only the most demanding kind to involve 'signals' in their strict sense. Their key idea is summarized below, and illustrated by the table in (39):

"We define a signal as any action or structure that causes a reaction in another organism, where it is the function of both action and reaction to play these particular roles in the interaction (...). If these conditions are satisfied, then the action is a signal; the reaction is a response; and the overall interaction is communicative. If only the reaction is functional in this way, then the action is a cue; and if only the action is functional in this way, then it is coercive." (Scott-Phillips et al. 2012)

(39) **Definitions of signals, responses, cues and coercion** (from Scott-Phillips et al. 2012, Table 1)

	function of action to affect the receiver?	function of reaction to be affected by action?
cue	no	yes
coercion	yes	no
communication ('signals')	yes	yes

Thus for Scott-Phillips et al. (as well as for the literature), a cue arises in production for non-communicative reasons: "for example, the use of urine to mark territory may have begun as a marker of fear, produced by animals at the periphery of territory in which they felt safe, which other animals used as the cue of the focal individual's presence (Maynard Smith and Harper 2003)." A further example might be the waggle dance of bees, if indeed it initially arose for non-communicative reasons, as we discussed in Section 3.2.2. In a second step, receivers evolve to understand this cue. Since we do not distinguish between these two steps, there is no direct counterpart of this in our analysis. But we can think of cues as cases in which overall message production cost is non-positive ( $c \leq 0$ ): the message doesn't have production costs, but it does have non-communicative benefits for the sender. If  $c < 0$ , one might expect the behavior to already be present in the resident population, but if  $c = 0$ , this need not be so. Our condition  $b^- \geq c$  will be trivially satisfied (where  $b^-$  is, as in (5) in the main text, the 'pre-comprehension' benefit obtained by a mutant signaler interacting with a resident). In the end, the resident population will be invaded by mutants that produce and understand the signal, obtaining in the interaction the full benefit rather than just  $b^-$ . Furthermore, on the assumption that  $c = 0$ , it will be because the message transmits information that it will come to be produced systematically in the end.

For Scott-Phillips et al., coercion arises when message production takes advantage of a pre-existing comprehension mechanism. For instance, "many mating displays may have begun as scenarios in which a preference for objects of a certain colour allowed the behaviour of potential receivers to be manipulated by others". Our analysis of roughness and nonlinearities in Section 3.3 is an example as well; but so are *all* our examples in which message cost is positive ( $c > 0$ ) and invasion occurs because receivers understand enough of the message to satisfy  $b^- \geq c$ . This also applies to ritualized woodpecker drumming and ape gestures as we conceive of them: when studied at the evolutionary stage in which they have lost any efficacy for foraging or for action, their cost is clearly strictly positive, and thus the

condition  $b^* \geq c$  is crucial to explain their emergence in a population that didn't originally use these signals.<sup>30</sup>

By distinguishing between the evolution of action (message production) and reaction (message comprehension), Scott-Phillips et al. 2012 can give a definition of communication that involves *both* evolutionary steps. Since we do not distinguish between these steps in our analysis, we cannot provide a similar definition. Still, the sign of the  $c$  parameter could offer another way to carve the space of possibilities in what might correspond to their second step.

---

<sup>30</sup> For Scott-Phillips et al. 2012, instances of ritualization fall under 'cues', but this is because they consider ritualized behaviors that are still efficacious and thus could satisfy  $c \leq 0$ . This does not apply to non-efficacious ritualized signals.



### Appendix III. Other Evolutionary Paths for the Informativity Principle

In Section 6 of the main text, we proposed an evolutionary path for pragmatic principles that starts with production; once the principle is established in production, it can emerge in comprehension. But could there be other orders of emergence? Here we restrict attention to the Informativity Principle, which has given rise to more empirical applications than the Urgency Principle. We consider two scenarios, one in which the Informativity Principle emerges in comprehension first and in production second; and one in which mutants simultaneously apply it in production and in comprehension. We show that, unlike the 'production first' scenario we discussed in the main text, these alternatives paths put severe constraints on the parameters of the model.

#### A. Emergence of the Informativity Principle in comprehension first and in production second

We use the same hypotheses and notations as in Section 6, except for modifications that are mentioned below.

##### A1. Conditions on which the Informativity Principle can arise in comprehension

###### □ Assumptions

We consider a (resident) population No-Inf that applies the Informativity Principle neither in production nor in comprehension, and a (mutant) population Inf that applies the principle in comprehension but not in production, as summarized in (40).

- (40) **Production and comprehension rules for Inf and No-Inf**
- a. Production: for Inf and No-Inf alike, only produce true calls, and if  $A^+$  and  $A^-$  are both true, use  $A^-$  at rate  $r$ , and  $A^+$  at rate  $(1-r)$
  - b. Comprehension:
    - No-Inf: interpret  $A^+$  as  $A^+$  and  $A^-$  as  $A^-$  (no strengthening)
    - Inf: interpret  $A^+$  as  $A^+$  and  $A^-$  as  $A^-$  and not  $A^+$  (strengthening).

We assume that there are four possible payoffs, as outlined in (41): When communicated,  $A^+$  will always be true and will give rise to a payoff of  $b^+ > b^-$ . When communicated,  $A^-$  without strengthening will always be true and will give rise to benefit  $b^-$ . With strengthening to  $A^-$  and not  $A^+$ , it will sometimes be true, giving rise to a payoff  $b^* > b^-$ ; and it will sometimes be false, yielding a (likely negative) payoff  $-e$  owing to the communication error.

##### (41) Benefits of communicating different meanings

Assumptions:

- $b^+ > b^-$
- $b^* > b^-$
- $e > 0$

Information inferred	Actual truth value	Benefit
$A^-$ (without strengthening)	True (successful communication)	$b^-$
<b><math>A^-</math> and not <math>A^+</math></b>	True (successful communication)	$b^*$
<b><math>A^-</math> and not <math>A^+</math></b>	False (unsuccessful communication)	$-e$
$A^+$	True (successful communication)	$b^+$

###### □ Payoffs

We will now motivate the distribution of payoffs summarized in (42).

##### (42) Summary of payoffs

Sender	No-Inf	Inf	No-Inf	Inf
--------	--------	-----	--------	-----

Receiver	No-Inf	Inf	Inf	No-Inf
<b>FF</b> (A <sup>-</sup> : false, A <sup>+</sup> : false)	0	0	0	0
<b>TT</b> (A <sup>-</sup> : true, A <sup>+</sup> : true)	$r b^- + (1-r)b^+$	$-re + (1-r)b^+$	$-re + (1-r)b^+$	$r b^- + (1-r)b^+$
<b>TF</b> (A <sup>-</sup> : true, A <sup>+</sup> : false)	$b^-$	$b^*$	$b^*$	$b^-$
<b>Weighted shared payoff</b>	$p_{TT}[r b^- + (1-r)b^+] + p_{TF} b^-$	$p_{TT}[-re + (1-r)b^+] + p_{TF} b^*$	$p_{TT}[r(b^- - e) + 2(1-r)b^+]/2 + p_{TF}[b^* + b^-]/2$	

The two strategies do not differ in their production, and thus only the type of the Receiver matters. Hence, the first and fourth columns are the same, and the second and third columns are the same. (As a result, in the last row, the last column is the average of the previous two.)

**FF:** A<sup>-</sup>: false, A<sup>+</sup>: false

In all cases, neither call can be produced, and all parties get a payoff of 0.

**TT:** A<sup>-</sup>: true, A<sup>+</sup>: true

In all cases, the sender fails to apply the Informativity Principle in production, and thus produces A<sup>-</sup> at rate  $r$  and A<sup>+</sup> at rate  $(1-r)$ .

- When the receiver is No-Inf, no strengthening is applied in comprehension and the payoff is  $r b^- + (1-r)b^+$ .
- When the receiver is Inf, strengthening is incorrectly applied when the signal is A<sup>-</sup> (strengthened to A<sup>+</sup> and not A<sup>+</sup>), hence a negative payoff of  $-e$  in this case. Overall, the payoff is  $-re + (1-r)b^+$ .

**TF:** A<sup>-</sup>: true, A<sup>+</sup>: false

In all cases, the sender produces A<sup>-</sup>.

- When No-Inf is the receiver, it doesn't strengthen the message and yields a payoff of  $b^-$ .
- When Inf is the receiver, it correctly strengthens the message to A<sup>-</sup> and not A<sup>+</sup>, yielding a payoff of  $b^*$ .

#### □ Evolutionary stability of No-Inf

As before, No-Inf is evolutionarily stable relative to Inf just in case:

Condition (1): payoff of No-Inf vs No-Inf > payoff of Inf vs No-Inf, or

Condition (2): these payoffs are equal, but payoff of No-Inf vs Inf > payoff of Inf vs Inf.

Condition (1) can be successively rewritten as follows:

$$\begin{aligned}
 (43) \quad & p_{TT}[r b^- + (1-r)b^+] + p_{TF} b^- > p_{TT}[r(b^- - e) + 2(1-r)b^+]/2 + p_{TF}[b^* + b^-]/2 \\
 & p_{TT} 2[r b^- + (1-r)b^+] + p_{TF} 2b^- > p_{TT}[r(b^- - e) + 2(1-r)b^+] + p_{TF}[b^* + b^-] \\
 & p_{TT} [2r b^- + 2(1-r)b^+ - r b^- + r e - 2(1-r)b^+] + p_{TF} [2b^- - b^* - b^-] > 0 \\
 & p_{TT} [r b^- + r e] + p_{TF} [b^- - b^*] > 0 \\
 & p_{TT} r [b^- + e] > p_{TF} [b^* - b^-]
 \end{aligned}$$

This can be interpreted as follows. In all cases, interaction is with No-Inf. Focusing on the receiver role, No-Inf has an advantage over Inf whenever one is in a TT situation and A<sup>-</sup> is produced, which happens at rate  $r$ ; the magnitude of the advantage is the difference between what No-Inf gets, namely  $b^-$ , and what Inf gets, namely  $e$ , hence a differential of  $b^- + e$ . Still focusing on the receiver role, Inf has an advantage over No-Inf whenever one is in a TF situation, for in this case A<sup>-</sup> is produced, Inf correctly strengthens it to A<sup>-</sup> and not A<sup>+</sup>, yielding payoff of  $b^*$ , whereas No-Inf, which doesn't perform the strengthening, only yields  $b^-$ . The magnitude of the advantage is thus  $b^* - b^-$ . The last line of (43) says that the (weighted) magnitude of No-Inf's advantage in TT situations more than outweighs the (weighted) magnitude of Inf's advantage in TF situations. Some values of the parameters will ensure

that No-Inf is stable and others will not. It follows that invasion of No-Inf by Inf can certainly not be guaranteed for all reasonable values of the parameters.

In sum, applying the Informativity Principle in comprehension first has mixed results, because it leads to useful strengthenings in some cases, and to communication errors in others.

We note that, in this particular case, the conditions for total invasion are exactly the same. The reason comes from the fact that production strategies are the same and that utilities are shared between sender and receiver. As a result, the utility obtained in mixed encounters are the same for the two types of players ( $U_{I-J} = U_{J-I}$ ), and it is the average of the utilities obtained in homogeneous encounters ( $U_{J-I} = (U_{I-I} + U_{J-J})/2$ ). Hence, it lies in between the utilities obtained homogeneous encounters. And so  $U_{I-I} \geq U_{J-I}$  and  $U_{I-J} \geq U_{J-J}$  and one inequality is strict, is equivalent to  $U_{I-I} > U_{J-I}$ .

## A2. Conditions on which the Informativity Principle can arise in production once it has emerged in comprehension

By contrast, we will now see that on the (non-trivial) assumption that the Informativity Principle has emerged in comprehension, it will emerge in production in a second step.

### □ Assumptions

We consider a (resident) population No-Inf\* that applies the Informativity Principle in comprehension but not in production, and a (mutant) population Inf\* that applies the principle both in production and in comprehension, as summarized in (40).

#### (44) Production and comprehension rules for Inf\* and No-Inf\*

- a. Production: only produce true calls, and in addition, if  $A^+$  and  $A^-$  are both true,  
No-Inf\*: produce  $A^-$  at rate  $r$ , and  $A^+$  at rate  $(1-r)$ ;  
Inf\*: only produce  $A^+$
- b. Comprehension: for Inf and No-Inf alike, interpret  
 $A^+$  as  $A^+$  and  $A^-$  as  $A^-$  and not  $A^+$  (strengthening).

We assume that the benefits of communicating different meanings are the same as in (41).

### □ Payoffs

We will now motivate the distribution of payoffs summarized in (45).

#### (45) Summary of payoffs

Sender	No-Inf*	Inf*	No-Inf*	Inf*
Receiver	No-Inf*	Inf*	Inf*	No-Inf*
<b>FF</b> ( $A^-$ : false, $A^+$ : false)	0	0	0	0
<b>TT</b> ( $A^-$ : true, $A^+$ : true)	$-re + (1-r)b^+$	$b^+$	$-re + (1-r)b^+$	$b^+$
<b>TF</b> ( $A^-$ : true, $A^+$ : false)	$b^*$	$b^*$	$b^*$	$b^*$
<b>Weighted shared payoff</b>	$p_{TT} [-re + (1-r)b^+] + p_{TF} b^*$	$p_{TT} b^+ + p_{TF} b^*$	$p_{TT} [-re + (2-r)b^+]/2 + p_{TF} b^*$	

**FF:**  $A^-$ : false,  $A^+$ : false

In all cases, neither call can be produced, and all parties get a payoff of 0.

**TT:**  $A^-$ : true,  $A^+$ : true

- When No-Inf\* is the sender, it produces  $A^-$  at rate  $r$ , and  $A^+$  at rate  $(1-r)$ . Whoever the receiver is, it incorrectly strengthens  $A^-$  to  $A^-$  and not  $A^+$ , yielding a payoff of  $-e$ , whereas  $A^+$  yields a payoff of  $b^+$ . Overall, the payoff is  $-re + (1-r)b^+$ .

- When Inf\* is the sender, it always produces  $A^+$ , yielding a payoff of  $b^+$ .

**TF:**  $A^-$ : true,  $A^+$ : false

The sender always produces  $A^-$ , and the receiver always applies the Informativity Principle in comprehension, which yields a payoff of  $b^*$ .

#### □ *Evolutionary stability of No-Inf\**

As before, No-Inf\* is evolutionarily stable relative to Inf\* just in case:

Condition (1): payoff of No-Inf\* vs No-Inf\* > payoff of Inf\* vs No-Inf\*, or

Condition (2): these payoffs are equal, but payoff of No-Inf\* vs Inf\* > payoff of Inf\* vs Inf\*.

By inspecting the table in (45), it can be checked that rows FF and TF are the same for all encounters and can thus be ignored. Condition (1), payoff of No-Inf\* vs No-Inf\* > payoff of Inf\* vs No-Inf\*, thus reduces to  $p_{TT}(-re + (1-r)b^+ - b^+) > 0$ , i.e.  $p_{TT}r(-e - b^+) > 0$ . Assuming  $r > 0$ ,  $e > 0$  and  $b^+ > 0$ , this will never hold. And clearly condition (2) can never hold either. Thus No-Inf\* is not evolutionarily stable relative to Inf\*.

#### □ *Total Invasion of No-Inf\* by Inf\**

For similar reasons as above (shared payoff between sender and receiver, and a similar strategy in comprehension), the conditions for total invasion of Inf\* are the same as the conditions of non-evolutionary stability of No-Inf\*. Hence, No-Inf\* will be invaded by Inf\*.

#### □ *Stepping back*

The scenario on which the Informativity Principle arises first in comprehension and second in production may or may not get off the ground because the conditions for the first step are non-trivial. However once the principle has emerged in comprehension, there is a clear path to its emergence in production.

### **B. Simultaneous emergence of the Informativity Principle in production and in comprehension**

We turn to the case in which a resident population No-Inf applies the Informativity Principle neither in production nor in comprehension, and mutants Inf simultaneously apply it in production and in comprehension.

#### □ *Assumptions*

We consider a (resident) population No-Inf that applies the Informativity Principle neither in production nor in comprehension, and a (mutant) population Inf that applies the principle both in production and in comprehension, as summarized in (40).

#### (46) **Production and comprehension rules for Inf and No-Inf**

a. Production: for Inf and No-Inf alike, only produce true calls, and if  $A^+$  and  $A^-$  are both true,

No-Inf: use  $A^-$  at rate  $r$ , and  $A^+$  at rate  $(1-r)$ ;

Inf: use  $A^+$

b. Comprehension:

No-Inf: interpret  $A^+$  as  $A^+$  and  $A^-$  as  $A^-$  (no strengthening)

Inf: interpret  $A^+$  as  $A^+$  and  $A^-$  as  $A^-$  and not  $A^+$  (strengthening).

We assume that the benefits of communicating different meanings are the same as in (41).

#### □ *Payoffs*

We will now motivate the distribution of payoffs summarized in (45).

#### (47) **Summary of payoffs**

Sender	No-Inf	Inf	No-Inf	Inf
Receiver	No-Inf	Inf	Inf	No-Inf
<b>FF</b> ( $A^-$ : false, $A^+$ : false)	0	0	0	0
<b>TT</b> ( $A^-$ : true, $A^+$ : true)	$r b^- + (1-r) b^+$	$b^+$	$-r e + (1-r) b^+$	$b^+$
<b>TF</b> ( $A^-$ : true, $A^+$ : false)	$b^-$	$b^*$	$b^*$	$b^-$
<b>Weighted shared payoff</b>	$p_{TT} [r b^- + (1-r) b^+] + p_{TF} b^-$	$p_{TT} b^+ + p_{TF} b^*$	$p_{TT} [-r e + (2-r) b^+]/2 + p_{TF} [b^* + b^-]/2$	

**FF:**  $A^-$ : false,  $A^+$ : false

In all cases, neither call can be produced, and all parties get a payoff of 0.

**TT:**  $A^-$ : true,  $A^+$ : true

–When No-Inf is the sender, it produces  $A^-$  at rate  $r$ , and  $A^+$  at rate  $(1-r)$

- When No-Inf is the receiver, it doesn't strengthen anything and the overall payoff is  $r b^- + (1-r) b^+$ .
- When Inf is the receiver, it wrongly strengthens  $A^-$  to  $A^-$  *and not*  $A^+$ , hence an overall payoff of  $-r e + (1-r) b^+$ .

–When Inf is the sender, it only produces  $A^+$ , yielding in all cases a payoff of  $b^+$ .

**TF:**  $A^-$ : true,  $A^+$ : false

In all cases,  $A^-$  is produced.

- When No-Inf is the receiver, it does not apply any strengthening, which yields a payoff of  $b^-$ .
- When Inf is the receiver, it correctly strengthens  $A^-$  to  $A^-$  *and not*  $A^+$ , which yields a payoff of  $b^*$ .

#### □ *Evolutionary stability of No-Inf*

As before, No-Inf is evolutionarily stable relative to Inf just in case:

Condition (1): payoff of No-Inf vs No-Inf > payoff of Inf vs No-Inf, or

Condition (2): these payoffs are equal, but payoff of No-Inf vs Inf > payoff of Inf vs Inf.

Condition (1) can be successively rewritten as follows:

$$\begin{aligned}
 (48) \quad & p_{TT} [r b^- + (1-r) b^+] + p_{TF} b^- > p_{TT} [-r e + (2-r) b^+]/2 + p_{TF} [b^* + b^-]/2 \\
 & p_{TT} 2[r b^- + (1-r) b^+] + p_{TF} 2b^- > p_{TT} [-r e + (2-r) b^+] + p_{TF} [b^* + b^-] \\
 & p_{TT} [2r b^- + (2-2r-2+r) b^+ + r e] + p_{TF} [2b^- - b^* - b^-] > 0 \\
 & p_{TT} [2r b^- - r b^+ + r e] + p_{TF} [b^- - b^*] > 0 \\
 & p_{TT} r [2b^- - b^+ + e] > p_{TF} [b^* - b^-]
 \end{aligned}$$

Some values of the parameters will ensure that No-Inf is stable and others will not. It follows that invasion of No-Inf by Inf can certainly not be guaranteed for all reasonable values of the parameters.

## References

- Alexander, J., Skyrms, B., & Zabell, S. L. (2012). Inventing new signals. *Dynamic Games and Applications*, 2, 129-145.
- Arnal, L.H., A. Flinker, A. Kleinschmidt, A.L. Giraud, and D. Poeppel. 2015. Human screams occupy a privileged niche in the communication soundscape. *Current Biology* 25 (15): 2051–2056.
- Arnal, L. H., Kleinschmidt, A., Spinelli, L., Giraud, A. -L., and Mégevand, P. (2019). The rough sound of salience enhances aversion through neural synchronisation. *Nat. Commun.* 10, 1–12. doi: 10.1038/s41467-019-11626-7
- Arnold, K., Price, T.: 2015, The Effect of Playback Distance on Male Putty-Nosed Monkey Alarm Calling Behaviour. Manuscript, University of St Andrews.
- Arnold, Kate and Klaus Zuberbühler, 2012. Call combinations in monkeys: Compositional or idiomatic expressions? *Brain and Language* 120(3): 303-309.
- Aubin, T.: 1991 Why do distress calls evoke interspecific responses? An experimental study applied to some species of birds. *Behavioural Processes*, 23(2), 103–111. [https://doi.org/10.1016/0376-6357\(91\)90061-4](https://doi.org/10.1016/0376-6357(91)90061-4)
- Aubin, T. & Brémond, J-C.: 1989, Parameters used for recognition of distress calls in two species: *Larus argentatus* and *Sturnus vulgaris*. *Bioacoustics* 2(1), 23-33.
- Barrett, J. A., & Skyrms, B.: 2017, Self-assembling games. *The British Journal for the Philosophy of Science*.
- Beckers, G. J., Huybregts, M. A., Everaert, M. B., & Bolhuis, J. J.: 2024, No evidence for language syntax in songbird vocalizations. *Frontiers in Psychology*, 15, 1393895.
- Benitez-Quiroz C, Wilbur R, Martinez A.: 2016, The Not face: a Grammaticalization of Facial Expressions of Emotion. *Cognition* 150: 77-84
- Blumstein, D. T., Bryant, G. A. & Kaye, P.: 2012, The sound of arousal in music is context-dependent. *Biology Letters*, 8, 744-747.
- Blumstein, D. T., & Récapet, C.: 2009, The sound of arousal: The addition of novel non-linearities increases responsiveness in marmot alarm calls. *Ethology*, 115(11), 1074-1081.
- Brochhagen, T., Franke, M., & van Rooij, R.: 2018, Coevolution of lexical meaning and pragmatic use. *Cognitive Science*, 42(8), 2757-2789.
- Byrne RW, Cartmill E, Genty E, Graham KE, Hobaiter C, Tanner J.: 2017, Great Ape gestures: Intentional Communication with a Rich Set of Innate Signals. *Animal Cognition* 20(4):755–69.
- Dautriche I, Buccola B, Berthet M, Fagot J, Chemla E.: 2022, Evidence for compositionality in baboons (*Papio papio*) through the test case of negation. *Sci Rep.* 12(1):19181. doi: 10.1038/s41598-022-21143-1. PMID: 36357450; PMCID: PMC9649700.
- Douglas, P. and Moscovice, L.: 2015, Pointing and pantomime in wild apes? Female bonobos use referential and iconic gestures to request genito-genital rubbing. *Scientific Reports* 5, 13999. <https://doi.org/10.1038/srep13999>
- Engel, M. S.: 2006, A giant honey bee from the middle Miocene of Japan (Hymenoptera: Apidae). *Am. Mus. Novit.* 3504, 1–12. doi: 10.1206/00030082(2006)504[0001:AGHBFT]2.0.CO;2
- Fitch WT, Hauser MD: 1995, Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on ‘honest’ advertisement. *J Comp Psychol* 37:191–219.
- Fitch, W. T., Neubauer, J., & Herzl, H.: 2002, Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour* 63(3), 407–418. <https://doi.org/10.1006/anbe.2001.1912>
- Fuller JL, Cords M.: 2019, Versatility in a loud call: Dual affiliative and agonistic functions in the blue monkey boom. *Ethology* 126, 1:10-23
- Fuller, James. 2013. *Diversity of form, content, and function in the vocal signals of adult male blue monkeys (Cercopithecus mitis stuhlmanni): An evolutionary approach to understanding a signal repertoire*. PhD thesis, Columbia University.
- Gardner, Kathryn; Seeley, Thomas; Calderone, Nicholas: 2008, Do honeybees have two discrete dances to advertise food sources?. *Animal Behaviour* 75, 1291-1300
- Garcia M, Theunissen F, Sèbe F, Clavel J, Ravignani A, Marin-Cudraz T et al.: 2020, Evolution of communication signals and information during species radiation. *Nat. Commun.* 2020

- Graham KE, Hobaiter C, Ounsley J, Furuichi T, Byrne RW: 2018, Bonobo and chimpanzee gestures overlap extensively in meaning. *PLoS Biology* 16(2): e2004825. <https://doi.org/10.1371/journal.pbio.2004825>
- Gautier, J. P.: 1971, Etude morphologique et fonctionnelle des annexes extralaryngées des cercopithecinae; liaison avec les cris d'espacement. *Biologia Gabonica*, 7, 230–267.
- Gautier, Jean-Pierre: 1988, Interspecific affinities among guenons as deduced from vocalizations. In Gautier-Hion, A., Bourlière, F., Gautier, J.P. & Kingdon (Eds.), *A Primate Radiation - Evolutionary Radiation of the African Guenons* (pp. 194–226). Cambridge University Press.
- Gautier-Hion, Annie, Marc Colyn & Jean-Pierre Gautier. 1999. Histoire naturelle des primates d'Afrique Centrale. Ecofac editions. 162 pages. Libreville, Gabon: Ecofac Editions.
- Hasson, O. 1994. Cheating signals. *Journal of Theoretical Biology* 167:223–238.
- Hechavarría, J. C., Beetz, M. J., Garcia-Rosales, F., and Kössl, M.: 2019, Superfast periodicities in distress vocalizations emitted by bats. *bioRxiv* [Preprint] doi: 10.1101/734640
- Hechavarría, J.C., Jerome Beetz, M., García-Rosales, F. et al.: 2020, Bats distress vocalizations carry fast amplitude modulations that could represent an acoustic correlate of roughness. *Sci Rep* 10, 7332 (2020). <https://doi.org/10.1038/s41598-020-64323-7>
- Hedtke, S.M., Patiny, S. & Danforth, B.N.: 2013, The bee tree of life: a supermatrix approach to apoid phylogeny and biogeography. *BMC Evol Biol* 13, 138. <https://doi.org/10.1186/1471-2148-13-138>
- Hewitt, G., MacLarnon, A., & Jones, K.: 2002, The functions of laryngeal air sacs in primates: A new hypothesis. *Folia Primatologica* 73(2-3), 70–94.
- Hobaiter C, Byrne RW: 2011, The gestural repertoire of the wild chimpanzee. *Anim Cognit* 14(5):745–767
- Krebs, J. R. & Dawkins, R.: 1984, Animal signals: mindreading and manipulation. In *Behavioural ecology: an evolutionary approach* (eds J. R. Krebs & N. B. Davies), pp. 380–402, 2nd edn. Oxford, UK: Blackwell.
- Hobaiter C, Byrne R: 2011, The gestural repertoire of the wild chimpanzee. *Anim Cogn* 14:745–767.
- Hofbauer, J., Sigmund, K.: 1988, *The Theory of Evolution and Dynamical Systems*. Cambridge UK: Cambridge University Press.
- Hofbauer, J., Sigmund, K. 1998. *Evolutionary Games and Population Dynamics*. Cambridge UK: Cambridge University Press.
- Hofbauer, J., Huttegger, S., 2008. Feasibility of communication in binary signaling games. *Journal of Theoretical Biology* 254, 843–849.
- Hofbauer, J., Schuster P., Sigmund, K., 1979. A note on evolutionarily stable strategies and game dynamics. *Journal of Theoretical Biology* 81: 609–612.
- Hurford, J., 1989. Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua* 77, 187–222.
- Huttegger, S., Zollman, K., 2011. Signaling games: dynamics of evolution and learning. In: Benz, A., Ebert, C., Jäger, G., van Rooij, R. (Eds.), *Language, Games, and Evolution. Trends in Current Research on Language and Game Theory*. Springer.
- Huttegger, S., B. Skyrms, P. Tarres, and E. Wagner. 2014. Some dynamics of signaling games. *Proceedings of the National Academy of Sciences*, 111 (Supplement 3), 10873–10880. 49
- Jäger, G., 2008. Evolutionary stability conditions for signaling games with costly signals. *Journal of Theoretical Biology* 253, 131–141.
- l'Anson Price, R., & Grüter, C.: 2015, Why, when and where did honey bee dance communication evolve? *Frontiers in Ecology and Evolution*, 3, 125.
- Leinonen, L., Laakso, M., Carlson, S., and Linnankoski, I. (2003). Shared means and meanings in vocal expression of man and macaque. *Logoped. Phoniatr. Vocol.* 28, 53–61. doi: 10.1080/14015430310011754
- Lewis, D.: 1969, *Convention: A Philosophical Study*. Harvard University Press, Cambridge, MA.
- Nowak, M.A., Komarova, N.L., Niyogi, P.: 2002, Computational and evolutionary aspects of language. *Nature* 417, 611–617.
- Lindauer, M.: 1956, Über die Verständigung bei indischen Bienen. *Z. Vergleichende Physiol.* 38, 521–557. doi: 10.1007/BF00341108

- Lindauer, Martin. *Communication among social bees*. Cambridge, Massachusetts: Harvard University Press, 1971
- Magrath, R. D., Haff, T. M. & Igic, B.: 2020, Interspecific Communication: Gaining Information from Heterospecific Alarm Calls. In T. Aubin & N. Mathevon (eds), *Coding Strategies in Vertebrate Acoustic Communication*, Springer.
- Maynard Smith, J. & Price, G. R. (1973), The Logic of Animal Conflict. *Nature*, 246(5427), pp.15–18.
- Maynard Smith, J. & Harper, D. G. C. 2003 *Animal Signals*. Oxford, UK: Oxford University Press.
- Menzel R (2019) The waggle dance as an intended flight: a cognitive perspective. *Insects* 10(12), 424.
- Mumm, C. A. S., and Knörnschild, M.: 2017, Territorial choruses of giant otter groups (*Pteronura brasiliensis*) encode information on group identity. *PLoS One* 12:e0185733. doi: 10.1371/journal.pone.0185733
- 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)
- Nowak, M.A., Krakauer, D.C.: 1999, The evolution of language. *Proc. Nat. Acad. Sci. USA* 96, 80288033.
- Nowak, M.A., Plotkin, J.B., Krakauer, D.C.: 1999, The evolutionary language game. *Journal of Theoretical Biology* 200, 147–162.
- Ouattara Karim, Lemasson Alban and Zuberbühler Klaus: 2009, Campbell's monkeys use affixation to alter call meaning. *PLoS ONE* 4 (11): e7808.
- Patel-Grosz, P.: 2023, The search for universal primate gestural meanings. To appear in *Proceedings of Sinn und Bedeutung*, 27. <https://ling.auf.net/lingbuzz/007212>
- Pawlowsch, C.: 2007, Finite populations choose an optimal language. *Journal of Theoretical Biology* 249, 606–616.
- Pawlowsch, C.: 2008, Why evolution does not always lead to an optimal signaling system. *Games and Economic Behavior* 63, 203–226.
- Pawlowsch, C., Metrikopoulos, P., Ritt, N. 2011. Neutral stability, drift, and the differentiation of languages. *Journal of Theoretical Biology* 287: 1–12.
- Postal O, Dupont T, Bakay W, Dominique N, Petit C, Michalski N, Gourévitch B. Spontaneous Mouse Behavior in Presence of Dissonance and Acoustic Roughness. *Front Behav Neurosci*. 2020 Oct 8;14:588834. doi: 10.3389/fnbeh.2020.588834. PMID: 33132864; PMCID: PMC7578920.
- Salis, Ambre; Badaire, Flavien; Coye, Camille; Leroux, Maël; Lengagne, Thierry; Schlenker, Philippe; Chemla, Emmanuel: to appear, Mechanisms of mobbing call recognition: Exploring featural decoding in great tits. *Animal Behaviour*.
- Salis, Ambre; Ryder, Robin; Molina, Axel; Schlenker, Philippe; Chemla, Emmanuel (in preparation). Phylogenetic reconstruction of a call combination in tits and chickadees. Manuscript, Institut Jean-Nicod.
- Schlenker, P., Chemla, E., Arnold, K., Lemasson, A., Ouattara, K., Keenan, S., Stephan, C., Ryder, R. & Zuberbühler, K.: 2014, Monkey semantics: two 'dialects' of Campbell's monkey alarm calls. *Linguistics and Philosophy* 37(6), 439–501.
- Schlenker, P., Chemla, E., Arnold, K. & Zuberbühler, K.: 2016a, Pyow-Hack Revisited: Two Analyses of Putty-nosed Monkey Alarm Calls. *Lingua* 171, 1–23.
- Schlenker, P., Chemla, E., 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.: 2016b, Formal Monkey Linguistics. *Theoretical Linguistics* 42(1–2), 1–90. DOI: 10.1515/tl-2016-0001
- Schlenker, Philippe; Coye, Camille; Leroux, Maël; Chemla, Emmanuel: 2023, The ABC-D of Animal Linguistics: Are Syntax and Compositionality for Real? *Biological Reviews* 98, 4:1142–1159.
- Schlenker, P., Salis, A., Leroux, M., Coye, C., Rizzi, L., Steinert-Threlkeld, S., & Chemla, E.: 2024, Minimal Compositionality versus Bird Implicatures: two theories of ABC-D sequences in Japanese tits. *Biological Reviews*.
- Schlenker, Philippe; Coye, Camille; Salis, Ambre; Steinert-Threlkeld, Shane; Ravaux, Lucie; Chemla, Emmanuel: to appear, Anti-Babel: Three Degrees of Interspecies Comprehension. *Mind & Language*.



- Schneider C, Call J, Liebal K.: 2010, Do bonobos say NO by shaking their head? *Primates*. 51(3):199-202. doi: 10.1007/s10329-010-0198-2.
- Scott-Phillips TC, Blythe RA, Gardner A, West SA. How do communication systems emerge? *Proc Biol Sci*. 2012 May 22;279(1735):1943-9. doi: 10.1098/rspb.2011.2181. Epub 2012 Jan 4. PMID: 22217724; PMCID: PMC3311886.
- Searcy, W. A. and Nowicki, S.: 2007, *The Evolution of Animal Communication*. Princeton, NJ: Princeton University Press
- 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.
- Suzuki, T. N. AND 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>
- Steinert-Threlkeld, Shane; Schlenker, Philippe; Chemla, Emmanuel: 2021, Referential and General Calls in Primate Semantics. *Linguistics & Philosophy* 44:1317–1342.
- Taylor, P. D., & Jonker, L. B.: 1978, Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40 (1), 145–156.
- Tinbergen, N. 1952. "Derived activities"; their causation, biological significance, origin, and emancipation during evolution. *Q. Rev. Biol.* 27:1-32.
- Tomasello, M., Call, J.: 2019, Thirty years of great ape gestures. *Anim Cogn* 22, 461–469. <https://doi.org/10.1007/s10071-018-1167-1>
- Trapa, P.E., Nowak, M.A.: 2000, Nash equilibria for an evolutionary language game. *Journal of Mathematical Biology* 41, 172–188.
- von Frisch, K.: 1967, *The Dance Language and Orientation of Bees*. Harvard University Press: Boston, MA, USA.
- Wang, Léo: 2024, *Optimal Lexica*. MA thesis, Ecole Normale Supérieure-PSL University, Paris, France.
- Waser P. M. & Waser M. S.: 1977, Experimental studies of primate vocalization: specializations for long-distance propagation. *Z. Für Tierpsychol.* 43, 239–263.
- Weibull, J. W. 1995. *Evolutionary Game Theory*. MIT Press.
- Wilkins, H. D. & Ritchison, G.: 1999, Drumming and tapping by red-bellied woodpeckers: description and possible causation. *J. Field Ornithol.* 70, 578–586