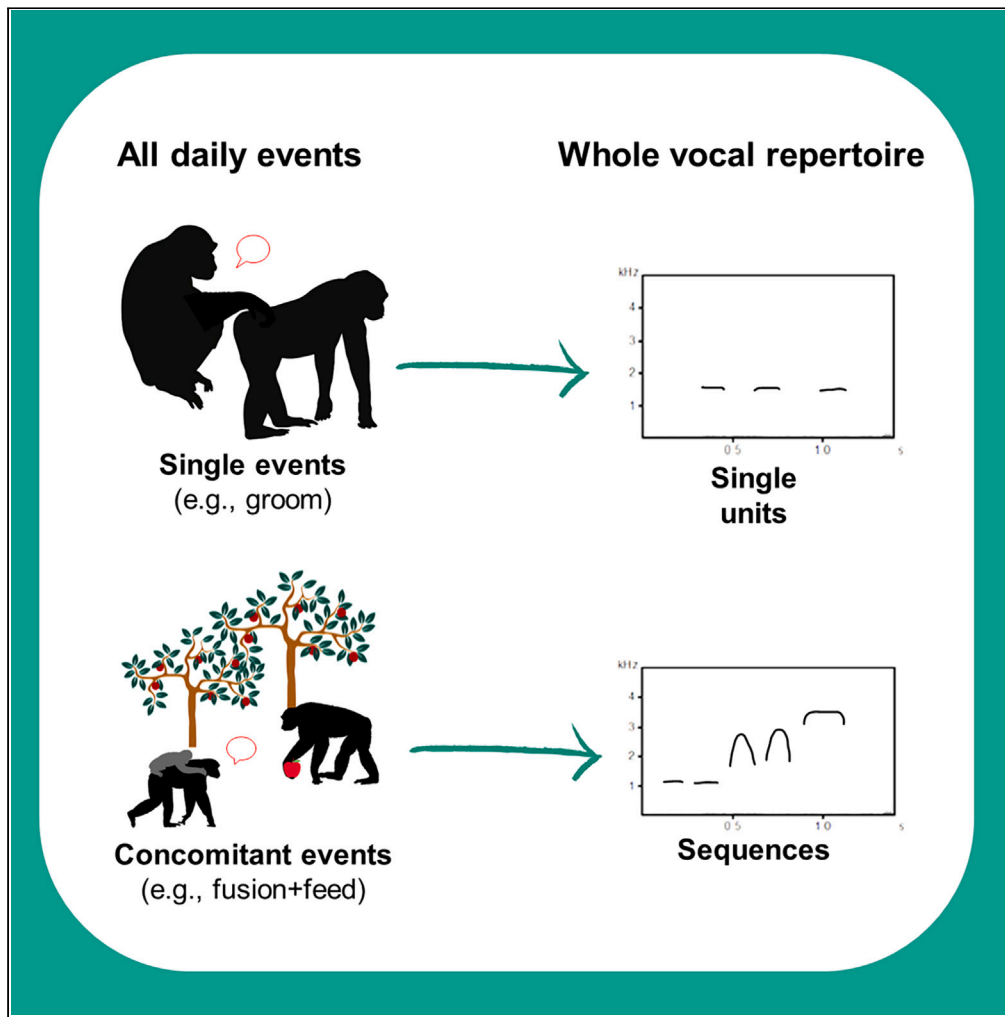


Article

# Chimpanzees show the capacity to communicate about concomitant daily life events



Tatiana Bortolato,  
Angela D.  
Friederici, Cédric  
Girard-Buttoz,  
Roman M. Wittig,  
Catherine  
Crockford

tatiana\_bortolato@eva.mpg.de  
(T.B.)  
crockford@isc.cnrs.fr (C.C.)

**Highlights**

Chimpanzees present key  
preconditions for the  
evolution of compound  
communication

They use sequences across  
many daily life events, not  
just in alarm contexts

Utterance length and  
diversity correlate with  
concomitant events and  
their diversity

They have the potential to  
express concomitant daily  
events through vocal  
sequences

Bortolato et al., iScience 26,  
108090  
November 17, 2023 © 2023 The  
Authors.  
[https://doi.org/10.1016/  
j.isci.2023.108090](https://doi.org/10.1016/j.isci.2023.108090)



## Article

## Chimpanzees show the capacity to communicate about concomitant daily life events

Tatiana Bortolato,<sup>1,2,3,6,\*</sup> Angela D. Friederici,<sup>4</sup> Cédric Girard-Buttoz,<sup>1,2,3,5</sup> Roman M. Wittig,<sup>1,2,3,5</sup> and Catherine Crockford<sup>1,2,3,5,\*</sup>

## SUMMARY

**One universal feature of human language is its versatility in communicating about juxtapositions of everyday events. Versatile combinatorial systems of communication can be selected for if (a) several vocal units are flexibly combined into numerous and long vocal sequences and (b) vocal sequences relate to numerous daily life events. We propose (b) is more likely during simultaneous or serial (concomitant) events than single events. We analyzed 9,391 vocal utterances across the repertoire of wild chimpanzees and their events of production. Chimpanzees used vocal sequences across a range of daily life events and twice as often during concomitant than single events. Also, utterance diversity correlated positively with event diversity. Our results show the potential of chimpanzee vocal sequences to convey combined information about numerous daily life events, a step from which generalized combinatoriality could have evolved.**

## INTRODUCTION

One major feature of human speech and language is the ability to combine meaning-bearing words into phrases by means of syntactic rules that allow us to communicate about everyday events.<sup>1</sup> We communicate in a wide array of social situations about an unlimited range of everyday events, including those that occur simultaneously or serially, using a relatively limited sound set.<sup>2</sup> The sound set of many non-human animals (hereafter, animals), including non-human primates (hereafter, primates), is also limited. Since many animals also combine single signals into signal sequences,<sup>3,4</sup> the question remains how and why do such sequences emerge and are selected for? It is reasonable to think that selection pressure to favor the transition from elementary (i.e., single signals) to compound (i.e., sequences) communication to flexibly describe daily events should require prior emergence of two preconditions: (a) vocal sequences must be numerous (i.e., incorporating call types from across most of the repertoire), long (i.e., surpassing the level of single units or bigrams), and flexibly combined across different types of vocal units and (b) vocal sequences must disambiguate not only rare events but also daily life events and, as such, must combine individual signals such that they refer to numerous events. These preconditions are in line with the mathematical model proposed by Nowak and Komarov<sup>5,6</sup> for the evolution of compound signaling, where the total number of messages has to exceed a critical value and individual components must occur in many different messages.<sup>5,6</sup>

Determining whether species other than humans share the above preconditions would help to understand the evolutionary drivers leading to the emergence of combinatorial forms of communication (i.e., the combination of meaning-bearing single signals into meaningful sequences) and help us retrace how complex combinatorial communication systems could have emerged. Many animals combine vocal signals into sequences.<sup>3,4</sup> A recent study in chimpanzees has established aforementioned precondition (a) that sequences are structurally numerous, long, and flexibly combined across the repertoire.<sup>7</sup> In relation to aforementioned (b), regarding meaning-bearing units, several species have “compositional” communication (i.e., the meaning of the sequence can be derived from its individually meaningful vocal units).<sup>8–12</sup> However, examples per species are few and short (usually, combinations of two “meaning-bearing units”), and generally in a limited range of events, such as during alarm.<sup>8,13,14</sup> Nonetheless, it is unclear whether there is a genuine lack of combinatorial signaling potential in the animal kingdom or if such discoveries are constrained by research effort since for example in the vocal domain establishing the meaning of each vocal sequence using the gold standard playback experiments often requires years of work. Thus, studies have not yet examined combinatoriality (i.e., related to the meaning of its parts<sup>3</sup>) in longer utterances, nor whether there is an animal signaling system that supports vocal combinatoriality across numerous daily events.

<sup>1</sup>The Ape Social Mind Lab, Institut des Sciences Cognitives, CNRS UMR 5229, Bron 69500, France

<sup>2</sup>Department of Human Behaviour, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany

<sup>3</sup>Tai Chimpanzee Project, Centre Suisse des Recherches Scientifiques, Abidjan 1303, Ivory Coast

<sup>4</sup>Department of Neuropsychology, Max Planck Institute for Human Cognitive Sciences, 04103 Leipzig, Germany

<sup>5</sup>These authors contributed equally

<sup>6</sup>Lead contact

\*Correspondence: [tatiana\\_bortolato@eva.mpg.de](mailto:tatiana_bortolato@eva.mpg.de) (T.B.), [crockford@isc.cnrs.fr](mailto:crockford@isc.cnrs.fr) (C.C.)

<https://doi.org/10.1016/j.isci.2023.108090>



**Table 1. Data summary per group**

Group	East group	North group	South group	Total
N° individuals in the group	38	25	40	103
N° individuals included in the analyses (focal and/or ad-lib data)	35 (18 adults, 17 immatures)	24 (12 adults, 12 immatures)	39 (19 adults, 20 immatures)	98
N° individuals with focal data +/- ad-lib data	29	20	34	83
Total focal hours	552.5	251.1	494.8	1298.4
Mean focal hours per individual	18.87	12.6	14.55	15.34
Range focal hours across individuals	4.4–30.6	4.8–22.6	5.1–26.9	4.4–30.6
N° individuals w/ad-lib data only	6	4	5	15

Here, we offer a global approach to assess the communicative preconditions of (b) in an animal species using the whole repertoire and across daily life events. While (b) could arise through a number of processes, we propose and examine one type of scenario that (b) is more likely during simultaneous or serial events than single events. A fundamental premise of this approach is that daily life events do not necessarily occur neatly, serially, spaced one after another. Rather several events can occur either at the same time (such as *approaching* a conspecific while *encountering a threat*) or in rapid sequential order (such as *approaching* and then *affiliating* with a conspecific) or as a mixture of both (such as *approaching* and then *receiving aggression* while *encountering a threat*). We call such events that are linked in time “concomitant events”. The idea is that if there is a benefit to communicating information about numerous daily events, and particularly the juxtaposition of those events, selection should favor the production of vocal sequences rather than single calls, particularly when experiencing concomitant rather than single events. Despite many species experiencing concomitant events, few are documented to produce vocal sequences in response to them.<sup>8–12</sup>

Although we may be limited in our capacity to infer the real goals and intentions of animals,<sup>15</sup> playback studies across taxa overwhelmingly demonstrate that acoustically distinct vocalizations that are highly context specific convey event-related information to others. Such calls, that are emitted predominantly during a particular event (like predator encounters or feeding), and convey information about the event, are deemed “meaning-bearing units”.<sup>16–18</sup> While playback methods assess the meaning conveyed by calls,<sup>19,20</sup> they are not feasible across the entire repertoire, especially in species uttering a large diversity of vocal sequences. Hence, to assess general meaning *potential*, it is reasonable to assume that calls emitted during a particular event have the potential to convey related information to others and have the potential to be meaning bearing.

If animal repertoires can encode information relating to concomitant events within vocal sequences and across most daily life events, what do we expect regarding the relationship between vocal utterances and events? While there are several ways to conceive of this relationship, we formulate three predictions. First, we predicted that if animals regularly combine single vocal units into vocal sequences to communicate about more than one event per utterance, individuals should be more likely to produce vocal sequences when exposed to concomitant rather than to single events (prediction 1). For example, if vocal units A, B, and C are emitted during fusion (meeting), approach, and feed respectively, when these three events are combined, we could reasonably expect a vocal utterance containing A + B + C. Second, we predicted that the diversity of utterances emitted by each individual reflects the diversity of events and event combinations that elicit utterances (prediction 2). Here, individuals emitting utterances during different multiple events might reasonably emit differently structured vocal sequences.

The third prediction addresses whether communicating about concomitant events is critical for the species or is rather a by-product of the method used, given the lack of playback experiments. If the former is true, we expect a weaker correlation between the diversity of utterances and events when the event diversity measure includes only events assumed to be single rather than concomitant events (prediction 3).

We studied wild chimpanzees (*Pan troglodytes verus*), a species that not only satisfy the structural preconditions laid out in (a) in the first paragraph, by flexibly combining vocal units into hundreds of vocal sequences up to a length of 10 units,<sup>7</sup> but also demonstrates context-specific single vocal units across the vocal repertoire, including during activities such as feed, rest, travel, hunt, and social interactions such as groom, greet, and play events.<sup>21,22</sup> Moreover, some of their vocal combinations are potentially compositional,<sup>10–12</sup> where vocal units (e.g., A and B) produced singly appear to hold the information they convey when combined with other units into vocal sequences (e.g., A + B) to communicate about concomitant events. Chimpanzees have some vocal control and ability to modify their vocal behavior as a function of their social environment.<sup>12,23–28</sup> Taken together, they are a suitable species for examining the potential of precondition (b), which, to our knowledge, has never been assessed in animals.

We collected 1,807.3 h of behavioral and vocal data (1,298.4 focal hours + 508.9 *ad libitum* hours) from three neighboring groups (East, North, and South) of wild chimpanzees in the Tai National Park, Ivory Coast.<sup>29</sup> We obtained a total of 9,391 vocal utterances across the whole repertoire from 98 wild chimpanzees, ranging from 0 to 55 years old. For each vocal utterance, we assessed its contexts of production, by considering all the events that occurred at the time of vocalizing (singly or as concomitant events).

## RESULTS

Using focal animal sampling, we recorded all occurrences of vocal utterances emitted by the focal individual,<sup>30</sup> and *ad libitum* vocal utterances emitted by identified individuals around the focal (Table 1). We noted the identity of the signaler, and the events occurring at the moment of

**Table 2. Definition of the vocal and event terms used**

Term	Definition	Includes
Utterance	A bout of vocal emission (single units and sequences)	Single vocal unit: 1 or more vocal units of the same call type, with less than 2-s between units (e.g., a grunt or a series of grunts)  Sequence: a combination of at least two successive different vocal units with less than 1-s between call types (e.g., grunt + bark)
Event	Context or situation in which a vocal utterance was emitted (i.e., activity, social interaction, environmental change)	Single event: only one event occurred during utterance emission  Concomitant event: numerous events occurred during utterance emission, either simultaneously or in rapid sequential order or both.

vocal production. We identified 13 single vocal units (Table 2, see STAR methods), which were variously combined, resulting in 570 unique utterances in our sample. We also identified 24 single events that elicited vocal utterances (Table 3). Events were defined as those occurring at the moment each vocal utterance was emitted (Table 2) and could be i) the activity of the caller (feed, rest, or travel), ii) a social interaction in which the caller engaged (e.g., affiliation), and/or iii) an environmental change which the caller paid attention to/looked toward (e.g., animal encounter). 22 of 24 single events have previously been described to elicit chimpanzee vocal utterances, including 11 single events reported to elicit context-specific vocal utterances (noted with “*ar*” on Table 3). Counting each different event and concomitant events (such that “approach+affiliation” is different from “approach” only) led to 131 unique total events (hereafter, single and concomitant events are termed “events”, Figure 1; Table S2). Not only do many of these events occur daily<sup>31</sup> but chimpanzees also vocalize during these events on most days. For instance, chimpanzees vocalized during approach events in 91.1% of all study days and vocalized during feeding in 90.5%, travel in 84.2%, and fusion in 78.5% of study days. This contrasts with vocalizing during alarm events (here noted as “animal encounter”) which occurred in 22.8% of study days, where alarm events have been the focus of most previous studies on compound calling. We tested the relationship between the vocal repertoire and the events of vocal production, at two levels: first, at the level of the utterance, whether the likelihood to produce a vocal sequence rather than a single vocal unit increased when the caller was exposed to concomitant events, and second, at the level of the individual chimpanzee, whether the diversity of utterances (or “repertoire size”, i.e., number of unique utterances) emitted relates to the diversity of events in which each individual vocalized.

### Vocal sequences were emitted during concomitant rather than single events

Based on language properties (i.e., a mixture of compositional syntax and non-compositional syntax [e.g., idiomatic expressions]<sup>56</sup>), we do not always expect a 1:1 relation between the length of utterances (i.e., the number of different vocal units within an utterance, such as hoos + grunts + panted grunts = 3) and events, such that one vocal unit denotes one event. Therefore, we investigated if the overall likelihood to produce vocal sequences rather than single vocal units increased when chimpanzees were exposed to concomitant events versus single events. To model this, we determined for each vocal utterance whether it was a single vocal unit (0) or a sequence (1) (binomial response), and whether it was emitted during a single or a concomitant event (categorical predictor). We used a binomial generalized linear mixed model (GLMM), and we controlled for group identity (model 1a). In support of prediction 1, we found that chimpanzees were twice as likely (0.188 vs. 0.391) to utter vocal sequences when exposed to concomitant events than when exposed to single events (full-null model comparison:  $\chi^2 = 77.697$ , degree of freedom [df] = 1,  $p < 0.001$ , Figure 2A; Table 4, model 1a). The marginal and conditional  $R^2$  were 0.031 and 0.116, respectively (model 1a). Group identity did not affect the response ( $\chi^2 = 3.575$ ,  $p = 0.167$ , Table 4).

### Positive relationship between the diversity of utterances and the diversity of events

We measured the number of unique utterances (including single vocal units and sequences) and the number of unique events (including single and concomitant events) produced per individual and field season. The sample size varied between individuals (mean: 61.78, range: 1 to 214 utterances) and was strongly correlated with both measures of utterance and event diversity. To circumvent this issue, we conducted an analysis in two steps. First, we fitted two linear regressions, one with the log number of unique utterances as a function of sample size (logged) and another with the log number of unique events also as a function of sample size (logged) (Figure S1, See STAR methods). Then, we used the residuals of those two models to fit a Gaussian linear mixed model (LMM) (model 2a). In this model, we also controlled for group identity. In support of prediction 2, we found that the residual number of unique utterances emitted per chimpanzee clearly increased with the residual number of unique events (full-null model comparison:  $\chi^2 = 54.904$ ,  $df = 1$ ,  $p < 0.001$ , Figures 3A and 3B; Table 4, model 2a). Group identity did not affect the response ( $\chi^2 = 2.344$ ,  $p = 0.101$ , Table 4).

**Table 3. Single events in which vocal utterances were emitted, ordered by frequency of occurrence**

Event type	Single events	Code	N total utterances	Description	Studies that reported vocalizations during the specified event
Activity	Feed <sup>a</sup>	FE	1943	Arrive at a feeding area, eat, crack nuts or search for or collect fruits, tree leaves, nuts, meat, honey, insects, etc.	Kalan and Boesch; Fedurek and Slocombe <sup>32,33</sup>
	Travel <sup>a</sup>	TR	1085	Locomotion. Excludes walking within a feeding site, traveling during patrols or intergroup encounters. If the caller is resting and right after the call starts traveling, the activity is coded as travel	Fedurek et al.; Crockford et al. <sup>34,35</sup>
	Rest <sup>a</sup>	RE	593	Sit or lie down on the ground, or tree branch, without food.	Crockford et al. <sup>35</sup>
Social interaction	Approach <sup>a</sup>	AP	2074	Move toward another individual or receive an approach within the party. Includes submissive greetings. "Approach" followed by "give aggression" is coded only as "give aggression"	Fedurek et al.; Luef and Pika <sup>36,37</sup>
	Play <sup>a</sup>	PL	924	Play with another individual. Includes tickling, chasing, wrestling, etc. Solo play is not included	Luef and Pika; Davila-Ross et al. <sup>37,38</sup>
	Receive aggression <sup>a</sup>	RA	745	Receive aggression with or without body contact. Includes attack, bite, displace, display, chase away, hit, slap, threaten, etc.	Fedurek et al.; Slocombe and Zuberbühler <sup>36,39</sup>
	Affiliation	AF	676	Give or receive an affiliative gesture. Includes hug, arm reach, kiss, touch genitals, etc (but not groom).	van Lawick-Goodall <sup>40</sup>
	Groom <sup>a</sup>	GR	353	Groom or being groomed by another individual. Solo grooming is not included	Fedurek et al.; Watts <sup>41,42</sup>
	Beg	BE	482	Request food or tools from the possessor, or request to be breastfed or carried by the mother	Levréro and Mathevon <sup>43</sup>
	Give aggression	GA	220	Give aggression with or without body contact. Includes attack, bite, displace, display, chase away, hit, slap, threaten, etc.	Fedurek et al.; Slocombe and Zuberbühler <sup>36,44</sup>
	Food share	SH	16	Receive or give food from/to another individual or right after the share of it	
	Copulate <sup>a</sup>	CO	14	Copulate with another individual	Nishida; Townsend et al. <sup>45,46</sup>
	Solicit copulation	SC	10	Show a gesture and posture directed to an estrous female by a male or vice versa, which normally leads to copulation	Nishida <sup>45</sup>
Change in the environment	Distress <sup>a</sup>	DI	1026	Show signs of distress (e.g., rapid searching behavior, retreat, temper tantrum) to an event external or related to self, such as startle to external events, mother moves away, food exclusion, isolation, being groomed roughly	Levréro and Mathevon; Dezechache et al. <sup>43,47</sup>
	Fusion	FU	735	Join a party after at least 30 min of absence, or be joined by one or several others after 30 min of absence.	Fedurek et al. <sup>34</sup>

(Continued on next page)

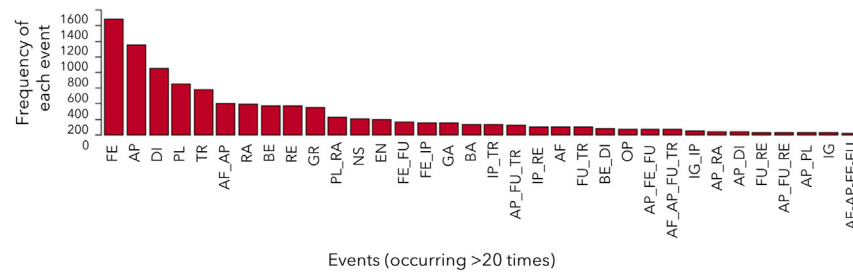
Table 3. Continued

Event type	Single events	Code	N total utterances	Description	Studies that reported vocalizations during the specified event
	Inter-party communication	IP	571	Reply to a chimpanzee calling from outside the party, within 30 s <sup>48</sup> after hearing the call. Includes chorusing with other individuals of the party if it is in reply to an outside party call	Fedurek et al.; Mitani and Nishida <sup>34,49</sup>
	Nest time	NS	244	Call given while in either a day or night nest, while on the ground but looking at individuals already in the nests, or in the evening, while resting or traveling before the nest, while looking up into the trees	Boesch and Crockford <sup>50</sup>
	Animal encounter <sup>a</sup>	EN	204	Look at a threat or signs of threat (snake, leopard, leopard scat, big animal, dead animal, etc.) or hear alarm calls (from conspecifics or other species)	Crockford et al.; Girard-Buttoz et al.; Dezecache et al. <sup>24,25,35,51</sup>
	Bystander to aggression	BA	149	Call given as a bystander to aggression between other individuals.	
	Outside party noise	OP	94	Call immediately after hearing a noise beyond visibility (within 30 s), <sup>48</sup> due to a potential threat (aeroplane noise, branches moving due to unseen individual, either human, animal or another chimpanzee, or due to falling trees, weather, such as thunder, lightning, start raining, or strong wind; etc.).	Crockford et al. <sup>24,35</sup>
	Intergroup encounter	IG	79	Visual or vocal encounter with a neighboring group, looking in the direction of the neighbor group, even if the caller at the moment is resting or running	Herbinger et al. <sup>52</sup>
	Hunt <sup>a</sup>	HU	20	Stalking, pursuing, capturing, and killing monkeys. Once eating starts, becomes “feed” context	Mitani and Watts; Crockford and Boesch <sup>53,54</sup>
	Bystander to copulation	BC	15	Look and/or push in between the mating pair (“Interfere copulation” by Nishida). Mostly by immature individuals and associated with aggression or affiliation	Nishida <sup>45</sup>
	Human directed	HD	15	Look at the observer, mostly in reaction to unexpected, sudden movements from the human observer. A behavior only observed in immature individuals	Hopkins et al. <sup>55</sup>

<sup>a</sup>Events with highly context-specific vocalizations previously reported.

To test the relevance of considering concomitant events rather than single events on vocal production, we fitted a model similar to [model 2a](#) where we only included single events as a predictor ([model 2b](#), [Table S4](#)). For instance, consider an individual vocalizing on three separate occasions, first during an affiliation event, then during an approach, and finally during a concomitant event of affiliation + approach. When we count the concomitant event in this example as distinct, we consider the individual has vocalized in three unique events ([model 2a](#)). However, when we break down the concomitant event into its respective units, we consider the individual has vocalized in two unique events ([model 2b](#)). In support of prediction 3, a model comparison showed a substantial increase in explanatory power with respect to utterance diversity when event diversity included concomitant events rather than only single events ([model 2a](#): concomitant events, Akaike information criterion [AIC] = 100, marginal  $R^2 = 0.361$ ; [model 2b](#): single events, AIC = 124.55, marginal  $R^2 = 0.219$ ).

A further point to assess relates to the inclusion of individuals of all ages in each model. Previous research suggests that until close to adulthood, individuals continue to increase the length and diversity of their utterances with age (up to 10 years old<sup>57</sup>). Moreover, unlike adults, immatures might not engage as often in concomitant events, nor communicate about several events at the same time. Therefore, to ensure



**Figure 1. Frequency of each event eliciting vocalizations across the dataset**  
Concomitant events are ordered alphabetically, not in the sequential order of the event. See Table 3 for definitions of two-letter codes.

that the pattern found between utterances and events reflects a functional relationship rather than independent co-expansion of both parameters, due to for instance developmental processes, we repeated both analyses only with adult individuals. In both models, we included, again, group identity as a control variable. When assessing only adults, we found the same pattern as in model 1a between the length of the utterances and events (full-null model comparison:  $\chi^2 = 71.945$ ,  $df = 1$ ,  $p < 0.001$ , Figure 2B; Table S3, model 1b). The marginal and conditional  $R^2$  were 0.048 and 0.068, respectively (model 1b). Regarding diversity, compared with model 2a, the number of unique utterances also clearly increased with the number of events in adults (full-null model comparison:  $\chi^2 = 12.462$ ,  $df = 1$ ,  $p < 0.001$ , Table S5, model 2c). The marginal and conditional  $R^2$  were 0.242 and 0.305, respectively (model 2c).

We found a certain degree of variability within each analysis. In Model 1 the probability to produce sequences was not 0 in single events, nor 1 in concomitant events. Accordingly in Model 2, we did not find a 1:1 relation between the diversity of the utterances and events. When looking at specific events we also did not find a 1:1 relation with utterance length. Rather, we found a general pattern of increasing utterance length when more events occurred at the same time (Figure 4).

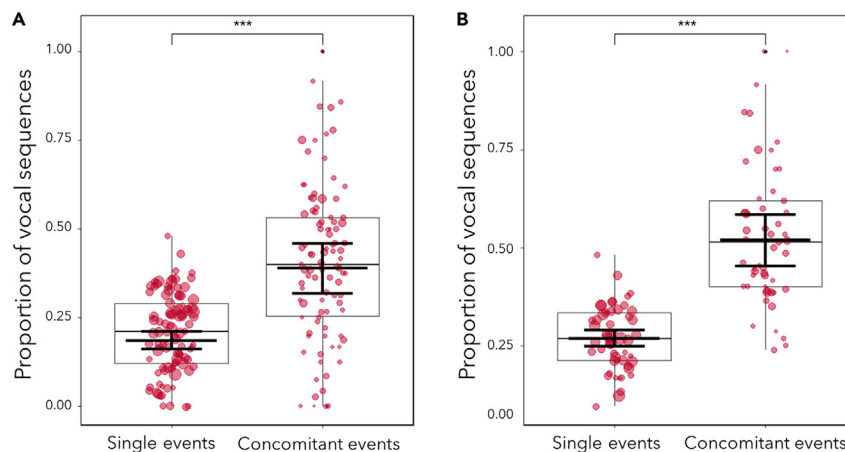
## DISCUSSION

Evidence for the capacity of vocal sequences to express a wide range of daily life events—and in particular concomitant events—remains limited in animal communication research.<sup>8–12</sup> Given the heavy constraints on research effort in terms of determining information conveyed by vocal repertoires, it is unclear to what extent the results of previous work are actually representative of animal communicative capacities. With this in mind, we assessed the global potential of an animal’s vocal system to express concomitant events in their vocal sequences, to disambiguate not only rare events but also daily life events.

In the chimpanzee vocal repertoire, we found support for utterance-event relationships, in line with our three predictions. Overall, our results demonstrate that the length and diversity of utterances are influenced by both the occurrence and diversity of concomitant events, respectively, that signalers are exposed to while vocalizing. Moreover, this pattern holds across numerous daily life events. These results indicate that chimpanzees have at least some of the foundations needed to use a small vocal repertoire to flexibly combine single meaning-bearing vocal units in order to substantially expand the number of messages that can be conveyed well beyond the limitations imposed by the size of the repertoire itself. In our sample, chimpanzees used 13 single vocal units during 24 single events. However, when taking combinations of vocal units and concomitant events into account, these numbers increased to 570 distinct utterances across 131 events. Thus, our results offer an empirical example of a species with the structural and contextual potential to engage in generalized compound communication. This resonates with Nowak and Komarova’s computationally modeled evolutionary theory of the transition from elementary to generalized compound communication.<sup>5,6</sup> Such a capacity constitutes a critical step in the evolutionary path to language. We could likewise argue this to be the case for other complex communication systems; however to our knowledge, no other meaning-rich and repertoire-wide combinatorial communication system apart from language is known to exist.

Even though the number of single vocal units and single events is not equal, and the same single vocal unit can be produced in more than one event, we found a positive relation between utterances and events. This suggests that at least some utterances in the chimpanzees’ repertoire are candidates for “compositional combination”.<sup>56</sup> Hence, chimpanzees might combine two or more vocal units that keep the meaning of their parts to denote concomitant events. This is in line with recent studies suggesting compositionality in specific sequences of the chimpanzee’s repertoire emitted during the concomitant events of fusion and greeting,<sup>12</sup> the arrival of a higher-ranking individual at the party while feeding,<sup>10</sup> and when recruiting conspecifics during alarm events.<sup>11</sup> Animal precedents for this exist; for example two bird species combine calls that singly convey information about alarm or recruitment contexts and when emitted as a vocal sequence convey information about both alarm and recruitment.<sup>8,9</sup>

The degree of variability we found in our results could exist for several reasons. Vocal sequences may be longer than event sequences. A first example of this is when sequences may follow non-compositional rules, such as the “pyow-hack sequence” in putty-nose monkeys.<sup>58</sup> “Pyows” and “hacks” produced singly are alarm calls to general disturbances and eagles, respectively. When combined, the sequence elicits group movement.<sup>58</sup> Hence, the utterance has a length of two but refers to a single event. Second, our method could lead to underestimating relevant events that elicit vocal utterances, especially when the call function refers to an event we did not detect because it was unrelated to



**Figure 2. Proportion of vocal sequences as a function of single or concomitant events**

(A and B) (A) Includes individuals of all ages (0–55 years;  $n = 9,391$  utterances, 98 individuals) and (B) only adults (>10 years;  $n = 5,399$  utterances, 53 individuals). The y axis depicts the likelihood of the utterance to be a sequence (a combination of different vocal units, as opposed to a single vocal unit). Shown are medians (thin horizontal lines), quartiles (25% and 75%; boxes), percentiles (2.5% and 97.5%; vertical lines), as well as the fitted model and its 95% confidence intervals (thick horizontal lines). Each dot represents the value per individual under each category, and its area is proportionate to the total number of utterances recorded per chimpanzee and predictor category (mean: 47.91, range: 1–188). \*\*\* $p < 0.001$ ,  $p$  values were determined using a binomial GLMM test.

any current change in circumstance, also leading to utterances being longer than events. Chimpanzee recruitment calls, for example, where receivers respond by arriving within the visibility of the caller some minutes after a call,<sup>32</sup> fall outside our definition of “event”.

Vocal sequences may be shorter than event sequences. Here, signaler age, particularly of immatures, may lead to under-representation of vocal communication during concomitant events. Although infants are exposed to all events that mothers are exposed to, developmental studies show that behavioral participation, and sometimes vocal participation, in certain events emerges later in development, such as grooming,<sup>59</sup> hunting,<sup>60</sup> or vocalizing to external threats,<sup>51</sup> potentially leading to utterances being shorter than coded events. Finally, our method could lead to overestimating events in circumstances where only the most salient events may be communicated. For example, during fusion at a feeding patch, in some cases, it might be relevant to communicate only about the food, and not the union (e.g., if the separation period was short).

Drawing parallels with other species is difficult, as we lack comparable studies analyzing concomitant events of vocal production across the whole vocal repertoire. However, many species other than chimpanzees combine vocal units in sequences;<sup>61</sup> therefore, it is reasonable to expect similar results in at least some other species when considering the unfolding of concomitant events during vocal sequence production. Similar results might also be expected in the gestural or multimodal domain.<sup>62–65</sup> Intriguingly, however, in species where vocal sequence production has been mapped across the vocal repertoire, few vocal sequences are emitted per species, particularly by adults (red-capped mangabeys,<sup>66,67</sup> de brazza’s monkeys,<sup>66</sup> campbell’s monkeys,<sup>66</sup> blue monkeys,<sup>68</sup> chacma baboons,<sup>69</sup> olive baboons,<sup>69</sup> guinea baboons,<sup>69,70</sup> red- and gray-shanked douc langurs<sup>71</sup>). These species also routinely face multiple simultaneous or sequential events in their daily life, such as greeting in a food patch, or an approach followed by affiliation. If they would routinely use vocal sequences in concomitant daily life events, we would expect a large diversity of vocal sequences in such repertoires, which is not what has been reported so far. To date, reported patterns of vocal sequence usage across species rather rule out the possibility that this is a common phenomenon in the animal kingdom.

Critical theories state that complex signaling should evolve in species that live in complex societies.<sup>72</sup> This might particularly apply to fission-fusion species<sup>73</sup> where coordination and cooperation are challenges in noisy and low-visibility habitats, with dispersed, ephemeral food sources, of which chimpanzees are one example. Accordingly, in complex societies, the need to communicate about more than one event at a time, to transmit a broader diversity of information, and thus reduce uncertainty<sup>74</sup> to solve social and ecological challenges, might drive the ability to produce complex signaling systems.<sup>72</sup> In such situations, species could produce novel vocal units to denote concomitant events, until the required messages exceed a critical threshold making it more economic to reuse existing vocal units in sequences or combinations than produce new ones.<sup>5</sup> Given that many species have a limited repertoire of single vocal units suggesting constraints on versatile signal production,<sup>75</sup> including humans, combining single units into sequences to denote concomitant events should be a common strategy to expand the range of information conveyed.

While we do not address syntax here, it seems likely that a precondition from which syntax could have evolved would be the production of numerous and long vocal sequences to convey numerous messages about daily events. Here, syntax is defined as a computational ordering system of meaning-bearing units with rules that specify how the units relate to each other to convey a certain meaning.<sup>5,6</sup> In a system where few and short sequences (e.g., two-units long) are used, there is limited scope for computational ordering. If there is no or limited selection pressure to combine meaning-bearing units into longer than two-unit utterances, it seems unlikely that there would be strong selection for a computational ordering system. Another likely driver of the evolution of syntax is thought to be the emergence of communication about past



**Table 4. Results of models comparing vocal utterances and events of vocal production for 98 individuals (0–55 years) (Models 1a and 2a)**

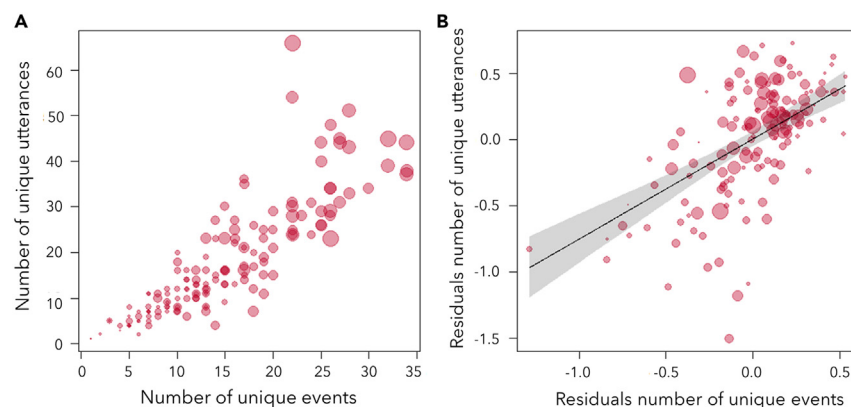
Model	Response	Predictor	Estimate	SE	Lower 95% CI	Upper 95% CI	X <sup>2</sup>	P
1a	Binomial: Single vocal unit (0) vs. sequence (1)	Intercept	−1.258	0.137	−1.502	−1.012		
		Concomitant event (yes)	1.023	0.084	0.867	1.189	77.697	<0.001
		Group (North)	−0.259	0.213	−0.647	0.146	3.575	0.167
		Group (South)	−0.368	0.196	−0.705	−0.030		
2a	Residuals number unique utterances	Intercept	0.080	0.049	−0.015	0.169		
		Residuals number unique events	0.752	0.091	0.572	0.934	68.328	<0.001
		Group (North)	−0.097	0.085	−0.267	0.062	2.343	0.101
		Group (South)	−0.147	0.068	−0.272	−0.014		

In bold statistically significant results ( $p \leq 0.05$ ). CI, confidence interval. The coded level for each categorical predictor is indicated in parentheses. X<sup>2</sup> degrees of freedom = 1.

or future as well as current events, when markers of “before”, “now”, or “later” become relevant.<sup>76</sup> Communicating about two or more concomitant events in the present offers one evolutionary scenario from which later on the capacity to communicate about concomitant events in the past or future may have emerged, such as in humans.

To conclude, our results resonate with Nowak and Komarova’s evolutionary theory of the transition from elementary to generalized compound communication.<sup>5,6</sup> Particularly, we demonstrate that one of our closest living relatives, the chimpanzee, presents some of the critical preconditions for the emergence of complex linguistic processes. Specifically, chimpanzees have the potential to flexibly generate vocal sequences to routinely convey combined information about daily life events. How the increase in utterance length and diversity corresponds to an expansion in the information conveyed remains to be studied. Nonetheless, we argue that through empirical demonstration of potential, we open the possibility that animal communication systems may include critical preconditions from which highly combinatorial, and eventually hierarchically structured, signaling systems could have evolved.

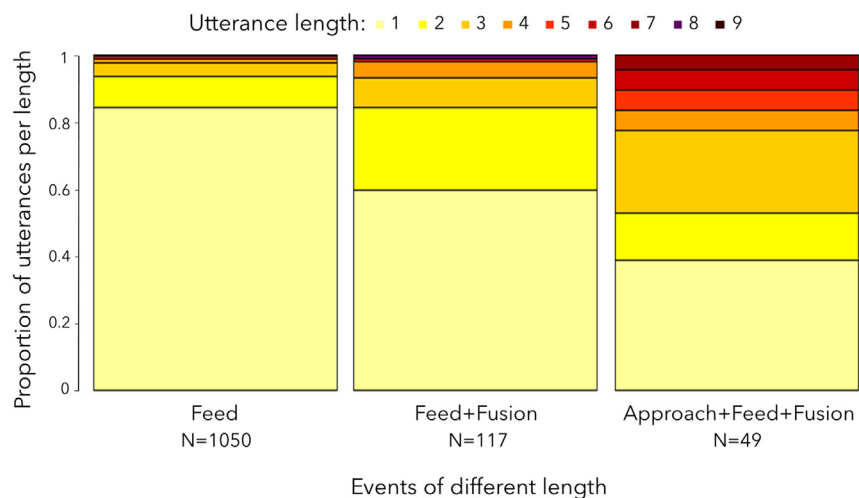
Here, we want to highlight the relevance of considering concomitant events, which unfold during vocal sequence production, in animal communication research, with the aim to encourage similar studies in other species that produce vocal sequences. Compound communication facilitates the expression of new messages that have not been encountered before, increasing the potential of a limited vocal system to encode many messages. Ultimately, at the structural level, language requires the production of sequences with three or more vocal units to form hierarchical structures. At the level of meaning, language requires the capacity to combine meaning-bearing units into sequences



**Figure 3. Relation between the diversity of utterances and events**

(A and B) (A) Relation between diversity of utterances and diversity of events per chimpanzee. (B) Modelled relation between the residuals of each variable (on log scale), fitted against logged sample size (the total number of utterances per chimpanzee and field season) for individuals of all ages (model 2a, N = 152) (see Figure S1 for details).

The line represents the fitted model and the polygon its 95% confidence limit. Each dot depicts an individual per season and their area is proportionate to the total number of utterances per chimpanzee and field season (mean: 61.78, range: 1–214).



**Figure 4. An example case of the proportion of utterances of each length produced during three different events: a single event (“feed”), two and three concomitant events of different lengths**

Concomitant events are ordered alphabetically, not in the sequential order of the event. The color on the bars represents the number of different vocal units per utterance. N represents the total number of utterances per event in the dataset.

drawing flexibly on sounds (phonemes) from across the vocal repertoire. Accordingly, we propose using a whole-repertoire approach as a critical empirical route to draw inferences about the evolutionary path of complex communication systems, such as language. This could be complemented with a further and more focused approach, matching vocal sequences and concomitant events in a 1:1 relationship to quantify the information content of sequences.<sup>77,78</sup>

### Limitations of the study

We studied the vocal repertoire only from the production side, basing our analyses on the structural differences of the utterances, which might not always be functionally relevant for the chimpanzees. However, we guided our analyses according to findings from previous studies on single vocal units and their context specificity. Establishing independent functional relevance for each utterance requires time-consuming playback experiments for each utterance and hence is not feasible at the repertoire level, especially with a species prolific in vocal sequence diversity like chimpanzees which utter several hundred sequences.<sup>7</sup> Therefore, a crucial initial step that we took in this study was to determine if data patterns are broadly consistent with functional relevance by assessing whether utterance diversity relates to event diversity. Last, we caution that this is a correlational study; thus, we cannot infer causality between the behavioral events and the vocal output. However, the logical chain of causality is unlikely to be reversed (e.g., spotting a predator will likely elicit an alarm call but not the opposite).

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Audio and event data collection
  - Vocal coding
  - Event coding
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Model 1a
  - Model 1b
  - Model 2a
  - Model 2b
  - Model 2c
  - Implementation

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.108090>.

## ACKNOWLEDGMENTS

We thank the Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, the Ministère de Eaux et Forêts in Côte d'Ivoire, and the Office Ivoirien des Parcs et Réserves for permitting the study (Research permit: Wittig/008/MESRS/DGRI). Special thanks to Christophe Boesch for creating and nurturing the Tai Chimpanzee Project, the Centre Suisse de Recherches Scientifiques and field assistants of the Tai Chimpanzee Project for their support. We thank Kayla Kolff for her assistance with data collection, Roger Mundry for his invaluable help with the analysis and comments on the manuscript, and Antoine Valet for coding the vocal data for the reliability test. This study was funded by the Hominoid Brain Connectomics Project through the Max Planck Society (M.IF.NEPF8103 and M.IF.EVAN8103).

## AUTHOR CONTRIBUTIONS

Conceptualization of meta-project, A.D.F., C.C., & R.M.W.; Conceptualization of this particular study, C.C., C.G-B. & T.B.; Methodology, C.C., C.G-B. & T.B.; Data collection, T.B.; Investigation, T.B.; Formal analysis, C.G-B. & T.B.; Writing – Original Draft, T.B.; Writing – Review & Editing, A.D.F., C.C., C.G-B. & R.M.W.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: March 3, 2023

Revised: June 3, 2023

Accepted: September 26, 2023

Published: September 28, 2023

## REFERENCES

- Fitch, W.T. (2020). Animal cognition and the evolution of human language: why we cannot focus solely on communication. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375, 20190046. <https://doi.org/10.1098/rstb.2019.0046>.
- Moran, S., McCloy, D., and Wright, R. (2012). Revisiting population size vs. phoneme inventory size. *Language* 88, 877–893.
- Townsend, S.W., Engesser, S., Stoll, S., Zuberbühler, K., and Bickel, B. (2018). Compositionality in animals and humans. *PLoS Biol.* 16, e2006425.
- Zuberbühler, K. (2020). Syntax and compositionality in animal communication. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375, 20190062. <https://doi.org/10.1098/rstb.2019.0062>.
- Nowak, M.A., Plotkin, J.B., and Jansen, V.A. (2000). The evolution of syntactic communication. *Nature* 404, 495–498.
- Nowak, M.A., and Komarova, N.L. (2001). Towards an evolutionary theory of language. *Trends Cogn. Sci.* 5, 288–295.
- Girard-Buttoz, C., Zaccarella, E., Bortolato, T., Friederici, A.D., Wittig, R.M., and Crockford, C. (2022). Chimpanzees produce diverse vocal sequences with ordered and recombinatorial properties. *Commun. Biol.* 5, 410. <https://doi.org/10.1038/s42003-022-03350-8>.
- Suzuki, T.N., Wheatcroft, D., and Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nat. Commun.* 7, 10986. <https://doi.org/10.1038/ncomms10986>.
- Engesser, S., Ridley, A.R., and Townsend, S.W. (2016). Meaningful call combinations and compositional processing in the southern pied babbler. *Proc. Natl. Acad. Sci. USA* 113, 5976–5981. <https://doi.org/10.1073/pnas.1600970113>.
- Leroux, M., Bosshard, A.B., Chandia, B., Manser, A., Zuberbühler, K., and Townsend, S.W. (2021). Chimpanzees combine pant hoots with food calls into larger structures. *Anim. Behav.* 179, 41–50. <https://doi.org/10.1016/j.anbehav.2021.06.026>.
- Leroux, M., Schel, A.M., Wilke, C., Chandia, B., Zuberbühler, K., Slocombe, K.E., and Townsend, S.W. (2023). Call combinations and compositional processing in wild chimpanzees. *Nat. Commun.* 14, 2225.
- Girard-Buttoz, C., Bortolato, T., Laporte, M., Grampp, M., Zuberbühler, K., Wittig, R.M., and Crockford, C. (2022). Population-specific call order in chimpanzee greeting vocal sequences. *iScience* 25, 104851. <https://doi.org/10.1016/j.isci.2022.104851>.
- Engesser, S., Holub, J.L., O'Neill, L.G., Russell, A.F., and Townsend, S.W. (2019). Chestnut-crowned babbler calls are composed of meaningless shared building blocks. *Proc. Natl. Acad. Sci. USA* 116, 19579–19584. <https://doi.org/10.1073/pnas.1819513116>.
- Quattara, K., Lemasson, A., and Zuberbühler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proc. Natl. Acad. Sci. USA* 106, 22026–22031.
- Zuberbühler, K., and Neumann, C. (2017). Referential communication in nonhuman animals. In *APA handbook of comparative psychology: Basic concepts, methods, neural substrate, and behavior*, J. Call, G.M. Burghardt, I.M. Pepperberg, C.T. Snowdon, and T. Zentall, eds. (American Psychological Association), pp. 645–661. <https://doi.org/10.1037/0000011-031>.
- Crockford, C., Wittig, R.M., and Zuberbühler, K. (2017). Vocalizing in chimpanzees is influenced by social-cognitive processes. *Sci. Adv.* 3, e1701742. <https://doi.org/10.1126/sciadv.1701742>.
- Gill, S.A., and Sealy, S.G. (2004). Functional reference in an alarm signal given during nest defence: seet calls of yellow warblers denote brood-parasitic brown-headed cowbirds. *Behav. Ecol. Sociobiol.* 56, 71–80. <https://doi.org/10.1007/s00265-003-0736-7>.
- Kirchhof, J., and Hammerschmidt, K. (2006). Functionally referential alarm calls in tamarins (*Saguinus fuscicollis* and *Saguinus mystax*) – evidence from playback experiments. *Ethology* 112, 346–354. <https://doi.org/10.1111/j.1439-0310.2006.01165.x>.
- Cheney, D.L., and Seyfarth, R.M. (2008). *Baboon Metaphysics: The Evolution of a Social Mind* (University of Chicago Press). <https://doi.org/10.7208/9780226102429>.
- Zuberbühler, K., and Wittig, R.M. (2011). Field experiments with non-human primates: a tutorial. *Field Lab. Methods Primatol. Pract. Guide*, 207–224.
- Bouchard, A., and Zuberbühler, K. (2022). An intentional cohesion call in male chimpanzees

- of Budongo Forest. *Anim. Cogn.* 25, 853–866. <https://doi.org/10.1007/s10071-022-01597-6>.
22. Crockford, C. (2019). Why does the chimpanzee vocal repertoire remain poorly understood and what can be done about it? In *The Chimpanzees of the Tai Forest*, C. Boesch, R. Wittig, C. Crockford, L. Vigilant, T. Deschner, and F. Leendertz, eds. (Cambridge University Press), pp. 394–409. <https://doi.org/10.1017/9781108674218.025>.
  23. Fischer, J., and Hage, S.R. (2019). Primate vocalization as a model for human speech: scopes and limits. *Hum. Lang. Genes Brains Behav.* 639–656.
  24. Crockford, C., Wittig, R.M., Mundry, R., and Zuberbühler, K. (2012). Wild chimpanzees inform ignorant group members of danger. *Curr. Biol.* 22, 142–146. <https://doi.org/10.1016/j.cub.2011.11.053>.
  25. Girard-Buttoz, C., Surbeck, M., Samuni, L., Tkaczynski, P., Boesch, C., Fruth, B., Wittig, R.M., Hohmann, G., and Crockford, C. (2020). Information transfer efficiency differs in wild chimpanzees and bonobos, but not social cognition. *Proc. Biol. Sci.* 287, 20200523. <https://doi.org/10.1098/rspb.2020.0523>.
  26. Liebal, K., Waller, B.M., Slocombe, K.E., and Burrows, A.M. (2014). *Primate Communication: A Multimodal Approach* (Cambridge University Press).
  27. Schel, A.M., Townsend, S.W., Machanda, Z., Zuberbühler, K., and Slocombe, K.E. (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PLoS One* 8, e76674. <https://doi.org/10.1371/journal.pone.0076674>.
  28. Slocombe, K.E., Lahiff, N.J., Wilke, C., and Townsend, S.W. (2022). Chimpanzee vocal communication: what we know from the wild. *Curr. Opin. Behav. Sci.* 46, 101171. <https://doi.org/10.1016/j.cobeha.2022.101171>.
  29. Wittig, R.M. (2018). Tai chimpanzees. In *Encyclopedia of animal cognition and behavior*, J. Vonk and T. Shackelford, eds. (Springer), pp. 6849–6855.
  30. Altmann, J. (1974). Observational study of behavior: sampling methods. *Behaviour* 49, 227–267.
  31. Lehmann, J., Korstjens, A.H., and Dunbar, R.I.M. (2007). Fission–fusion social systems as a strategy for coping with ecological constraints: a primate case. *Evol. Ecol.* 21, 613–634. <https://doi.org/10.1007/s10682-006-9141-9>.
  32. Kalan, A.K., and Boesch, C. (2015). Audience effects in chimpanzee food calls and their potential for recruiting others. *Behav. Ecol. Sociobiol.* 69, 1701–1712. <https://doi.org/10.1007/s00265-015-1982-1>.
  33. Fedurek, P., and Slocombe, K.E. (2013). The social function of food-associated calls in male chimpanzees: Chimpanzee food-associated calls. *Am. J. Primatol.* 75, 726–739. <https://doi.org/10.1002/ajp.22122>.
  34. Fedurek, P., Donnellan, E., and Slocombe, K.E. (2014). Social and ecological correlates of long-distance pant hoot calls in male chimpanzees. *Behav. Ecol. Sociobiol.* 68, 1345–1355. <https://doi.org/10.1007/s00265-014-1745-4>.
  35. Crockford, C., Gruber, T., and Zuberbühler, K. (2018). Chimpanzee quiet hoo variants differ according to context. *R. Soc. Open Sci.* 5, 172066. <https://doi.org/10.1098/rsos.172066>.
  36. Fedurek, P., Tkaczynski, P.J., Hobaiter, C., Zuberbühler, K., Wittig, R.M., and Crockford, C. (2021). The function of chimpanzee greeting calls is modulated by their acoustic variation. *Anim. Behav.* 174, 279–289.
  37. Luef, E.M., and Pika, S. (2019). Social relationships and greetings in wild chimpanzees (*Pan troglodytes*): use of signal combinations. *Primates* 60, 507–515. <https://doi.org/10.1007/s10329-019-00758-5>.
  38. Davila-Ross, M., Allcock, B., Thomas, C., and Bard, K.A. (2011). Aping expressions? Chimpanzees produce distinct laugh types when responding to laughter of others. *Emotion* 11, 1013–1020. <https://doi.org/10.1037/a0022594>.
  39. Slocombe, K.E., and Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proc. Natl. Acad. Sci. USA* 104, 17228–17233. <https://doi.org/10.1073/pnas.0706741104>.
  40. van Lawick-Goodall, J. (1986). *The Chimpanzees of Gombe: Patterns of Behavior* (Belknap P.).
  41. Fedurek, P., Slocombe, K.E., Hartel, J.A., and Zuberbühler, K. (2015). Chimpanzee lip-smacking facilitates cooperative behaviour. *Sci. Rep.* 5, 13460. <https://doi.org/10.1038/srep13460>.
  42. Watts, D.P. (2016). Production of grooming-associated sounds by chimpanzees (*Pan troglodytes*) at Ngogo: variation, social learning, and possible functions. *Primates* 57, 61–72. <https://doi.org/10.1007/s10329-015-0497-8>.
  43. Levréro, F., and Mathevon, N. (2013). Vocal signature in wild infant chimpanzees. *Am. J. Primatol.* 75, 324–332.
  44. Slocombe, K.E., and Zuberbühler, K. (2005). Agonistic screams in wild chimpanzees (*Pan troglodytes schweinfurthii*) vary as a function of social role. *J. Comp. Psychol.* 119, 67–77. <https://doi.org/10.1037/0735-7036.119.1.67>.
  45. Nishida, T. (1997). Sexual behavior of adult male chimpanzees of the Mahale Mountains national park, Tanzania. *Primates* 38, 379–398. <https://doi.org/10.1007/BF02381879>.
  46. Townsend, S.W., Deschner, T., and Zuberbühler, K. (2008). Female chimpanzees use copulation calls flexibly to prevent social competition. *PLoS One* 3, e2431. <https://doi.org/10.1371/journal.pone.0002431>.
  47. Dezechache, G., Zuberbühler, K., Davila-Ross, M., and Dahl, C.D. (2021). A machine learning approach to infant distress calls and maternal behaviour of wild chimpanzees. *Anim. Cogn.* 24, 443–455.
  48. Baker, K.C., and Aureli, F. (1996). The neighbor effect: other groups influence intragroup agonistic behavior in captive chimpanzees. *Am. J. Primatol.* 40, 283–291.
  49. Mitani, J.C., and Nishida, T. (1993). Contexts and social correlates of long-distance calling by male chimpanzees. *Anim. Behav.* 45, 735–746. <https://doi.org/10.1006/anbe.1993.1088>.
  50. Boesch, C., and Crockford, C. (2005). Call combinations in wild chimpanzees. *Beyond Behav.* 142, 397–421. <https://doi.org/10.1163/1568539054012047>.
  51. Dezechache, G., Crockford, C., and Zuberbühler, K. (2019). The development of communication in alarm contexts in wild chimpanzees. *Behav. Ecol. Sociobiol.* 73, 104. <https://doi.org/10.1007/s00265-019-2716-6>.
  52. Herbinger, I., Papworth, S., Boesch, C., and Zuberbühler, K. (2009). Vocal, gestural and locomotor responses of wild chimpanzees to familiar and unfamiliar intruders: a playback study. *Anim. Behav.* 78, 1389–1396.
  53. Mitani, J.C., and Watts, D.P. (1999). Demographic influences on the hunting behavior of chimpanzees. *Am. J. Phys. Anthropol.* 109, 439–454. [https://doi.org/10.1002/\(SICI\)1096-8644\(199908\)109:4<439::AID-AJPA2>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1096-8644(199908)109:4<439::AID-AJPA2>3.0.CO;2-3).
  54. Crockford, C., and Boesch, C. (2003). Context-specific calls in wild chimpanzees, *Pan troglodytes verus*: analysis of barks. *Anim. Behav.* 66, 115–125. <https://doi.org/10.1006/anbe.2003.2166>.
  55. Hopkins, W.D., Tagliatalata, J., and Leavens, D.A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Anim. Behav.* 73, 281–286. <https://doi.org/10.1016/j.anbehav.2006.08.004>.
  56. Berthet, M., Coyse, C., Dezechache, G., and Kuhn, J. (2023). Animal linguistics: a primer. *Biol. Rev. brv.* 98, 81–98. <https://doi.org/10.1111/brv.12897>.
  57. Bortolato, T., Mundry, R., Wittig, R.M., Girard-Buttoz, C., and Crockford, C. (2023). Slow development of vocal sequences through ontogeny in wild chimpanzees (*Pan troglodytes verus*). *Dev. Sci.* 26, e13350. <https://doi.org/10.1111/desc.13350>.
  58. Arnold, K., and Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Curr. Biol.* 18, R202–R203. <https://doi.org/10.1016/j.cub.2008.01.040>.
  59. Bründl, A.C., Tkaczynski, P.J., Nohon Kohou, G., Boesch, C., Wittig, R.M., and Crockford, C. (2021). Systematic mapping of developmental milestones in wild chimpanzees. *Dev. Sci.* 24, e12988. <https://doi.org/10.1111/desc.12988>.
  60. Boesch, C. (2002). Cooperative hunting roles among Tai chimpanzees. *Hum. Nat.* 13, 27–46.
  61. Engesser, S., and Townsend, S.W. (2019). Combinatoriality in the vocal systems of nonhuman animals. *Wiley Interdiscip. Rev. Cogn. Sci.* 10, e1493.
  62. Genty, E., Clay, Z., Hobaiter, C., and Zuberbühler, K. (2014). Multi-Modal Use of a Socially Directed Call in Bonobos. *PLoS One* 9, e84738. <https://doi.org/10.1371/journal.pone.0084738>.
  63. Genty, E., Neumann, C., and Zuberbühler, K. (2015). Complex patterns of signalling to convey different social goals of sex in bonobos, *Pan paniscus*. *Sci. Rep.* 5, 16135. <https://doi.org/10.1038/srep16135>.
  64. Liebal, K., Call, J., and Tomasello, M. (2004). Use of gesture sequences in chimpanzees. *Am. J. Primatol.* 64, 377–396. <https://doi.org/10.1002/ajp.20087>.
  65. Zuberbühler, K. (2018). Combinatorial capacities in primates. *Curr. Opin. Behav. Sci.* 21, 161–169. <https://doi.org/10.1016/j.cobeha.2018.03.015>.
  66. Bouchet, H., Blois-Heulin, C., and Lemasson, A. (2013). Social complexity parallels vocal complexity: a comparison of three non-human primate species. *Front. Psychol.* 4, 390. <https://doi.org/10.3389/fpsyg.2013.00390>.
  67. Bouchet, H., Pellier, A.-S., Blois-Heulin, C., and Lemasson, A. (2010). Sex differences in the vocal repertoire of adult red-capped mangabeys (*Cercocebus torquatus*): a multi-level acoustic analysis. *Am. J. Primatol.* 72, 360–375.
  68. Fuller, J.L. (2014). The vocal repertoire of adult male blue monkeys (*Cercopithecus mitis stuhlmanni*): a quantitative analysis of acoustic structure. *Am. J. Primatol.* 76, 203–216.
  69. Hammerschmidt, K., and Fischer, J. (2019). Baboon vocal repertoires and the evolution of primate vocal diversity. *J. Hum. Evol.* 126,

- 1–13. <https://doi.org/10.1016/j.jhevol.2018.10.010>.
70. Kemp, C., Rey, A., Legou, T., Boë, L.-J., Berthommier, F., Becker, Y., and Fagot, J. (2017). Vocal repertoire of captive Guinea baboons (*Papio papio*). *Orig. Hum. Lang. Contin. Discontinuities Nonhum. Primates* **15**.
  71. Riondato, I., Marco, G., Gamba, M., and Giacoma, C. (2013). Vocalization of red-and grey-shanked douc langurs (*Pygathrix nemaeus* and *P. cinerea*). *J. Primatol.* **2**, 75–82.
  72. Freeberg, T.M., Dunbar, R.I.M., and Ord, T.J. (2012). Social complexity as a proximate and ultimate factor in communicative complexity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 1785–1801. <https://doi.org/10.1098/rstb.2011.0213>.
  73. Aureli, F., Schaffner, C., Boesch, C., Bearder, S., Call, J., Chapman, C., Connor, R., Fiore, A., Dunbar, R., Henzi, S., et al. (2008). Fission-Fusion dynamics: New research frameworks. *Curr. Anthropol.* **49**, 627–654. <https://doi.org/10.1086/586708>.
  74. Seyfarth, R., and Cheney, D. (2018). Pragmatic flexibility in primate vocal production. *Curr. Opin. Behav. Sci.* **21**, 56–61. <https://doi.org/10.1016/j.cobeha.2018.02.005>.
  75. Seyfarth, R.M., and Cheney, D.L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain Lang.* **115**, 92–100. <https://doi.org/10.1016/j.bandl.2009.10.003>.
  76. Bickerton, D. (2018). *Language and Species* (University of Chicago Press).
  77. Elie, J.E., and Theunissen, F.E. (2016). The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. *Anim. Cogn.* **19**, 285–315. <https://doi.org/10.1007/s10071-015-0933-6>.
  78. Garcia, M., Theunissen, F., Sèbe, F., Clavel, J., Ravignani, A., Marin-Cudraz, T., Fuchs, J., and Mathevon, N. (2020). Evolution of communication signals and information during species radiation. *Nat. Commun.* **11**, 4970. <https://doi.org/10.1038/s41467-020-18772-3>.
  79. Steventon, J. (2002). *Cybertracker V. 3.284* (Cybertracker Softw).
  80. Clark, A.P. (1993). Rank differences in the production of vocalizations by wild chimpanzees as a function of social context. *Am. J. Primatol.* **31**, 159–179.
  81. van Hooff, J. (1973). A structure analysis of the social behaviour of a semi-captive group of chimpanzees. *Soc. Commun. Mov.* **75–162**.
  82. Van Lawick-Goodall, J. (1968). The Behaviour of Free-living Chimpanzees in the Gombe Stream Reserve. *Anim. Behav. Monogr.* **1**, 161–IN12. [https://doi.org/10.1016/S0066-1856\(68\)80003-2](https://doi.org/10.1016/S0066-1856(68)80003-2).
  83. Boersma, P., and Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.05) (Computer program). Retrieved May 1, 2009.
  84. Arnold, K., and Zuberbühler, K. (2006). The alarm-calling system of adult male putty-nosed monkeys, *Cercopithecus nictitans martini*. *Anim. Behav.* **72**, 643–653. <https://doi.org/10.1016/j.anbehav.2005.11.017>.
  85. Seyfarth, R.M., Cheney, D.L., and Marler, P. (1980). Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Anim. Behav.* **28**, 1070–1094. [https://doi.org/10.1016/S0003-3472\(80\)80097-2](https://doi.org/10.1016/S0003-3472(80)80097-2).
  86. Baayen, R.H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R* (Cambridge University Press).
  87. Barr, D.J., Levy, R., Scheepers, C., and Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68**, 255–278.
  88. Forstmeier, W., and Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behav. Ecol. Sociobiol.* **65**, 47–55.
  89. Dobson, A.J., and Barnett, A.G. (2002). *Exponential Family and Generalized Linear Models* (Introd. Gen. Linear Models Boca Raton FL Chapman HallCRC).
  90. Luke, S.G. (2017). Evaluating significance in linear mixed-effects models in R. *Behav. Res. Methods* **49**, 1494–1502.
  91. R Core Team (2013). *R: A language and environment for statistical computing*.
  92. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. Preprint at ArXiv.
  93. Quinn, G.P., and Keough, M.J. (2002). *Experimental Design and Data Analysis for Biologists* (Cambridge University Press).
  94. Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (Sage).

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Analyzed data	This paper	<a href="https://doi.org/10.6084/m9.figshare.22214188">https://doi.org/10.6084/m9.figshare.22214188</a>
Software and algorithms		
R software	Team, R.C. (2013). R: A language and environment for statistical computing.	<a href="https://www.r-project.org/">https://www.r-project.org/</a>
CyberTracker software	Steventon, J. (2002). Cybertracker v. 3.284. Cybertracker Softw.	<a href="https://cybertracker.org/">https://cybertracker.org/</a>
Praat software	Boersma, P., and Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.05)[Computer program]. Retrieved May 1, 2009.	<a href="https://praat.softonic.com/">https://praat.softonic.com/</a>

## RESOURCE AVAILABILITY

## Lead contact

Tatiana Bortolato ([tatiana\\_bortolato@eva.mpg.de](mailto:tatiana_bortolato@eva.mpg.de)).

## Materials availability

This study did not generate new unique reagents.

## Data and code availability

- The data reported in this study have been deposited at 'figshare' repository and are publicly available as of the date of publication <https://doi.org/10.6084/m9.figshare.22214188>.
- The code used is available in this paper's [STAR methods](#).
- Any additional information required to reanalyse the data reported in this paper is available from the [lead contact](#) (Bortolato Tatiana) upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

We conducted 9 months of focal animal sampling<sup>30</sup> resulting in 1,807.3 h of vocal recordings (1,298.4 focal hours +508.9 *ad libitum* hours). We studied 98 habituated wild chimpanzees (0–55 years old) from three neighboring groups (East, North and South) living in the Taï National Park, Ivory Coast.<sup>29</sup> Each day, we conducted two half-day (6 h) focal follows from dawn to dusk, resulting in vocal data from 157 days (69 days East, 21 days North, 67 days South), distributed over two field seasons (January–May 2019 and December 2019–March 2020). Each individual was sampled over a two-month period each field season. We conducted focal samples on all available adults (49 total), and 34 of 49 immatures. For the 11 individuals under 1.4-year-olds, we collected *ad libitum* vocal utterances whilst focalling their mothers, as at this age, they are permanently with 2 m of their mother. For another 4 individuals (between 1.5 and 5.6 years old), we collected only *ad libitum* vocalizations as they died before we could do focal follows on them. See [Table 1](#) for detailed information about number of individuals and focal hours per group. Focal individuals were randomly chosen, and sampling was balanced between a.m. and p.m. per individual. We did not sample 5 additional individuals present in the groups for the following reasons: two individuals (an immature in North group and an adult in South group) died before we could get any vocalization from them, two twin infants from East group could not be individually identified, and one new immigrant female in East group was not well enough habituated to sample. See [Dataset S2](#) for the list of ages, sexes, focal hours and utterances recorded during focals for each individual. The study used only observational, non-invasive procedures. The Ministère de l'Enseignement Supérieur et de la Recherche Scientifique of Cote d'Ivoire (Wittig/008/MESRS/DGRI) and the Office Ivoirien des Parcs et Réserves approved our study. International, national, and institutional ethical guidelines were followed. Researchers followed a strict hygiene protocol, including a five-day quarantine, wearing surgical masks, and keeping 7m distance from the chimpanzees to prevent the transmission of pathogens.

## METHOD DETAILS

## Audio and event data collection

All data was collected in real-time, based on a protocol designed before the start of data collection. 85.8% (8,056 utterances) of the data was collected by T.B. across groups and seasons, and 14.2% (1,335 utterances) of the data was collected by a dedicated and experienced research

assistant (RA) (in East group during the second season). We recorded all occurrences of vocal utterances from the focal individual and from identified *ad libitum* individuals using a Sennheiser ME67 microphone and Tascam DR-40X digital recorder (48 kHz sampling rate and 24-bit sampling depth). Vocal data was recorded in real-time in the field and all utterances were coded by T.B. after the field season. For each vocal utterance, we also collected, in real-time, detailed behavioral and environmental data to determine the event that elicited the vocal utterance, using a smartphone and Cybertracker software<sup>79</sup> (CT). When events unfolded rapidly, they were spoken into the microphone as events unfolded and added to the CT app later.

### Vocal coding

Even though vocal elements present distinctive acoustic and visual features, most vocal units ('call types') can grade into other vocal units. Hence, the chimpanzee's vocal repertoire is not exempt from the 'lumping-splitting' problem.<sup>22</sup> In particular, the number of single vocal units reported in chimpanzees' repertoire varies between 12 and 32 call types depending on the researcher.<sup>7,50,80–82</sup> Here, we took a conservative approach considering only 13 single vocal units ('call types')<sup>57</sup> as the building blocks of the repertoire and 'lumping' the different acoustically distinct variants of the same call types<sup>35</sup> together; this will ultimately under-represent utterance length and diversity, and hence the complexity of the repertoire. We defined a single vocal unit as either a vocal unit (or call type) produced in isolation or repeated with less than 2-s pause in between repetitions. We defined a vocal sequence as the combination of at least two different vocal units emitted within less than 1-s of each other<sup>7,57</sup> (for a sample of 314 sequences, the average interval duration between different units was  $0.23 \pm 0.04$  s (mean  $\pm$  SE)<sup>57</sup>). We used the term 'utterance' as any bout of vocal emission, as such encompassing both single vocal units and vocal sequences (Table 2). The 13 single vocal units we identified and their frequency in the dataset are: barks 'BK' (549 utterances), grunts 'GR' (3,570 utterances), hoos 'HO' (2,268 utterances), non-vocal sounds 'NV' (344 utterances), screams 'SC' (867 utterances), whimpers 'WH' (1,274 utterances), pants 'PN' (830 utterances), panted barks 'PB' (735 utterances), panted grunts 'PG' (986 utterances), panted hoos 'PH' (1,214 utterances), panted roars 'PR' (1 utterance), panted screams 'PS' (489 utterances), and panted whimpers 'PW' (17 utterances). We found panted whimpers in immature individuals only and one panted roar in an adult individual. T.B. coded the types of vocal units present in each utterance using a combined approach of listening to the utterance while visually examining the spectrogram, using PRAAT software.<sup>83</sup> From a total of 11,397 vocal utterances, T.B. classified 10,304 of them. 1,093 utterances, 9.6%, were discarded due to bad quality (overlap with other chimpanzees' vocalizations or other forest noise, poor sound-to-noise ratio) or unclassifiable call type (due to the graded call system) (see Bortolato et al.<sup>57</sup> for details on the coding procedure and chimpanzee's repertoire). A further 913 utterances were discarded from the dataset as we were unable to assign an event (8.86% - see below), resulting in a final dataset of 9,391 utterances. We had 6,857 utterances composed of single vocal units and 2,534 utterances composed of vocal sequences. We had an average of 61.78 vocal utterances per individual (range: 1–214). To ensure replicability, we conducted inter-rater reliability (IRR) with a blind coder, using combined visual and auditory information from audio recordings and respective spectrograms. To ensure that all call types were assessed, T.B. selected utterances randomly within each call type across the dataset, whether they occurred singly or in sequences. A blind coder then listened to the audio recordings and simultaneously visually assessed spectrograms of vocal utterances judging the call type of each vocal element present in the utterance, as well as the actual utterance (Figures S2 and S3). T.B. and the blind coder scored a subsample of 1361 vocal elements across 203 utterances, reaching substantial agreement on categorising the vocal sample at three different levels. The first level tested agreement on the coding of vocal elements, hence the number of individual elements of each type per utterance (Cohen's Kappa score  $\pm$  SE:  $0.913 \pm 0.008$ ) (Figure S2; Table S1). The second level tested agreement on whether an utterance was a single vocal unit or a vocal sequence (i.e., comprising more than 1 vocal unit) (Cohen's Kappa score  $\pm$  SE:  $0.811 \pm 0.045$ ) (Figure S2). This second level of analyses corresponds to the data that entered the first set of statistical analyses testing the likelihood of producing vocal sequences rather than single vocal units when chimpanzees were exposed to concomitant events versus single events (Models 1a–1b). The third level tested agreement on the coding of utterance type, i.e., the different vocal units within an utterance (Cohen's Kappa score  $\pm$  SE:  $0.77 \pm 0.06$ ) (Figure S2). The utterance type was essential to calculate utterance diversity per individual which entered our second set of statistical analyses (Models 2a–2b–2c). These different levels of testing IRR complement each other to ensure replicability when coding the vocal repertoire (Figure S3). The subset of data used in the IRR test comprised 71.9% of single vocal units and 28.1% of vocal sequences which is representative of the whole dataset which comprised 73.1% of single vocal units and 26.9% of vocal sequences. Due to the time-consuming coding methods (an average of 150 utterances coded per day for an expert) and the large dataset, it was not feasible to include a bigger sample size in the IRR test. However, our IRR test included all call types and was representative of the distribution of the repertoire.

### Event coding

Although how to assign 'meaning' to vocalisations is an open debate,<sup>15</sup> playback experiments overwhelmingly demonstrate that information from a vocalisation conveyed to conspecifics is related to the event in which the vocalisation is emitted.<sup>17,84,85</sup> Hence, we defined the event relevant to the utterance emitted as (i) what the signaler was doing whilst vocalising, specifically the current activity (feed, rest or travel) or (ii) any current social interaction engaged in (e.g., groom, receive aggression, play) and/or (ii) attention/looking toward a change in the environment, the direction in which the signaler continues looking whilst signaling (e.g., third-party aggression, animal encounter, inter-party communication). In our sample, this led to 24 events that occurred during vocal production (Table 3). We limited the event classification to obvious changes that could be easily and reliably identified at the moment a vocal utterance was emitted. We excluded changes related to the vocal utterances that were detached in time (e.g., recruitment of conspecifics from long distances, resulting in others joining minutes later<sup>32</sup>), since those could not be reliably traced back to a given single vocalisation. To ensure the replicability of our event coding scheme, inter-rater

reliability (IRR) testing on ID and event of vocal production was conducted in the field during a standard half-day focal follow between T.B. and the RA. This particular IRR test based on 30 vocal events provided a Cohen's Kappa of 0.96 for chimpanzee ID and for events of vocal production a Cohen's Kappa of 0.853.

Without detailed analyses of utterance meaning, for the purposes of this study, we made several assumptions about which events related to utterances. First, considering that vocal sequences may keep the combined meaning of their single vocal units,<sup>8,11,12</sup> we assumed that utterances emitted during concomitant events potentially conveyed information to conspecifics about those events. A recent playback study on chimpanzees demonstrated that the information conveyed in the combination of two vocal units may be additive, hence, the sequence A + B retain the information of both units emitted singly, animal encounter and recruitment, respectively.<sup>11</sup> Therefore, when the emission of a vocal utterance corresponded to more than one co-occurring or sequential event we noted each event, leading to so-called 'concomitant events' (Table 2).

Concomitant events can co-occur, such as *approaching* a conspecific while *encountering a threat*, and can also occur sequentially, in rapid successions, such as *approaching* and then *affiliating* with a conspecific. Finally, concomitant events can both co-occur and occur sequentially, such as *approaching* and then *receiving aggression* while *encountering a threat*. For several of the concomitant events, there is no sequential order of the events, hence, we did not make a distinction between co-occurring and sequential events, coding the concomitant events alphabetically. We considered concomitant events as those that unfold whilst a vocal sequence continued to be emitted. For instance, an individual approaches a partner and immediately after, hugs the partner, with no intervening event. If during this sequence of events vocal emission continues, we consider this as a concomitant event (approach + affiliation). If vocal emission is produced only during the approach but not during the following affiliation, the approach alone is considered as the event defining the context of vocal emission.

Second, during many social interactions and environmental changes, individuals will be also either resting, traveling or feeding at the same time, such as approaching another whilst feeding, or playing whilst traveling. In order not to over-inflate the number of concomitant events, we assumed that social interactions are more salient than activity. Thus, we coded 'approach whilst eating' simply as 'approach'. Chimpanzees frequently add event-specific calls, like feed-associated calls, to, for example, contact calls, during inter-party exchanges or fusions (Personal Observations and Leroux et al.<sup>10</sup>). Hence, we exceptionally included the activity as an additional event when signallers were visually separated from receivers (during inter-party communication and fusions). The rationale here is that the chimpanzees might then be likely to communicate about the activity as well as long-distance since the receiver does not see what they are currently doing. In these situations, we coded concomitant events as 'fusion/inter-party communication + activity'.

Even though we took this conservative approach regarding the concomitant events, we repeated our analyses from Models 1a and 2a but this time considering two 'extreme' situations: (1) we always kept the activity in combination with another event so that each situation had one activity plus sometimes other events and (2) we removed activity in combination with other events in all situations including inter-party calling and fusions. When repeating model 1a, we found the same pattern, independent of the situation considered. Chimpanzees were more likely to produce single vocal units in single events, and vocal sequences in concomitant events when considering the 'always activity situation' (full-null model comparison:  $\chi^2 = 22.215$ ,  $df = 1$ ,  $p < 0.001$ ) and 'never activity situation' (full-null model comparison:  $\chi^2 = 48.231$ ,  $df = 1$ ,  $p < 0.001$ ). However, in the 'conservative situation', the model fitted the data better than in the other two extreme situations (conservative approach  $R^2m = 0.031$ , 'always situation approach'  $R^2m = 0.016$ , 'never activity situation'  $R^2m = 0.021$ ). For model 2a, we also found the same pattern of the 'conservative situation', with the 'always activity situation' (full-null model comparison:  $\chi^2 = 37.39$ ,  $df = 1$ ,  $p < 0.001$ ) and 'never activity situation' (full-null model comparison:  $\chi^2 = 38.784$ ,  $df = 1$ ,  $p < 0.001$ ). Here as well, the model fitted the data better in the 'conservative situation' ( $R^2m = 0.361$ ) than in the two extreme situations ('always situation approach'  $R^2m = 0.247$ , 'never activity situation'  $R^2m = 0.269$ ).

Whilst in some cases calls or call variants are acoustically specific to certain events (such as pant-grunts during approaches), other calls can also be produced in these events, and two different events could evoke the same vocal output. As our aim was to study the potential for utterances to relate to events, we only included events which elicited vocal utterances in our sample. Hence, behavioral events such as 'patrol', 'self-groom' or 'solo-play', which rarely elicit vocal utterances, were not included in the analysis. We are also aware that we consider the order of the vocal utterances (i.e., grunt+scream and scream+grunt as distinct), but do not consider the order of events (i.e., we consider affiliation+aggression and aggression+affiliation as the same). This constraint may underestimate the correlation between these two parameters. The main reason for the latter is that the sequential temporal order of the different events cannot always be determined, such as in co-occurring situations. From the 24 single events, only immature individuals (<10 years) vocalised in 'human directed' and 'bystander to copulation' events. We only included utterances where the event of production was unequivocal. The Tai forest is a quite open primary rainforest, which allows us to see callers, interaction partners and the events as they unfold for a vast majority of the recordings, as demonstrated by high IRR scores (see above). 913 of 10,304 utterances were excluded as we were unable to assign an event. 80.72% (N = 737) utterances were collected *ad libitum*, where visibility of what the vocaliser could see was sometimes occluded. The remaining 176 utterances with unassigned events were from focal individuals, 156 utterances of which were due to unclear 'Inter-party communication' context - as several individuals inside and outside of the party were vocalising at the same time.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Model 1a

To test our first prediction, we assessed whether chimpanzees were more likely to utter a vocal sequence when exposed to concomitant events as compared to single events using a binomial Generalized Linear Mixed Model (GLMM)<sup>86</sup> with log link function. The response in this model was whether the utterance was a vocal sequence (y/n) (no denoting single calls and yes denoting vocal sequences, i.e., utterance



with at least two different vocal units being combined). We used a categorical predictor corresponding to concomitant events (y/n), that is, whether the vocal utterance was emitted during a single or a concomitant event.

We used such a statistical approach since at the repertoire level, we did not expect a 1:1 matching of the number of events and the number of vocal units in the sequence. This was because, first, whilst we had up to 5 concomitant events, we recorded sequences of up to 10 vocal units. Second, some sequences may follow non-compositional rules, resulting in sequences with more units than the number of events. Third, if some events are more salient than others, such that not all concomitant events elicit vocal utterances, there may be cases where there are more events than vocal units in the utterance. In our statistical model, we controlled for group identity as a fixed effect. As we have several utterances per individual, we included caller identity as a random effect to avoid pseudo-replication. Moreover, for half of the individuals, we have data from two field seasons, collected one year apart. Hence, we also included a dummy variable `season_ID` as another random factor to account for repeated sampling of the same individual during the same field season, as the age of individuals changes from one season to the next and hence the events individuals engage in and their vocal repertoires might also change. To keep the type I error rate at the nominal level of 5% we included all identifiable random slopes.<sup>87</sup> Particularly, we included a random slope of our predictor variable (concomitant event y/n) within caller identity and within field season. To avoid concomitant testing, we conducted a full-null model comparison,<sup>88</sup> where the null model only lacked the predictor concomitant event (y/n), but included group identity as a control predictor and was otherwise identical to the full model. This comparison was based on a likelihood ratio test.<sup>89</sup>

### Model 1b

To study the effect of age on the relationship between the length of utterances and events, we fitted another model with the same structure as [model 1a](#), but with a subsample of the data containing only adults (*Model 1b*). We considered individuals to be immature when <10 years, given that individuals at this age have reached the adult vocal repertoire level regarding the length of utterances and diversity of unique utterances (i.e., the number of distinct utterances uttered).<sup>57</sup> Each vocal utterance constituted a data point. The sample for [model 1a](#) encompasses 9391 utterances from 98 individuals (0–55 years old) and the sample for *Model 1b* includes 5399 utterances from 53 adult individuals (10–55 years). In both models, we included data from the three groups.

The R script formula for Model 1(a-b): `glmer (vocal sequence y/n ~ concomitant event y/n + group identity + (1 + concomitant event y/n | caller identity) + (1 | field season), family = binomial)`.

### Model 2a

With the second model we wanted to evaluate the relation between the diversity of utterances (or “repertoire size”) and the diversity of events per individual; that is the number of unique or different vocal utterances that an individual produced (response) and the number of unique events in which each individual vocalised (predictor) ([Figure S1A](#)). As both variables were highly correlated with sample size (total number of utterances for a given chimpanzee and field season, [Figure S1A](#); [Figure 3A](#)), we first had to control for this effect. To do so, we fitted two linear regressions one for each of the desired variables (log-transformed) as a function of the sample size (log-transformed) and extracted the respective residuals (the difference of each observation of the respective response estimated fitted value) ([Figures S1C](#) and [S1D](#)). The R formula for the two initial models was:

Residuals diversity of utterances = `as.vector (residuals (lm (log (number unique utterances) ~ log (sample size))))`

Residuals diversity of events = `as.vector (residuals (lm (log (number unique events) ~ log (sample size))))`

As a second step, we modeled the relationship between the residuals of each model using a Linear Mixed Model (LMM)<sup>86</sup> with the residual diversity of utterances as the response and the residual diversity of events as the predictor variable. We used a Gaussian error distribution and identity link function ([Figure S1B](#)). We controlled for group identity as a fixed effect and for caller identity as a random intercept. We used the Satterthwaite approximation for the model estimated using restricted maximum likelihood REML (R package `lmerTest`).<sup>90</sup> We conducted a full-null model comparison,<sup>88</sup> where the null model lacked the predictor ‘residuals of the diversity of events’ but was otherwise identical to the full model. For this model, we used one data point per chimpanzee per field season which resulted in 152 data points from 98 individuals (0–55 years old) out of 3 groups.

The R script formula for model 2a: `lmerTest::lmer (Residuals of the diversity of utterances ~ 1 + Residuals of the diversity of events + group identity + (1 | caller identity), REML = T)`.

### Model 2b

To compare the relevance of considering the impact of concomitant events on utterance diversity, we fitted a new model similar to [model 2a](#), which only included the number of unique single events as a predictor. In this case, we broke down the concomitant events into their building units. For instance, when considering concomitant events as distinct, an individual vocalising in the following events: affiliation, approach and one concomitant event affiliation+approach, was coded as vocalising in three unique events ([model 2a](#)). When not considering concomitant events as distinct, in this example, the individual would be coded as vocalising in two unique events (*Model 2b*). All the other parameters and procedures of the model remained the same as in [model 2a](#). The sample for this model encompassed again 152 data points from 98 individuals (0–55 years old). We compared the effect size ( $R^2$ ) and AIC values of both models (2a and 2b) to assess which model fitted the data better.

### Model 2c

To analyze the effect of age on the strength of the relationship between the diversity of utterances and the diversity of events, we fitted an additional model exactly as for [model 2a](#), but with a subsample of the data including only adults (>10 years) (*Model 2c*). For this model, we used 73 data points from 51 individuals (10–55 years old), from 3 chimpanzee groups.

### Implementation

We fitted the models in R 4.1.1<sup>91</sup> using the functions `glmer` and `lmer` (R package `lme4`).<sup>92</sup> Before fitting the models, we tested collinearity between our predictor variables using the function `vif` (R package `car`), which did not reveal any issue (Variance Inflation Factor of all predictors <1.12).<sup>93</sup> After fitting the models, we assessed model stability for all models by excluding each level of each random effect one at a time (function provided by Roger Mundry); all models were stable.<sup>94</sup> For Model 2 (Linear Mixed Model), we also visually checked the assumptions of normal distribution and homogeneous residuals using QQ-plots of residuals and residuals plotted against fitted values.<sup>93</sup> The plots did not reveal obvious deviations from the assumptions. We assessed the significance of each predictor variable using an LRT between the full model and a reduced model comprising all the variables except the one being evaluated using the function `drop1`.<sup>92</sup> We calculated effect sizes using the function `r.squaredGLMM` (R package `MuMin`). We derived 95% confidence intervals by means of parametric bootstraps (function provided by Roger Mundry). We did not include sex and rank as predictors in any model, as it has been shown that these variables do not affect the response variables used here.<sup>57</sup>