https://doi.org/10.35662/unine-thesis-3213

PhD thesis in comparative cognition

(Doctorat ès sciences en Cognition Comparée)

Submitted to the Faculty of Science, Institute of Biology, University of Neuchâtel, Switzerland and the Doctoral School n°476 Neurosciences et Cognition, University Lyon 1 Claude Bernard, France

CALL COMBINATIONS IN SOOTY MANGABEYS

Auriane LE FLOCH

Defended on August 27th, 2025

Thesis directors

Catherine CROCKFORD / Klaus ZUBERBÜHLER

Thesis co-supervisors

Roman WITTIG / Steven MORAN

Expert committee composition

Mélissa BERTHET, University of Rennes, France

Judith BURKART, University of Zurich, Switzerland

Marion RICHARD, University of Lyon 1, France

Robert SEYFARTH, University of Pennsylvania, USA

Simon TOWNSEND, University of Zurich, Switzerland

unine
Université de Neuchâtel
Faculté des sciences

Rue Emile-Argand 11
CH-2000 Neuchâtel
doctorat.sciences@unine.ch

IMPRIMATUR POUR THESE DE DOCTORAT

La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par

Madame Auriane LE FLOCH

Titre :

"Call combinations in sooty mangabeys"

sur le rapport des membres du jury composé comme suit :

- **Prof. Klaus Zuberbühler,** directeur de thèse, Université de Neuchâtel, Suisse
- **Dr Catherine Crockford,** directrice de thèse, CNRS, France
- **Prof. ass. Steven Moran,** co-directeur de thèse, Université de Neuchâtel, Suisse
- **Dr Mélissa Berthet,** Université de Rennes, France
- **Dr Marion Richard,** Université de Lyon 1, France
- **Prof. Judith Burkart,** Université de Zurich, Suisse
- **Prof. ass. Simon Townsend,** Université de Zurich, Suisse
- **Prof. Robert Seyfarth,** University of Pennsylvania, USA

Neuchâtel, le 21 octobre 2025

Le Doyen, Prof. P. Brunner

Summary

Human language is built on the ability to generate infinite new meanings by combining smaller units according to structured rules. Understanding the evolutionary origins of this capacity requires a comparative approach examining how nonhuman animals use vocal combinations. Across taxa, many species produce vocal sequences by combining calls, and these combinations can generate specific meanings, often through mechanisms such as call order or reoccurrence. However, how widespread call combinations are within species' vocal systems—and what mechanisms support them—remains poorly understood. To date, only a few studies, mostly in apes, have investigated these processes at the scale of full vocal repertoires.

With this thesis, I address this gap by investigating how vocal sequences contribute to meaning in the communication system of sooty mangabeys *(Cercocebus atys)*, using a combination of naturalistic observations and experimental approaches. I focus on two central questions: (1) What mechanisms are used to generate meaning through call combinations, such as combining distinct calls, altering their order, or reoccurrence of specific call types? and (2) How widespread is this use of combination for meaning across the vocal repertoire?

First, I conducted a full-system analysis of all vocal sequences observed in sooty mangabeys. This revealed that while many sequences are rare or unstructured, a subset followed consistent organisational patterns, including fixed call order, iteration and affixation-like structures. These findings show that sooty mangabeys use rule-based vocal sequencing. Second, I tested whether these sequences carry potential meaning. Natural observations showed that order-sensitive bigrams—such as 'twitter_grunt'—were more context-specific than their component calls, and that iteration further modulated contextual use. Third, predator model experiments revealed that the alarm call 'shrill' could be produced alone or in an affix-like structure with an added 'hoo' call. These vocal utterances appeared to encode specific threat-related information, supporting the idea that combinations may enrich communicative content even in the absence of predator-specific alarm calls.

These findings provide a detailed account of how vocal combinations are used to generate meaning in a nonhuman species. They show that different combinatorial mechanisms—ordering, iteration, and affixation—are present within a single species and are associated with specific contexts of production: affiliative contexts, alarm

contexts, and, in the case of aggression, suggest the presence of sequences with no consistent structure. This likely reflects ecological and social pressures such as predation risk and the demands of managing social interactions. Compared to apes, sooty mangabeys combine calls across their repertoire less extensively and rely on a more restricted subset of rule-based sequences. This suggests that extensive compositionality across the vocal repertoire may have evolved after the evolutionary split between catarrhine monkeys and great apes, rather than being shared across the primate lineage. More broadly, this thesis provides new insights into the diversity, function and evolutionary origins of vocal combinatoriality in animal communication.

Keywords: primate vocal communication; meaning; vocal sequences; language evolution

Résumé en français

Le langage humain repose sur la capacité à générer une infinité de nouveaux sens par la combinaison d'unités plus petites selon des règles structurées. Comprendre les origines évolutives de cette capacité nécessite une approche comparative examinant comment les animaux non humains utilisent des combinaisons vocales. De nombreuses espèces, dans divers groupes taxonomiques, produisent des séquences vocales par combinaison de cris, et ces combinaisons peuvent générer des significations spécifiques, souvent via des mécanismes comme l'ordre ou la récurrence des cris. Toutefois, l'ampleur de ces combinaisons dans les systèmes vocaux des espèces, et les mécanismes qui les sous-tendent, restent mal compris. À ce jour, seules quelques études, principalement chez les grands singes, ont abordé ces questions à l'échelle de répertoires vocaux complets.

Cette thèse vise à combler cette lacune en explorant la manière dont les séquences vocales contribuent à la signification dans le système de communication des mangabeys fuligineux *(Cercocebus atys)*, en combinant observations naturalistes et approches expérimentales. Deux questions principales sont abordées : (1) Quels mécanismes permettent de générer du sens par combinaison de cris, tels que l'enchaînement de cris distincts, la modification de leur ordre ou la récurrence de certains types de cris ? et (2) Dans quelle mesure cette combinaison à visée communicative est-elle utilisée à travers le répertoire vocal ?

Premièrement, j'ai réalisé une analyse systématique de toutes les séquences vocales observées chez les mangabeys fuligineux. Cette analyse montre que, bien que de nombreuses séquences soient rares ou non structurées, un sous-ensemble suit des schémas organisationnels constants, incluant un ordre fixe des cris, la réitération et des structures proches de l'affixation. Ces résultats révèlent l'usage de séquences vocales basées sur des règles. Deuxièmement, j'ai examiné si ces séquences véhiculent une signification. Les observations ont montré que certains bigrammes sensibles à l'ordre – comme 'twitter_grunt' – étaient plus spécifiques à un contexte donné que les cris pris isolément, et que la réitération modulait davantage l'usage contextuel. Troisièmement, des expériences avec des modèles de prédateurs ont révélé que le cri d'alarme 'shrill' pouvait être émis seul ou dans une structure de type affixe avec l'ajout du cri 'hoo'. Ces émissions semblaient encoder des informations

spécifiques liées à la menace, suggérant que les combinaisons enrichissent le contenu communicatif même en l'absence de cris d'alarme spécifiques aux prédateurs.

Ces résultats fournissent une description détaillée de l'usage des combinaisons vocales pour générer du sens chez une espèce non humaine. Ils montrent que divers mécanismes combinatoires – ordre, réitération, affixation – coexistent au sein d'une même espèce et sont associés à des contextes de production distincts : contextes affiliatifs, contextes d'alarme, et, en cas d'agression, présence de séquences sans structure régulière. Cela reflète probablement des pressions écologiques et sociales telles que le risque de prédation et les exigences liées aux interactions sociales. Comparés aux grands singes, les mangabeys fuligineux combinent leurs cris de manière moins étendue et reposent sur un sous-ensemble plus restreint de séquences basées sur des règles. Cela suggère que la compositionalité étendue du répertoire vocal pourrait être un trait évolutif spécifique aux grands singes, plutôt qu'une caractéristique commune à l'ensemble des primates. Plus largement, cette thèse apporte un nouvel éclairage sur la diversité, la fonction et les origines évolutives de la combinatoire vocale dans la communication animale.

Mots-clés: communication vocale des primates; sens; sequences vocales; évolution du language

Résumé substantiel en français

Pour mieux comprendre comment les combinaisons de cris contribuent à la signification dans la communication animale non humaine, cette thèse adopte une approche comparative et systémique en proposant une évaluation quantitative détaillée du système vocal du mangabeys fuligineux *(Cercocebus atys)*. En analysant non seulement les cris isolés, mais aussi leurs séquences et leur contexte de production, l'objectif est d'identifier les mécanismes par lesquels le sens est généré ou modifié, ainsi que l'étendue de ces mécanismes dans le répertoire vocal. Ce travail contribue ainsi à mieux retracer les origines évolutives du langage.

Plus précisément, cette thèse explore comment les séquences vocales participent à la transmission de sens dans le système de communication du mangabeys fuligineux, un singe forestier vivant dans le Parc National de Taï, en Côte d'Ivoire. Grâce à une approche exhaustive et fondée sur les données, j'examine comment différents mécanismes combinatoires, tels que l'ordre des cris, leur récurrence (itération) et l'ajout d'éléments semblables à des affixes, sont utilisés, seuls ou en combinaison, pour générer ou modifier le sens. Les données ont été collectées sur plusieurs saisons de terrain, totalisant 15 mois, auprès de deux groupes sauvages habitués à la présence humaine, ce qui a permis une étude détaillée de leur comportement vocal naturel, à la fois par observation et par expérimentation.

Le premier chapitre propose une caractérisation complète de la structure des séquences vocales à travers le répertoire du mangabeys fuligineux. À partir d'observations naturelles et de modèles bayésiens, il quantifie la fréquence et la diversité des séquences, et identifie des régularités structurelles systématiques, incluant des règles d'ordre, des schémas itératifs, et des cas d'organisation hiérarchique. L'analyse montre que si de nombreuses séquences sont rares ou semblent non structurées, un sous-ensemble limité est produit fréquemment et de manière systématique, selon des arrangements non aléatoires et fondés sur des règles. Ces résultats suggèrent que les mangabeys fuligineux produisent régulièrement des séquences vocales, mais reposent principalement sur un ensemble restreint de combinaisons structurées, chacune pouvant remplir des fonctions communicatives distinctes.

Le deuxième chapitre s'intéresse au potentiel sémantique des combinaisons de cris, en analysant la distribution contextuelle des séquences les plus fréquemment

produites en observation naturelle — celles composées des appels ‘grunt’ et ‘twitter’. En utilisant des analyses de distances euclidiennes, l’étude teste si les bigrammes transmettent une information plus spécifique au contexte que les appels pris isolément, et si l’itération — la réapparition d’un type d’appel dans une séquence — modifie cette spécificité contextuelle. Les résultats indiquent que certaines séquences, en particulier ‘twitter_grunt’, sont plus étroitement associées à des contextes sociaux spécifiques que leurs appels composants, et que l’itération peut renforcer cette association. Ces observations soutiennent l’idée que les séquences vocales des mangabeys peuvent fonctionner de manière compositionnelle, avec un sens qui émerge à la fois de la structure et de la répétition.

Le troisième chapitre examine les réponses vocales à des expériences avec des modèles de prédateurs, pour évaluer si les séquences d’alarme véhiculent des informations supplémentaires par rapport aux cris pris individuellement. En s’appuyant sur des modèles bayésiens, l’analyse se concentre sur le cri ‘shrill’, souvent produit seul mais pouvant aussi apparaître avec un élément supplémentaire, le ‘hoo’, jouant un rôle d’affixe. Ces séquences semblent liées à l’information sur les prédateurs, possiblement en signalant le type ou le degré d’urgence de la menace. Ces résultats suggèrent que les mangabeys fuligineux utilisent des structures similaires à l’affixation pour enrichir de manière flexible le contenu des signaux dans les contextes d’alarme, même en l’absence de cris d’alarme spécifiques à chaque prédateur.

Ensemble, les trois chapitres adoptent une approche basée sur l’ensemble du répertoire vocal pour explorer la structure et le sens dans un système de communication primate. En combinant des observations naturelles détaillées avec des données expérimentales, ce travail examine à la fois les comportements vocaux fréquents et des contextes plus rares tels que les rencontres avec des prédateurs. Cette perspective intégrée et large dépasse les études centrées sur un seul type de cri ou de contexte, en offrant une compréhension plus complète du système vocal des mangabeys fuligineux. La présence de multiples mécanismes générateurs de sens, ordre, itération, affixation et compositionnalité, suggère que les fondements cognitifs de la combinatoire vocale sont diversifiés et dépendants du contexte. Les résultats montrent également que, comparés aux grands singes, les mangabeys fuligineux combinent leurs cris de manière moins étendue et reposent sur un sous-ensemble plus restreint de combinaisons, ce qui soutient l’hypothèse d’une diversification et d’une flexibilité compositionnelles accrues dans l’évolution des grands singes. Dans

l’ensemble, ma thèse apporte de nouveaux éclairages sur le fonctionnement des combinaisons vocales dans la communication animale, et sur ce qu’elles révèlent des bases évolutives du langage.

Remerciements/Acknowledgements

I would like to thank Cathy, Roman, Klaus, and Steven for giving me this amazing opportunity, supporting and guiding.

Je remercie le Ministère des Eaux et Forêts et l'Office Ivoirien des Parcs et Réserves en Côte d'Ivoire de m'avoir permis de prendre des données dans le parc national de Taï. Special thanks to Christophe and Hedwig Boesch for establishing and nurturing the Taï Chimpanzee Project, Anderson Bitty, the Centre Suisse de Recherches Scientifiques, field assistants, and all members of the Taï Chimpanzee and Taï Monkey Projects for their invaluable support.

I am very grateful to Florence Lévrero and Liuba Papeo for their invaluable feedback and support as members of my Comité de suivi individuel de thèse. I also wish to thank the members of my viva committee: Simon Townsend, Robert Seyfarth, Judith Burkart, Marion Richard, and Mélissa Berthet, for evaluating my work and for their insightful feedback and stimulating discussion.

Je voudrais remercier ma mère Martine et mon père Hervé, Manon, Clément, Maël, Bekzat, Ahmed, Nathalie, Yvette ainsi que Chloé et Alice et toute ma famille pour leur soutien. Merci à Tanit et Natacha, mes comparses mangabeys. Merci Natacha également pour les magnifiques photos que tu as prise en forêt, dont celle sur la photo de couverture de cette thèse.

I would also like to thank all forest friends and colleagues on whom I could rely for advice and support: Elé, Carme, Anaïs, Ane, Tushna, Tina, Mathieu, Kim, Ryan, Fanny, Constant, Narcisse, Simon, Colby, Paterson, Boddo, Ernest, Giuliana, Oscar, Antoine, Bosco, Giacomo, and Marine.

I would like to thank primatologists who advised me and or supported me during the preparation of this work: Cédric, Alex, Mélissa, Maël, Julian, Mathilde, Tatiana, Derry, Esa, Caroline, Lara, Larissa, Noémie, Adrian, Mathieu, Sarah, Emilie, Bastien, and the primatologists from Zurich.

Enfin, je souhaite remercier tous et toutes mes am.i.e.s de Toulouse, Lyon, Neuchâtel et Paris.

Table of content

General introduction

Non-human animals and the origins of language

Humans possess a unique form of communication, the language, that is unmatched in its structure and expressive power in other animal species. This system relies on two core levels of organisation: phonology, which governs how meaningless sound units (phonemes) are combined into meaningful ones (morphemes), and morphosyntax, which organises these morphemes into words and sentences (Hurford, 2012). For instance, the phonemes /c/, /a/, and /t/ combine to form the morpheme “cat,” which carries meaning, and can then be used in more complex structures like “The cat sleeps.” These rules are hierarchical and generative, enabling humans to produce an unbounded number of utterances from a finite set of elements. One of their most distinctive features is recursion, the ability to embed units within other units to create potentially infinite combinations of sounds and meanings (Boeckx, 2011; Chomsky, 1956).

Because no fossils of language exist, comparative studies of animal communication systems are essential for investigating the origins and evolution of human language (Bergman et al., 2019; Darwin, 1981; Fitch, 2010; Hauser, 2002; Leroux & Townsend, 2020; Townsend et al., 2018; Zuberbühler, 2017). Such studies focus not only on our closest relatives, nonhuman primates (hereafter referred to as primates), but also on more distantly related species that have independently evolved combinatorial systems. This broader comparative approach allows researchers to investigate both shared ancestry and convergent evolution, shedding light on the evolutionary pressures and constraints that gave rise to vocal communication systems involving syntactic-like organisations.

Interestingly, some human languages use as few as 11 phonemes, demonstrating that speech can rely on a limited number of sounds (Maddieson, 1984; Moran et al., 2012). This range overlaps with the size of many primate vocal repertoires (McComb & Semple, 2005) and suggests that the number of available sounds is not a constraint for generating new meanings, or even language. Instead, the key lies in the capacity to flexibly combine vocal units into a wide array of utterances that create meanings. To understand how such combinatorial capacities may have evolved, we must investigate how other species utilise vocal elements in combination and ask whether these combinations, like those found in language, serve to generate meaning.

Building on 40 years of research into the meaning of single calls in animal communication, the past two decades have increasingly focused on whether combinations of calls can also convey meaning (Berthet et al., 2023; Zuberbühler, 2019). This line of research began with studies on vervet monkey alarm calls during predator encounters, where acoustically distinct calls were mapped onto specific external threats (Struhsaker, 1967). Playback experiments later confirmed that, even in the absence of a predator, these calls elicited distinct behavioural responses in receivers, indicating that listeners extracted different meanings (Seyfarth et al., 1980a, 1980b). Similar findings have since been reported in other species, including other monkeys (Zuberbühler, 2001; Zuberbühler et al., 1997), mongooses (Manser, 2001), and birds (Cunningham & Magrath, 2017; Grieves et al., 2014; Suzuki, 2012), prompting broader questions about whether and how meaningful combinations of calls occur across taxa.

A growing number of studies now suggest that combining calls into vocal sequences is a widespread feature of animal communication (see Engesser & Townsend, 2019 for review). These combinations can generate specific meanings, often through mechanisms such as call order or reoccurrence within the sequence (Berthet et al., 2023). Observational and experimental work has shown that some combinations are associated with specific contexts (e.g., chimpanzees (Girard-Buttoz et al., 2025), elephants (Hedwig & Kohlberg, 2024), meerkats (Collier et al., 2017), Campbell's monkeys (Ouattara et al., 2009a), and complementary playback experiments indicate that receivers can make use of such signals to obtain information (Berthet et al., 2019; Clay & Zuberbühler, 2011; Engesser et al., 2016). However, we still know very little about how widespread these mechanisms are across vocal repertoires. Most research to date has focused on a limited number of call combinations within a given species, often restricted to a single context such as alarm calling (e.g., olive colobus (Gallot et al., 2024), titi monkeys (Berthet et al., 2018, 2019), Japanese great tits (Suzuki et al., 2016)). While these studies have yielded key insights into how animals use sequences to convey meaning, the extent to which such mechanisms operate across a broader range of contexts remains largely unexplored, leaving open the question of whether meaningful call combinations are rare exceptions or an integral part of vocal communication systems. Notably, only a handful of studies, primarily on great apes, have investigated vocal combinations at the level of the full repertoire. System-level approaches have been applied in gorillas (Hedwig et al., 2014), bonobos (Berthet et al., 2025), and chimpanzees (Girard-Buttoz et al., 2022, 2025), as well as in a few non-

ape species such as marmosets (Bosshard et al., 2024), meerkats (Collier et al., 2017), and whales (Sharma et al., 2024) revealing that call combinations tend to occur across a wide range of contexts. Beyond these examples, however, comprehensive studies that link combination structure to production context across the entire vocal system remain scarce.

Understanding the scope and mechanisms of vocal combination across full repertoires is essential for several reasons. First, it allows us to assess whether combinations occur in isolated instances or reflect a general strategy for encoding meaning. Second, it helps us compare species using quantitative measures and explore how ecological and social factors shape communicative strategy. Finally, it provides a necessary empirical foundation for evaluating whether aspects of human language, such as the use of combinatorial rules to create meaning, have evolutionary parallels in other species.

This thesis addresses these gaps by investigating how vocal sequences contribute to meaning in a nonhuman animal system, using a full-repertoire, data-driven approach. In the next section, I introduce the mechanisms by which animals may generate meaning through vocal sequences, ranging from compositionality to call order, reoccurrence, and affixation, and explain how each contributes to communicative function.

Mechanisms for meaning generation in vocal sequences

Vocal sequences in animals can convey information not only through the presence of specific calls, but also through how those calls are combined (Suzuki, 2025). A range of mechanisms may underlie the generation or modification of meaning in vocal combinations, including call order, reoccurrence patterns, and affixation (Berthet et al., 2023). I also report hierarchical organisation, even though it is not clear yet whether such structures can also generate meaning. All such mechanisms for meaning generation are not mutually exclusive. They may operate compositionally, where the meaning of a sequence derives from the meanings of its components and their arrangement, or non-compositionally, where a sequence conveys a holistic meaning distinct from its elements (Arnold & Zuberbühler, 2012). Importantly, whether a given combination can generate specific meaning also depends on how widely it is used across different contexts. If a combination is used flexibly across a wide range of

situations, then it is less likely that listeners can infer a specific meaning from the combination alone, making contextual cues necessary for interpretation. In contrast, when a combination occurs consistently in a restricted set of contexts, there is greater potential for listeners to attribute meaning to the combination itself (Berthet et al., 2023). Below, I outline the main types of mechanisms identified in the animal communication literature.

Call order

The order in which calls appear in a sequence can influence its meaning. In some species, rearranging the same elements leads to distinct communicative outcomes. For example, chestnut-crowned babblers produce two functionally distinct vocalisations, flight calls ('AB') and prompt calls ('BAB'), by combining the same acoustic elements in different orders (Engesser et al., 2015). Similarly, olive colobus monkeys arrange the same units ('A', 'B', 'C') in different sequences depending on the type of danger (Gallot et al., 2024). In chimpanzees, 'grunt' + 'hoo' sequences occur more often during travel and fusion events, whereas 'hoo' + 'grunt' sequences are more typical of resting contexts (Girard-Buttoz et al., 2025). Ordering can also occur through gradation, where vocal units are arranged in a specific order, creating a pattern that moves from one type to another. For example, meerkats use graded sequences to signal increasing danger levels, starting with less urgent calls and moving to more intense ones as the threat escalates (Rauber et al., 2020). Similarly, titi monkeys begin their sequences with high-urgency calls when facing potential danger and then transition to less urgent social calls, providing both immediate warning and additional social information (Narbona Sabaté et al., 2022).

Call reoccurrence

The reoccurrence of specific call types may also shape sequence meaning. This can involve adjacent repetition (e.g., ABBB) and/or non-adjacent iterations (e.g., ABA). For instance, in bonobos, playback experiments suggest that such reoccurrences may influence meaning, as sequences containing multiple instances of specific call types (e.g., 'barks' and 'peeps') elicited stronger foraging responses from receivers (Clay & Zuberbühler, 2011). However, this effect might also be explained by broader structural

factors, such as the overall proportion of high-preference call types within a sequence (Clay & Zuberbühler, 2009) or the presence of specific bigrams embedded within longer call sequences (Zuberbühler, 2020). Conversely, pied babblers produce call types defined by the number of repeated vocal elements, each conveying different information about food (Engesser et al., 2017). Chickadees and great tits vary the number of vocal elements in their mobbing calls to signal information about predator presence or urgency (Freeberg, 2008; Salis et al., 2022). Similarly, male putty-nosed monkeys increase the number of 'pyow' and 'hack' calls to indicate the distance the group should travel (Arnold & Zuberbühler, 2012). Encoding meaning can also be achieved through the overall proportion of call types within a sequence. Titi monkeys, for example, use combinations of 'A' and 'B' calls to probabilistically encode information about predator type and location, with the proportion of 'B' call pairs influencing the conveyed meaning (Berthet et al., 2019). Similarly, Amboli bush frogs vary the proportion of two call types depending on the presence of territorial rivals (Bhat et al., 2022).

Call affixation

A well-known example of affixation comes from Campbell's monkeys, where males add a suffix '-oo' to alarm calls to broaden their meaning: for instance, 'krak' refers to leopards, while 'krak-oo' indicates a more general disturbance (Coye et al., 2015; Ouattara et al., 2009b). Similar structures are found in social contexts. In Campbell's monkeys, as well as Diana monkeys and mongooses, calls are occasionally combined with an additional unit that refines or emphasises identity-related information, rather than altering the core meaning of the call (Candiotti et al., 2012; Coye et al., 2018; Jansen et al., 2012).

Compositionality

Compositionality refers to the principle by which individual calls are combined to form sequences that convey information different from that expressed by each composing call when produced in isolation and has been suggested to operate in several animal vocal systems (Engesser & Townsend, 2019; Gontier et al., 2024; Townsend et al., 2018; Zuberbühler, 2020). Interestingly, growing evidence suggests that

compositionality may occur in a range of distantly related species, although it has not always been explicitly tested, particularly through playback experiments which are traditionally used to confirm whether receivers infer meaning.

Studies using playback experiments have provided more direct evidence that receivers interpret combinations compositionally. In chimpanzees, the ‘alarm-huu’ + ‘waa-bark’ sequence is produced during snake encounters, whereas the individual calls occur in broader contexts (‘alarm-huus’ in response to potential threats and ‘waa-barks’ for general recruitment). Such combination appears to be interpreted by receivers as a recruitment call in a dangerous situation (Leroux et al., 2023). Similarly, in southern pied babblers, alert calls to low-urgency threats and recruitment calls used during locomotion are combined into a ‘mobbing sequence’ when encountering terrestrial predators, eliciting appropriate response in receivers (i.e., quickly approaching the speaker) (Engesser et al., 2016). In Japanese great tits, ‘ABC’ calls trigger predator-scanning and ‘D’ calls attract conspecifics, but when combined as ‘ABC–D’, receivers both scan and approach (Suzuki et al., 2016). In Campbell’s monkeys, ‘hok’ calls are produced in response to eagles and ‘krak’ calls to leopards, but when combined with a ‘hoo’ suffix or preceded by a ‘boom’ call, these calls occur in more general, less urgent alarm contexts, also showing appropriate responses in receivers (Coye et al., 2015; Ouattara et al., 2009b; Zuberbühler, 2002). In putty-nosed monkeys, groups travel farther in response to ‘pyow-hack’ sequences than to either ‘pyow’ calls, which are general alarms, or ‘hack’ calls, which signal eagle presence (Arnold & Zuberbühler, 2008). In dwarf mongooses, aerial and terrestrial alarm calls are produced in response to specific threats, while their combination signals a general alarm (Collier et al., 2020).

Other studies, although lacking playback evidence, suggest compositionality based on contextual changes in production. In male Humayun frogs, single ‘AN’ notes are produced by callers when alone and are combined with a ‘DN’ notes more frequently in the presence of a rival male (Bhat et al., 2022). In chimpanzees, ‘pant-hoots’ are used across various contexts, while ‘food-calls’ occur during feeding; the combination of both is more frequent in high-value feeding contexts, such as when large food patches are available or when a high-ranking individual joins the party (Leroux et al., 2021). In bonobos, the combination ‘whistles’ + ‘high-hoots’ is more strongly associated with the caller travelling to a new party than when ‘high-hoots’ are produced alone (Schamberg et al., 2016), and ‘low-hoots’ + ‘high-hoots’ occurs more often than ‘high-hoots’ alone before travel events. ‘Low-hoots’ + ‘high-hoots’ is also more likely to elicit

inter-party recruitment, with individuals from other parties approaching the caller (Schamberg et al., 2017). In forest elephants, single ‘rumbles’ are produced across several social contexts, while ‘broadband’ calls occur mainly in resource-related situations; combined ‘rumbles’ and ‘broadband’ calls are observed more broadly and particularly in high-stakes contexts (Hedwig & Kohlberg, 2024).

Taken together, these examples demonstrate that a call used singly and in combination can occur in different contexts, supporting the idea that combinations can function as meaning modifiers, altering or expanding the information conveyed to receivers.

Idioms

Another well-known example is in male putty-nosed monkeys, where a sequence of two distinct alarm calls, ‘pyow’ and ‘hack’, conveys a third, distinct meaning: “let’s move” that is not reducible to the meanings of either call alone (Price et al., 2009; Schlenker et al., 2016). Such combinations resemble idioms in human language (e.g., “kick the bucket”), which require listeners to override the literal meanings of component units. The cognitive mechanisms needed to interpret such non-compositional sequences appear to partially overlap with those used to process compositional utterances (Suzuki et al., 2020). Indeed, understanding idioms entails accessing the meanings of individual components but choosing to disregard them in favour of a third, holistic meaning (Peterson et al., 2001).

Hierarchical structures

In human language, hierarchical structure refers to the recursive embedding of elements within larger constituents, allowing increasingly complex expressions to be built (e.g., “the dog bites the man” → “who owns a cat” → “which has a black tail”) (Berwick & Chomsky, 2016; Chomsky, 1995). In animal vocal communication, hierarchical structures have been proposed when combinations of vocal units follow layered rules, such that the result of one combination becomes the starting point for another (Berthet et al., 2023). In other words, the output of one rule feeds into a second rule, creating a sequence composed of smaller, organised building blocks. However, most meaningful call combinations described in animals involve only two units arranged in sequence, interpreted either linearly (e.g., ‘A’ then ‘B’ means X) or through

variation in repetition (e.g., more ‘A’ than ‘B’ in an ‘AB’ sequence means X, whereas the reverse means Y), neither of which constitutes hierarchical structure as defined in human language (Suzuki et al., 2020). Nonetheless, the animal literature reports three mechanisms analogous to those described in human language, which may involve some degree of hierarchical structuring. As their role in meaning production has not yet been demonstrated, however, such interpretations should be treated with caution.

The first is recursion, where vocal units are repeated in a nested way to create longer, more complex sequences (Boeckx, 2011; Chomsky, 1956; Martins, 2012). For example, a sequence pattern ‘AB’ can expand recursively into ‘AB-ABC-AB’, where part of the pattern is repeated within itself, a mechanism only observed in orangutan vocal sequence production so far, without clear implications on the meaning (Lameira et al., 2024) (but see Bosshard et al., 2024 for hierarchical nesting of sequence in marmosets).

The second is recombination, where sequences are expanded (e.g., ‘AB’ becomes ‘ABC’), and has, to my knowledge, been formally assessed only in chimpanzees (Girard-Buttoz et al., 2022) and marmosets (Bosshard et al., 2024). Although not explicitly framed this way in their respective studies, recombination may be present in putty-nosed monkey combinations of ‘pyow’ and ‘hack’ calls that appear chunked together in larger structures (Mehon et al., 2024) or underlie the differences between ‘ABC’ and ‘ABC-D’ sequences in tits (Suzuki et al., 2017), as well as between ‘RU-RO’ and ‘RU-RO-RU’ sequences in elephants (Hedwig et al., 2021), though the latter could also be interpreted as iteration.

The third is long-distance dependencies, where a relationship between two vocal units in a sequence are separated by intervening vocal units, and whose connection cannot be accounted for by adjacent combinations alone (Wilson et al., 2020). For instance, one unit in a sequence influences the choice of a later unit (e.g., ‘A-AB’ becomes ‘A-AB-A’, while ‘B-AB’ becomes ‘B-AB-C’). This mechanism has only been found in bird songs (Markowitz et al., 2013; Searcy et al., 2022) and marmosets (Bosshard et al., 2024). Interestingly, although vocal sequences in animals sometimes follow hierarchical structures, their implications for meaning remain unclear. Yet such rules could play a crucial role in enhancing communicative flexibility.

Taken together, these mechanisms of ordering, reoccurrence, affixation, compositionality, idiom, and possibly hierarchical structure, illustrate that vocal combinations can modify, refine, or expand meaning in multiple ways. Understanding how different species use these strategies, and under what ecological and social conditions, can help identify the building blocks of more complex communication systems such as human language.

Sooty mangabey as a model species

To better understand how vocal combinations function and the extent to which different combinatorial mechanisms are widespread, it is crucial to compare species that are both ecologically and phylogenetically close. Sooty mangabeys *(Cercocebus atys)* offer an especially valuable case for such a comparative approach with apes, for whom vocal sequence systems have already been characterised in detail (Berthet et al., 2025; Girard-Buttoz et al., 2022, 2025).

They are cercopithecine monkeys, an evolutionary branch that, while distinct from apes, remains closely related within the primate lineage (Pozzi et al., 2014). Comparing them with apes can therefore shed light on which vocal capacities are likely to be evolutionarily shared across primates, and which might be specific to apes. This comparison is particularly relevant given that sooty mangabeys and apes, especially chimpanzees and bonobos, share several key socio-ecological traits, as described below.

Sooty mangabeys live in large multi-male, multi-female social groups of up to 100 individuals (Range et al., 2007). They are highly terrestrial, spending most of their time on the ground in dense West African rainforest (Gba et al., 2020). This makes their daily experience strikingly similar to that of chimpanzees and bonobos, who also live in large, complex groups and can occupy similar forest environments (Boesch & Boesch-Achermann, 2000; Furuichi & Thompson, 2008). As a result, sooty mangabeys face comparable communicative challenges: maintaining cohesion in visually obstructed habitats, engaging in frequent social interactions within large groups, and responding effectively to multiple predator types (Ehardt, 1988; Fruteau et al., 2011; Gust, 1995; Gust & Gordon, 1993; León et al., 2023; Mielke, Crockford, et al., 2020; Mielke et al., 2019; Mielke, Preis, et al., 2020; Range et al., 2007; Range & Noë, 2002).

The need to generate various meanings through different vocal utterances, possibly including sequences, has been suggested for species facing specific social and ecological pressures, such as (1) living in large social groups where managing social interactions is critical (Bouchet et al., 2013; McComb & Semple, 2005), (2) inhabiting dense forests where visibility is limited and vocal signals are used for maintaining cohesion (Coye et al., 2018; Marler, 1975), and (3) being subject to predation pressures that demand threat differentiation (Stephan & Zuberbühler, 2008; Zuberbühler, 2001). Sooty mangabeys exemplify all three conditions: they occupy dense West African rainforest, live in large social groups, and face significant predation pressure from leopards, eagles, chimpanzees, and humans (McGraw et al., 2007). Moreover, they are known to produce vocal sequences relatively widely across their repertoire (Range & Fischer, 2004; Sigmundson et al., 2025), making them a valuable model for studying meaning generation in sequences using compositionality and rules.

Taken together, these features make sooty mangabeys an ideal species for examining the structure and meaning of vocal sequences within a full vocal repertoire. They are ecologically comparable to apes, being closely related phylogenetically, which allows for inferences about the evolution of mechanisms that support meaning generation in primates more broadly.

Thesis aims and structure

To better understand how call combinations contribute to meaning in nonhuman animal communication, this thesis adopts a comparative and system-level approach, offering a detailed quantitative assessment of the sooty mangabey vocal system. By analysing not just isolated calls but their combinations and contexts, it aims to identify the mechanisms through which meaning is generated or modified, and how widespread these mechanisms are across the repertoire. In doing so, this work contributes to broader efforts to trace the evolutionary roots of language.

Specifically, I investigate how vocal sequences contribute to meaning in the communication system of sooty mangabeys (*Cercocebus atys*), a forest-dwelling cercopithecine monkey inhabiting Taï National Park, Ivory Coast. Using a full-repertoire, data-driven approach, I examine how different combinatorial mechanisms such as call order, iteration, and affixation are used, alone or combined, to generate or modify meaning. Data were collected over multiple field seasons in 2022 and 2023,

spanning a total of 15 months, from two wild groups habituated to human presence, allowing detailed observational and experimental study of their vocal behaviour.

The first chapter provides a comprehensive characterisation of vocal sequence structure across the sooty mangabey repertoire. Using natural observations and Bayesian models, it quantifies the frequency and diversity of vocal sequences and identifies consistent rules, including call order, iterative patterns, and possible instances of hierarchical organisation. The analysis reveals that while many sequences are rare or appear unstructured (i.e., without rule), a limited subset is produced frequently and systematically, following non-random, rule-based arrangements. These findings suggest that while sooty mangabeys regularly produce in vocal sequence they primarily rely on a restricted set of rule-based combinations, each potentially serving distinct communicative functions.

The second chapter focuses on the potential for meaning in call combinations by analysing the contextual distribution of the sequences in natural observations. Using Euclidean distance analyses, the study examines whether bigrams or iterations (i.e., the reoccurrence of a call type within a sequence) vary systematically depending on context, suggesting that contextual factors influence the structure and usage of vocal sequences. Results indicate that the bigram ‘twitter_grunt’ is more tightly linked to specific social contexts than its component calls, and that iteration may further influence sequence meaning. These findings support the idea that sooty mangabey sequences can function compositionally, with meaning emerging from both order and structure.

The third chapter investigates vocal responses to predator model experiments to explore whether alarm call sequences encode additional information beyond individual calls. Using Bayesian models, the analysis focuses on the vocal utterance ‘shrill’, which is often produced alone but can also appear with an additional element, the ‘hoo’, functioning as an affix-like structure. These sequences appear to relate to information about predators, possibly indicating differences in the type or urgency of threat. These findings suggest that sooty mangabeys use affixation-like patterns to flexibly enrich signal content in alarm contexts, even in the absence of predator-specific alarm calls.

Together, the three chapters adopt a comprehensive vocal repertoire approach to investigate the structure and meaning of a primate vocal system. By combining detailed natural observations with experimental data, this work examines both frequent vocal

behaviours and rarer observational contexts such as predator encounters. This broad, integrated perspective moves beyond studies focused on single call types or contexts, offering a more complete understanding of the sooty mangabey vocal system. The presence of multiple mechanisms that possibly generate meaning (ordering, iteration, affixation, and compositionality) suggests that the cognitive underpinnings of vocal combinatoriality are diverse and context dependent. These findings also show that, compared to apes, sooty mangabeys combine calls less extensively and rely on a more restricted subset of combinations, supporting the idea of an evolutionary shift in compositional diversification and flexibility between catarrhine monkeys and great apes. Overall, my thesis contributes new insights into how vocal combinations function in animal communication and what they reveal about the evolutionary foundations of language.

References

Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology*, *18*(5), R202–R203. https://doi.org/10.1016/j.cub.2008.01.040

Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: Compositional or idiomatic expressions? *Brain and Language*, *120*(3), 303–309. https://doi.org/10.1016/j.bandl.2011.10.001

Bergman, T. J., Beehner, J. C., Painter, M. C., & Gustison, M. L. (2019). The speech-like properties of nonhuman primate vocalizations. *Animal Behaviour*, *151*, 229–237. https://doi.org/10.1016/j.anbehav.2019.02.015

Berthet, M., Coye, C., Dezecache, G., & Kuhn, J. (2023). Animal linguistics: A primer. *Biological Reviews*, *98*(1), 81–98. https://doi.org/10.1111/brv.12897

Berthet, M., Mesbahi, G., Pajot, A., Cäsar, C., Neumann, C., & Zuberbühler, K. (2019). Titi monkeys combine alarm calls to create probabilistic meaning. *Science Advances*, *5*(5), eaav3991. https://doi.org/10.1126/sciadv.aav3991

Berthet, M., Neumann, C., Mesbahi, G., Cäsar, C., & Zuberbühler, K. (2018). Contextual encoding in titi monkey alarm call sequences. *Behavioral Ecology and Sociobiology*, *72*(1), 8. https://doi.org/10.1007/s00265-017-2424-z

Berthet, M., Surbeck, M., & Townsend, S. W. (2025). Extensive compositionality in the vocal system of bonobos. *Science*, *388*(6742), 104–108. https://doi.org/10.1126/science.adv1170

Berwick, R. C., & Chomsky, N. (2016). *Why only us: Language and evolution*. MIT press.

Bhat, A. S., Sane, V. A., Seshadri, K. S., & Krishnan, A. (2022). Behavioural context shapes vocal sequences in two anuran species with different repertoire sizes. *Animal Behaviour*, *184*, 111–129. https://doi.org/10.1016/j.anbehav.2021.12.004

Boeckx, C. (Ed.). (2011). *The Oxford Handbook of Linguistic Minimalism* (1st ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199549368.001.0001

Boesch, C., & Boesch-Achermann, H. (2000). *The Chimpanzees of the Taï Forest: Behavioural Ecology and Evolution*. Oxford University Press. https://books.google.ch/books?id=6drtXIjlPZgC

Bosshard, A. B., Burkart, J. M., Merlo, P., Cathcart, C., Townsend, S. W., & Bickel, B. (2024). Beyond bigrams: Call sequencing in the common marmoset (Callithrix jacchus) vocal system. *Royal Society Open Science*, *11*(11), 240218. https://doi.org/10.1098/rsos.240218

Bouchet, H., Blois-Heulin, C., & Lemasson, A. (2013). Social complexity parallels vocal complexity: A comparison of three non-human primate species. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00390

Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012). Context-related call combinations in female Diana monkeys. *Animal Cognition*, *15*(3), 327–339. https://doi.org/10.1007/s10071-011-0456-8

Chomsky, N. (1956). Three models for the description of language. *IEEE Transactions on Information Theory*, *2*(3), 113–124. https://doi.org/10.1109/TIT.1956.1056813

Chomsky, N. (1995). Language and nature. *Mind*, *104*(413), 1–61.

Clay, Z., & Zuberbühler, K. (2009). Food-associated calling sequences in bonobos. *Animal Behaviour*, *77*(6), 1387–1396. https://doi.org/10.1016/j.anbehav.2009.02.016

Clay, Z., & Zuberbühler, K. (2011). Bonobos Extract Meaning from Call Sequences. *PLoS ONE*, *6*(4), e18786. https://doi.org/10.1371/journal.pone.0018786

Collier, K., Radford, A. N., Stoll, S., Watson, S. K., Manser, M. B., Bickel, B., & Townsend, S. W. (2020). Dwarf mongoose alarm calls: Investigating a complex non-human animal call. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1935), 20192514. https://doi.org/10.1098/rspb.2019.2514

Collier, K., Townsend, S. W., & Manser, M. B. (2017). Call concatenation in wild meerkats. *Animal Behaviour*, *134*, 257–269. https://doi.org/10.1016/j.anbehav.2016.12.014

Coye, C., Ouattara, K., Arlet, M. E., Lemasson, A., & Zuberbühler, K. (2018). Flexible use of simple and combined calls in female Campbell's monkeys. *Animal Behaviour*, *141*, 171–181. https://doi.org/10.1016/j.anbehav.2018.05.014

Coye, C., Ouattara, K., Zuberbühler, K., & Lemasson, A. (2015). Suffixation influences receivers' behaviour in non-human primates. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1807), 20150265. https://doi.org/10.1098/rspb.2015.0265

Cunningham, S., & Magrath, R. D. (2017). Functionally referential alarm calls in noisy miners communicate about predator behaviour. *Animal Behaviour*, *129*, 171–179. https://doi.org/10.1016/j.anbehav.2017.05.021

Darwin, C. (1981). *The descent of man, and selection in relation to sex*. Princeton University Press.

Ehardt, C. L. (1988). Affiliative behavior of adult female sooty mangabeys (Cercocebus atys). *American Journal of Primatology*, *15*(2), 115–127. https://doi.org/10.1002/ajp.1350150205

Engesser, S., Crane, J. M. S., Savage, J. L., Russell, A. F., & Townsend, S. W. (2015). Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System. *PLOS Biology*, *13*(6), e1002171. https://doi.org/10.1371/journal.pbio.1002171

Engesser, S., Ridley, A. R., & Townsend, S. W. (2016). Meaningful call combinations and compositional processing in the southern pied babbler. *Proceedings of the National Academy of Sciences*, *113*(21), 5976–5981. https://doi.org/10.1073/pnas.1600970113

Engesser, S., Ridley, A. R., & Townsend, S. W. (2017). Element repetition rates encode functionally distinct information in pied babbler ‘clucks’ and ‘purrs.’ *Animal Cognition*, *20*(5), 953–960. https://doi.org/10.1007/s10071-017-1114-6

Engesser, S., & Townsend, S. W. (2019). Combinatoriality in the vocal systems of nonhuman animals. *Wiley Interdisciplinary Reviews: Cognitive Science*, e1493. https://doi.org/10.1002/wcs.1493

Fitch, W. T. (2010). *The evolution of language* (Cambridge University Press).

Freeberg, T. M. (2008). Complexity in the chick-a-dee call of carolina chickadees (Poecile carolinesis): Associations of context and signlater behavior to call structure. *The Auk*, *125*(4), 896–907. https://doi.org/10.1525/auk.2008.07025

Fruteau, C., van de Waal, E., van Damme, E., & Noë, R. (2011). Infant access and handling in sooty mangabeys and vervet monkeys. *Animal Behaviour*, *81*(1), 153–161. https://doi.org/10.1016/j.anbehav.2010.09.028

Furuichi, T., & Thompson, J. (Eds.). (2008). *The Bonobos: Behavior, Ecology, and Conservation*. Springer New York. https://doi.org/10.1007/978-0-387-74787-3

Gallot, Q., Depriester, C., Moran, S., & Zuberbühler, K. (2024). A non-human primate combinatorial system for long-distance communication. *iScience*, *27*(11). https://doi.org/10.1016/j.isci.2024.111172

Gba, B. C., Bene, J.-C. K., Gone Bi, Z. B., Mielke, A., & Kone, I. (2020). Within-group spatial position and activity budget of wild sooty mangabeys (*Cercocebus atys*) in Taï National Park, Côte d'Ivoire. *International Journal of Biological and Chemical Sciences*, *13*(7), 2991–3008. https://doi.org/10.4314/ijbcs.v13i7.2

Girard-Buttoz, C., Neumann, C., Bortolato, T., Zaccarella, E., Friederici, A. D., Wittig, R. M., & Crockford, C. (2025). Versatile use of chimpanzee call combinations promotes meaning expansion. *Science Advances*, *11*(19), eadq2879. https://doi.org/10.1126/sciadv.adq2879

Girard-Buttoz, C., Zaccarella, E., Bortolato, T., Friederici, A. D., Wittig, R. M., & Crockford, C. (2022). Chimpanzees produce diverse vocal sequences with ordered and recombinatorial properties. *Communications Biology*, *5*(1), 410. https://doi.org/10.1038/s42003-022-03350-8

Gontier, N., Hartmann, S., Pleyer, M., & Rodrigues, E. D. (2024). Combinatoriality and Compositionality in Communication, Skills, Tool Use, and Language. *International Journal of Primatology*, s10764-024-00416–00418. https://doi.org/10.1007/s10764-024-00416-8

Grieves, L. A., Logue, D. M., & Quinn, J. S. (2014). Joint-nesting smooth-billed anis, Crotophaga ani, use a functionally referential alarm call system. *Animal Behaviour*, *89*, 215–221. https://doi.org/10.1016/j.anbehav.2014.01.008

Gust, D. A. (1995). Moving up the dominance hierarchy in young sooty mangabeys. *Animal Behaviour*, *50*(1), 15–21. https://doi.org/10.1006/anbe.1995.0216

Gust, D. A., & Gordon, T. P. (1993). Conflict resolution in sooty mangabeys. *Animal Behaviour*, *46*(4), 685–694. https://doi.org/10.1006/anbe.1993.1245

Hauser, M. D. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Hedwig, D., Hammerschmidt, K., Mundry, R., Robbins, M. M., & Boesch, C. (2014). Acoustic structure and variation in mountain and western gorilla close calls: A syntactic approach. *Behaviour*, *151*(8), 1091–1120.

Hedwig, D., & Kohlberg, A. (2024). Call combination in African forest elephants Loxodonta cyclotis. *PLOS ONE*, *19*(3), e0299656. https://doi.org/10.1371/journal.pone.0299656

Hedwig, D., Poole, J., & Granli, P. (2021). Does Social Complexity Drive Vocal Complexity? Insights from the Two African Elephant Species. *Animals*, *11*(11), 3071. https://doi.org/10.3390/ani11113071

Hurford, J. R. (2012). Linguistics From an Evolutionary Point of View. In *Philosophy of Linguistics* (pp. 477–502). Elsevier. https://doi.org/10.1016/B978-0-444-51747-0.50014-7

Jansen, D. A., Cant, M. A., & Manser, M. B. (2012). Segmental concatenation of individual signatures and context cues in banded mongoose (Mungos mungo) close calls. *BMC Biology*, *10*(1), 97. https://doi.org/10.1186/1741-7007-10-97

Lameira, A. R., Hardus, M. E., Ravignani, A., Raimondi, T., & Gamba, M. (2024). Recursive self-embedded vocal motifs in wild orangutans. *eLife*, *12*, RP88348. https://doi.org/10.7554/eLife.88348.3

León, J., Thiriau, C., Crockford, C., & Zuberbühler, K. (2023). Comprehension of own and other species' alarm calls in sooty mangabey vocal development. *Behavioral Ecology and Sociobiology*, *77*(5), 56. https://doi.org/10.1007/s00265-023-03318-6

Leroux, M., Bosshard, A. B., Chandia, B., Manser, A., Zuberbühler, K., & Townsend, S. W. (2021). Chimpanzees combine pant hoots with food calls into larger structures. *Animal Behaviour*, *179*, 41–50. https://doi.org/10.1016/j.anbehav.2021.06.026

Leroux, M., Schel, A. M., Wilke, C., Chandia, B., Zuberbuhler, K., Slocombe, K. E., & Townsend, S. W. (2023). Call combinations and compositional processing in wild chimpanzees. *Nature Communications*, *14*, 2225. https://doi.org/10.1038/s41467-023-37816-y

Leroux, M., & Townsend, S. W. (2020). Call combinations in great apes and the evolution of syntax. *Animal Behavior and Cognition*, *7*(2), 131–139. https://doi.org/10.26451/abc.07.02.07.2020

Maddieson. (1984). *Patterns of Sounds* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511753459

Manser, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 11.

Markowitz, J. E., Ivie, E., Kligler, L., & Gardner, T. J. (2013). Long-range Order in Canary Song. *PLoS Computational Biology*, *9*(5), e1003052. https://doi.org/10.1371/journal.pcbi.1003052

Marler, P. (1975). On the origin of speech from animal sounds. *The Role of Speech in Language*, *11*(37).

Martins, M. D. (2012). Distinctive signatures of recursion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 2055–2064. https://doi.org/10.1098/rstb.2012.0097

McComb, K., & Semple, S. (2005). Coevolution of vocal communication and sociality in primates. *Biology Letters*, *1*(4), 381–385. https://doi.org/10.1098/rsbl.2005.0366

McGraw, W. S., Zuberbuhler, K., & Noë, R. (2007). *Monkeys of the Taï forest: An African primate community* (Vol. 51). Cambridge University Press.

Mehon, F. G., Zuberbühler, K., & Stephan, C. (2024). Population differences in putty-nosed monkey (Cercopithecus nictitans) call order. *Primates*. https://doi.org/10.1007/s10329-024-01155-3

Mielke, A., Crockford, C., & Wittig, R. M. (2019). Snake alarm calls as a public good in sooty mangabeys. *Animal Behaviour*, *158*, 201–209. https://doi.org/10.1016/j.anbehav.2019.10.001

Mielke, A., Crockford, C., & Wittig, R. M. (2020). Predictability and variability of association patterns in sooty mangabeys. *Behavioral Ecology and Sociobiology*, *74*(4), 46. https://doi.org/10.1007/s00265-020-2829-y

Mielke, A., Preis, A., Samuni, L., Gogarten, J. F., Lester, J., Crockford, C., & Wittig, R. M. (2020). *Consistency of social interactions in sooty mangabeys and chimpanzees* [Preprint]. Animal Behavior and Cognition. https://doi.org/10.1101/2020.07.10.196949

Moran, S., McCloy, D., & Wright, R. (2012). Revisiting population size vs. Phoneme inventory size. *Language*, *88*(4), 877–893. https://doi.org/10.1353/lan.2012.0087

Narbona Sabaté, L., Mesbahi, G., Dezecache, G., Cäsar, C., Zuberbühler, K., & Berthet, M. (2022). Animal linguistics in the making: The Urgency Principle and titi

monkeys' alarm system. *Ethology Ecology & Evolution*, *34*(3), 378–394. https://doi.org/10.1080/03949370.2021.2015452

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009a). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, *106*(51), 22026–22031. https://doi.org/10.1073/pnas.0908118106

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009b). Campbell's Monkeys Use Affixation to Alter Call Meaning. *PLoS ONE*, *4*(11), e7808. https://doi.org/10.1371/journal.pone.0007808

Peterson, R. R., Burgess, C., Dell, G. S., & Eberhard, K. M. (2001). Dissociation between syntactic and semantic processing during idiom comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1223–1237. https://doi.org/10.1037/0278-7393.27.5.1223

Pozzi, L., Hodgson, J. A., Burrell, A. S., Sterner, K. N., Raaum, R. L., & Disotell, T. R. (2014). Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution*, *75*, 165–183. https://doi.org/10.1016/j.ympev.2014.02.023

Price, T., Arnold, K., Zuberbühler, K., & Semple, S. (2009). Pyow but not hack calls of the male putty-nosed monkey (Cercopithcus nictitans) convey information about caller identity. *Behaviour*, *146*(7), 871–888. https://doi.org/10.1163/156853908X396610

Range, F., & Fischer, J. (2004). Vocal Repertoire of Sooty Mangabeys (Cercocebus torquatus atys) in the Tai National Park. *Ethology*, *110*(4), 301–321. https://doi.org/10.1111/j.1439-0310.2004.00973.x

Range, F., Forderer, T., Storrer-Meystre, Y., Benetton, C., & Fruteau, C. (2007). The structure of social relationships among sooty mangabeys in Tai. In *Monkeys of the Tai Forest: An African primate community* (pp. 109–130). Cambridge University Press.

Range, F., & Noë, R. (2002). Familiarity and dominance relations among female sooty mangabeys in the Taï National Park: Relations Among Female Sooty Mangabeys. *American Journal of Primatology*, *56*(3), 137–153. https://doi.org/10.1002/ajp.1070

Rauber, R., Kranstauber, B., & Manser, M. B. (2020). Call order within vocal sequences of meerkats contains temporary contextual and individual information. *BMC Biology*, *18*(1), 119. https://doi.org/10.1186/s12915-020-00847-8

Salis, A., Léna, J.-P., & Lengagne, T. (2022). Which acoustic parameters modify the great tit's response to conspecific combinatorial mobbing calls? *Behavioral Ecology and Sociobiology*, *76*(4), 46. https://doi.org/10.1007/s00265-022-03157-x

Schamberg, I., Cheney, D. L., Clay, Z., Hohmann, G., & Seyfarth, R. M. (2016). Call combinations, vocal exchanges and interparty movement in wild bonobos. *Animal Behaviour*, *122*, 109–116. https://doi.org/10.1016/j.anbehav.2016.10.003

Schamberg, I., Cheney, D. L., Clay, Z., Hohmann, G., & Seyfarth, R. M. (2017). Bonobos use call combinations to facilitate inter-party travel recruitment. *Behavioral Ecology and Sociobiology*, *71*(4), 75. https://doi.org/10.1007/s00265-017-2301-9

Schlenker, P., Chemla, E., Arnold, K., & Zuberbühler, K. (2016). Pyow-hack revisited: Two analyses of Putty-nosed monkey alarm calls. *Lingua*, *171*, 1–23. https://doi.org/10.1016/j.lingua.2015.10.002

Searcy, W. A., Soha, J., Peters, S., & Nowicki, S. (2022). Long-distance dependencies in birdsong syntax. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1967), 20212473. https://doi.org/10.1098/rspb.2021.2473

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980a). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, *210*(4471), 801–803.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980b). Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Animal Behaviour*, *28*(4), 1070–1094. https://doi.org/10.1016/S0003-3472(80)80097-2

Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. *Nature Communications*, *15*(1), 3617. https://doi.org/10.1038/s41467-024-47221-8

Sigmundson, R., Girard-Buttoz, C., Le Floch, A., Azaiez, T. S., McElreath, R., Zuberbühler, K., Wittig, R. M., & Crockford, C. (2025). Vocal sequence diversity and length remain stable across ontogeny in a catarrhine monkey (Cercocebus atys). *Communications Biology*, *8*(1), 465. https://doi.org/10.1038/s42003-025-07922-2

Stephan, C., & Zuberbühler, K. (2008). Predation increases acoustic complexity in primate alarm calls. *Biology Letters*, *4*(6), 641–644. https://doi.org/10.1098/rsbl.2008.0488

Struhsaker, T. T. (1967). Auditory communication among vervet monkeys (Cercopithecus aetthiops). *Social Communication among Primates*.

Suzuki, T. N. (2012). Referential mobbing calls elicit different predator-searching behaviours in Japanese great tits. *Animal Behaviour*, *84*(1), 53–57. https://doi.org/10.1016/j.anbehav.2012.03.030

Suzuki, T. N. (2025). How to study animal syntax. In L. Raviv & C. Boeckx (Eds.), *The Oxford Handbook of Approaches to Language Evolution* (1st ed., pp. 593–606). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780192886491.013.30

Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nature Communications*, *7*(1), 10986. 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. (2020). The syntax–semantics interface in animal vocal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20180405. https://doi.org/10.1098/rstb.2018.0405

Townsend, S. W., Engesser, S., Stoll, S., Zuberbühler, K., & Bickel, B. (2018). Compositionality in animals and humans. *PLOS Biology*, *16*(8), e2006425. https://doi.org/10.1371/journal.pbio.2006425

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., Van Der Kant, A., Smith, K., & Rey, A. (2020). Non-adjacent Dependency Learning in Humans and Other Animals. *Topics in Cognitive Science*, *12*(3), 843–858. https://doi.org/10.1111/tops.12381

Zuberbühler, K. (2001). Predator-specific alarm calls in Campbell's monkeys, Cercopithecus campbelli. *Behavioral Ecology and Sociobiology*, *50*(5), 414–422. https://doi.org/10.1007/s002650100383

Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, *63*(2), 293–299. https://doi.org/10.1006/anbe.2001.1914

Zuberbühler, K. (2017). The Primate Roots of Human Language. In R. M. Quam, M. A. Ramsier, R. R. Fay, & A. N. Popper (Eds.), *Primate Hearing and Communication* (Vol.

63, pp. 175–200). Springer International Publishing. https://doi.org/10.1007/978-3-319-59478-1_7

Zuberbühler, K. (2019). Evolutionary roads to syntax. *Animal Behaviour*, *151*, 259–265. https://doi.org/10.1016/j.anbehav.2019.03.006

Zuberbühler, K. (2020). Syntax and compositionality in animal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20190062. https://doi.org/10.1098/rstb.2019.0062

Zuberbühler, K., Noë, R., & Seyfarth, R. M. (1997). Diana monkey long-distance calls: Messages for conspecifics and predators. *Animal Behaviour*, *53*(3), 589–604. https://doi.org/10.1006/anbe.1996.0334

Chapter 1: Rule-based sequences in sooty mangabey vocal communication

This chapter is based on a manuscript currently under review for publication as: Le Floch, A., Girard-Buttoz, C., Azaiez, T. S., Bande, N., Wittig, R. M., Moran, S., Zuberbühler, K., & Crockford, C. (2025). Rule-based sequences in sooty mangabey vocal communication. Royal Society Open Science (submitted). *The PhD candidate (Le Floch, A.) was responsible for study design, data collection, data coding, analysis and manuscript preparation.*

Disclosure: A subset of the data presented in this chapter were collected and used by Tanit Souha Azaiez for her master thesis, deposed at the University of Neuchâtel. All data reused for this thesis were analysed differently.

Abstract

Investigating how non-human animals produce call sequences provides key insights into the evolutionary origins of meaning in vocal communication, including syntax. Many species combine calls into structured sequences, often following specific rules, yet most studies focus on only a few sequences per species. This limits our understanding of their ability to combine calls and convey meaning through sequences. Our study addresses this gap by documenting the vocal sequence repertoire and underlying rules of sooty mangabeys *(Cercocebus atys)*, a West African monkey species. We collected and analysed 1,672 recordings from two wild groups in the Taï National Park, Ivory Coast. Sooty mangabeys frequently combine calls but rely on a limited set of sequences. Within these, we identified rules of call ordering and reoccurrence. Notably, they produced non-adjacent dependencies (i.e., relationship between the first and last call of a sequence irrespective of intervening calls), which are rarely reported in animal communication. Our findings suggest that sooty mangabeys also use sequences without apparent rules, likely encoding different types of information. While context of production remains essential for determining meaning, our results highlight how a whole-repertoire approach can reveal rule-based sequences and their potential for meaning expansion beyond the number of individual calls available in the repertoire.

Introduction

One key feature of language is syntax, allowing humans to flexibly combine sounds into words and words into sentences, thus generating a potentially infinite number of structure-meaning relations (Hockett & Hockett, 1960; Hurford, 2012; Marler, 1998). The evolution of this ability, presumably unique to humans, is challenging to reconstruct (Fitch, 2010). One powerful way to explore the problem is to study vocal signals in non-human animals (Hauser, 2002; Moran & Bickel, 2020) and how combining them into sequences can convey meaning (Berthet et al., 2025; Girard-Buttoz et al., 2025). Relevant here is that, although humans are capable of producing and perceiving hundreds of phonemes (i.e., smallest units of sound in a language), all languages only utilise a fraction of them, with some languages using as few as 11 phonemes (Maddieson & Precoda, 1989; Moran et al., 2012). This demonstrates that a small number of sounds is sufficient to sustain a language. The evolution of a communicative system capable of conveying a substantial range of meanings is thus likely less about producing increasingly rich vocal repertoires and more about evolving the ability to combine sounds in structured ways. This study investigates the vocal sequence repertoire of a *catarrhine* monkey, focusing on the rules that govern sequence structuring, independent of whether they convey specific meanings, to explore the potential for meaning generation through principles of sequence organisation.

Much research effort is currently being devoted to understanding the early origins of syntax and how it is integrated into the semantic system (Suzuki et al., 2020). This has led to a growing body of literature documenting how various species—including primates, birds, mongooses, marine mammals, elephants, hyraxes, bats, and amphibians—combine acoustically discrete sound units into larger sequences (Bhat et al., 2022; Engesser & Townsend, 2019; Hedwig & Kohlberg, 2024). Although not all sequences convey context-specific information, recent research reveals that some sequences do, spanning a diverse range of species e.g., apes (Berthet et al., 2025; Clay & Zuberbühler, 2009, 2011; Girard-Buttoz, Bortolato, et al., 2022; Girard-Buttoz et al., 2025; Hedwig et al., 2015; Leroux et al., 2021, 2022, 2023; Schamberg et al., 2016, 2017), bats (Zhang et al., 2019), birds (Engesser et al., 2015, 2016, 2017, 2019; Freeberg, 2008; Landsborough et al., 2019; Salis et al., 2022; Spiess et al., 2022; Suzuki et al., 2016; Toshitaka N Suzuki, 2014), elephants (Hedwig & Kohlberg, 2024), frogs (Bhat et al., 2022), mongooses (Collier et al., 2020; Rauber et al., 2020), monkeys (Arnold & Zuberbühler, 2006b, 2006a, 2008, 2012; Berthet et al., 2019;

Candiotti et al., 2012; Coye et al., 2015, 2018; Mehon et al., 2024; Narbona Sabaté et al., 2022; Ouattara et al., 2009b, 2009a; Torti et al., 2013; Wich et al., 2003; Zuberbühler, 2002).

However, due to the time-consuming nature of this work, typically conducted through playback experiments (Berthet et al., 2023), only few such sequences have been reported per species. A promising approach to address this limitation is to first analyse entire vocal repertoires, focusing on whether sequences show rule-based organisation, irrespective of their semantic associations. One key metric in such analyses is flexibility—the extent to which calls in a vocal repertoire can be combined freely to form sequences (Girard-Buttoz, Zaccarella, et al., 2022). Flexibility can increase by expanding both the length of sequences and their diversity, i.e., the number of distinct sequences uttered. Diversity can also arise from adhering to specific rules that define how sequences in a vocal system are typically organised, including patterns of reoccurrence and ordering (Table 1). Overall, few studies have investigated the natural vocal production of rule-based sequences in animals at the level of the entire repertoire, and even fewer have explored this in relation to associated meaning (see Table 1). Assessing vocal repertoire flexibility and diversity, therefore, provides a critical lens to further explore whether sequence structure may generate new meaning or simply reflect organisational variation without semantic implications.

Table 1. Meaning-related rules in animal vocal sequences

References for animal species were selected based on their relevance to meaning. Note that, while classified into different categories, these rules are not necessarily mutually exclusive, and the table is not intended to be exhaustive. An asterisk (*) denotes rules that were analysed in this study.

Category	**Rule**	**Description**	**Animal species**
Ordering	Positional bias*	The tendency for a particular call type to occur at a specific position within a sequence.	Chestnut-crowned babblers (Engesser et al., 2015), Campbell's monkeys (Zuberbühler, 2002), Japanese great tits (Suzuki et al., 2016,

			2017), Chimpanzees(Girard-Buttoz et al., 2025)
	Gradation	Call types are arranged in a particular order within a sequence, forming a pattern that transitions from one type to another. For example, meerkats (Rauber et al., 2020) use graded sequences, progressing from less urgent to more urgent calls as danger increases.	Meerkats (Rauber et al., 2020), black-fronted titi monkeys (Narbona Sabaté et al., 2022)
Reoccurrence	Iteration*	Repeated occurrence of a call type within a sequence, interspersed with other call types, such as in the pattern ABA, where ‘A’ is repeated with ‘B’ in between.	Bonobos (Clay & Zuberbühler, 2011)
	Repetition*	Consecutive occurrence of the same call type at	Humayun’s night frogs (Bhat et al., 2022), bonobos (Clay & Zuberbühler, 2011), Carolina

		least two times, such as in the pattern ABBB, where the call 'A' is followed by multiple repetitions of 'B'.	chickadees (Freeberg, 2008), great tits (Salis et al., 2022), pied babblers (Engesser et al., 2017), black-and-white colobus (Schel et al., 2009), putty-nosed monkeys (Arnold & Zuberbühler, 2012), Thomas langurs (Wich et al., 2003), indris (Torti et al., 2013)
	Proportion	The overall number of times a call type appears in a sequence compared to the occurrence of one or other call types.	Amboli bush frogs (Bhat et al., 2022), titi monkeys (Berthet et al., 2019)

The main hypothesis for explaining species differences in flexibility is the amount of information that a species needs to convey (Nowak et al., 2002). For this reason, many animal communication studies are concerned with the impact of structure on meaning, such as whether the meaning of a sequence is derived or distinct from the meaning of its constituent parts, e.g., apes (Berthet et al., 2025; Girard-Buttoz, Bortolato, et al., 2022; Girard-Buttoz et al., 2025; Hedwig et al., 2015; Leroux et al., 2021, 2023; Schamberg et al., 2016), birds (Engesser et al., 2016; Freeberg, 2008), monkeys (Arnold & Zuberbühler, 2008; Zuberbühler, 2002), elephants (Hedwig & Kohlberg, 2024), mongooses (Collier et al., 2020). Several distinct patterns have been found across the species that have been analysed. First, in some species one call type functions as a kind of affix that can be merged with an otherwise stand-alone call (e.g., 'A' becomes Ab), which then modifies the initial meaning, as observed in Campbell's monkeys (Coye et al., 2015; Ouattara et al., 2009b), mongooses (Jansen et al., 2012) and Diana monkeys (Candiotti et al., 2012) and as suggested for some chimpanzee

call combinations (Girard-Buttoz et al., 2025). Another mechanism is the production of distinct structures, i.e., repeating a specific call in a sequence, as in bonobos (Clay & Zuberbühler, 2011) or titi monkeys (Berthet et al., 2019), or producing different call orders depending on event type, for example in meerkats (Rauber et al., 2020), titi monkeys (Narbona Sabaté et al., 2022), Chestnut-crowned babblers (Engesser et al., 2015) and Campbell's monkeys (Zuberbühler, 2002).

Understanding the evolutionary drivers and constraints of flexibility, therefore, is likely to advance our understanding of what exactly led to language in humans. This requires a systematic, whole-repertoire approach, which involves assessing all possible sequences within a repertoire along with their flexibility, including the potential rules that govern their organisation—an approach that has rarely been explored, e.g., chimpanzees (Girard-Buttoz, Zaccarella, et al., 2022; Leroux et al., 2022), marmosets (Bosshard et al., 2024). Such an approach is a critical first step to assess the potential of a given repertoire to expand meanings through call combinations. Here, we addressed this gap with an observational study of wild sooty mangabey vocal behaviour, a primate with a call repertoire comparable in size to great apes (Moran, Maiolini, et al., 2022; Range & Fischer, 2004) and suggested containing several call sequences (Grampp et al., 2023; Range & Fischer, 2004). We were particularly interested in identifying which calls from the repertoire were combined, how long and diverse the resulting sequences were, and whether it was possible to discern rules in their composition. To this end, we adopted a whole-repertoire perspective, examining all sequence types for which sufficient recordings were available (see Limitations).

Methods

Data collection

ALF, NB and TSA collected data on two groups of sooty mangabeys habituated to human presence, the Taï Monkey Project (TMP) group from May to July 2022 and the Taï Chimpanzee Project (TCP) group from February to October 2023 at the Taï National Park (Ivory Coast). We used half-day continuous focal animal sampling (J. Altmann, 1974) on adult individuals from dawn to midday or from midday to dusk. We recorded all the calls produced by the focal individuals on an all-occurrence basis, as well as calls produced by known adult individuals in visual range on an ad-libitum basis, using directional microphones (Sennheiser models K6, MK 316, MKH 416 P48 and

MKH 416T-F) and recorders (Marantz Solid State models PMD661 and PMD661 MKII), digitised at a 44,1 kHz sampling rate and 16-bit sampling depth in a '.wav' format. We collected all vocal utterances for which we could attribute a known caller, without overlap with other callers.

This work was conducted under the research permit for the Taï Chimpanzee Project (Wittig/006/MESRS/DRI) from the Ivorian Ministry of Higher Education and Scientific Research and the Ivorian Office of Parks and Reserves for access to the Taï National Park. This study was purely observational with audio recordings and as such approved by the "Ethikrat der Max Planck Gesellschaft". The use of wild animals in this research adheres to the guidelines set forth by the Association for the Study of Animal Behaviour and the International Union for Conservation of Nature.

Audio recording coding

ALF, NB and TSA coded audio recordings using Raven Pro (ALF v1.6.5; NB and TSA v1.6.4) (K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology, 2024). We visually inspected spectrograms (FFT size 1024, Hann window, hope size 256 samples) while listening to the recordings to identify and manually annotate each sound element with start and end times (see glossary for definition of an element). Drawing from the sooty mangabey vocal repertoire (Range & Fischer, 2004) we identified 10 of the 13 described call types in our sample: 'grunt', 'twitter', 'growl', 'scream', 'grumble', 'vibrato' (referred to as 'copulation call' in Range & Fischer's vocal repertoire), 'whoop-gobble', 'wau', 'shrill' (the first high-frequency sound elements of the alarm call in Range & Fischer's vocal repertoire) and 'hoo' (referring to both the 'hoo' call described in Range & Fischer's vocal repertoire and the last, optional low-frequency sound element that was initially described as part of the alarm call). The 'intense threat' calls were not identified in our sample; these are likely rare, and the original repertoire does not report their frequency. Similarly, the three alarm call types previously defined by predator type (eagle, leopard, viper) were not acoustically characterised and could not be reliably distinguished by ear, so we classified all such calls as 'shrill' to focus on acoustic properties rather than context. We systematically collected all utterances that could be recorded during focal follows (i.e., except when multiple callers where vocalising at the same time, preventing subsequent accurate coding of the recordings), supplemented with ad libitum sampling

(see Limitations). All call types used in this study are detailed in the supplementary materials together in the section Document S1 with spectrographic examples and corresponding definitions from Range & Fischer's vocal repertoire.

Following previous work on sooty mangabey vocal sequences, we kept the same criteria to classify vocal elements, calls and sequences (Sigmundson et al., 2025). To define a call, we applied a 2-second threshold for intervals between elements of the same type. This rule ensured that continuous emissions of the same call type were not artificially split into multiple calls, avoiding overinflation of the call count in our sample.. Of the 8,259 intervals measured across 1,693 utterances containing repetitions of single call types, 98.18% were shorter than 2 seconds, with a median of 0.09 seconds and a standard deviation of 1.82 seconds, confirming that our 2 second threshold captured typical within-call intervals. By using this threshold, we ensured that only intervals longer than 2 seconds were treated as breaks between utterances of the same call type.

Sequences were defined separately, as groups of distinct calls occurring in rapid succession. Based on the distribution of 1,122 intervals between calls across 619 utterances, most intervals were short (median = 0.27 s; 69.4% <1 s; SD = 6.39 s). We therefore used a 1-second maximum interval to group calls into sequences. This conservative threshold captures typical temporal clustering of different calls while excluding longer pauses, ensuring that sequence counts were not overinflated.

Importantly, the call and sequence thresholds were applied in order: first to identify calls, then to group these calls into sequences. This procedure avoids inconsistencies, as intervals longer than 2 seconds are always considered breaks between calls, whereas shorter intervals may still define sequences if multiple call types occur in close succession.

To assess inter-rater reliability on call classification, ALF independently recoded a randomly selected subset of 320 calls from NB (14% of the sample) and 116 calls from TSA (21%). Agreement was high, with Cohen's kappa values of 0.96 (NB) and 0.95 (TSA).

Glossary

Term	Definition
Vocal element	A continuous sound with a distinct beginning and end. In rare cases, two different types of elements may occur without a clear silent gap between them; in such cases, they are distinguished using their specific acoustic properties (Kershenbaum et al., 2016) (see supplementary materials for details on acoustic properties of elements).
Call type / Vocal unit	A call consisting of one or multiple elements of the same type, produced in succession with intervals of less than two seconds between them.
Repetitions	Refer to the number of vocal elements present within a given call type.
Vocal sequence	A series of different consecutive call types (or vocal units), where each call type follows the previous one with an interval of less than one second between them. We represent sequences using their composing call types but without including the repetitions of vocal elements (e.g., AAB and AABB are treated as AB).
Iterations	Refer to the repeated occurrence of a call type within a sequence, interspersed with other call types, such as ABA where the call type A isiterated and occurring again after B.
X-call type sequence	A vocal sequence containing X distinct call types, which may or may not include one or more iterations. For example, a two-call type sequence can be AB, ABA, or ABABA.
Sequence length	The number of call types in a vocal sequence was measured in two ways: including iterations (e.g., ABA) to capture the raw sequence length, and excluding iterations (e.g., ABA becomes AB) to

	determine how many different call types are combined. Here the number of elements within call types is not included.
Combination pair	Two call types produced consecutively within a sequence, regardless of order. For example: the sequence ABC contains two pairs: A and B, and B and C; the sequences AB and BA hold the same combination pair: A and B.
Sequence diversity	The number of unique sequences, defined by the specific order of call types, and without accounting for adjacent element repetitions. With iterations considered, sequences like AB and ABA are unique. Without iterations, a unique sequence (e.g., AB) includes all sequences starting with AB, such as AB, ABA, and ABAB.
Vocal utterance	Either a single call type (vocal unit) produced in isolation (not part of a sequence) or a vocal sequence.

Analyses

We first visually inspected the sequence flexibility by looking at (1) the number of different call types in the repertoire that can be combined, (2) the sequence length with and without iterations (i.e., non-adjacent repetitions of calls, see Table 1), (3) the number of different possible combination pairs and (4) the sequence diversity with and without iterations (see glossary for definitions). We then focused on a subset of the data for which we had sufficient sample size to analyse their rules: ordering was investigated on 2-call type sequences (93% of the sequence production) and reoccurrence (i.e., both adjacent and non-adjacent repetitions, see Table 1) on sequences composed only of ‘grunts’ and ‘twitter’ (61% of the sequences, see Tables S1, S2, S3 and S4 for details).

Ordering

To analyse ordering in 2-call type sequences (see glossary), we examined positional bias based on the first occurrence of each call type (i.e., whether they were more likely

to occur in the first or second position within the sequence). 3-call type and 4-call type sequences constituted only 7% of the total sequences produced (*N* = 26 sequences) and were excluded from the analysis due to insufficient sample size to analyse their structure (Figure 3, see Tables S1 and S3 for details). Out of the 10 call types we could distinguish in our sample, only six appeared in two-call type sequences (see glossary) and could be examined: 'shrill', 'grunt', 'growl', 'scream', 'twitter', and 'hoo'. Four were formally analysed for positional bias, while the positions of 'shrill' and 'hoo' were reported descriptively, as they consistently occurred in the same position (first and second, respectively). The remaining call type appearing in 2-call type sequence, 'grumble', was excluded due to low sample size (*N* = 4). We used sequences produced by both sexes for the analysis. Bayesian logistic regression models were employed to examine call position (first/second) as a binary outcome variable. This resulted in four models being constructed for the following call types: 'grunt', 'growl', 'scream' and 'twitter'. These models included the group (TMP/TCP) as fixed effect (for models with at least five sequences sampled from each group: model 'twitter', model 'grunt', model 'growl', model 'scream') and the identity of the caller as random intercept (for models with at least 50% of the callers sampled that produced at least two sequences: models 'twitter' and 'grunt') to account for repeated samples of the same individual in the same group.

Reoccurrence

We started by analysing iterations in sequences composed of only 'grunts' and 'twitters' (these sequences with iterations represented 44.5% of all the 'grunt' and 'twitter' sequences). Specifically, we aimed to identify which call type ('grunt' or 'twitter') was more likely to occupy the final position, depending on which call type initiated the sequence. To do so, we fitted a Bayesian logistic regression model with the last call type ('grunt'/'twitter') as the binary outcome variable. The model included the call type that initiated the sequence ('grunt'/twitter') and the group (TMP/TCP) as fixed effects. Caller identity was included as a random intercept to control for repeated samples from the same individuals. To control for the possibility that iteration patterns were driven by short sequences, we ran a second analogous model only with the sequences longer than three calls (*N* = 49 sequences, 48.5% of all iterative sequences composed of 'grunt' and 'twitter' calls).

We then examined the repetitions of elements within calls in sequences composed only of ‘grunts’ and ‘twitters’, depending on the call position within the sequence. For this analysis, we used sequences both with and without iterations. We used Bayesian negative binomial regression models to analyse the number of repetitions of elements in each call as a function of their position within the sequence. Separate models were run for each call type (‘grunt’/‘twitter’).

To analyse positional patterns, we classified each occurrence of a call within a sequence as *first*, *middle*, or *last*, relative to other calls of the same type. For example, for ‘grunt’ calls, the *first* ‘grunt’ call was defined as the initial ‘grunt’ call in the sequence, even if a ‘twitter’ call occurred beforehand. Calls appearing only once in a sequence were treated as *first*. The *last* ‘grunt’ call was the final ‘grunt’ call in the sequence (regardless of whether a ‘twitter’ call occurred after), and *middle* ‘grunt’ calls included any grunt calls occurring between the first and last, which could consist of one or several calls depending on sequence length. This classification applied to all sequences, including short sequences such as twitter_grunt (one first ‘twitter’ call and one first ‘grunt’ call) or twitter_grunt_twitter (one first ‘twitter’ call, one first ‘grunt’ call, and one last ‘twitter’ call. The same analysis was applied to the model analysing ‘twitter’ calls, considering the first, middle, and last ‘twitter’ calls within sequences. For both models, we threated the position (first/middle/last) as fixed factor. We controlled for the length of the sequence (i.e., the total number of calls in the sequence, such as grunt_twitter_grunt = 3 calls), the type of call initiating the sequence (‘grunt’/’twitter’) and the group (TMP/TCP) by threating them as fixed factors and accounted for individual variability through a random intercept. After fitting these two models, we performed pairwise comparisons between the levels of the variable position (first/middle/last) using the posterior distribution of the estimated number of repetitions for each position (i.e., estimated marginal means approach).

Model specifications

For all our models, we used the default flat priors from the *brms* package version 2.20.4 (Bürkner, 2015): the intercept had a weakly informative normal prior centered at 0 with a standard deviation of 10, the fixed effect for the predictor (group) a normal prior centered at 0 with a standard deviation of 2.5, and the standard deviation of the random intercepts and a Student-t prior with 3 degrees of freedom, centered at 0, and with a

scale parameter of 10). We ran the models using four Markov chains, each with 4,000 iterations, including a warm-up of 2,000 iterations per chain. This resulted in a total of 8,000 post-warm-up samples. We manually set the *adapt delta* parameter either to 0.95 (positional bias models 'twitter' and 'grunt'; element repetition models), 0.90 (positional bias models 'growl' and 'scream'), or 0.99 for the models exploring call iterations, ensuring thorough exploration of the posterior distribution and minimising the likelihood of divergent transitions. Convergence of the Markov chains was evaluated using the potential scale reduction factor R-hat, with values of 1 indicating adequate convergence. We also visually inspected trace plots for each parameter to ensure stable and well-mixed chains. Posterior predictive checks were conducted to assess the fit of the models to the data.

The analyses were conducted using R version 4.3.3 (R Core Team, 2021) with the *brms* package version 2.20.4 (Bürkner, 2015). The comparisons were conducted using the *emmeans* package version 1.10.0 (Lenth, 2017).

Results

We conducted continuous focal recordings on adult sooty mangabeys. Counting only for periods when the focal individual was in direct sight of the observer, this resulted in TCP females: *N* = 17 individuals, *N* = 232h and 27min (mean per individual = 13h and 40min, SE = 0h and 56min); TMP females: *N* = 18 individuals, *N* = 262 hours and 45 minutes (mean per individual = 14h and 35min, SE = 3h and 28min); TCP males: *N* = 8 individuals, *N* = 122h and 21min (mean per individual = 15h and 17min, SE = 1h and 39min); TMP males: *N* = 4 individuals, *N* = 10h and 30min (mean per individual = 2h and 37min, SE = 0h and 24min). We also recorded vocal utterances from adlib individuals (i.e., who were not focal individuals): *N* = 4 TCP females; *N* = 2 TMP females; *N* = 5 TMP males. From this, we collected and analysed a total of *N* = 1,672 recordings comprising *N* = 2,023 vocal utterances, the majority consisting of only one call type (females: *N* = 1,364, 82.9% of all utterances; males: *N* = 287, 75.9% of all utterances; Figure 3).

Call types

Females used seven of eight call types in sequences. Among these, ‘grunts’ and ‘twitters’ were the most frequently found in sequences, accounting for 42.4% and 37.1% of all sequences, respectively (Table 2 and Figure 1). These were followed by ‘growls’ (7.9%), ‘hoos’ (5.2%), ‘screams’ (3.7%), ‘shrills’ (2%) and ‘grumbles’ (1.7%).

Males also used seven of eight call types in sequences, with a large dominance of ‘hoos’ (39.5%) and ‘shrills’ (38.7%). Other, less frequently combined call types were ‘growls’ (13%), ‘screams’ (3.1%), ‘grunts’ (2.3%), ‘grumbles’ (1.9%) and ‘waus’ (1.5%) (Table 2 and Figure 1).

Although both sexes combined a similar number of call types into sequences, females used most of their calls both singly and in sequence, with only two of seven call types exclusively used in sequences. In contrast, males primarily used calls in sequence, with five of seven call types used exclusively in sequences (Table 2 and Figure 2). Note that while one ‘hoo’ call was produced singly, this occurred only once, so we classified this call as being produced exclusively in sequence.

Table 2. Occurrence of calls used singly (single calls) and in sequences (combined calls) for each call type and sex.

The columns % combined indicate the percentage of calls for calls produced in sequence only (i.e., excluding calls produced singly) for each call type and sex.

	Female			**Male**		
Call type	***N* single calls**	***N* combined calls**	**% combined**	***N* single calls**	***N* combined calls**	**% combined**
grunt	968	391	42.4	225	6	2.3
twitter	133	342	37.1	-	-	-
shrill	35	18	2	-	101	38.7

hoo	-	48	5.2	1	103	39.5
growl	69	73	7.9	57	34	13
scream	49	34	3.7	-	8	3.1
grumble	-	16	1.7	-	5	1.9
wau	-	-	-	-	4	1.5
vibrato	110	-	-	-	-	-
whoopgobble	-	-	-	4	-	-

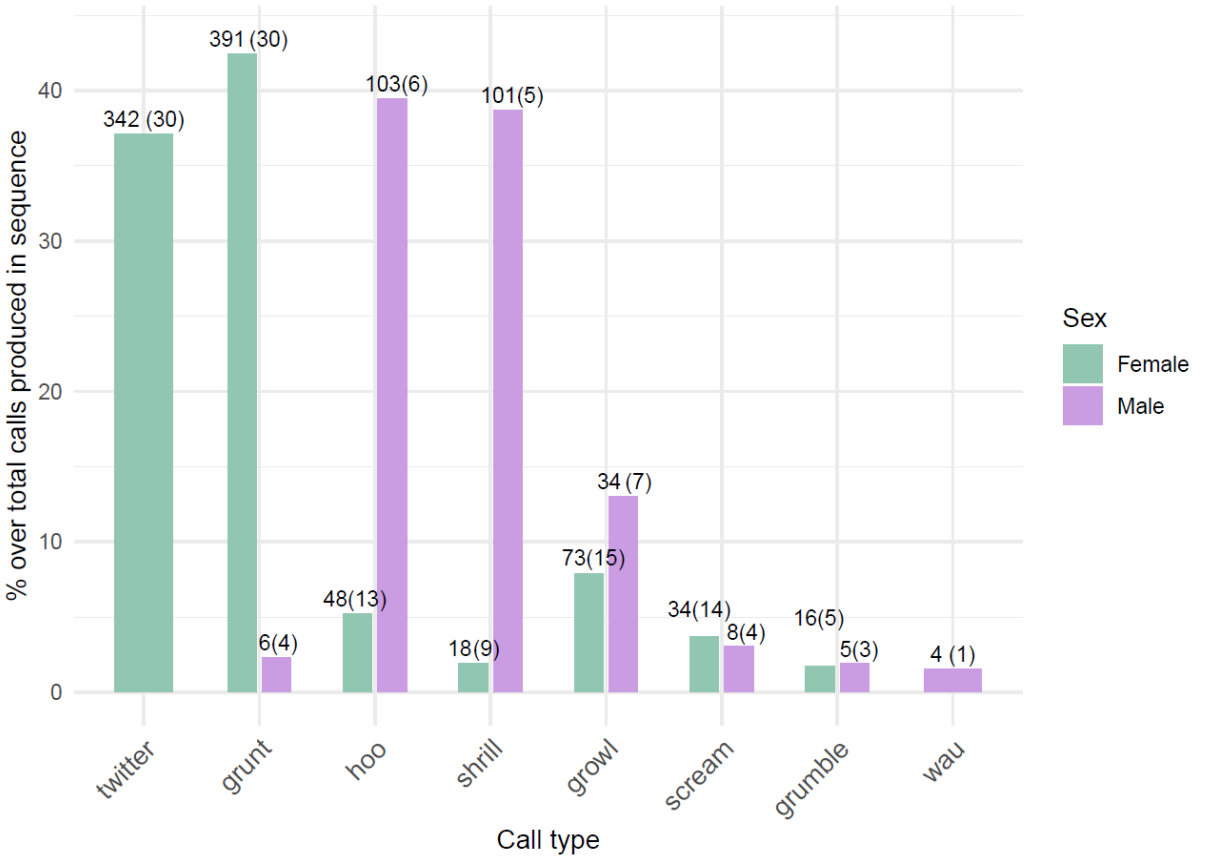


Figure 1. Frequency of call types used in sequences

These numbers represent all occurrences of calls found in sequence, i.e., including iterations (see glossary). Above each bar is indicated the sample size with the number of calls recorded next to the number of individuals producing them in brackets.

Combination pairs

In females, most combination pairs (78.2%) were between ‘grunts’ and ‘twitters’, (Figure 2a and Figure S1). Combination pairs between ‘growls’ and ‘hoos’ accounted for another 7.2%, between ‘growls’ and ‘screams’ 4.8% and between ‘growls’ and ‘grumbles’ 3.4%. In males, most combination pairs were between ‘shrills’ and ‘hoos’ (71.2%). These were followed by combination pairs between ‘growls’ and ‘screams’ (6.5%), ‘growls’ and ‘hoos’ (5.3%), ‘growls’ and ‘grumbles’ (4.7%) and ‘growls’ and ‘grunts’ (4.1%) (Figure 2b and Figure S1).

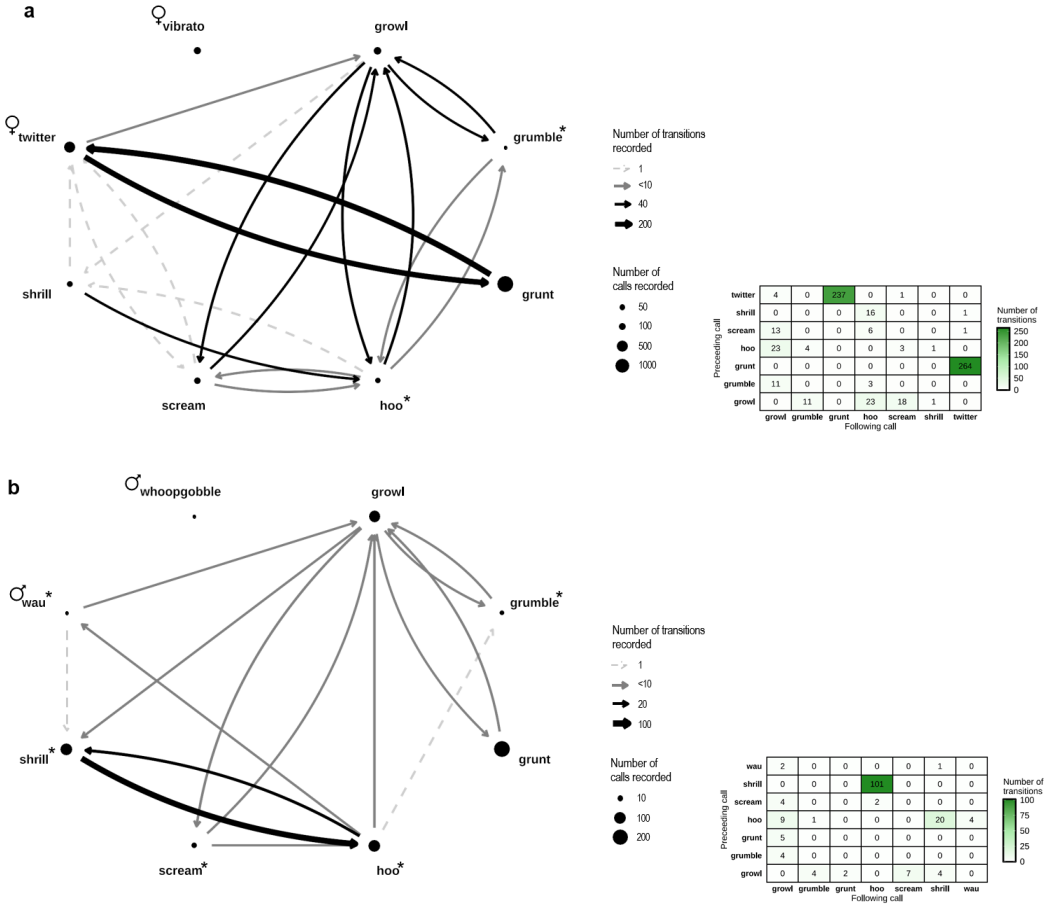


Figure 2. Vocal repertoire of adult sooty mangabey (a) females and (b) males.

These networks represent all recorded call types and transitions between the different call types in sequences, including iterations. * = call type produced exclusively in sequence with another call type (i.e., never or idiosyncratic -maximum one occurrence-singly). ♀ = call type produced only by females. ♂ = call type produced only by males. Node (circle) size is proportional to the occurrence of call for each call type and edge (line) thickness and colour are proportional to the number of recorded transitions. The

associated heatmaps provide an alternative representation of these transitions. See Tables S2 and S4 for detailed sample sizes.

Sequence length

The most frequent sequences were composed of two call types for both sexes (Figure 3). Not counting iterations, the maximum sequence length consisted of four different call types for males (although these were infrequent; three sequences produced by two individuals) and three for females (Figure 3a). For sequences with iterations (i.e., repeated occurrences of one call type), the most frequent length was still two calls, for both sexes. In females, typical sequence length ranged up to nine calls, exceptionally consisting of 38 calls, though these were observed in single individuals. In males, sequences ranged up to six calls (Figure 3b).

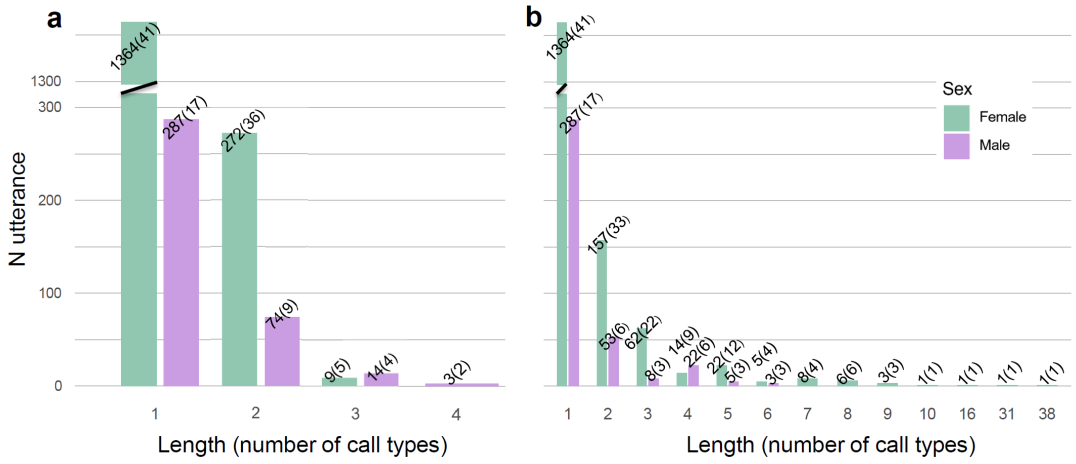


Figure 3. Number of utterances recorded per length and sex (a) without iteration and (b) with iteration

Above each bar is indicated the sample size with the number of utterances recorded next to the number of individuals producing sequences of that length in brackets. Refer to the glossary for definitions of iterations and sequence length.

Sequence diversity

Regarding sequence diversity (see glossary), without iterations, females produced 17 different sequences. The most frequent sequence was grunt_twitter (54.4%) followed

by twitter_grunt (26.3%). With iterations, females produced 41 different sequences. Here again, the most frequent was grunt_twitter (29.2%) followed by grunt_twitter_grunt (15.7%) and twitter_grunt (15.7%).

Males, on the other hand, produced 11 unique sequences without iterations. The most frequent among these was shrill_hoo (72.5%) followed by shrill_hoo_growl (6.6%). When accounting for iterations, males produced 21 unique sequences. The most frequent sequence was shrill_hoo (53.8%) followed by shrill_hoo_shrill_hoo (17.6%). For further details, including sample size, please refer to Tables S1, S2, S3 and S4.

Rules

Ordering - positional bias in 2-call type sequences

We found four call types with positional bias in sequence and two that did not have consistent bias (Table 3 and Figure 4). All ‘shrills’ were found in first position and all ‘hoos’ in second position. ‘Shrills’ only occurred in shrill_hoo sequences whereas ‘hoos’ could be combined, although less frequently, with other call types. We found consistent statistical support for ‘grunt’ to be biased towards the first position (mean probability and 95% CI for ‘grunt’ to be in the first position 0.65, 95%-CI[0.52, 0.79]) and most ‘grunts’ (98.7%) were part of grunt_twitter and twitter_grunt sequences (Table 3 and Figure 4). We found some support, though with greater uncertainty, for ‘twitters’ to be biased toward the second position (mean probability and 95% CI for ‘twitter’ to be in the first position 0.37, 95%-CI[0.23, 0.51]). Similar to ‘grunts’, most ‘twitters’ (99.7%) were found in grunt_twitter and twitter_grunt sequences (Table 3 and Figure 4). We did not find consistent support for a positional bias for ‘growls’ and ‘screams’ (Table 3 and Figure 4; see Table S5 for detailed results and Figure S2 for posterior predictive checks of the models).

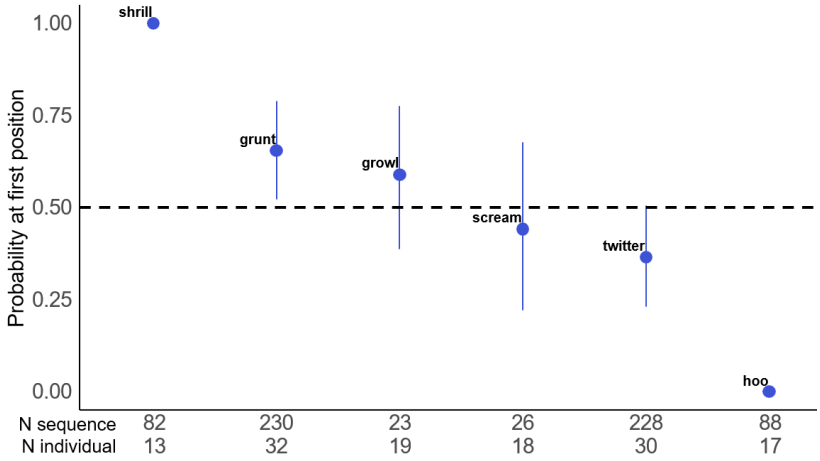


Figure 4. Probability of call types to occur in the first position in 2-call type sequences

The probability is calculated using either (1) the mean posterior distribution of each model ('grunt'/'growl'/'scream'/'twitter') or (2) the mean occurrence of the raw data ('shrill'/'hoo'). The bars represent the 95% credible interval.

Table 3. Positional bias in 2-call type sequences

Shown here are the mean posterior distributions and confidence intervals (CI) for the probability that each call type appears in the first position in a two-call-type sequence; probabilities for the second position are complementary (1 minus the first-position probability). Call types marked with an asterisk (*) are those showing a statistically significant bias. Sequence types are presented here without iterations, so we understand which call type of the 2-call type sequences comes first and which comes second. Sequence types and sample sizes for 'shrill' and 'hoo' are included, although these were not analysed through models.

Call type	**Mean posterior distribution**	**95% CI range**	**Sequence type**	***N* sequence**
shrill	-	-	shrill_hoo	82
Hoo	-	-	shrill_hoo growl_hoo	82 3

			scream_hoo	3
grunt*	0.65	0.52, 0.79	grunt_twitter twitter_grunt grunt_growl	153 74 3
twitter*	0.37	0.23, 0.51	grunt_twitter twitter_grunt twitter_growl	153 74 1
growl	0.59	0.39, 0.78	growl_scream scream_growl growl_grumble growl_hoo grunt_growl grumble_growl twitter_growl	16 7 3 3 3 1 1
scream	0.44	0.22, 0.68	growl_scream scream_growl scream_hoo	16 7 3

Reoccurrence - iterations of calls in sequences composed of ‘grunts’ and ‘twitters’

Considering sequences composed only of ‘grunts’ and ‘twitters’ occurring in iteration and starting with a ‘grunt’, we found consistent support that the last call in a sequence was more likely a ‘grunt’ than a ‘twitter’ (mean probability and 95% CI for ‘grunt’ to be last 0.97, 95%-CI[0.86, 1]). However, for sequences starting with ‘twitter’, we did not find that one call type was consistently more likely to be the last call than the other (mean probability and 95% CI for ‘grunt’ and ‘twitter’ to be last 0.57, 95%-CI[0.12, 0.94]

vs 0.43 95%-CI[0.06, 0.88]) (Figure 5). Restricting the analysis to sequences longer than three calls (N = 49 sequences, 48.5% of all iterative grunt–twitter sequences) produced similar results. For sequences starting with a 'grunt', the last call was still more likely to be a 'grunt' than a 'twitter' (mean probability = 0.85, 95%-CI [0.43, 1.00], 89%-CI [0.55, 0.99); for sequences starting with a 'twitter', there was still no consistent tendency for one call type to occur last (mean probability for 'grunt' to be last = 0.74, 95%-CI [0.20, 0.99], mean probability for 'twitter' to be last = 0.26, 95%-CI [0.01, 0.80]) (see Table S6 for detailed model summaries and Figure S3 for posterior predictive checks of the models).

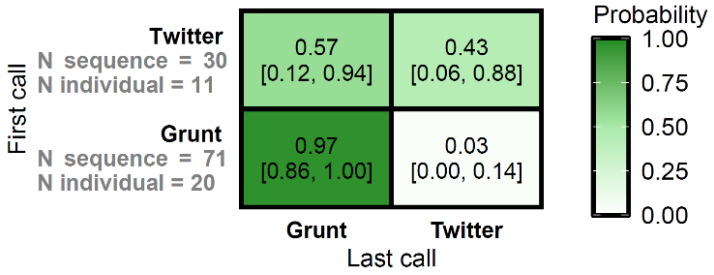


Figure 5. Probability that an iterative sequence ends with a 'grunt' or 'twitter' call (Last call), depending on the call type initiating the sequence (First call)

Forest green shades indicate probability strength (darker = higher). Numbers show the probabilities calculated using the mean posterior distribution of the model, with the 95% credible intervals in brackets.

Reoccurrence - repetition of elements within calls in sequences composed of 'grunts' and 'twitters'

In sequences composed only of 'grunts' and 'twitters' the last 'grunt' call contained more repetition of elements than the first and the middle 'grunt' calls (first 'grunt' calls mean posterior distribution = 2.29, 95%-CI[1.70, 3.04]; middle 'grunt' calls mean posterior distribution = 2.75, 95%-CI[1.87, 3.97]; last 'grunt' calls mean posterior distribution = 5.55, 95%-CI[4.03, 7.56]) (Figure 6a). Pairwise comparisons confirmed these patterns, with the number of elements within the last 'grunt' differing from both the first (mean posterior difference = -0.89, 95% CI [–1.10, –0.66]) and the middle

'grunt' (mean posterior difference = -0.70, 95% CI [-1.01, -0.40]), while the first and middle 'grunts' did not differ (mean posterior difference = -0.18, 95% CI [–0.49, 0.13]).

In these same sequences, we found consistent support for the first 'twitter' call to have more element repetitions than the middle and the last 'twitter' calls (first 'twitter' calls mean posterior distribution = 5.12, 95%-CI[4.29, 6.09]; middle 'twitter' calls mean posterior distribution = 2.52, 95%-CI[1.81, 3.48]; last 'twitter' calls mean posterior distribution = 2.48, 95%-CI[1.91, 3.21]) (Figure 6b). Here again, pairwise comparisons confirmed these patterns: the first 'twitter' call contained more repeated elements than both the middle (posterior mean difference = 0.71 [95% CI 0.40, 1.03]) and last 'twitter' calls (0.72 [0.52, 0.96]), whereas middle and last 'twitter' calls did not differ (0.01 [–0.33, 0.34]).

For both 'grunt' and 'twitter' calls, element repetitions were slightly higher when the sequence started with a 'twitter' compared with a 'grunt' (*model grunt* mean posterior distribution = 3.55 [95% CI 2.61–4.81] vs. 2.29 [1.70–3.04]; *model twitter* mean posterior distribution = 6.75 [5.51–8.13] vs. 5.12 [4.29–6.09]). Moreover, sequence length had a small negative effect on repetitions (*model grunt* mean posterior slope = –0.04 [–0.07, –0.01]; *model twitter* mean posterior slope = –0.04 [–0.06, –0.02]), indicating slightly fewer repeated elements in longer sequences (see Table S7 for detailed summaries of models and Figure S4 for posterior predictive checks).

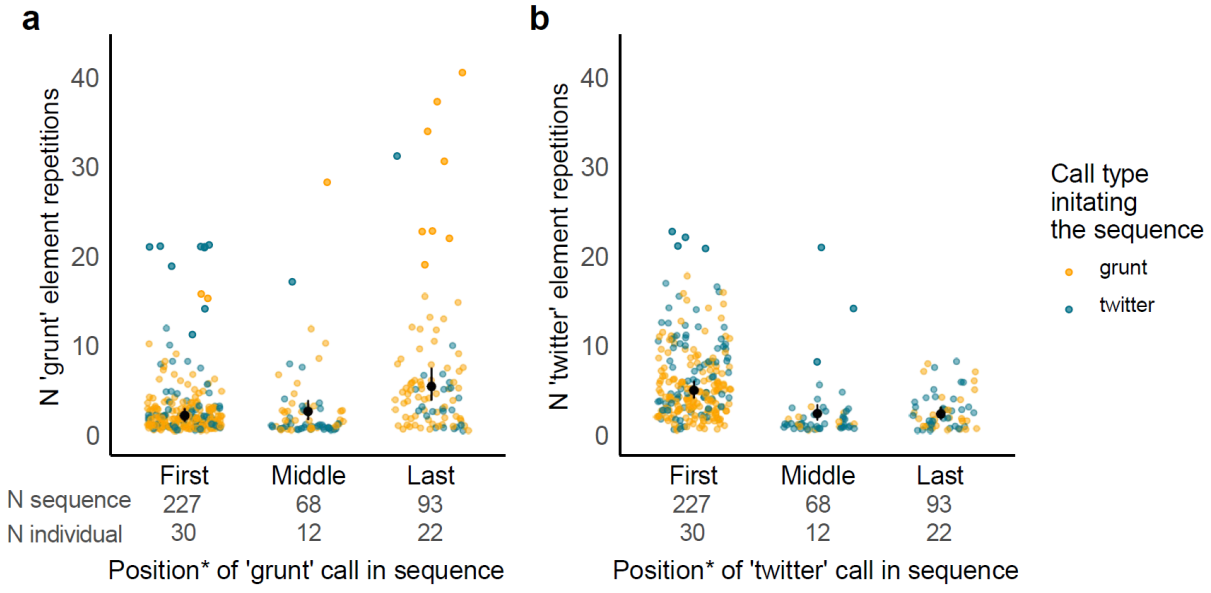


Figure 6. Number of element repetitions for (a) 'grunt' calls and (b) 'twitter' calls, depending on their position* in the sequence

*Position refers to the order of each call type among calls of the same type, regardless of whether the other call type occurs earlier in the sequence (e.g., the “first” ‘grunt’ is the first ‘grunt’ produced, even if a ‘twitter’ precedes it). This includes short sequences (e.g., twitter_grunt, twitter_grunt_twitter), and calls appearing only once in a sequence are treated as “first.” The “middle” position can include one or several calls of the same type, depending on sequence length. Black points and error bars represent the mean posterior distributions and 95% credible intervals of the models. Coloured points show individual calls, with yellow indicating calls from sequences starting with a grunt and blue with a twitter.

Discussion

This study assessed the potential for a *catarrhine* monkey vocal repertoire to produce new meaning through sequences. We quantified sooty mangabey sequence types and analysed their rules of ordering and reocurrence, from which we identified non-adjacent dependencies. Whether these rules change the meaning conveyed to receivers compared to when each call is emitted alone remains to be tested. We discuss these rules and compare the sooty mangabey flexibility in sequence production with that of other species.

Rules

Structural patterns in the two dominant combination pairs—’shrills’ with ‘hoos’ and ‘grunts’ with ‘twitters’—suggest different mechanisms. ‘Shrills’ and ‘hoos’ showed strong positional bias, with ‘shrill’ first and ‘hoo’ second in 100% of sequences. ‘Shrills’ occurred alone, but not ‘hoos’ (one occurrence recorded). One explanation for this could be a similar mechanism asSympatric-living Campbell’s monkeys that add a suffix 'oo' to an alarm call changing the meaning from a high-urgency to a lower-urgency threat (Coye et al., 2015; Ouattara et al., 2009b). Both ‘shrills’ and ‘hoos’ are produced in alarm contexts (León et al., 2023; Range & Fischer, 2004), but confirming whether specific predator information is conveyed to conspecifics requires experimental studies. Sequences composed of ‘grunts’ and ‘twitters’ also showed strong positional bias, with ‘grunts’ being more likely in first position. Unlike shrill_hoo sequences, reverse sequences were also relatively prevalent, with ‘twitter’ in the first position in

32.6% of ‘grunt’ and ‘twitter’ sequences (Table 3). Sooty mangabey sequences of ‘grunts’ and ‘twitter’ resemble those described in other primates, such as chimpanzees (Girard-Buttoz et al., 2025), bonobos (Berthet et al., 2025) and marmosets (Bosshard et al., 2024) as most calls can be produced independently, likely carry their own meaning (Range & Fischer, 2004), and may even change meaning when combined into a specific order (although this remains to be tested in sooty mangabeys). In contrast, more distantly related species—such as bats, songbirds, hyraxes and whales—produce rule-based sequences, but their composing units are typically not produced independently and do not carry meaning on their own (Berwick et al., 2011; Engesser & Townsend, 2019; Kershenbaum et al., 2012; Walmsley et al., 2020; Zhang et al., 2019).

When specifically analysing sequences composed only of ‘grunts’ and ‘twitters’, additional rules of the reoccurrence of each call were evident. Sequences starting with a ‘grunt’ call more likely ended with a ‘grunt’, which was not the case for sequences starting with a ‘twitter’ call (Figure 5). This rule (i.e., iterated sequences starting with a ‘grunt’ tend to end with a ‘grunt’) appears consistent with non-adjacent dependencies in which a relationship between two vocal units cannot be explained solely by adjacent pairings (Wilson et al., 2020). For example, a non-adjacent dependency rule might specify that a sequence starting with ‘A’ must end with ‘C’, regardless of the call types occurring in between (e.g., ABBC is consistent with the rule, whereas ABBA violates it). Such rules have only been found so far in canaries, sparrows and marmosets (Bosshard et al., 2024; Markowitz et al., 2013; Searcy et al., 2022), and it remains unclear if they relate to meaning. Overall, sooty mangabey vocal sequences are built from calls that can be emitted individually or combined into short sequences of two call types, or longer iterative sequences containing the same two call types (e.g., grunt_twitter), which can follow rules of non-adjacent dependencies resulting in predictable structures. Such a system is constrained in that there are limitations in the call types that can combine together into an utterance. It might nonetheless provide several different predictable structures to which meaning could potentially be bound. Thus, such a system holds the capacity to expand messaging well beyond the number of utterances available if two calls combined in only one way and suggests that large repertoires of distinct sequences are not the only option for generating a wide variety of utterances. Whether these predictable sequences carry different meanings remains to be tested.

Regarding element repetitions in sequences composed only of ‘grunts’ and ‘twitters’, the first ‘twitter’ call of the sequence had more element repetitions compared to ‘twitter’ calls produced later in the sequence, and likewise for the last ‘grunt’ call in a sequence (Figure 6a&b). Such variation in repetition according to position may relate to the information conveyed. For instance, we could speculate that greater ‘grunt’ repetitions could signal the end of a sequence, whereas stronger ‘twitter’ repetitions, due to its greater amplitude, could function to capture recipient attention at the start of a sequence. Moreover, both calls had more repetitions when the sequence was initiated by a ‘twitter’, which could potentially reflect a higher-arousal context of production. These possibilities remain to be tested. Additionally, longer sequences (i.e., those with more calls) showed overall slightly fewer element repetitions. This reflects a pattern consistent with Menzerath’s linguistic law (G. Altmann, 1980; Menzerath, 1954) (i.e., the tendency for larger linguistic constructions to have shorter constituents), which has been observed in many primate vocalisations (Moran, Gallot, et al., 2022) and suggested to relates to information transfer efficiency (Gustison et al., 2016). However, confirming that sooty mangabey vocal utterances can adhere to Menzerath’s law require more rigorous testing.

Taken together, our results indicate that sooty mangabeys use rule-based iteration and repetition, but it is worthy to note that these are subject to some inter-individual variability. It remains possible that the joint configuration of iteration and repetition also contributes to meaning or reflects internal factors such as arousal. Investigating this interaction will require analyses that integrate repetition counts with sequence order, ideally in combination with production-context information and with sufficiently large samples per individual.

Sequence flexibility

Sooty mangabeys primarily produced single calls (82.9% of female vocal production and 75.9% of male vocal production) and both sexes can combine most of their calls (nine out of 10 call types, respectively) into sequences. However, some call types—’hoo’, ‘grumble’, and ‘wau’—were almost exclusively produced in sequences. Calls restricted to sequences occur widely across taxa though functional explanations remain unclear (e.g., apes (Girard-Buttoz, Zaccarella, et al., 2022), bats (Zhang et al., 2019), birds (Engesser et al., 2015; Freeberg, 2008; Spiess et al., 2022; Suzuki et al.,

2016; Toshitaka N Suzuki, 2014), cetaceans (Luís et al., 2019; Riesch et al., 2008; Selbmann et al., 2023; Walmsley et al., 2020), frogs (Bhat et al., 2022), monkeys (Bouchet et al., 2010, 2012, 2013; Candiotti et al., 2012; Coye et al., 2015, 2018; Fuller, 2014; Nadhurou et al., 2016; Schel et al., 2009; Wich et al., 2003)), with only a few exceptions (e.g., mongooses (Jansen et al., 2012), Campbell's monkeys (Ouattara et al., 2009b)).

Without iterations, sooty mangabeys produced sequences of two to four distinct call types, consistent with patterns observed across many species e.g., apes (Girard-Buttoz, Bortolato, et al., 2022; Hedwig et al., 2015; Lameira et al., 2024; Leroux et al., 2021, 2023; Schamberg et al., 2016, 2017), birds (Engesser et al., 2015, 2016, 2019, 2024; Freeberg, 2008; Hailman et al., 1987; Spiess et al., 2024; Suzuki et al., 2016; Toshitaka N Suzuki, 2014; Walsh et al., 2019), cetaceans (Selbmann et al., 2023; Walmsley et al., 2020), elephants (Hedwig et al., 2021; Hedwig & Kohlberg, 2024; Pardo et al., 2019), frogs (Bhat et al., 2022; Deng et al., 2024), mongooses (Collier et al., 2017, 2020; Jansen et al., 2012), monkeys (Arnold & Zuberbühler, 2006b, 2006a; Batist et al., 2023; Berthet et al., 2019; Bezerra et al., 2011; Bouchet et al., 2010, 2012, 2013; Candiotti et al., 2012; Cleveland & Snowdon, 2010; Coye et al., 2015, 2018; Hending et al., 2020; Le Floch et al., 2021; Mandl et al., 2019; Mehon et al., 2024; Ouattara et al., 2009b, 2009a; Řeháková-Petrů et al., 2012; Robinson, 1984; Schel et al., 2009; Torti et al., 2013; Wich et al., 2003; Zuberbühler, 2002), although most sequences comprised two calls. Longer sequences combining six to eight different call types are less frequently documented (possibly due to methodological constraints) but have been reported in Amboli bush frogs (Bhat et al., 2022), geladas (Gustison et al., 2016), greater horseshoe bats (Zhang et al., 2019), meerkats (Rauber et al., 2020), and chimpanzees (Girard-Buttoz, Zaccarella, et al., 2022). When iterations are included, sooty mangabeys showed sequence lengths comparable to those of chimpanzees (up to 10 call types) and marmosets (up to nine call types) (Bosshard et al., 2024; Girard-Buttoz, Zaccarella, et al., 2022).

The frequencies of individual call types used in sequences, the types of combination pairs and unique sequences, were unevenly distributed (Figure 1, Figure S1, Tables S1, S2, S3 & S4). All call types were found in sequences produced by at least eight individuals with a minimum of 20 occurrences, except 'waus' that were found in four sequences produced by one individual, 'vibratos' and 'whoop-gobbles' that were only produced singly. Females predominantly used 'grunts' and 'twitters' (79.5% of recorded

calls in sequences), while males primarily used ‘shrills’ and ‘hoos’ (78.2% of recorded calls in sequences) (Figure 1). Five sequences with iterations were predominant, requiring only four call types: grunt_twitter, twitter_grunt and grunt_twitter_grunt in females and shrill_hoo and shrill_hoo_shrill_hoo in males. Notably, more than half of the unique sequences with iterations occurred only once in the dataset (21/41 in females and 14/21 in males) (Tables S2 & S4). To facilitate comparison across species, we focus on two metrics contributing to vocal system flexibility and that are the most frequently reported in the literature: the overall number of call types that are combined and the number of distinct sequences.

Sooty mangabeys—who can combine eight out of their 10 call types—are at the upper end of the range of call type usage in sequences documented in animals (although methodological differences mean such comparisons should be interpreted with caution), alongside chimpanzees (Girard-Buttoz, Zaccarella, et al., 2022; Leroux et al., 2022), common marmosets (Bosshard et al., 2024), black-and-white ruffed lemurs (Batist et al., 2023), Japanese great tits (Toshitaka N Suzuki, 2014), cotton-top tamarins (Cleveland & Snowdon, 2010) and orcas (Selbmann et al., 2023). This is higher than species like Campbell’s monkeys (Bouchet et al., 2013) and Philippine tarsiers (Řeháková-Petrů et al., 2012) which combine about half of their call types, as well as De Brazza’s monkeys (Bouchet et al., 2012), Sahamalaza sportive lemurs (Mandl et al., 2019), and mongoose lemurs (Nadhurou et al., 2016), which combine less than a third of their call types. Surprisingly, closely related red-capped mangabeys combine fewer of their calls (seven out of 12) (Bouchet et al., 2010), although this study was conducted in captivity where social and environmental factors may influence call and sequence use.

The second metric useful for comparative purposes, the number of possible sequences (excluding iterations), shows that sooty mangabeys produced 28 different sequences. Compared to species with a similar vocal repertoire size, sooty mangabeys exhibit greater sequence diversity than Japanese great tits with 19 reported sequences (Toshitaka N Suzuki, 2014), De Brazza’s monkeys—1—(Bouchet et al., 2012), Campbell’s monkeys—3—(Bouchet et al., 2013), red-capped mangabeys—8—(Bouchet et al., 2010) and sahamalaza sportive lemurs—1—(Mandl et al., 2019), though they fall well below chimpanzees—282—(Girard-Buttoz, Zaccarella, et al., 2022) and marmosets—7 different bigrams and 56 different trigrams reported—(Bosshard et al., 2024). While sooty mangabeys can produce a considerable variety of

sequences and are flexible in the number of different call types they can combine, the majority of their sequences use few call types overall, focusing mainly on sequences of two call types. This suggests they primarily use a limited set of common sequences, while other sequences occur less frequently but remain noteworthy.

Despite their ability to combine most calls, sooty mangabeys do not do so frequently, which requires explanation. In Shannon's theory of information, systems with repetitive use of few signals have low information entropy, conveying less information compared to systems with greater signal diversity (Shannon, 1948). Sooty mangabeys might also use less common, non-rule-based sequences to convey meanings. Less common sequences often involve the same calls but arranged in different orders, specifically 'scream', 'hoo', 'grumble' and 'growl' usually emitted in aggression contexts (Range & Fischer, 2004) (Tables S1, S2, S3 & S4). These sequences could result from call production during rapid changes of events (Bortolato et al., 2023) (e.g., in aggressive contexts where social exchanges can unfold at speed with different levels of aggressive behaviours or bystander individuals joining the fight), offering a read-out of rapid social shifts to non-visible conspecifics. However, confirming this hypothesis would require further investigation into the contexts of sequence production. Another possibility is under-sampling of these sequences, particularly in contexts that pose challenges for data collection (see limitations).

Sex differences

Overall, females but never males produced sequences composed of 'grunts' and 'twitters' (note that adult males do not produce 'twitters'). This aligns with findings in closely related red-capped mangabeys where females produce more sequence types and at higher rates than males, likely because their social role is to mediate intragroup social relationships (Bouchet et al., 2010). This contrasts with chimpanzees, which share vocal sequences across sexes (Crockford & Boesch, 2004), suggesting usage constraints in *catarrhine* monkeys compared to apes.

Limitations

Despite careful observations, some vocal utterances in some contexts were difficult to capture and hence may be underrepresented in the dataset, like events occurring in

tall trees, at night, or during rapidly unfolding aggression contexts where vocalisations from different signallers were often simultaneous. Thus, only frequently emitted and well-observed sequences could be analysed in our study. Moreover, call types were classified into broad categories, which may overlook finer acoustic variations within each type. For instance, ‘twitters’ are suggested to have acoustic variations depending on the context of production (Range & Fischer, 2004). If such variations carry specific meaning, they could impact sequence structures, which we did not consider in our analyses.

Conclusion

Sooty mangabeys produce rule-based sequences that include non-adjacent dependencies which have been rarely documented in animal communication. They also combine some calls without using specific rule, with half of their sequence types recorded being idiosyncratic. These mechanisms are likely promoted by different contexts. An advantage of different rule-based sequences using the same call types, such a twitter_grunt and grunt_twitter, may be to convey different meanings depending on the order of production, element repetitions, or on non-adjacent dependencies across extended sequences. Such sequences would for instance allow unambiguous messaging even in low-visibility dense forest habitats in which sooty mangabeys live. Indeed, ‘grunts’ and ‘twitters’ are produced in a wide range of contexts, from daily activities such as moving and feeding to affiliative behaviours (Quintero et al., 2022; Range & Fischer, 2004). This contextual variation will allow future research to determine whether emitting these calls either singly and in differently structured sequences conveys distinct meanings to receivers. In contrast, non-rule-based sequences may be promoted by rapid changes of events, offering for instance a flexibly produced readout of the unfolding of rapid social shifts during aggressive contexts. These findings highlight the importance of a whole-repertoire approach to assess sequence usage and associated rules, as this is key to understanding how vocal systems can potentially expand meanings beyond the number of vocalisations.

References

Altmann, G. (1980). Prolegomena to Menzerath's law. In *Glottometrika* (Vol. 2, pp. 1–10).

Altmann, J. (1974). Observational Study of Behavior: Sampling Methods. *Behaviour*, *49*(3–4), 227–266. https://doi.org/10.1163/156853974X00534

Arnold, K., & Zuberbühler, K. (2006a). Semantic combinations in primate calls. *Nature*, *441*(7091), 303–303. https://doi.org/10.1038/441303a

Arnold, K., & Zuberbühler, K. (2006b). The alarm-calling system of adult male putty-nosed monkeys, Cercopithecus nictitans martini. *Animal Behaviour*, *72*(3), 643–653. https://doi.org/10.1016/j.anbehav.2005.11.017

Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology*, *18*(5), R202–R203. https://doi.org/10.1016/j.cub.2008.01.040

Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: Compositional or idiomatic expressions? *Brain and Language*, *120*(3), 303–309. https://doi.org/10.1016/j.bandl.2011.10.001

Batist, C. H., Razafindraibe, M. N., Randriamanantena, F., & Baden, A. L. (2023). Bioacoustic characterization of the black-and-white ruffed lemur (Varecia variegata) vocal repertoire. *Primates*, *64*(6), 621–635. https://doi.org/10.1007/s10329-023-01083-8

Berthet, M., Coye, C., Dezecache, G., & Kuhn, J. (2023). Animal linguistics: A primer. *Biological Reviews*, *98*(1), 81–98. https://doi.org/10.1111/brv.12897

Berthet, M., Mesbahi, G., Pajot, A., Cäsar, C., Neumann, C., & Zuberbühler, K. (2019). Titi monkeys combine alarm calls to create probabilistic meaning. *Science Advances*, *5*(5), eaav3991. https://doi.org/10.1126/sciadv.aav3991

Berthet, M., Surbeck, M., & Townsend, S. W. (2025). Extensive compositionality in the vocal system of bonobos. *Science*, *388*(6742), 104–108. https://doi.org/10.1126/science.adv1170

Berwick, R. C., Okanoya, K., Beckers, G. J. L., & Bolhuis, J. J. (2011). Songs to syntax: The linguistics of birdsong. *Trends in Cognitive Sciences*, *15*(3), 113–121. https://doi.org/10.1016/j.tics.2011.01.002

Bezerra, B. M., Souto, A. S., Radford, A. N., & Jones, G. (2011). Brevity is not always a virtue in primate communication. *Biology Letters*, *7*(1), 23–25. https://doi.org/10.1098/rsbl.2010.0455

Bhat, A. S., Sane, V. A., Seshadri, K. S., & Krishnan, A. (2022). Behavioural context shapes vocal sequences in two anuran species with different repertoire sizes. *Animal Behaviour*, *184*, 111–129. https://doi.org/10.1016/j.anbehav.2021.12.004

Bortolato, T., Friederici, A. D., Girard-Buttoz, C., Wittig, R. M., & Crockford, C. (2023). Chimpanzees show the capacity to communicate about concomitant daily life events. *iScience*, *26*(11), 108090. https://doi.org/10.1016/j.isci.2023.108090

Bosshard, A. B., Burkart, J. M., Merlo, P., Cathcart, C., Townsend, S. W., & Bickel, B. (2024). Beyond bigrams: Call sequencing in the common marmoset (Callithrix jacchus) vocal system. *Royal Society Open Science*, *11*(11), 240218. https://doi.org/10.1098/rsos.240218

Bouchet, H., Blois-Heulin, C., & Lemasson, A. (2012). Age- and sex-specific patterns of vocal behavior in De Brazza's monkeys (*Cercopithecus neglectus*). *American Journal of Primatology*, *74*(1), 12–28. https://doi.org/10.1002/ajp.21002

Bouchet, H., Blois-Heulin, C., & Lemasson, A. (2013). Social complexity parallels vocal complexity: A comparison of three non-human primate species. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00390

Bouchet, H., Pellier, A., Blois-Heulin, C., & Lemasson, A. (2010). Sex differences in the vocal repertoire of adult red-capped mangabeys (*Cercocebus torquatus*): A multi-level acoustic analysis. *American Journal of Primatology*, *72*(4), 360–375. https://doi.org/10.1002/ajp.20791

Bürkner, P.-C. (2015). *brms: Bayesian Regression Models using "Stan"* (p. 2.22.0) [Dataset]. https://doi.org/10.32614/CRAN.package.brms

Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012). Context-related call combinations in female Diana monkeys. *Animal Cognition*, *15*(3), 327–339. https://doi.org/10.1007/s10071-011-0456-8

Clay, Z., & Zuberbühler, K. (2009). Food-associated calling sequences in bonobos. *Animal Behaviour*, *77*(6), 1387–1396. https://doi.org/10.1016/j.anbehav.2009.02.016

Clay, Z., & Zuberbühler, K. (2011). Bonobos Extract Meaning from Call Sequences. *PLoS ONE*, *6*(4), e18786. https://doi.org/10.1371/journal.pone.0018786

Cleveland, J., & Snowdon, C. T. (2010). The Complex Vocal Repertoire of the Adult Cotton-top Tamarin (Saguinus oedipus oedipus)1). *Zeitschrift Für Tierpsychologie*, *58*(3), 231–270. https://doi.org/10.1111/j.1439-0310.1982.tb00320.x

Collier, K., Radford, A. N., Stoll, S., Watson, S. K., Manser, M. B., Bickel, B., & Townsend, S. W. (2020). Dwarf mongoose alarm calls: Investigating a complex non-human animal call. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1935), 20192514. https://doi.org/10.1098/rspb.2019.2514

Collier, K., Townsend, S. W., & Manser, M. B. (2017). Call concatenation in wild meerkats. *Animal Behaviour*, *134*, 257–269. https://doi.org/10.1016/j.anbehav.2016.12.014

Coye, C., Ouattara, K., Arlet, M. E., Lemasson, A., & Zuberbühler, K. (2018). Flexible use of simple and combined calls in female Campbell's monkeys. *Animal Behaviour*, *141*, 171–181. https://doi.org/10.1016/j.anbehav.2018.05.014

Coye, C., Ouattara, K., Zuberbühler, K., & Lemasson, A. (2015). Suffixation influences receivers' behaviour in non-human primates. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1807), 20150265. https://doi.org/10.1098/rspb.2015.0265

Crockford, C., & Boesch, C. (2004). Call combinations in wild chimpanzees. *Behaviour*, 25.

Deng, K., He, Q.-L., Wang, T.-L., Wang, J.-C., & Cui, J.-G. (2024). Network analysis reveals context-dependent structural complexity of social calls in serrate-legged small treefrogs. *Current Zoology*, *70*(2), 253–261. https://doi.org/10.1093/cz/zoac104

Engesser, S., Crane, J. M. S., Savage, J. L., Russell, A. F., & Townsend, S. W. (2015). Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System. *PLOS Biology*, *13*(6), e1002171. https://doi.org/10.1371/journal.pbio.1002171

Engesser, S., Holub, J. L., O'Neill, L. G., Russell, A. F., & Townsend, S. W. (2019). Chestnut-crowned babbler calls are composed of meaningless shared building blocks. *Proceedings of the National Academy of Sciences*, *116*(39), 19579–19584. https://doi.org/10.1073/pnas.1819513116

Engesser, S., Ridley, A. R., & Townsend, S. W. (2016). Meaningful call combinations and compositional processing in the southern pied babbler. *Proceedings of the National Academy of Sciences*, *113*(21), 5976–5981. https://doi.org/10.1073/pnas.1600970113

Engesser, S., Ridley, A. R., & Townsend, S. W. (2017). Element repetition rates encode functionally distinct information in pied babbler 'clucks' and 'purrs.' *Animal Cognition*, *20*(5), 953–960. https://doi.org/10.1007/s10071-017-1114-6

Engesser, S., Ridley, A. R., Watson, S. K., Kita, S., & Townsend, S. W. (2024). Seeds of language-like generativity in bird call combinations. *Proceedings of the Royal Society B: Biological Sciences*, *291*(2033), 20240922. https://doi.org/10.1098/rspb.2024.0922

Engesser, S., & Townsend, S. W. (2019). Combinatoriality in the vocal systems of nonhuman animals. *Wiley Interdisciplinary Reviews: Cognitive Science*, e1493. https://doi.org/10.1002/wcs.1493

Fitch, W. T. (2010). *The evolution of language* (Cambridge University Press).

Freeberg, T. M. (2008). Complexity in the chick-a-dee call of carolina chickadees (Poecile carolinesis): Associations of context and signlater behavior to call structure. *The Auk*, *125*(4), 896–907. https://doi.org/10.1525/auk.2008.07025

Fuller, J. L. (2014). The vocal repertoire of adult male blue monkeys (Cercopithecus mitis stulmanni): A quantitative analysis of acoustic structure. *American Journal of Primatology*, *76*(3), 203–216. https://doi.org/10.1002/ajp.22223

Girard-Buttoz, C., Bortolato, T., Laporte, M., Grampp, M., Zuberbühler, K., Wittig, R. M., & Crockford, C. (2022). Population-specific call order in chimpanzee greeting vocal sequences. *iScience*, *25*(9), 104851. https://doi.org/10.1016/j.isci.2022.104851

Girard-Buttoz, C., Neumann, C., Bortolato, T., Zaccarella, E., Friederici, A. D., Wittig, R. M., & Crockford, C. (2025). Versatile use of chimpanzee call combinations promotes meaning expansion. *Science Advances*, *11*(19), eadq2879. https://doi.org/10.1126/sciadv.adq2879

Girard-Buttoz, C., Zaccarella, E., Bortolato, T., Friederici, A. D., Wittig, R. M., & Crockford, C. (2022). Chimpanzees produce diverse vocal sequences with ordered and recombinatorial properties. *Communications Biology*, *5*(1), 410. https://doi.org/10.1038/s42003-022-03350-8

Grampp, M., Samuni, L., Girard-Buttoz, C., León, J., Zuberbühler, K., Tkaczynski, P., Wittig, R. M., & Crockford, C. (2023). Social uncertainty promotes signal complexity during approaches in wild chimpanzees (*Pan troglodytes verus*) and mangabeys (*Cercocebus atys atys*). *Royal Society Open Science*, *10*(11), 231073. https://doi.org/10.1098/rsos.231073

Gustison, M. L., Semple, S., Ferrer-i-Cancho, R., & Bergman, T. J. (2016). Gelada vocal sequences follow Menzerath's linguistic law. *Proceedings of the National Academy of Sciences*, *113*(19), E2750–E2758. https://doi.org/10.1073/pnas.1522072113

Hailman, J. P., Ficken, M. S., & Ficken, R. W. (1987). Constraints on the Structure of Combinatorial "Chick-a-dee" Calls. *Ethology*, *75*(1), 62–80. https://doi.org/10.1111/j.1439-0310.1987.tb00642.x

Hauser, M. D. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Hedwig, D., & Kohlberg, A. (2024). Call combination in African forest elephants Loxodonta cyclotis. *PLOS ONE*, *19*(3), e0299656. https://doi.org/10.1371/journal.pone.0299656

Hedwig, D., Mundry, R., Robbins, M. M., & Boesch, C. (2015). Contextual correlates of syntactic variation in mountain and western gorilla close-distance vocalizations: Indications for lexical or phonological syntax? *Animal Cognition*, *18*(2), 423–435. https://doi.org/10.1007/s10071-014-0812-6

Hedwig, D., Poole, J., & Granli, P. (2021). Does Social Complexity Drive Vocal Complexity? Insights from the Two African Elephant Species. *Animals*, *11*(11), 3071. https://doi.org/10.3390/ani11113071

Hending, D., Seiler, M., & Stanger-Hall, K. F. (2020). The Vocal Repertoire of the Northern Giant Mouse Lemur (Mirza zaza) in Captivity. *International Journal of Primatology*, *41*(5), 732–763. https://doi.org/10.1007/s10764-020-00181-4

Hockett, C. F., & Hockett, C. D. (1960). The Origin of Speech. *Scientific American*, *203*(3), 88–97. JSTOR.

Hurford, J. R. (2012). Linguistics From an Evolutionary Point of View. In *Philosophy of Linguistics* (pp. 477–502). Elsevier. https://doi.org/10.1016/B978-0-444-51747-0.50014-7

Jansen, D. A., Cant, M. A., & Manser, M. B. (2012). Segmental concatenation of individual signatures and context cues in banded mongoose (Mungos mungo) close calls. *BMC Biology*, *10*(1), 97. https://doi.org/10.1186/1741-7007-10-97

K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology. (2024). *Raven Pro: Interactive Sound Analysis Software* (Version 1.6.5) [Computer software]. The Cornell Lab of Ornithology. https://www.ravensoundsoftware.com/

Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., Bohn, K., Cao, Y., Carter, G., Cäsar, C., Coen, M., DeRuiter, S. L., Doyle, L., Edelman, S., Ferrer-i-Cancho, R., Freeberg, T. M., Garland, E. C., Gustison, M., Harley, H. E., … Zamora-Gutierrez, V. (2016). Acoustic sequences in non-human animals: A tutorial review and prospectus: Acoustic sequences in animals. *Biological Reviews*, *91*(1), 13–52. https://doi.org/10.1111/brv.12160

Kershenbaum, A., Ilany, A., Blaustein, L., & Geffen, E. (2012). Syntactic structure and geographical dialects in the songs of male rock hyraxes. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1740), 2974–2981. https://doi.org/10.1098/rspb.2012.0322

Lameira, A. R., Hardus, M. E., Ravignani, A., Raimondi, T., & Gamba, M. (2024). Recursive self-embedded vocal motifs in wild orangutans. *eLife*, *12*, RP88348. https://doi.org/10.7554/eLife.88348.3

Landsborough, B., Wilson, D. R., & Mennill, D. J. (2019). Variation in chick-a-dee call sequences, not in the fine structure of chick-a-dee calls, influences mobbing behaviour in mixed-species flocks. *Behavioral Ecology*, arz155. https://doi.org/10.1093/beheco/arz155

Le Floch, A., Bouchard, A., Gallot, Q., & Zuberbühler, K. (2021). Lesser spot-nosed monkeys coordinate alarm call production with associated Campbell's monkeys. *Behavioral Ecology and Sociobiology*, *75*(8), 112. https://doi.org/10.1007/s00265-021-03053-w

Lenth, R. V. (2017). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (p. 1.10.5) [Dataset]. https://doi.org/10.32614/CRAN.package.emmeans

León, J., Thiriau, C., Crockford, C., & Zuberbühler, K. (2023). Comprehension of own and other species' alarm calls in sooty mangabey vocal development. *Behavioral Ecology and Sociobiology*, *77*(5), 56. https://doi.org/10.1007/s00265-023-03318-6

Leroux, M., Bosshard, A. B., Chandia, B., Manser, A., Zuberbühler, K., & Townsend, S. W. (2021). Chimpanzees combine pant hoots with food calls into larger structures. *Animal Behaviour*, *179*, 41–50. https://doi.org/10.1016/j.anbehav.2021.06.026

Leroux, M., Chandia, B., Bosshard, A. B., Zuberbühler, K., & Townsend, S. W. (2022). Call combinations in chimpanzees: A social tool? *Behavioral Ecology*, arac074. https://doi.org/10.1093/beheco/arac074

Leroux, M., Schel, A. M., Wilke, C., Chandia, B., Zuberbuhler, K., Slocombe, K. E., & Townsend, S. W. (2023). Call combinations and compositional processing in wild chimpanzees. *Nature Communications*, *14*, 2225. https://doi.org/10.1038/s41467-023-37816-y

Luís, A. R., Alves, I. S., Sobreira, F. V., Couchinho, M. N., & Dos Santos, M. E. (2019). Brays and bits: Information theory applied to acoustic communication sequences of bottlenose dolphins. *Bioacoustics*, *28*(3), 286–296. https://doi.org/10.1080/09524622.2018.1443285

Maddieson, I., & Precoda, K. (1989). Updating UPSID. *The Journal of the Acoustical Society of America*, *86*(S1), S19–S19. https://doi.org/10.1121/1.2027403

Mandl, I., Schwitzer, C., & Holderied, M. (2019). *Sahamalaza Sportive Lemur, Lepilemur sahamalaza, Vocal Communication: Call Use, Context and Gradation*. https://doi.org/10.1159/000493939

Markowitz, J. E., Ivie, E., Kligler, L., & Gardner, T. J. (2013). Long-range Order in Canary Song. *PLoS Computational Biology*, *9*(5), e1003052. https://doi.org/10.1371/journal.pcbi.1003052

Marler, P. (1998). Animal communication and human language. *The Origin and Diversification of Language*, *1*, 19.

Mehon, F. G., Zuberbühler, K., & Stephan, C. (2024). Population differences in putty-nosed monkey (Cercopithecus nictitans) call order. *Primates*. https://doi.org/10.1007/s10329-024-01155-3

Menzerath, P. (1954). *Die Architektonik des deutschen Wortschatzes* (F. Dümmler).

Moran, S., & Bickel, B. (2020). Rejoinder to Huijbregts's: Biting into Evolution of Language. *Journal of Language Evolution*, *5*(2), 184–187. https://doi.org/10.1093/jole/lzaa005

Moran, S., Gallot, Q., Le Floch, Auriane, Maiolini, M., Soldati, A., & Zuberbühler, Klaus. (2022). Linguistic laws in primate vocalizations. *Proceedings of the Joint Conference on Language Evolution (JCoLE)*, 506–513.

Moran, S., Maiolini, M., & Lameira, Adriano. (2022). Great ape vocal repertoires are all similar in size: Now what? *Proceedings of the Joint Conference on Language Evolution*, 512--516. https://evolang.org/jcole2022/proceedings/jcole2022_proceedings.pdf

Moran, S., McCloy, D., & Wright, R. (2012). Revisiting population size vs. Phoneme inventory size. *Language*, *88*(4), 877–893. https://doi.org/10.1353/lan.2012.0087

Nadhurou, B., Gamba, M., Andriaholinirina, N. V., Ouledi, A., & Giacoma, C. (2016). The vocal communication of the mongoose lemur (Eulemur mongoz): Phonation mechanisms, acoustic features and quantitative analysis. *Ethology Ecology & Evolution*, *28*(3), 241–260. https://doi.org/10.1080/03949370.2015.1039069

Narbona Sabaté, L., Mesbahi, G., Dezecache, G., Cäsar, C., Zuberbühler, K., & Berthet, M. (2022). Animal linguistics in the making: The Urgency Principle and titi monkeys' alarm system. *Ethology Ecology & Evolution*, *34*(3), 378–394. https://doi.org/10.1080/03949370.2021.2015452

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature*, *417*(6889), 611–617. https://doi.org/10.1038/nature00771

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009a). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, *106*(51), 22026–22031. https://doi.org/10.1073/pnas.0908118106

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009b). Campbell's Monkeys Use Affixation to Alter Call Meaning. *PLoS ONE*, *4*(11), e7808. https://doi.org/10.1371/journal.pone.0007808

Pardo, M. A., Poole, J. H., Stoeger, A. S., Wrege, P. H., O'Connell-Rodwell, C. E., Padmalal, U. K., & De Silva, S. (2019). Differences in combinatorial calls among the 3

elephant species cannot be explained by phylogeny. *Behavioral Ecology*, *30*(3), 809–820. https://doi.org/10.1093/beheco/arz018

Quintero, F., Touitou, S., Magris, M., & Zuberbühler, K. (2022). The Evolution of Food Calls: Vocal Behaviour of Sooty Mangabeys in the Presence of Food. *Frontiers in Psychology*, *13*, 897318. https://doi.org/10.3389/fpsyg.2022.897318

R Core Team. (2021). *R: A language and environment for statistical computing* (Version 2021.09.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Range, F., & Fischer, J. (2004). Vocal Repertoire of Sooty Mangabeys (Cercocebus torquatus atys) in the Tai National Park. *Ethology*, *110*(4), 301–321. https://doi.org/10.1111/j.1439-0310.2004.00973.x

Rauber, R., Kranstauber, B., & Manser, M. B. (2020). Call order within vocal sequences of meerkats contains temporary contextual and individual information. *BMC Biology*, *18*(1), 119. https://doi.org/10.1186/s12915-020-00847-8

Řeháková-Petrů, M., Policht, R., & Peške, L. (2012). Acoustic Repertoire of the Philippine Tarsier (Tarsius syrichta fraterculus) and Individual Variation of Long-Distance Calls. *International Journal of Zoology*, *2012*(1), 602401. https://doi.org/10.1155/2012/602401

Riesch, R., Ford, J. K. B., & Thomsen, F. (2008). Whistle sequences in wild killer whales (*Orcinus orca*). *The Journal of the Acoustical Society of America*, *124*(3), 1822–1829. https://doi.org/10.1121/1.2956467

Robinson, J. G. (1984). Syntactic Structures in the Vocalizations of Wedge-Capped Capuchin Monkeys, Cebus Olivaceus. *Behaviour*, *90*(1–3), 46–78. https://doi.org/10.1163/156853984X00551

Salis, A., Léna, J.-P., & Lengagne, T. (2022). Which acoustic parameters modify the great tit's response to conspecific combinatorial mobbing calls? *Behavioral Ecology and Sociobiology*, *76*(4), 46. https://doi.org/10.1007/s00265-022-03157-x

Schamberg, I., Cheney, D. L., Clay, Z., Hohmann, G., & Seyfarth, R. M. (2016). Call combinations, vocal exchanges and interparty movement in wild bonobos. *Animal Behaviour*, *122*, 109–116. https://doi.org/10.1016/j.anbehav.2016.10.003

Schamberg, I., Cheney, D. L., Clay, Z., Hohmann, G., & Seyfarth, R. M. (2017). Bonobos use call combinations to facilitate inter-party travel recruitment. *Behavioral Ecology and Sociobiology*, *71*(4), 75. https://doi.org/10.1007/s00265-017-2301-9

Schel, A. M., Tranquilli, S., & Zuberbühler, K. (2009). The alarm call system of two species of black-and-white colobus monkeys (Colobus polykomos and Colobus guereza). *Journal of Comparative Psychology*, *123*(2), 136–150. https://doi.org/10.1037/a0014280

Searcy, W. A., Soha, J., Peters, S., & Nowicki, S. (2022). Long-distance dependencies in birdsong syntax. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1967), 20212473. https://doi.org/10.1098/rspb.2021.2473

Selbmann, A., Miller, P. J. O., Wensveen, P. J., Svavarsson, J., & Samarra, F. I. P. (2023). Call combination patterns in Icelandic killer whales (Orcinus orca). *Scientific Reports*, *13*(1), 21771. https://doi.org/10.1038/s41598-023-48349-1

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. The Bell System Technical Journal. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sigmundson, R., Girard-Buttoz, C., Le Floch, A., Azaiez, T. S., McElreath, R., Zuberbühler, K., Wittig, R. M., & Crockford, C. (2025). Vocal sequence diversity and length remain stable across ontogeny in a catarrhine monkey (Cercocebus atys). *Communications Biology*, *8*(1), 465. https://doi.org/10.1038/s42003-025-07922-2

Spiess, S., Mylne, H. K., Engesser, S., Mine, J. G., O'Neill, L. G., Russell, A. F., & Townsend, S. W. (2022). Syntax-like Structures in Maternal Contact Calls of Chestnut-Crowned Babblers (Pomatostomus ruficeps). *International Journal of Primatology*. https://doi.org/10.1007/s10764-022-00332-9

Spiess, S., Mylne, H. K., Engesser, S., Mine, J. G., O'Neill, L. G., Russell, A. F., & Townsend, S. W. (2024). Syntax-like Structures in Maternal Contact Calls of Chestnut-Crowned Babblers (Pomatostomus ruficeps). *International Journal of Primatology*, *45*(3), 543–562. https://doi.org/10.1007/s10764-022-00332-9

Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nature Communications*, *7*(1), 10986. 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. (2020). The syntax–semantics interface in animal vocal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20180405. https://doi.org/10.1098/rstb.2018.0405

Torti, V., Gamba, M., Rabemananjara, Z. H., & Giacoma, C. (2013). The songs of the indris (Mammalia: Primates: Indridae): contextual variation in the long-distance calls of a lemur. *Italian Journal of Zoology*, *80*(4), 596–607. https://doi.org/10.1080/11250003.2013.845261

Toshitaka N Suzuki. (2014). Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. *Animal Behaviour*, 7.

Walmsley, S. F., Rendell, L., Hussey, N. E., & Marcoux, M. (2020). Vocal sequences in narwhals (Monodon monoceros). *The Journal of the Acoustical Society of America*, *147*(2), 1078–1091. https://doi.org/10.1121/10.0000671

Walsh, S. L., Townsend, S. W., Morgan, K., & Ridley, A. R. (2019). Investigating the potential for call combinations in a lifelong vocal learner. *Ethology*, *125*(6), 362–368. https://doi.org/10.1111/eth.12860

Wich, S. A., Koski, S., de Vries, H., & van Schaik, C. P. (2003). Individual and Contextual Variation in Thomas Langur Male Loud Calls. *Ethology*, *109*(1), 1–13. https://doi.org/10.1046/j.1439-0310.2003.00837.x

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., Van Der Kant, A., Smith, K., & Rey, A. (2020). Non-adjacent Dependency Learning in Humans and Other Animals. *Topics in Cognitive Science*, *12*(3), 843–858. https://doi.org/10.1111/tops.12381

Zhang, K., Liu, T., Liu, M., Li, A., Xiao, Y., Metzner, W., & Liu, Y. (2019). Comparing context-dependent call sequences employing machine learning methods: An indication of syntactic structure of greater horseshoe bats. *Journal of Experimental Biology*, jeb.214072. https://doi.org/10.1242/jeb.214072

Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, *63*(2), 293–299. https://doi.org/10.1006/anbe.2001.1914

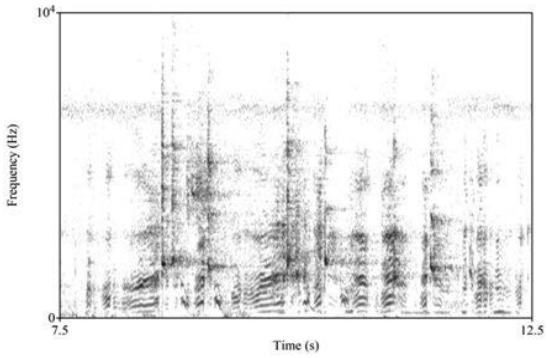
Supplementary material

Document S1. Call types

Range and Fischer (2004) have thoroughly described the complete vocal repertoire of sooty mangabeys, and readers seeking more detail should refer to their paper. For convenience, we provide a condensed version here including all the call types examined in our study.

Growl

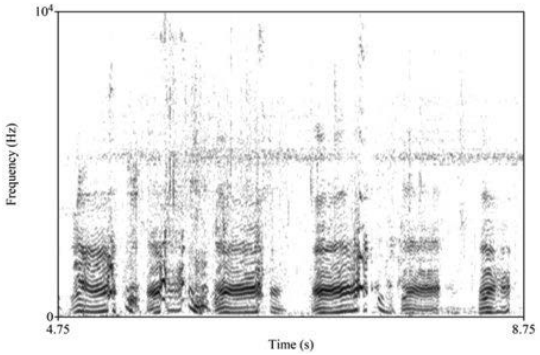
Growls are characterised by a low fundamental frequency band with many visible overtones, and they vary in both frequency and duration, often comprising a combination of shorter and prolonged elements.



Spectrographic examples of growls. y axis represents the frequency in Hz and x axis the time in s.

Grumble

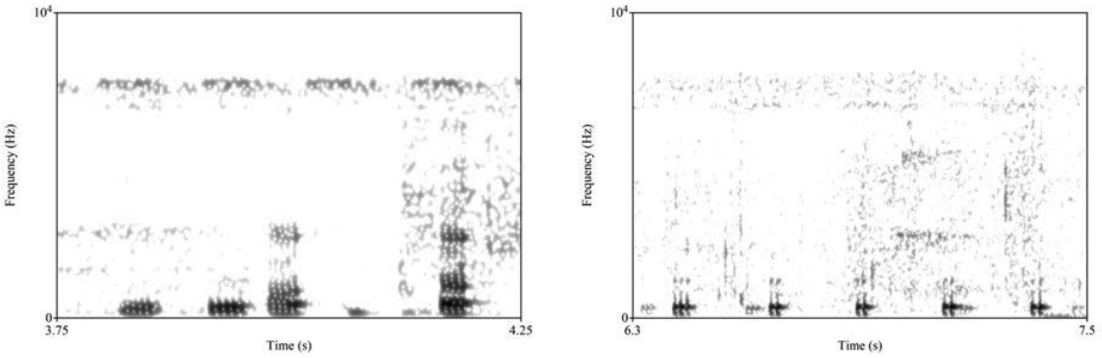
Grumbles are longer than growls, characterised by a low fundamental frequency and a rich harmonic structure with up to 12 overtones.



Spectrographic examples of grumbles interspersed with growls (shorter vocalisations). y axis represents the frequency in Hz and x axis the time in s.

Grunt

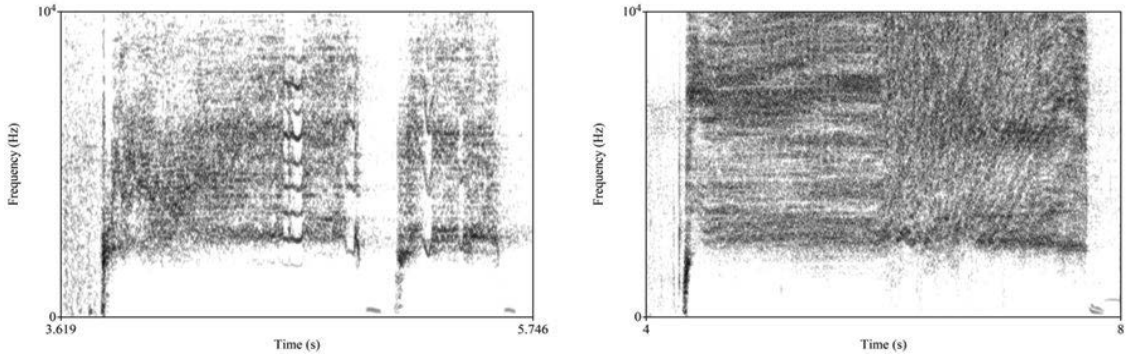
Grunts are low-frequency vocalisations, consistently short in duration (less than 200 milliseconds), and produced across a wide range of contexts. Adult females and males diverge in grunt production (see Range and Fischer).



Spectrograms of grunt calls, from an adult female (left) and an adult male (right). y axis represents the frequency in Hz and x axis the time in s.

Scream

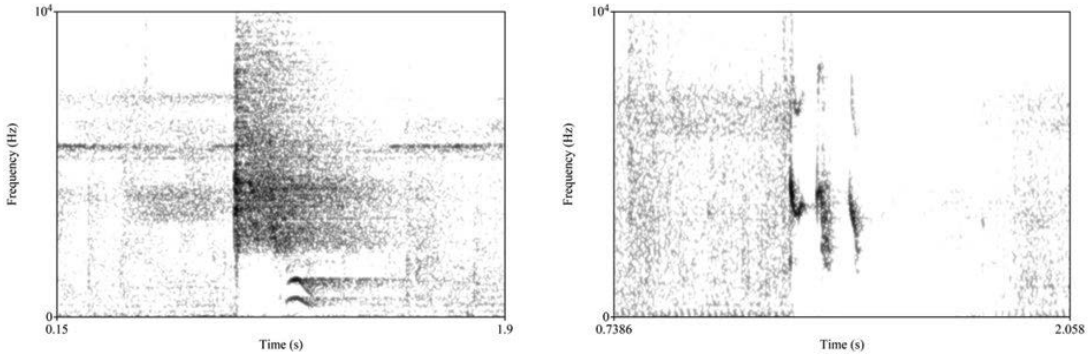
Screams are harsh calls characterised by wide-band noise extending up to 15 kHz, often lasting for several seconds.



Spectrographic examples of screams. y axis represents the frequency in Hz and x axis the time in s.

Shrill and hoo (referred to as alarm call in Range and Fischer, 2004)

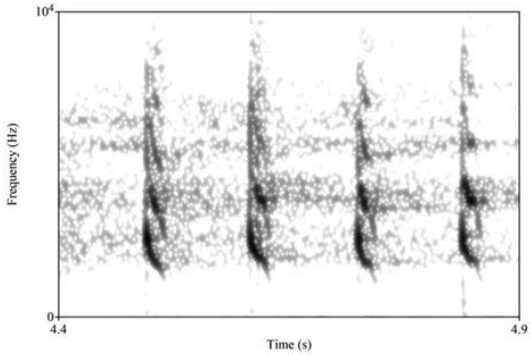
Shrill calls consist of several high-frequency elements, sometimes displaying a harmonic structure, though this is less common, and are frequently combined with a hoo call, which features a low fundamental frequency and a rich, readily visible harmonic structure.



Spectrograms of shrill calls combined with a hoo call (left) and shrill call produced singly (right). y axis represents the frequency in Hz and x axis the time in s.

Twitter

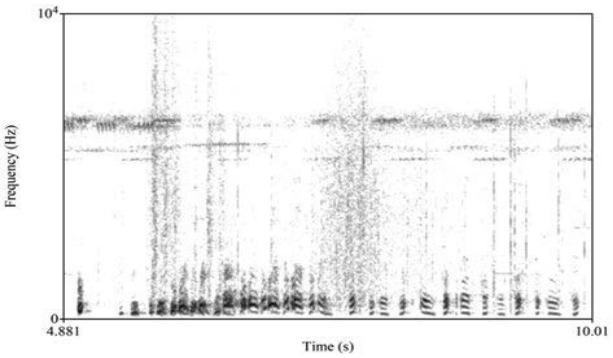
Twitters are short, high-frequency calls with a fundamental frequency around 1 kHz and many visible overtones extending up to 20 kHz. Adult males do not produce them.



Spectrographic example of twitter. y axis represents the frequency in Hz and x axis the time in s.

Vibrato (referred to as copulation call in Range and Fischer, 2004)

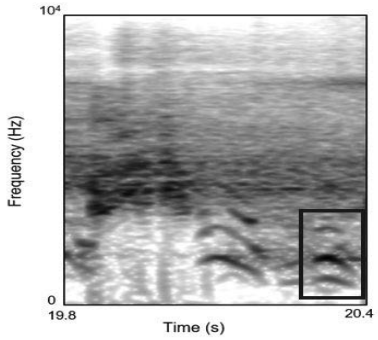
Complex call used exclusively by sexually mature females. It consists of many variable elements with a low fundamental frequency and a rich harmonic structure, with individual elements alternating between pitch shifts upwards and downwards, creating the characteristic "vibrato."



Spectrographic example of vibrato. y axis represents the frequency in Hz and x axis the time in s.

Wau

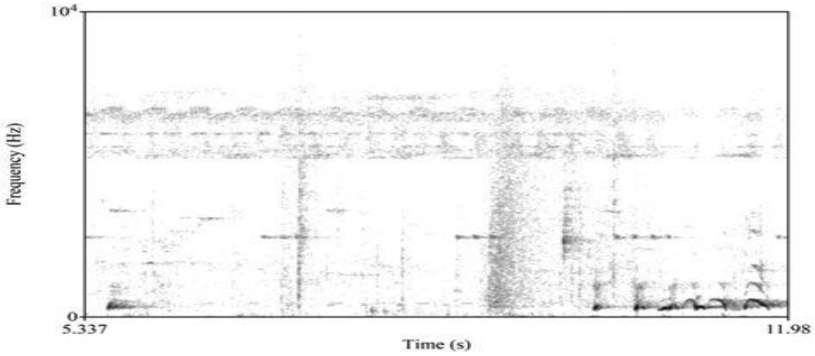
Short, low-frequency call with an ascending-descending pitch pattern and several visible harmonics.



Spectrographic example of wau (in the black box) preceded by shrill and hoo. y axis represents the frequency in Hz and x axis the time in s.

Whoop-gobble

Loud, low-frequency complex call that can be heard over long distances and produced by adult males. It begins with a single low-frequency element, followed by a pause that may last several seconds, before concluding with a series of repetitive, low-frequency elements. Although the pause between the first and final elements typically violates our usual rule for defining a call type, we treated it as a special case due to its consistent production in this form.



Spectrographic example of whoop-gobble. y axis represents the frequency in Hz and x axis the time in s.

Reference

Range, F., & Fischer, J. (2004). Vocal repertoire of sooty mangabeys (Cercocebus torquatus atys) in the Taï National Park. Ethology, 110(4), 301-321.

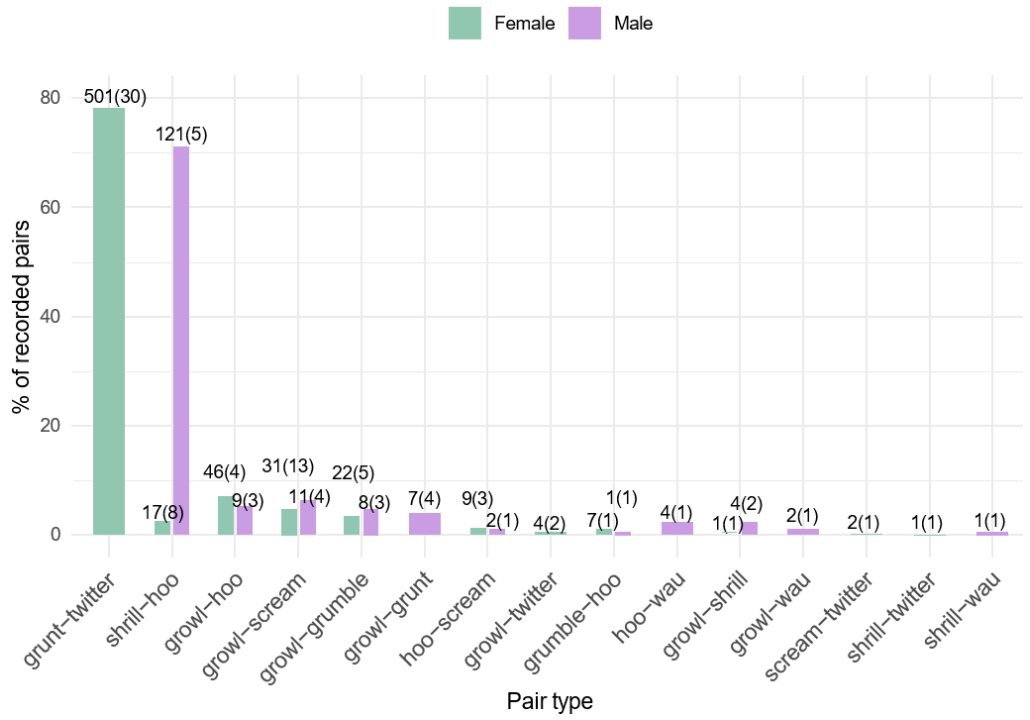


Figure S1. Percentage of combination pairs recorded per sex

Above each bar is indicated the sample size with the number of pair recorded next to the number of individuals producing them in brackets.

Table S1. Female sequence types without call iterations

Sequence type	N sequence	N individual	%
grunt_twitter	153	28	54.4
twitter_grunt	74	20	26.3
shrill_hoo	16	8	5.7
growl_scream	11	8	3.9
scream_growl	7	7	2.5
growl_grumble	3	3	1.1

growl_hoo	3	3	1.1
growl_hoo_grumble	3	1	1.1
scream_hoo	3	2	1.1
growl_scream_grumble	1	1	0.4
growl_scream_hoo	1	1	0.4
grumble_growl	1	1	0.4
grunt_twitter_growl	1	1	0.4
scream_growl_grumble	1	1	0.4
scream_twitter_growl	1	1	0.4
twitter_growl	1	1	0.4
twitter_growl_shrill	1	1	0.4

Table S2. Female sequence types with call iterations

Sequence type	**N sequence**	**N individual**	**%**
grunt_twitter	82	26	29.2
grunt_twitter_grunt	44	16	15.7
twitter_grunt	44	17	15.7
grunt_twitter_grunt_twitter_grunt	15	7	5.3

shrill_hoo	15	7	5.3
growl_scream	9	6	3.2
twitter_grunt_twitter_grunt	9	6	3.2
twitter_grunt_twitter	8	4	2.8
grunt_twitter_grunt_twitter_grunt_twitter_grunt	5	3	1.8
scream_growl	4	4	1.4
twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt	4	4	1.4
growl_grumble_growl	3	3	1.1
grunt_twitter_grunt_twitter	3	2	1.1
scream_growl_scream	3	3	1.1
growl_hoo_growl_hoo_growl	2	2	0.7
growl_scream_growl	2	2	0.7
grunt_twitter_grunt_twitter_grunt_twitter	2	2	0.7
grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt	2	2	0.7
twitter_grunt_twitter_grunt_twitter	2	2	0.7
twitter_grunt_twitter_grunt_twitter_grunt	2	2	0.7

growl_hoo_growl_grumble_growl	1	1	0.4
growl_hoo_growl_hoo_growl_hoo_growl	1	1	0.4
growl_hoo_growl_hoo_growl_hoo_growl_grumble	1	1	0.4
growl_hoo_growl_hoo_growl_hoo_growl_hoo_growl_hoo_grumble_hoo_grumble_growl_grumble_hoo_growl_hoo_grumble_growl_hoo_growl_hoo_grumble_hoo_growl_grumble_growl_hoo_growl_grumble_growl_hoo_growl_hoo_growl	1	1	0.4
growl_scream_growl_grumble	1	1	0.4
growl_scream_growl_scream_hoo_growl_scream_hoo	1	1	0.4
grumble_growl	1	1	0.4
grunt_twitter_grunt_twitter_grunt_twitter_growl	1	1	0.4
scream_growl_grumble_growl_grumble_growl	1	1	0.4
scream_hoo	1	1	0.4
scream_hoo_scream	1	1	0.4
scream_hoo_scream_hoo_scream	1	1	0.4
scream_twitter_scream_growl	1	1	0.4

shrill_hoo_shrill	1	1	0.4
twitter_growl	1	1	0.4
twitter_growl_shrill_twitter_growl	1	1	0.4
twitter_grunt_twitter_grunt_twitter_grunt_twitter	1	1	0.4
twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter	1	1	0.4
twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt	1	1	0.4
twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt	1	1	0.4
twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter_grunt_twitter	1	1	0.4

Table S3. Male sequence types without call iterations

Sequence type	**N sequence**	**N individual**	**%**
shrill_hoo	66	5	72.5
shrill_hoo_growl	6	1	6.6
growl_scream	5	4	5.5

growl_shrill_hoo	3	2	3.3
grunt_growl	3	2	3.3
shrill_hoo_wau	2	1	2.2
shrill_hoo_wau_growl	2	1	2.2
growl_grumble_grunt	1	1	1.1
growl_scream_hoo	1	1	1.1
grunt_growl_grumble	1	1	1.1
scream_hoo_grumble_growl	1	1	1.1

Table S4. Male sequence types with call iterations

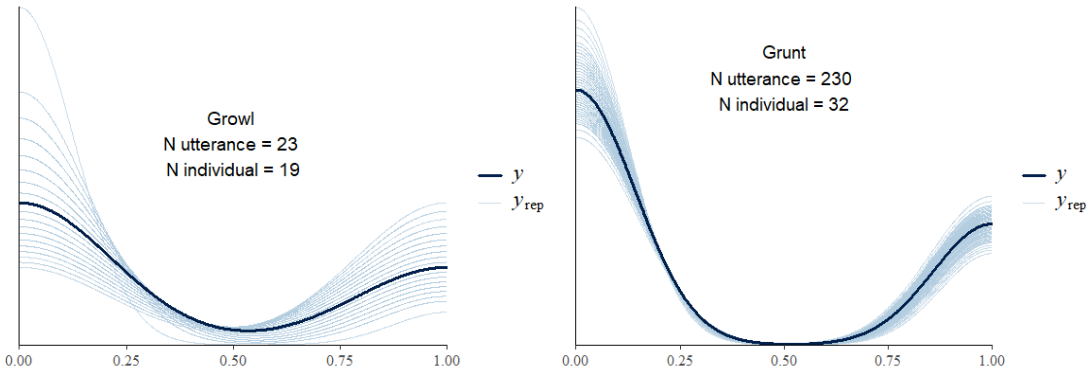
Sequence type	**N sequence**	**N individual**	**%**
shrill_hoo	49	4	53.8
shrill_hoo_shrill_hoo	16	4	17.6
shrill_hoo_growl	4	1	4.4
growl_scream	2	2	2.2
growl_scream_growl	2	2	2.2
growl_shrill_hoo_growl	2	2	2.2
grunt_growl	2	1	2.2

growl_grumble_growl_grunt	1	1	1.1
growl_scream_growl_scream_growl	1	1	1.1
growl_scream_hoo_growl	1	1	1.1
growl_shrill_hoo	1	1	1.1
grunt_growl_grumble_growl_grumble	1	1	1.1
grunt_growl_grunt_growl	1	1	1.1
scream_hoo_grumble_growl_grumble_growl	1	1	1.1
shrill_hoo_growl_shrill_hoo	1	1	1.1
shrill_hoo_shrill_hoo_growl	1	1	1.1
shrill_hoo_shrill_hoo_shrill_hoo	1	1	1.1
shrill_hoo_shrill_hoo_wau_growl	1	1	1.1
shrill_hoo_wau	1	1	1.1
shrill_hoo_wau_growl	1	1	1.1
shrill_hoo_wau_shrill_hoo	1	1	1.1

Table S5. Result models of positional bias analysis

Parameters	**Estimate**	**Error**	**Lower 95%CI**	**Upper 95%CI**	**Bulk ESS**	**Tail ESS**
Model - Growl						

Intercept	-0.38	0.44	-1.27	0.45	6426	4171
group	-0.8	0.83	-2.53	0.78	4711	4004
Model - Grunt						
sd(Intercept) - identity	0.98	0.34	0.44	1.73	3703	4269
Intercept	-0.66	0.31	-1.34	-0.07	3703	4269
group	-1.25	0.65	-2.55	-0.01	5039	5108
Model - Scream						
Intercept	0.25	0.52	-0.77	1.28	5973	5128
group	0.68	0.9	-1.01	2.56	5750	4681
Model - Twitter						
sd(Intercept) - identity	0.95	0.33	0.42	1.72	2472	3140
Intercept	0.56	0.31	-0.03	1.2	3753	5105
group	1.33	0.66	0.11	2.7	5554	5341



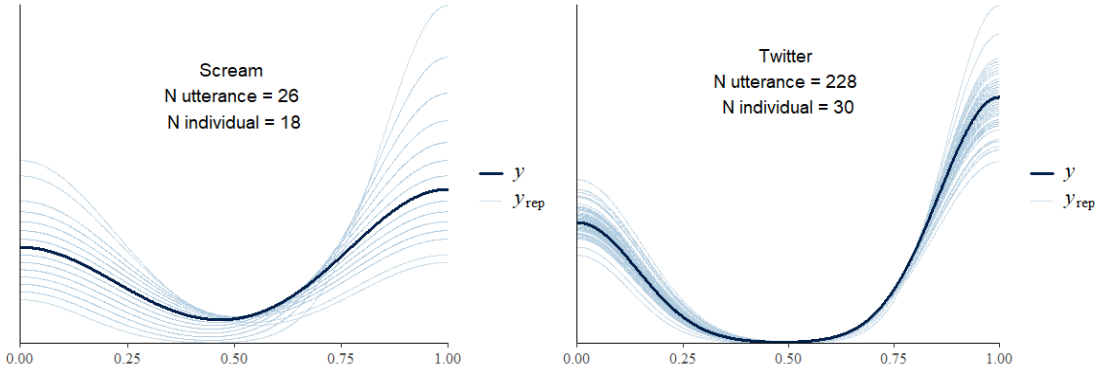


Figure S2. Posterior predictive of models for positional bias analysis

Table S6. Model summaries of call iteration analysis (grunt and twitter sequences)

Parameters	**Estimate**	**Error**	**Lower 95%CI**	**Upper 95%CI**	**Bulk ESS**	**Tail ESS**
Model – All iterated sequences						
sd(Intercept) identity	2.96	1.51	0.89	6.73	2213	3194
Intercept	-4.22	1.73	-8.48	-1.78	2811	2431
First call ‘grunt’/twitter’)	3.87	1.68	1.42	7.91	3645	2705
group	-0.28	2.48	-5.68	4.18	3688	3287
Model - Iterated sequences longer than 3 calls						
sd(Intercept) - identity	3.39	2.12	0.62	8.50	1991	2906
Intercept	-2.39	1.69	-6.51	0.28	3011	2519

First call ‘grunt’/twitter’)	0.98	1.71	-1.92	5.01	3800	2960
group	1.62	3.86	-5.62	10.55	3242	2821

Posterior predictive

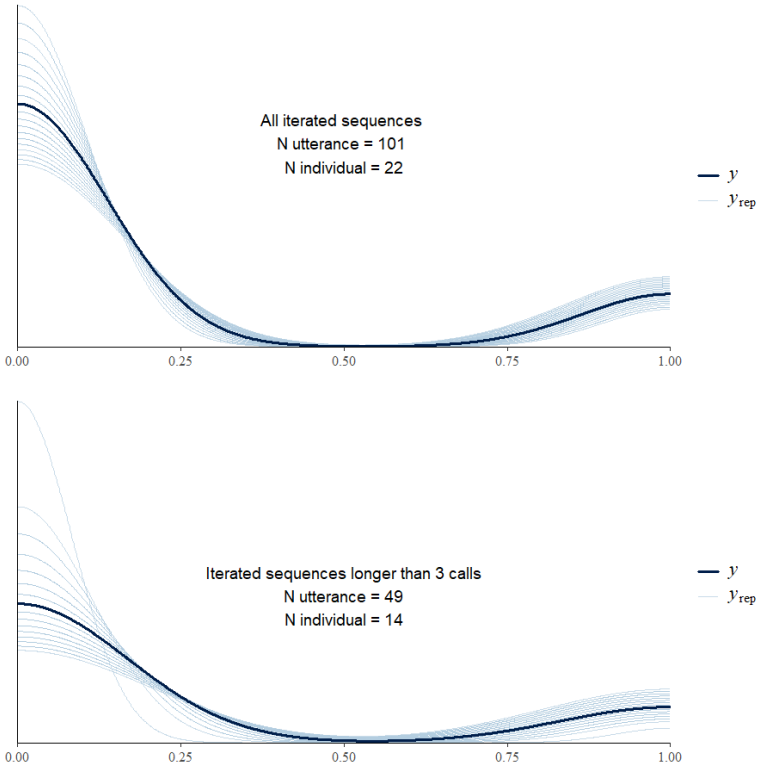


Figure S3. Posterior predictive of the call iteration models

Table S7. Result models of element repetition analysis (grunt and twitter sequences)

Parameters	**Estimate**	**Est.Error**	**l-95% CI**	**u-95% CI**	**Bulk ESS**	**Tail ESS**
Model - Grunt elements						
Caller’sid	0.48	0.11	0.30	0.71	2185	4317

Intercept	1.03	0.15	0.73	1.32	2353	4189
Middle position	0.18	0.16	-0.13	0.49	6452	6244
Last position	0.88	0.11	0.67	1.10	9821	9298
First call (‘grunt’/‘twitter’)	0.44	0.11	0.21	0.66	7630	6158
Sequence LIngth	-0.04	0.01	-0.07	-0.01	5949	5954
Group	0.15	0.24	-0.31	0.63	2423	6311
Model - Twitter elements						
Caller’s id	0.24	0.07	0.11	0.40	2530	3450
Intercept	1.85	0.09	1.67	2.03	4124	5153
Middle position	-0.71	0.16	-1.02	-0.40	7364	6210
Last position	-0.72	0.11	-0.95	-0.50	9296	6656
First call (‘grunt’/‘twitter’)	0.27	0.09	0.10	0.45	7274	5655
Sequence length	-0.04	0.01	-0.06	-0.02	6698	6225
Group	-0.18	0.15	-0.47	0.12	4050	5009

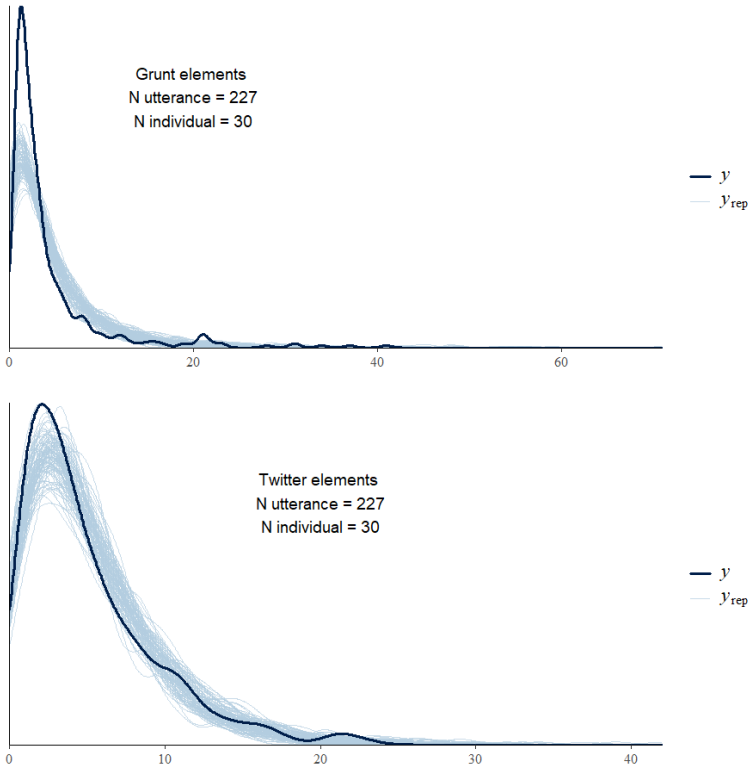


Figure S4. Posterior predictive of models for element repetition analysis

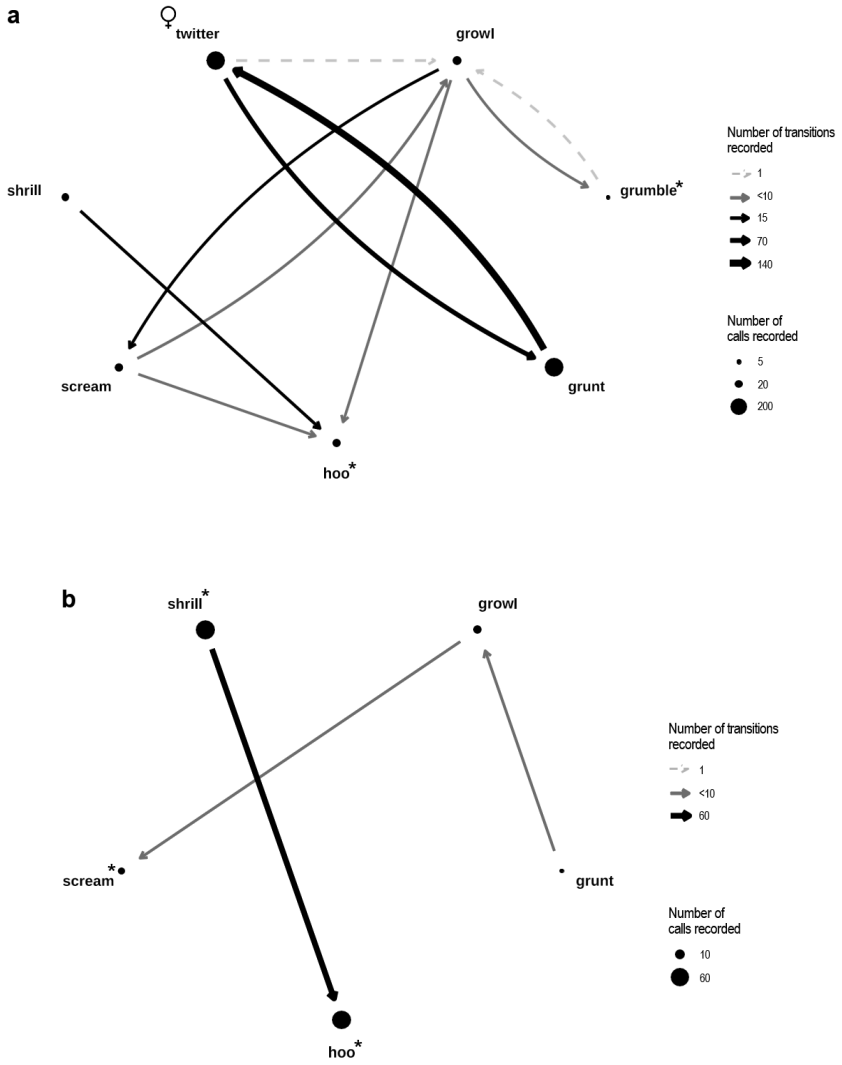


Figure S5. 2-call type sequences without iterations of adult sooty mangabey (a) females and (b) males

These networks represent all recorded call types and transitions between the different call types in sequences without iterations. * = call type produced exclusively in combination with another call type (i.e., never or idiosyncratic -maximum one occurrence- singly). ♀ = call type produced only by females. Node (circle) size is proportional to the occurrence of call for each call type and edge: (line) thickness and colour are proportional to the number of combinations.

Chapter 2: Call combination order and iterations may shift meaning in sooty mangabey vocal sequences

This chapter is based on a manuscript currently under review for publication as: Le Floch, A., Girard-Buttoz, C., Neumann, C., Azaiez, T. S., Bande, N., Wittig, R. M., Zuberbühler, K., & Crockford, C. (2025). Call combination order and iterations may shift meaning in sooty mangabey vocal sequences. BMC Biology (submitted). *The PhD candidate (Le Floch, A.) was responsible for study design, data collection, data coding, analysis and manuscript preparation.*

Disclosure: A subset of the data presented in this chapter were collected and used by Tanit Souha Azaiez for her master thesis, deposed at the University of Neuchâtel. All data reused for this thesis were analysed different

Abstract

Human language expands meaning through the structured combination of sounds, but such mechanisms remain rare in nonhuman animals, raising questions about their evolution. Shifts in meaning from single calls to combinations appear across a range of combinations and contexts in apes, but current evidence from other species is mainly restricted to alarm contexts. To address this, we applied a quantitative, whole-repertoire approach to assess meaningful combinatorial capacities in sooty mangabeys (*Cercocebus atys*), a forest-dwelling monkey. We recorded 1,751 vocal utterances from two groups in the Taï National Park, Ivory Coast. Using context of production as a proxy for potential meaning, we focused on two mechanisms: (1) bigrams—sequences of two different calls, and (2) iteration—reoccurrence of the same call type interspersed with others. Euclidean distance analyses revealed that, unlike its component single calls 'grunts' and 'twitter' or the bigram 'grunt_twitter' frequently used during feeding, 'twitter_grunt' occurred predominantly during infant-directed affiliations, suggesting order-sensitive meaning shifts. Iterative sequences of 'grunt' and 'twitter' calls also showed context-specific shifts, indicating a possible role for iteration in meaning generation. These findings suggest that meaningful combinatorial capacities extend beyond great apes and alarm contexts. However, repertoire-wide use of meaning-shifting bigrams remains unconfirmed outside hominids.

Introduction

Syntax, a key feature of language, allows humans to combine sounds into words and structured phrases, creating a potentially infinite range of meanings. The origins of this ability, unique to humans, remain difficult to reconstruct (Darwin, 1981; Fitch, 2010). Investigating whether rudimentary forms of syntax exist in non-human animals (hereafter animals) can provide insights into the evolutionary roots of our combinatorial communicative system (Bergman et al., 2019; Hauser, 2002; Moran & Bickel, 2020). Over the past four decades, research has successfully identified the meanings of many single calls across taxa such as primates and birds, as well as in mongooses (Arnold & Zuberbühler, 2006; Cunningham & Magrath, 2017; Fichtel & Kappeler, 2002; Grieves et al., 2014; Manser, 2001; Seyfarth et al., 1980; Struhsaker, 1967; Suzuki, 2012; Zuberbühler, 2001; Zuberbühler et al., 1997). Building on this foundation, further research has shown that vocal sequences can also encode semantic content (Berthet et al., 2023; Engesser & Townsend, 2019), likely reflecting evolutionary pressures to convey ecologically and socially relevant information (Bouchet et al., 2013; Freeberg et al., 2012; McComb & Semple, 2005; Stephan & Zuberbühler, 2008). Studies showed that call combinations are commonly used across a range of vertebrate species to signal the presence of predators (Berthet et al., 2019; Clarke et al., 2006; Engesser et al., 2016; Freeberg, 2008; Gallot et al., 2024; Ouattara et al., 2009a; Rauber et al., 2020; Salis et al., 2022). While much research has focused on alarm call combinations, sequences produced during other social contexts remain less explored. For instance, some sequences have been suggested to encode caller identity alongside other meanings such as external objects (e.g., food), greeting or travel calls that improve coordination during group movement in apes (Bortolato et al., 2023; Girard-Buttoz, Bortolato, et al., 2022; Leroux et al., 2021, 2022), monkeys (Arnold & Zuberbühler, 2012; Candiotti et al., 2012; Coye et al., 2018), meerkats (Jansen et al., 2012) and elephants (Hedwig & Kohlberg, 2024).

Animal vocal sequences can convey meaning through different mechanisms. A key distinction is whether a sequence is compositional or non-compositional (Townsend et al., 2018). In compositional sequences, the meaning of the sequence is derived from the meaning of at least one of the individual calls, such that the combination modifies or specifies the information conveyed compared to each call in isolation. Compositional sequences have been documented in bonobos (Berthet et al., 2025), chimpanzees (Girard-Buttoz et al., 2025; Leroux et al., 2023), Campbell's monkeys (Coye et al.,

2015; Ouattara et al., 2009b; Zuberbühler, 2002), Japanese tits (Suzuki et al., 2016), and pied babblers (Engesser et al., 2016). In contrast, non-compositional sequences carry a meaning distinct from their composing units, as suggested in putty-nosed monkeys and chimpanzees (Arnold & Zuberbühler, 2012; Girard-Buttoz et al., 2025; see Schlenker, Chemla, Arnold, et al., 2016 for a formal analysis of the linguistic phenomena underlying these examples). In addition to compositionality, rules can also influence meaning. The order in which vocal units are produced plays a role, as sequences such as AB and BA can have different meanings, as documented in Japanese tits (Suzuki et al., 2017), olive colobus monkeys (Gallot et al., 2024), or chimpanzees (Girard-Buttoz et al., 2025). Similarly, adjacent repetitions of specific unit types within a sequence can alter its meaning, with AB conveying different meaning than AAB, a phenomenon observed in diverse species such as Humayun's night frogs (Bhat et al., 2022), bonobos (Clay & Zuberbühler, 2011), Carolina chickadees (Freeberg, 2008), great tits (Salis et al., 2022), pied babblers (Engesser et al., 2017), Thomas langurs (Wich et al., 2003), and indris (Torti et al., 2013). Iteration, or the non-adjacent reoccurrence of specific unit types within a sequence, can also modify meaning, as in bonobos, where AB differs from ABA (Clay & Zuberbühler, 2011). Additionally, the overall proportion of a particular unit type within a sequence may contribute to meaning, as suggested by findings in Amboli bush frogs and titi monkeys (Berthet et al., 2019; Bhat et al., 2022). These rules are not mutually exclusive.

However, identifying such mechanisms has traditionally been time-consuming, as this often relies on playback experiments (Berthet et al., 2023; Zuberbühler & Wittig, 2011), which have limited the number of documented sequences per species. While playback experiments are critical for assessing whether utterances are meaningful to receivers, they typically test one combination at a time and therefore offer limited insight into the broader diversity of call combinations and the combinatorial mechanisms used by a given species. Such a global assessment is, however, crucial to understand the evolutionary processes that may have led to expansive meaning generation within a single repertoire, as is evident in human language. A promising way to augment experimental approaches is to use observational data to analyse all the vocal sequences produced by a given species across their entire vocal repertoire along with their associated contexts of production. To date, only a few wild species have been investigated in this way, primarily apes (Berthet et al., 2025; Girard-Buttoz et al., 2025; Hedwig et al., 2015), but recent work has also included meerkats and sperm whales (Collier et al., 2017; Sharma et al., 2024). Chimpanzees and bonobos frequently

produce bigrams (i.e., sequences composed of two different calls) across their vocal repertoires and that some of them appear to be compositional (Berthet et al., 2025; Girard-Buttoz et al., 2025; Girard-Buttoz, Zaccarella, et al., 2022; Leroux et al., 2023). In chimpanzees, call order within a bigram likely affect meaning; for example, the calls 'HO' and 'GR' are combined to form the bigram 'HO_GR', which occurs in different contexts than the reversed bigram 'GR_HO'. Conducting such an evaluation across the full repertoire of a non-ape primate species is essential to assess whether the capacity to produce a wide range of potentially meaningful sequences within a vocal system is evolutionarily ancient or a more recent development in the hominid lineage, and to compare it with more distantly related taxa to investigate possible convergent evolution.

Therefore, in our study, we investigated vocal sequence production across the full vocal repertoire of sooty mangabeys, a forest-dwelling monkey species known to produce call sequences (Le Floch et al., 2025; Sigmundson et al., 2025), acknowledging that constraints in behavioural data collection may prevent the recording of all possible utterances. To do so, we used the context of production as a proxy for potential meaning (Bouchard & Zuberbühler, 2022; Crockford et al., 2018; Gill & Sealy, 2004; Kirchhof & Hammerschmidt, 2006; Nellissen et al., 2024; Seyfarth et al., 1980). Context referred to the set of features characterising the circumstances surrounding the production of each utterance (Berthet et al., 2023), also called "events" in other studies (Girard-Buttoz et al., 2025). In our study and following the framework proposed by Berthet et al. 2023, contextual features include not only what is commonly referred to as context such as the type of activity (e.g., feeding) or the social interaction (e.g., affiliation) but also the attributes of the interaction partner (i.e., age class and/or sex). This was important as sooty mangabeys often vocalise during social interactions, and the usage of specific call types appears to vary depending on partner characteristics (Grampp et al., 2023; Range & Fischer, 2004). In particular, 'twitter' calls have been suggested to be produced by females interacting with infants (Range & Fischer, 2004). In a number of monkey species, adult females show intense interest towards infants, often seeking to handle or groom them (Dunayer & Berman, 2018), and specific vocalisations can facilitate such interactions (Bauers, 1993; Silk et al., 2016). Notably, rhesus macaques produce two distinct call types in this context, including one that has been suggested to be infant-directed (Whitham et al., 2007). As sooty mangabey females frequently interact with the infants of others (Fruteau et al., 2011), we aimed to disentangle whether, during interactions involving vocal production,

behaviours were directed towards the mother and/or towards her infant. This distinction allowed us to assess whether the production of specific calls and sequences was associated with characteristics of the actual recipient of the caller's social behaviours.

Adult sooty mangabeys can combine most of their call types in at least one combination (eight out of ten call types, with only 'vibratos' and 'whoop-gobbles' never recorded as part of a sequence) but they frequently rely on a limited subset of combinations, typically sequences composed of 'grunt' and 'twitter' calls or 'shrill' and 'hoo' calls (Le Floch et al., 2025). These combinations can include iterations, for example, a call type recurring after another, as in A_B_A, which may convey a different meaning than sequences without such call reoccurrences (e.g., A_B). In particular, sequences composed of 'grunt' and 'twitter' calls are only produced by females either as simple bigrams (e.g., 'grunt_twitter', 'twitter_grunt') or as longer sequences with iterations (e.g., 'grunt_twitter_grunt', 'twitter_grunt_twitter_grunt'), which raises the question of whether both order and/or iteration may influence meaning. See the glossary for our definitions of call, bigram, sequence and iteration.

We therefore aimed to determine whether sooty mangabey call sequences are compositional, that is, whether the meaning can be derived from the calls produced individually, but also whether ordering and iterations are associated with the potential meaning of the sequence. To investigate this, we focused on:

(1) bigrams (i.e., sequences composed of two distinct call types) following an analytical approach recently applied to a full repertoire analysis of chimpanzee vocal sequences (Girard-Buttoz et al., 2025). This approach allowed us to describe to what extent bigrams exhibited compositional or non-compositional structures, and whether the associated contextual features composing the context of production were associated with the order of calls.

(2) two-call-type sequences with iterations, to examine whether they differed in potential meaning from their component calls and bigrams, and whether the order of calls within these iterative sequences also influenced potential meaning (e.g., whether A_B_A differs from A, B, A_B, and/or B_A_B).

While adjacent repetitions (e.g., A_A_A) can alter sequence meaning (see above), our analysis focused on the type and order of calls within sequences. In this context, adjacent repetitions such as A_A_A are considered as three vocal elements forming one instance of the call type A (see glossary for definitions). This decision was

motivated by the considerable variation in repetition length across call types (Range & Fischer, 2004), which would have required distinct modelling strategies beyond the scope of the present study.

Methods

Data collection

For this study, ALF, NB and TSA collected acoustic and behavioural data on two groups of habituated sooty mangabeys: the Taï Monkey Project (TMP) group (May to July 2022) and the Taï Chimpanzee Project (TCP) group (February to October 2023) in the Taï National Park, Ivory Coast (McGraw et al., 2007; Wittig, 2018). All authors involved in data collection also passed inter-observer and intra-observer reliability tests for sooty mangabey identity recognition prior to beginning data collection. We employed half-day continuous focal animal sampling (Altmann, 1974) on adult individuals, conducting observations from dawn to midday and from midday to dusk. During these sessions, we recorded focal individual vocalisations, as well as calls from known adult individuals within visual range ad libitum, provided we could clearly determine caller identities. Based on prior familiarity with the vocal repertoire and training with experienced field assistants, we were able to make initial in-field identifications of most call types. To facilitate later identification and coding of focal individual calls in the recordings, we provided spoken notes to verbally indicate the presumed call types. Formal inter-observer reliability tests were conducted subsequently to confirm consistent classification (see section “audio recording coding” below). When other individuals produced calls, we verbally indicated it as well to avoid mistakenly coding non-focal calls later. Recordings were made using directional microphones (Sennheiser models K6, MK 316, MKH 416 P48, and MKH 416T-F) and Marantz Solid State recorders (PMD661 and PMD661 MKII), digitised at a 44.1 kHz sampling rate and 16-bit depth in .wav format.

For each recorded vocalisation, we noted the features of the context of production (i.e., the contextual features, defined as the set of features surrounding the vocal production) using a predefined ethogram (see supplementary Table S1), based on previously published work on sooty mangabey behaviours and contexts eliciting vocalisations (Grampp et al., 2023; León et al., 2022, 2023; Mielke et al., 2019;

Neumann & Zuberbühler, 2016; Quintero et al., 2022b, 2022a; Range & Fischer, 2004; Range & Noë, 2002), and recorded using Cybertracker on a smartphone.

Contextual features

We assessed 37 contextual features associated with vocal production, as detailed in Table S2. These included social interactions, defined as beginning when one individual initiated gaze or approached another, and ending when they moved apart or ceased social behaviours and gazing. When a vocal utterance occurred during such an interaction, we assigned two contextual features: (1) the type and direction of the social interaction involving the caller (e.g., give approach, give affiliation, receive aggression) and (2) the characteristics of the social partner, such as an adult male, subadult female, or mother carrying an infant (see Table S2 for classification details).

For social interactions involving a mother carrying an infant, we recorded the distance between the mother and the infant, categorising it as “carried on the belly”, “close” (less than one metre), or “far” (more than one metre). This distinction was important because female sooty mangabeys frequently vocalise while in social contexts with infants of other females (Range & Fischer, 2004). When the infant was more than one metre away from the mother, we treated the infant as an independent social partner (IN). When the infant was carried or close to the mother, we treated the mother and the infant as a unit (MI) unless the behaviour received was affiliative, in which case we could identify whether the interaction was specifically directed at the mother (XM), the infant (XI), or both (XX). This level of detail could only be applied to affiliative interactions as the receiver could be reliably determined. For faster-paced interactions, specifically during aggression or approaches, the specific target (i.e., mother, infant, or both) could not always be reliably determined and hence receiver was not assigned for these interactions. We also created a specific contextual feature for instances in which the caller directed gaze exclusively at the infant (IZ), without any accompanying behaviour. This applied regardless of whether the infant was close (< 1 metre) or far (> 1 metre) away from the mother.

When vocalisations occurred outside of social interactions, we assessed whether they were produced during environmental changes, such as encounters with other animal species, intergroup encounters with conspecifics, or after hearing alarm calls from

other monkey species, or during other specific contexts that could elicit vocal production, such as the arrival of an adult male in the vicinity.

If none of these above features applied, we assigned the activity of the caller as the contextual feature. As sooty mangabeys typically paused their activity to vocalise, making it difficult to assess what they were doing at the exact moment, we defined the activity based on the activity the caller was engaged in immediately before and after vocal production. If the caller engaged in the same activity before and after vocalisation, we categorised it as a stable activity (e.g., feeding, resting, moving). If the activity changed after vocalisation, we categorised it as a transitional activity (e.g., feeding, vocalising, then resting).

We conducted four to five separate inter-observer reliability tests with each data collector on focal behavioural data collection with Kappa scores ranging from 0.72 to 1 to assess the consistency of contextual feature classification (see Table S3 for details).

In total, our continuous focal observations on adult sooty mangabeys resulted in the following data: TCP females: $N = 17$ individuals, $N = 232$ h 27 min (mean per individual = 13 h 40 min, SE = 0 h 56 min); TMP females: $N = 18$ individuals, $N = 262$ h 45 min (mean = 14 h 35 min, SE = 3 h 28 min); TCP males: $N = 8$ individuals, $N = 122$ h 21 min (mean = 15 h 17 min, SE = 1 h 39 min); TMP males: $N = 4$ individuals, $N = 10$ h 30 min (mean = 2 h 37 min, SE = 0 h 24 min). In addition, we recorded vocal utterances from ad libitum individuals (i.e., individuals who were not focal subjects): $N = 4$ TCP females, $N = 2$ TMP females, and $N = 5$ TMP males.

The research was conducted under the Taï Chimpanzee Project research permit (Wittig/006/MESRS/DRI), issued by the Ivorian Ministry of Higher Education and Scientific Research and the Ivorian Office of Parks and Reserves, which granted access to the Taï National Park. This study employed purely observational methods, involving audio recordings, and was approved by the Ethikrat der Max Planck Gesellschaft. All research adhered to the ethical guidelines of the Association for the Study of Animal Behaviour and the International Union for Conservation of Nature.

Glossary

Following previous work on sooty mangabey sequences (Le Floch et al., 2025; Sigmundson et al., 2025), we used the following definitions:

Term	**Definition**
Vocal element	Continuous sound with a distinct beginning and end.
Call type	One or a succession of sound elements of the same type within an interval of less than two seconds.
Sequence	Series of different calls produced with intervals of less than one second between them.
Iteration	Repeated occurrence of a call type within a sequence, interspersed with other call types (e.g., ABA).
Bigram	Two different calls combined with an interval of less than one second. Can be a sequence by itself (e.g., AB) or part of a sequence (e.g., AB within ABA).
Vocal utterance	Either a single call type produced in isolation, or a sequence of different call types.

Audio recording coding

ALF, NB, and TSA analysed the audio recordings using Raven Pro (v1.6.5 and 1.6.4) (K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology, 2024). We manually inspected spectrograms (FFT size: 1024, Hann window, hop size: 256 samples) while simultaneously listening to the recordings to identify and annotate individual vocal elements (see glossary), marking their start and end times. We relied on the spoken vocal notes recorded in the field to identify the calls produced by the focal individuals. For classification, we referred to the established sooty mangabey vocal repertoire (Range & Fischer, 2004) and identified ten distinct call types: ‘grunt’,

‘twitter’, ‘growl’, ‘scream’, ‘grumble’, ‘vibrato’ (previously termed ‘copulation call’ in Range & Fischer’s repertoire), ‘whoop-gobble’, ‘wau’, ‘shrill’ (corresponding to the initial high-frequency elements of the alarm call as described in Range & Fischer’s repertoire), and ‘hoo’ (which includes both the ‘hoo’ call described by Range & Fischer and the final low-frequency sound element categorised initially as part of the alarm call). Additional details on each vocal element type, including spectrographic examples and definitions based on Range & Fischer’s repertoire, are provided in the supplementary materials (Document S1).

To assess inter-rater reliability, ALF independently recoded *N* = 320 calls from NB sample (14% of NB sample) and *N* = 116 from TSA sample (21% of TSA sample), randomly selected. Agreement was high (Cohen’s kappa: 0.96 for NB, 0.95 for TSA).

Vocal utterances studied

We separated adult male and female vocal utterances, as they do not share the same vocal repertoire of single calls and therefore do not produce comparable sequences (Range & Fischer, 2004; Sigmundson et al., 2025).

The most frequent sequences produced by females were composed exclusively of ‘grunt’ and ‘twitter’ calls (*N* = 215 out of 257 sequences; 84%), and these also occurred more often with a clearly identifiable context of production than other, less frequent sequences. They typically included at least one bigram—‘grunt_twitter’ (*N* = 73) or ‘twitter_grunt’ (*N* = 43)—or iterative forms—‘grunt_twitterX’ (e.g., ‘grunt_twitter_grunt’ or ‘grunt_twitter_grunt_twitter’) (*N* = 65) and ‘twitter_gruntX’ (*N* = 30) (Tables S5 and S6). Sequences composed of other call types were recorded only infrequently and even less often with a clearly identifiable context of production. Each occurred fewer than ten times (e.g., ‘growl_scream’, *N* = 9; ‘scream_growl’, *N* = 4; ‘shrill_hoo’, *N* = 9) and were excluded from quantitative analyses, though they are reported descriptively.

To analyse sequences composed of ‘grunts’ and ‘twitters’ while accounting for iteration, we grouped sequences longer than two calls into broader categories: ‘grunt_twitterX’ (e.g., ‘grunt_twitter_grunt’, ‘grunt_twitter_grunt_twitter’) and ‘twitter_gruntX’ (e.g., ‘twitter_grunt_twitter_grunt). These categories include non-adjacent repetitions of the same call types, ranging from three to seven calls for ‘grunt_twitterX’ and from three to 29 calls for ‘twitter_gruntX’. This approach allowed us to examine structural features

such as iteration and call order, while avoiding fine-grained categorisation based on the number of call reoccurrences, which varied across individuals (Table S5). The final categories used in female analyses were: ‘grunt_twitter’, ‘grunt_twitterX’, ‘twitter_grunt’, and ‘twitter_gruntX’, along with all calls produced singly: ‘grunt’, ‘twitter’, ‘growl’, ‘scream’, ‘shrill’, and ‘vibrato’.

For males, most utterances recorded for which we could identify the contextual feature of production consisted of single calls, and sequences involving multiple call types were infrequent and typically produced by only one to three individuals (Table S4). The most common sequence was ‘shrill_hoo’ (*N* = 19), recorded from *N* = 3 individuals; no single ‘shrill’ or ‘hoo’ calls were recorded in males, preventing testing the compositional nature of the ‘shrill_hoo’ sequence. All other sequences occurred five times or fewer. Given this distribution, we did not test for contextual differences between single calls, bigrams and longer sequences in males. However, we report sample sizes and associated contexts for the main vocal utterances (i.e., single ‘grunt’ and ‘growl’ calls and ‘shrill_hoo’ bigram) in the Results section.

Analyses

We compared the distribution of contextual features associated with each single call type (‘grunt’, ‘twitter’, ‘growl’, ‘scream’, ‘shrill’, and ‘vibrato’) to those associated with four sequence types: the bigrams ‘grunt_twitter’ and ‘twitter_grunt’, and their iterated forms ‘grunt_twitterX’ and ‘twitter_gruntX’, using the same analytical approach as a recent study on wild chimpanzees (Girard-Buttoz et al., 2025). Vocal utterances were assigned either one contextual feature (e.g., feeding, animal encounter) or two when involving social interactions, where both the type of social interaction and the characteristics of the interaction partner were recorded (e.g., give affiliation + adult female). Refer to the Methods section for details on data collection and contextual feature definitions.

Following Girard-Buttoz et al. 2025 we implemented a multinomial model in Stan via the ‘cmdstanr’ interface in R (Gabry et al., 2023; R Core Team, 2021; Stan Development Team, 2023). The input data consisted of a binary matrix with 1459 rows (utterances) and 37 columns (contextual features), where each cell indicated the presence (1) or absence (0) of each contextual feature during a given utterance. Utterance type (a factor with 10 levels: six single call types and four sequence types,

i.e., ‘twitter_grunt’, ‘grunt_twitter’, ‘twitter_gruntX’, ‘grunt_twitterX’) and caller identity were included as varying intercepts. The model was run with four chains, each with 500 iterations (and 1000 warm-up), yielding 2000 posterior draws.

To compute Euclidean distances between utterance types, we extracted from the multinomial model the posterior distribution of the probability vector of each contextual feature to occur for each utterance type, averaged across individuals (i.e., omitting the individual-level intercept). The estimated probabilities across the 37 contextual features for a given vocal utterance sum to 1. 35% of utterances (517 out of 1,549) occurred in two contextual features (i.e., social interaction + social partner attributes). To get a contextual feature distribution that mirrors this pattern, we multiplied each probability vector by the expected number of contextual features per recording of a given utterance type. For example, suppose utterance A was recorded three times: twice during give affiliation + adult female, and once during give approach + juvenile male. This yields raw contextual feature counts of 2 (affiliation), 2 (adult female), 1 (approach), and 1 (juvenile male), summing to 6 contextual features across 3 utterances, or 2.0 contextual features per occurrence. The corresponding raw probability vector is [2, 2, 1, 1]/6 = [0.33, 0.33, 0.17, 0.17]. This vector is then multiplied by the expected number of contextual features per utterance (2.0), yielding [0.66, 0.66, 0.33, 0.33] as the adjusted contextual feature distribution for utterance A. These adjusted vectors served as the basis for calculating the Euclidean distances between utterance types (see also Girard-Buttoz et al., 2025).

As the number of contextual features could vary depending on the utterance type, we modelled the “contextual feature length” (i.e., the number of contextual features associated with each utterance, one or two) using a zero-truncated Poisson likelihood, since each utterance had at least one associated contextual feature. Varying intercepts were included for utterance type and caller ID. From the posterior estimates of the model, we simulated 1,000 recordings per utterance type (where the contextual feature length was a random number drawn from a truncated Poisson distribution with the estimated mean for the current utterance type) and extracted which contextual feature(s) occurred. For each simulated recording, we computed the contextual feature-adjusted probability vector for the corresponding utterance type following the procedure outlined above. The outcome of these 1,000 simulations was then returned as a vector of means per contextual feature type. We repeated this procedure for all 10 utterance types, yielding a corresponding set of adjusted probability vectors. These

vectors were then used to compute Euclidean distances between each pair of utterance types.

To assess potential meaning shift between the single calls ‘twitters’ and ‘grunts’ and the four types of vocal sequences comprising these calls (‘grunt_twitter’, ‘twitter_grunt’, ‘grunt_twitterX’, ‘twitter_gruntX’), we compiled the posterior distribution of the proportion of association of each contextual feature with each utterance type from the multinomial model described above. We focused on the four contextual features most often associated with these vocal sequences: feed (N = 577 utterances, 36 individuals), give affiliation (N = 257 utterances, 27 individuals), receiver is an infant carried or close to their mother (N = 136 utterances, 25 individuals), and receivers are both infant and their nearby mothers (carried or close; N = 84, 21 individuals). We thus extracted four posterior distributions (one for each contextual feature) for each of the six utterance types (two single call types and four sequence types), resulting in 24 posteriors. To compare whether one utterance type occurred more often than another within the same contextual feature, we calculated the difference in the posterior samples of the proportion of that contextual feature for the two utterances. Specifically, for each posterior draw, we subtracted the posterior proportion of the contextual feature for utterance A from that for utterance B. We then determined the percentage of draws where this difference was positive, indicating how strongly the data supported utterance B being more frequent than utterance A in that contextual feature. This procedure generated a matrix of pairwise comparisons that reflected the differences in the strength of contextual feature association between each pair of utterances (Figure 4).

Results

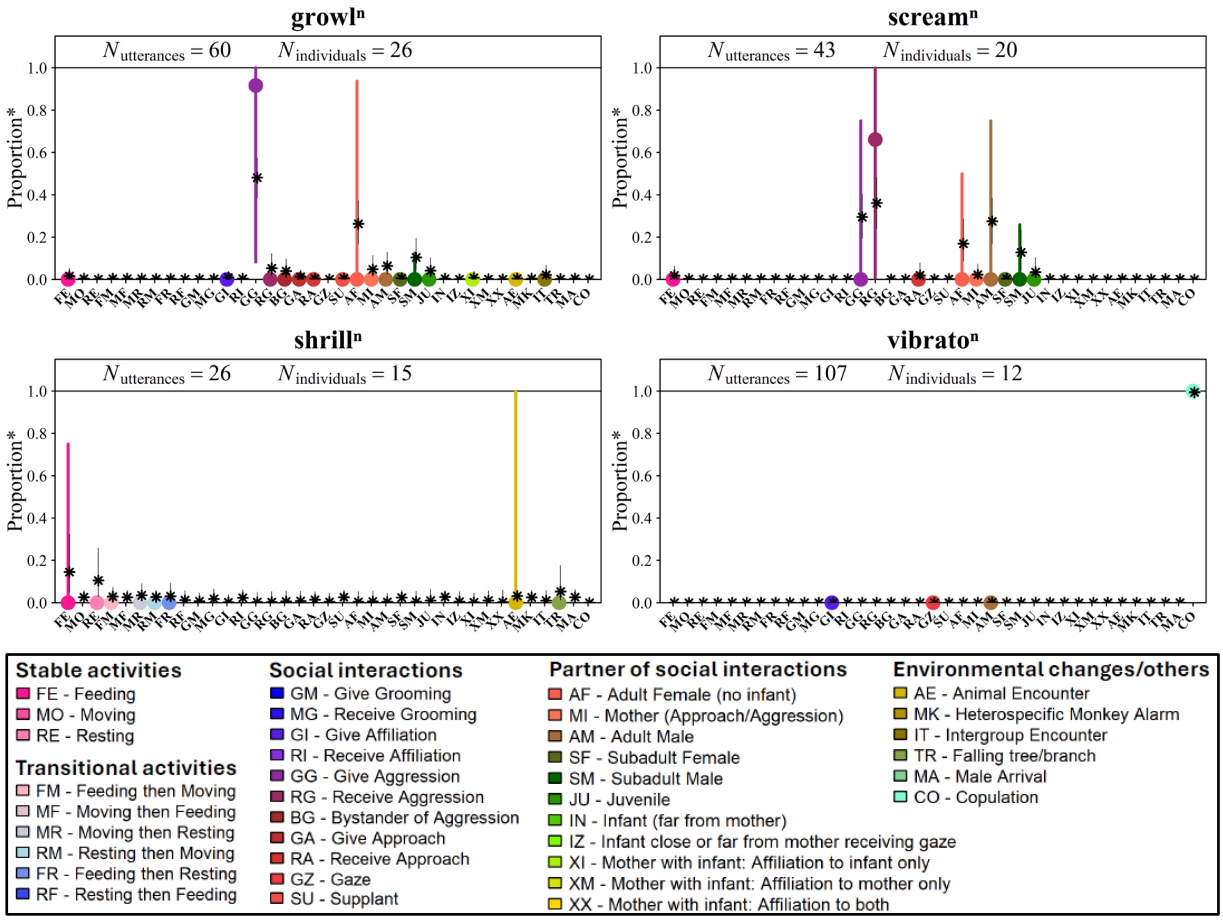
Our dataset comprised N = 2,023 vocal utterances, including N = 1,751 for which we could identify the set of contextual features associated with their production. In contexts where the caller was either giving or receiving a social interaction, we only retained the ones for which the contextual features corresponding to social interaction partner attributes (i.e., sex and age class) could be identified (see Tables S1 and S2 for details).

Females - utterance types recorded

We recorded 1501 utterances produced by females where we could document a clear context of production, including 1,245 single calls, 144 bigrams, and 112 sequences longer than two calls (range: 3–38 calls; sequences comprising more than five calls were rare, occurring fewer than ten times). The most frequent utterance type was ‘grunt’ (55.9%), followed by ‘twitter’ (7.5%), ‘vibrato’ (6.7%), ‘grunt_twitter’ (4.8%), ‘grunt_twitterX’ (4.1%), ‘growl’ (3.8%), ‘scream’ (2.7%), ‘twitter_grunt’ (2.7%), and ‘twitter_gruntX’ (1.9%). All other utterance types occurred fewer than nine times (see Table S6 for full details on sample sizes).

Females - contextual usage of single calls ‘growl’, ‘scream’, ‘shrill’ and ‘vibrato’

Female ‘growls’ were mainly produced when giving aggression (GG) to other adult females (AF) (Figure 1). In contrast, ‘screams’ were typically given when receiving aggression (RG), although they were also occasionally produced when giving it (GG). ‘Shrills’ occurred more frequently during feeding (FE) and when encountering animals other than monkeys (AE), while ‘vibrato’ calls were produced almost exclusively during copulation (CO).


growlⁿ
$N_{utterances}$ = 60
$N_{individuals}$ = 26
screamⁿ
$N_{utterances}$ = 43
$N_{individuals}$ = 20
shrillⁿ
$N_{utterances}$ = 26
$N_{individuals}$ = 15
vibratoⁿ
$N_{utterances}$ = 107
$N_{individuals}$ = 12
Proportion*
Stable activities
FE - Feeding
MO - Moving
RE - Resting
Transitional activities
FM - Feeding then Moving
MF - Moving then Feeding
MR - Moving then Resting
RM - Resting then Moving
FR - Feeding then Resting
RF - Resting then Feeding
Social interactions
GM - Give Grooming
MG - Receive Grooming
GI - Give Affiliation
RI - Receive Affiliation
GG - Give Aggression
RG - Receive Aggression
BG - Bystander of Aggression
GA - Give Approach
RA - Receive Approach
GZ - Gaze
SU - Supplant
Partner of social interactions
AF - Adult Female (no infant)
MI - Mother (Approach/Aggression)
AM - Adult Male
SF - Subadult Female
SM - Subadult Male
JU - Juvenile
IN - Infant (far from mother)
IZ - Infant close or far from mother receiving gaze
XI - Mother with infant: Affiliation to infant only
XM - Mother with infant: Affiliation to mother only
XX - Mother with infant: Affiliation to both
Environmental changes/others
AE - Animal Encounter
MK - Heterospecific Monkey Alarm
IT - Intergroup Encounter
TR - Falling tree/branch
MA - Male Arrival
CO - Copulation

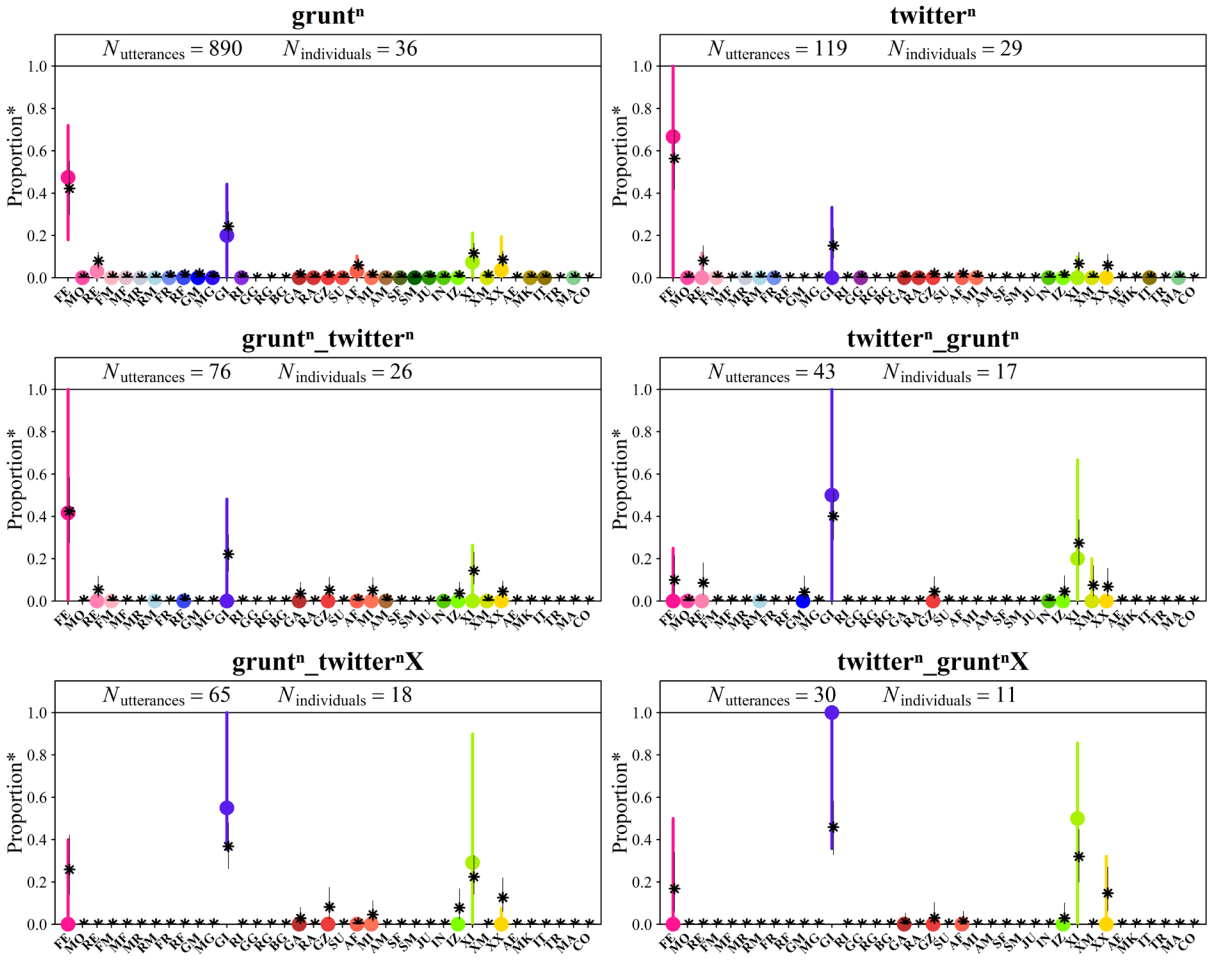

gruntⁿ
$N_{utterances}$ = 890
$N_{individuals}$ = 36
twitterⁿ
$N_{utterances}$ = 119
$N_{individuals}$ = 29
gruntⁿ_twitterⁿ
$N_{utterances}$ = 76
$N_{individuals}$ = 26
twitterⁿ_gruntⁿ
$N_{utterances}$ = 43
$N_{individuals}$ = 17
gruntⁿ_twitterⁿX
$N_{utterances}$ = 65
$N_{individuals}$ = 18
twitterⁿ_gruntⁿX
$N_{utterances}$ = 30
$N_{individuals}$ = 11
Proportion*

Figure 1. Distribution of contextual features of the most frequently recorded female vocal utterances. Coloured dots show the observed median individual proportion of occurrence of each utterance in each contextual feature. Vertical coloured lines indicate the 25th and 75th percentiles. A 25th percentile at proportion *P* means that 75% of individuals produced that utterance in the given contextual feature at a proportion ≥ *P*. Similarly, the median at *P* indicates that half of individuals produced the utterance at a proportion ≥ *P*, and the 75th percentile at *P* indicates that 25% produced it at a proportion ≥ *P*. Black stars and thin black lines show the posterior median and 89% credible interval estimated by our statistical model. Superscript *n* indicates that each call in the vocal utterances may include any number of vocal element repetitions (from zero upwards).

Note: "*Proportion**" refers to values calculated per contextual feature and may sum to more than one across all features because some utterances were assigned to multiple contextual features. For example, when a single utterance involved a social partner, it was counted for both the type of social interaction and recipient attribute (e.g., *give affiliation* + *adult female*). Each individual proportion remains ≤1, but the total across contextual features may exceed one as utterances can contribute to multiple feature types.

Females - contextual usage of 'grunt' and 'twitter' calls produced singly and in sequence

Main contextual features

Female 'grunts' were produced across a variety of contextual features but occurred most frequently during feeding (FE) and when giving affiliations (GI) that could be directed at infants only (XI) or both the mother-infant pair (XX), although in low proportions (<10%; Figure 1 and Table 1). These same contextual features also accounted for most occurrences of single 'twitters' (a call type not produced by adult males), with the addition of resting (RE) (Figure 1). Notably, 'grunt' and 'twitter' were closer to each other in Euclidean space than to any other single calls (Figure 2). When combined in sequences, they were also produced within the same contextual features, feeding (FE) and giving affiliations to infants (XI and XX), but with changing proportions depending on call order and the presence of iterations (Figure 1, Table 1 and see section below).

Table 1. Most frequent contextual features in which 'grunt' and 'twitter' calls were produced both singly and in sequences. Dark grey cells highlight the type of activity and social interaction (either "feeding" or "give affiliation") in which the highest number of utterances was recorded. When "give affiliation" is the most frequent, light grey cells indicate the most frequent type of recipient of affiliation ("to infant only" = affiliative behaviours directed at infant carried by or closed to their mother, "to infant + mother pair" = affiliative behaviours directed at both an infant and their mother). Numbers on top within each cell represent the observed median individual proportion of each utterance type in a given contextual feature, with the 25% and 75% quartiles in brackets. Values in italics below indicate the median proportion from the model's posterior distribution, with the 89% credible interval in brackets.

Utterance type	Feeding	Give affiliation	To infant only	To infant + mother pair
grunt	0.47 [0.18–0.72] *0.42 (0.30-0.55)*	0.20 [0.0–0.44] *0.24 (0.18-0.31)*	0.7 [0.0–0.21] *0.12 (0.8-0.16)*	0.4 [0.0–0.20] *0.8 (0.5-0.13)*
twitter	0.67 [0.0–1] *0.56 (0.42-0.70)*	0.0 [0.0–0.33] *0.15 (0.9-0.23)*	0.0 [0.0-0.10] *0.7 (0.3-0.12)*	0.0 [0.0-0.0] *0.6 (0.3-0.11)*
grunt_twitter	0.42 [0.0–1] *0.42 (0.28-0.58)*	0.0 [0.0–0.48] *0.22 (0.14-0.31)*	0.0 [0.0-0.26] *0.14 (0.8-0.23)*	0.0 [0.0-0.0] *0.5 (0.2-0.10)*
twitter_grunt	0.0 [0.0–0.25] *0.10 (0.4-0.22)*	0.50 [0–1] *0.40 (0.29-0.52)*	0.20 [0.0–0.67] *0.27 (0.17-0.39)*	0.0 [0.0-0.20] *0.7 (0.2-0.16)*
grunt_twitterX	0.0 [0.0–0.40]	0.55 [0.35–1]	0.29 [0.0–0.90]	0.0 [0.0-0.8] *0.13 (0.6-0.22)*

	0.26 (0.14-0.42)	*0.37 (0.26-0.48)*	*0.22 (0.14-0.33)*	
twitter_gruntX	0.0 [0.0–0.50] *0.17 (0.60-0.34)*	1 [0.36–1] *0.46 (0.33-0.58)*	0.50 [0.0–0.96] *0.32 (0.20-0.45)*	0.0 [0.0–0.32] *0.15 (0.6-0.27)*

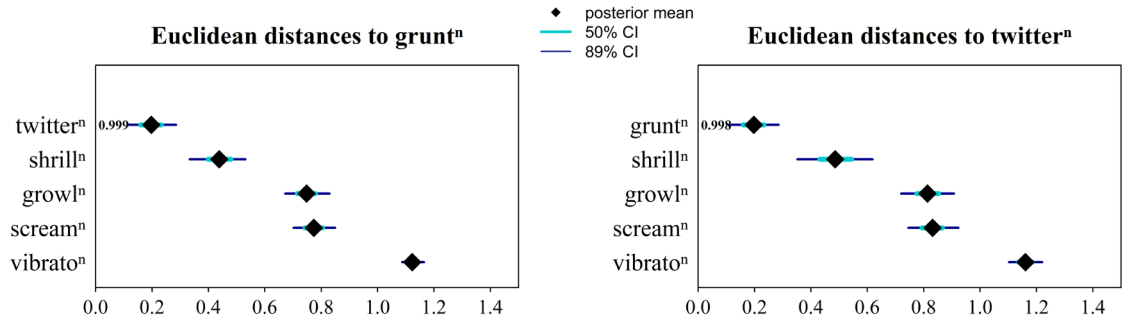


Figure 2. Euclidean distances used to measure differences in contextual features distribution between 'grunt' and 'twitter' to other call types produced in isolation. For each call, the mean distance and the 50% and 89% credible intervals relative to each comparison call are shown. Values on the Y-axis indicate the proportion of posterior samples in which the corresponding call was identified as the closest match to the focal call; only values greater than 0.01 are displayed. Superscript *n* indicates that each call can have any number of vocal element repetitions (from zero upwards).

Ordering effect in bigrams

Among bigrams, 'grunt_twitter' followed a similar distribution to its component calls, being used primarily during feeding (FE) and less so when giving affiliation (GI) (Figure 1 and Table 1). This was also supported by Euclidean distances indicating that the bigram 'grunt_twitter' was close in terms of context distribution to its component calls (Figure 3 and Table 2). In contrast, 'twitter_grunt' was clearly more frequent when giving affiliation (GI), especially when directed towards infants (XI and XX), and appeared less often uttered while feeding (FE) (Figure 1 and Table 1). Here, Euclidean distances showed less similarity to the contextual distribution of the component calls (Figure 3 and Table 2). This was further supported by the pairwise comparisons of the

posterior distributions of the proportions of vocal utterances during feeding, affiliative, and infant-directed contexts between vocal sequences and their constituent calls. We found 100% posterior support for the proportion of feeding to be higher when uttering single ‘twitters’ and single ‘grunts’ than when uttering the bigram ‘twiter_grunt’. Likewise, we found 99-100% posterior support for ‘twitter_grunt’ to be more often uttered during affiliation and directed at infants than single ‘grunts’ and single ‘twitters’ (Figure 4).

Table 2. **Euclidean distances between each target sequence and closest utterances, based on differences in contextual feature distribution**. The dominant contextual feature (i.e., most frequent contextual feature of production) is indicated for both target sequences and closest utterances. p.support indicates the proportion of posterior samples in which a given utterance was identified as the closest to the target sequence in terms of Euclidean distance. Only utterances with a p.support value greater than 0.01 are shown. Distances were calculated using all main utterance types (‘grunt’, ‘grunt_twitter’, ‘grunt_twitterX’, ‘twitter’, ‘twitter_grunt’, ‘twitter_gruntX’, ‘shrill’, ‘growl’, ‘scream’, ‘vibrato’).

Target sequence	**Target sequence dominant contextual feature**	**Closest utterances**	**Distance**	**p. support**	**Closest utterances dominant contextual feature**
grunt_twitter	Feeding	grunt	0.16	0.80	Feeding
		twitter	0.22	0.16	Feeding
		grunt_twitterX	0.30	0.04	Affiliation with infant
twitter_grunt	Affiliation with infant	twitter_gruntX	0.27	0.63	Affiliation with infant
		grunt_twitterX	0.30	0.34	Affiliation with infant

grunt_twitterX	Affiliation with infant	twitter_gruntX	0.26	0.48	Affiliation with infant
		twitter_grunt	0.30	0.17	Affiliation with infant
		grunt_twitter	0.30	0.24	Feeding
		grunt	0.31	0.10	Feeding
twitter_gruntX	Affiliation with infant	grunt_twitterX	0.26	0.56	Affiliation with infant
		twitter_grunt	0.27	0.43	Affiliation with infant

Comparing iterative vocal sequences to bigrams

Iterative sequences (‘grunt_twitterX’ and ‘twitter_gruntX’), regardless of whether they began with ‘grunt’ or ‘twitter’, showed greater contextual similarity to ‘twitter_grunt’ bigrams than to ‘grunt_twitter’ bigrams. Specifically, these sequences occurred more frequently during affiliative interactions (GI) with infants (XI and XX) than during feeding (FE) (Figure 1 & Table 1). Euclidean distance analyses showed that ‘twitter_gruntX’ was produced in very similar context to both ‘twitter_grunt’ and ‘grunt_twitterX’, yet remained more distant from its component calls (i.e., more dissimilar in terms of contextual usage). In contrast, ‘grunt_twitterX’ displayed a more intermediate pattern: it was contextually close not only to ‘twitter_gruntX’, but also to ‘twitter_grunt’, ‘grunt_twitter’, and ‘grunt’ (Figure 3 & Table 2). Pairwise comparisons of the posterior distribution of the proportion of utterance types by contexts (Figure 4) revealed:

97% and 81% posterior support for ‘grunt_twitterX’ to be more likely to occur during feeding than ‘twitter_grunt' and ‘twitter_gruntX’, respectively.

‘grunt_twitterX’ was less likely to occur when giving affiliations directed at infants than ‘twitter_grunt’ and ‘twitter_gruntX’, but these differences were supported by less than 85% of the pairwise posterior distribution comparison (Figure 4).

Thus, ‘grunt_twitterX’ iterative sequences were more similar to ‘twitter_grunt’ bigrams and ‘twitter_gruntX’, but were generally preferred during feeding and less so during infant-directed affiliations by comparison.

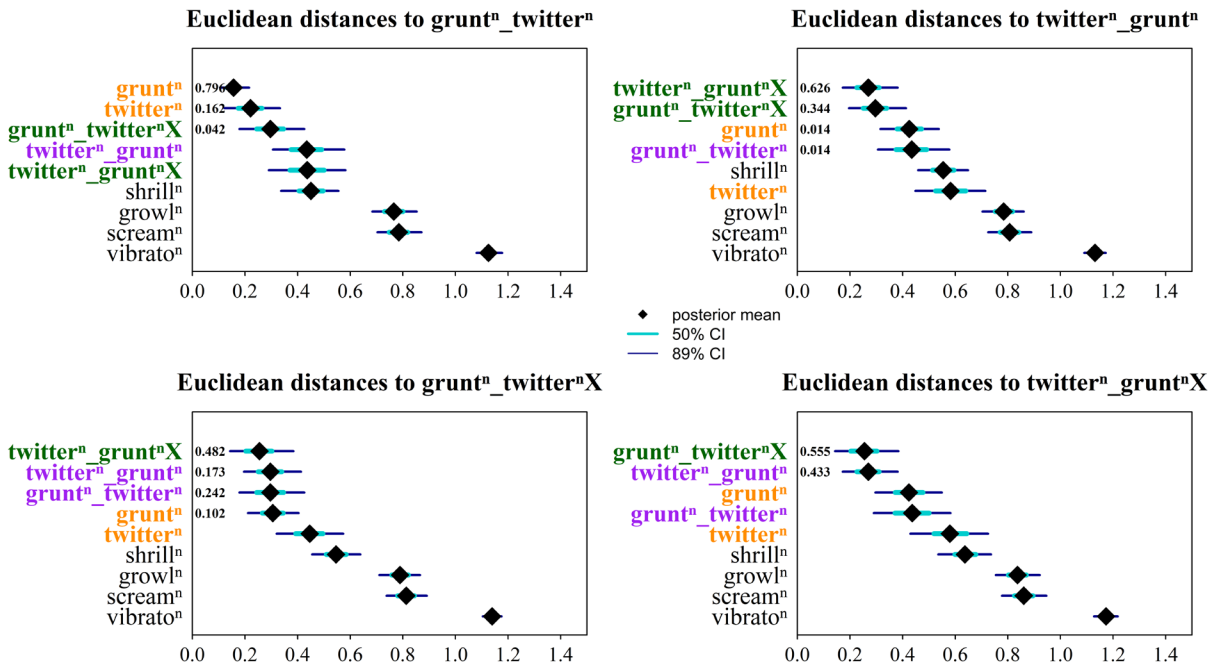


Figure 3. Euclidean distances used to measure differences in contextual feature distribution between each sequence and all other most common vocal utterances. For each sequence, the mean distance and the 50% and 89% credible intervals relative to each comparison utterance are shown. Component single calls are displayed in orange, component bigrams in purple and iterative sequences in green. Values on the Y-axis indicate the proportion of posterior samples in which the corresponding utterance was identified as the closest match to the focal sequence (p.support); only values greater than 0.01 are displayed. Superscript *n* indicates that each call in the vocal utterances can have any number of vocal element repetitions (from zero upwards).

Recipient of affiliation

We could record with certainty the recipient for all vocal utterances used during affiliative interactions. Those directed only towards infants were clearly dominated by ‘twitter_grunt’ and ‘twitter_gruntX’ sequences (Figures 1 and 4). In contrast, when affiliation was directed at both the infant and the mother, the distribution was more

balanced. We found between 83% and 100% posterior support for iterative sequences to be uttered more often in these types of affiliation than bigrams or single calls. Notably, single ‘grunts’ were also more likely than the ‘twitter_grunt’ bigram, but with weak posterior support (65%) for this difference (Figures 1 and 4). Generally, while single ‘grunts’ were produced towards a broader range of recipients, including adult females without infants (*N* = 12 ‘grunts’ from 10 individuals), they were still more frequently directed at females with infants, similar to single ‘twitters’ and their combinations (Figure 1 and Table S7 for details).

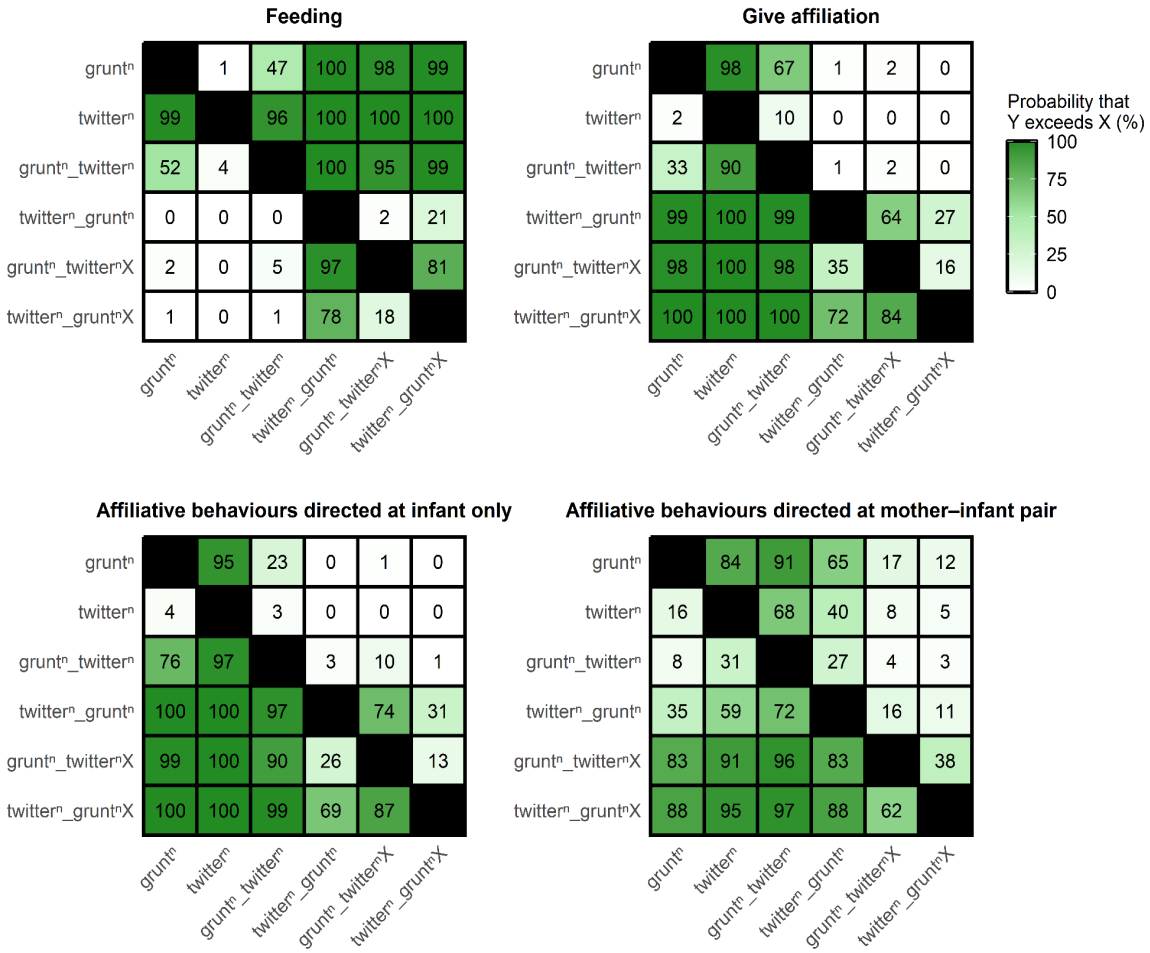


Figure 4. **Pairwise comparisons of the posterior distributions of the proportion of utterance occurrences produced in each contextual feature.** Each cell shows the percentage of posterior samples in which the utterance on the Y-axis was estimated to occur more frequently in the specified contextual feature (see matrix titles) than the utterance on the X-axis. Forest green shades indicate the strength of difference, with darker green representing higher percentages. Numbers inside cells are the rounded

percentage values. Superscript *n* indicates that each call in the vocal utterances can have any number of vocal element repetitions (from zero upwards).

Males - utterance types recorded

In male vocal utterances with identified contexts of production, we recorded *N* = 213 single calls, *N* = 21 bigrams, and *N* = 15 sequences longer than two calls (ranging from three to six calls). Among males, the predominant single calls with a clear production context were ‘grunts’, accounting for 83.6% of single calls, followed by ‘growls’ at 16%. The most frequently observed bigram was ‘shrill_hoo’ (*N* = 19), produced by three individuals. All other single calls and sequences (bigrams or longer) occurred five times or fewer (see Table S4 for complete sample sizes).

Males - contextual usage of single calls ‘grunt’ and ‘growl’ and bigram ‘shrill_hoo’

Like females, male ‘grunts’ occurred across a wide range of contextual features, with higher proportions recorded during feeding (FE), resting (RE), and affiliative interactions (GI), particularly those directed at adult females (AF) and other adult males (AM) (Figure 5). Note that males do not produce ‘twitters’. Male ‘growls’ were mostly produced during aggressive interactions (GG), especially towards other adult males (AM). ‘Shrill_hoo’ bigrams were more frequently given in response to monkey alarm calls (MK), and occasionally during encounters with non-monkey animals (AE) or as a bystander to aggression (BG) (Figure 5).

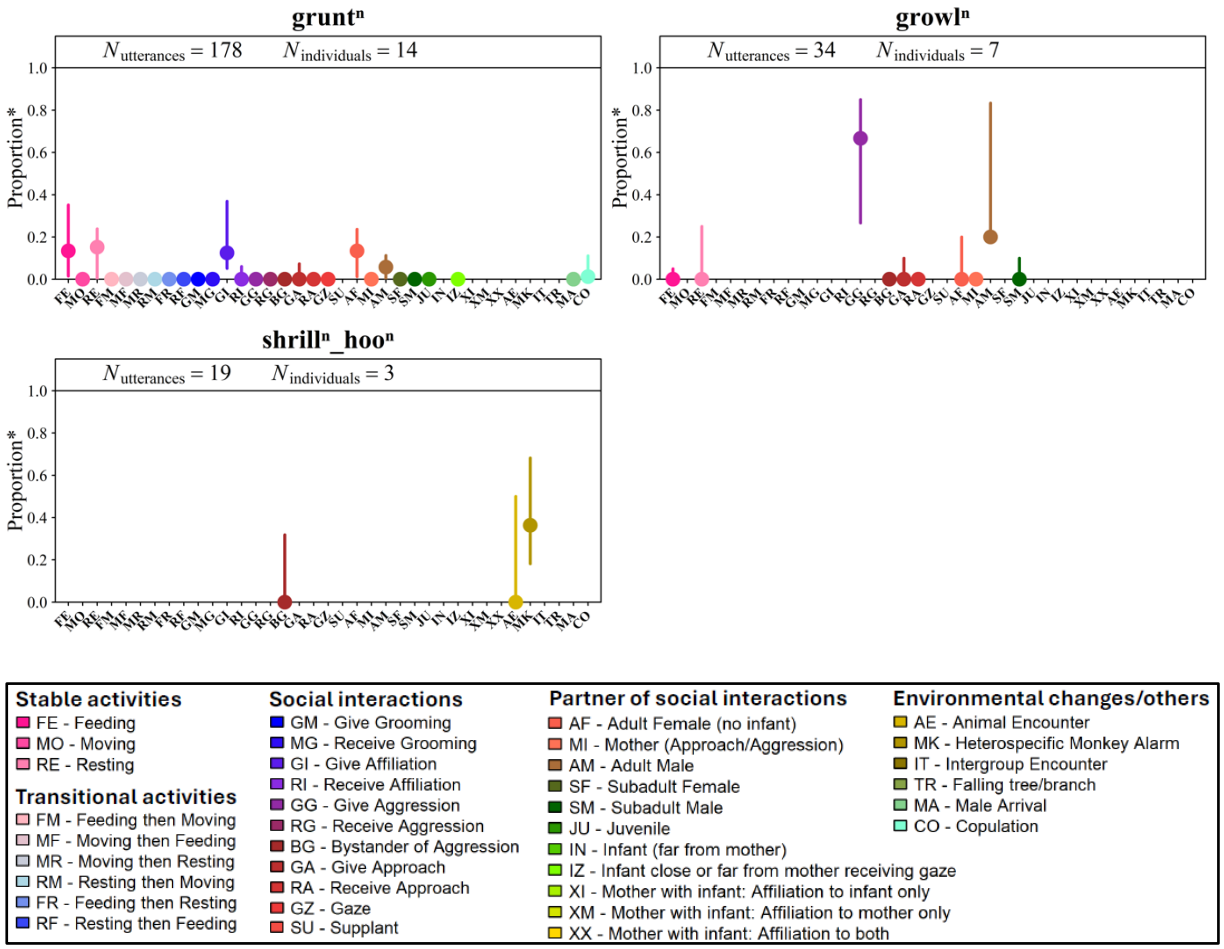


Figure 5. Distribution of contextual features of the most frequently recorded male vocal utterances. Coloured dots show the observed median individual proportion of occurrence of each utterance in each contextual feature. Vertical coloured lines indicate the 25th and 75th percentiles. A 25th percentile at proportion *P* means that 75% of individuals produced that utterance in the given contextual feature at a proportion ≥ *P*. Similarly, the median at *P* indicates that half of individuals produced the utterance at a proportion ≥ *P*, and the 75th percentile at *P* indicates that 25% produced it at a proportion ≥ *P*. Superscript *n* indicates that each call in the vocal utterances may include any number of vocal element repetitions (from zero upwards).

Note: "*Proportion**" refers to values calculated per contextual feature and may sum to more than one across all features because some utterances were assigned to multiple contextual features. For example, when a single utterance involved a social partner, it was counted for both the type of social interaction and recipient attribute (e.g., *give affiliation* + *adult female*). Each individual proportion remains ≤1, but the total across contextual features may exceed one as utterances can contribute to multiple feature types.

Discussion

We used a whole repertoire approach to map which natural contexts trigger sooty mangabey vocal utterances. Where vocal sequences occurred sufficiently frequently, we analysed potential for meaning shifts, for: (1) bigrams and their call components, and (2) bigrams and their iterative sequence forms. In females, criteria (1) and (2) were reached for 'twitter' and 'grunt' combinations, which typically occurred in feeding and affiliative contexts involving infants. This prompted a focused analysis of contextual shifts involving combinations of ‘twitters’ and ‘grunts’ in vocal sequences. Using Euclidean distance analyses to quantify contextual similarity between utterance types, we found a strong contextual shift in the usage of ‘twitter_grunt’ compared to either of its component calls. ‘Twitter_grunt’ was overwhelmingly emitted in affiliation to infants, whereas the component calls were mainly emitted during foraging. Although ‘twitter_grunt’ occurred in broadly overlapping contexts with its component calls, it differed in relative frequency, being more strongly associated with infant-directed affiliation than either individual call. This suggests a potential case of compositional structure in sooty mangabeys. Importantly, contextual usage influenced call order: the reversed bigram ‘grunt_twitter’ was used in contextual features more similar to those of its component calls than ‘twitter_grunt’, highlighting the role of ordering as a rule of sequence structure, promoting a shift in contextual usage. We further demonstrate that the iterative sequences ‘twitter_gruntX’ and ‘grunt_twitterX’ are produced in similar contexts to ‘twitter_grunt’, with only minor differences in contextual usage depending on which call initiate the sequence. In males, no vocal sequence reached the criteria for in depth contextual analysis. However, we discuss one vocal sequence emitted by males in alarm contexts, the ‘shrill_hoo’, although ‘hoos’ were not emitted independently.

Before examining the sequences, we confirmed that the contextual use of single calls across both sexes aligns well with descriptions from the previously established vocal repertoire, which focused solely on single calls (Range & Fischer, 2004). Below, we discuss the patterns and potential shifts in meaning observed in the vocal sequences.

Females - contextual usage of vocal sequences

Vocal sequences produced by females predominantly included just two call types, used in various configurations: ‘grunts’ and ‘twitters’. When emitted singly, these calls

overlapped contextually, primarily during feeding and affiliation involving infants, with feeding as the dominant context (Figure 1 and Table 1). Euclidean distance analyses confirmed their contextual similarity (Figure 2). Comparing the likelihood of each call within contexts, ‘grunts’ were more likely during affiliative interactions compared to ‘twitters’, while ‘twitters’ were more common during feeding compared to ‘grunts’ (Figure 4). These calls were also combined into bigrams, in either order, and into iterative sequences.

Ordering effect in bigrams

Our results show that combining two potentially meaning-bearing calls can lead to context-specific shifts in usage, depending on production order. Single ‘grunts’, ‘twitters’, and ‘grunt_twitter’ bigrams were mainly associated with feeding, and to a lesser extent, affiliative interactions involving infants (Figure 1 and Table 1). In contrast, the ‘twitter_grunt’ bigram showed a pronounced contextual shift, as it was almost exclusively used in affiliative contexts with infants (Figure 1 and Table 1). This order-dependent contrast suggests that while the individual calls have overlapping meanings, combining them in specific orders can clarify or narrow meaning.

Because both bigrams and single calls occurred in similar contextual features but with different probabilities, we can rule out idiomaticity, i.e., a meaning unrelated to parts (Price et al., 2009; Schlenker, Chemla, Arnold, et al., 2016). Instead, our findings suggest order-sensitive compositionality with clarification: the sequence meaning derives from its components, but this depends on which call initiates the sequence (Girard-Buttoz et al., 2025). Ordering effects have been described in chestnut-crowned babbler and olive colobus monkey sequences, though in those cases, the vocal units are not produced in isolation, so their meanings cannot be traced to individual components, which limits evidence of compositionality (Engesser et al., 2015; Gallot et al., 2024). In Campbell’s monkeys, sequences like ‘boom_krak’ are biased toward the meaning of the first call (‘boom’, a non-urgent threat), aligning with the “urgency principle” where critical information is placed first (Schlenker, Chemla, Schel, et al., 2016; Zuberbühler, 2002). In this sense, ‘twitter_grunt’ appears more comparable to order-sensitive bigrams in chimpanzees, which are produced in feeding and social contexts (Girard-Buttoz et al., 2025), or to the ‘alert–recruitment’ sequences of

Japanese tits, to which receivers respond only when presented in that specific order and not the reverse (Suzuki et al., 2016, 2017).

Compositionality involving non-alarm calls is rarely reported outside of great apes (Berthet et al., 2025; Girard-Buttoz et al., 2025), but see recent work in elephants (Hedwig & Kohlberg, 2024). More commonly, bigrams in socio-positive contexts feature affix-like elements added to stand-alone calls (e.g., Diana monkeys (Candiotti et al., 2012), Campbell's monkeys (Coye et al., 2018), mongooses (Jansen et al., 2012)), typically conveying signaller identity rather than refining context-specific meaning. Given that 'twitter_grunt' bigrams overwhelmingly target infants, identity-related information may be encoded in the calls. However, if identity marking were the sole function, we would not expect such clear shifts in contextual usage across 'grunt', 'twitter', 'grunt_twitter', and 'twitter_grunt'. Instead, our findings suggest a case of compositionality in a socially benign context, specifically during affiliation involving infants or contact calling during feeding.

Comparing iterative vocal sequences to bigrams

Iterative sequences starting with either 'grunt' or 'twitter' were primarily used in infant-directed affiliative contexts, similar to the 'twitter_grunt' bigram (Figure 1 and Table 1). Euclidean distances showed that 'twitter_grunt' and 'twitter_gruntX' were contextually similar but distinct from their components and from 'grunt_twitter', suggesting that 'twitter_grunt' may extend into 'twitter_gruntX' without a shift in meaning (Figure 3 and Table 2). In contrast, 'grunt_twitterX' had a more mixed profile. Though overlapping with 'twitter_grunt' and 'twitter_gruntX', it also aligned with 'grunt_twitter' and single 'grunt' calls (Figure 3), likely reflecting its more frequent use in feeding contexts (Figure 4). This suggests 'grunt_twitterX' occupies an intermediate position, with broader contextual use than 'twitter'-starting sequences.

Such patterns raise the possibility that the number of occurrences of a specific call type within a sequence could modulate meaning. Indeed, in several species, including bonobos (Clay & Zuberbühler, 2009, 2011), chickadees (Freeberg, 2008), great tits (Salis et al., 2022), putty-nosed monkeys (Arnold & Zuberbühler, 2012), titi monkeys (Berthet et al., 2019), and frogs (Bhat et al., 2022), the number of repetitions (whether adjacent or not) of specific call types or vocal elements within sequences has been linked to semantic content. However, these studies typically report the overall number

of occurrences without specifying whether these were consecutive repetitions of adjacent vocal elements composing a call (as defined in our study) or iterations of call types interspersed with other call types within a sequence (e.g., ABA). Nevertheless, they highlight that patterns of call reoccurrence within sequences likely contribute to meaning.

Interestingly, ‘grunt_twitterX’ sequences appear more likely to end with ‘grunt’ calls (Le Floch et al., 2025), possibly reflecting a sequence composed of one initial ‘grunt’ followed by one or more ‘twitter_grunt’ bigrams. In such a case, this may indicate that the caller embeds an infant-directed or infant-related message (‘twitter_grunt’) within a broader feeding context, clarified by the terminal ‘grunt’. In contrast, ‘twitter_gruntX’ does not appear to show a consistent terminal pattern, possibly because starting with ‘twitter’ already signals infant-directed affiliation. However, the idea that bigrams are embedded into iterative sequences remains speculative and warrants further investigation.

Recipient of affiliation

Given the variation in utterance usage within affiliative contexts, we explored whether the recipient influenced this variation. Female single ‘grunts’ in affiliation mainly were directed toward infants near or carried by their mothers, and less frequently towards other individuals. Conversely, male ‘grunts’ tended to target adult females without infants or other adult males (Figure 1). This likely reflects sex differences in social roles, with females being particularly interested in affiliative interactions involving infants (Fruteau et al., 2011). This pattern was more apparent when calls were combined into ‘twitter_grunt’, ‘twitter_gruntX’ and ‘grunt_twitterX’ sequences. Notably, when affiliative behaviours were directed solely at infants, females mainly used ‘twitter_grunt’ and ‘twitter_gruntX’. When directed at both mothers and infants, iterative sequences were more common, suggesting the recipient influences call combination use.

In many primates, calls before or during infant interactions signal benign intent and may facilitate access to infants by reducing maternal aggression (Bauers, 1993; Silk et al., 2000, 2003). For instance, in chacma and olive baboons, vocal production during infant handling is influenced by kinship and dominance between handler and mother (Silk et al., 2016, 2018), and Guinea baboons produce more ‘grunts’ when social bonds with mothers are weak (Faraut et al., 2019). In sooty mangabeys, females handle

infants gently, vary in interest based on infant age and availability, and show weaker kin biases than other primates (Fruteau et al., 2011). Grooming is often used to gain access, especially by lower-ranking females (Fruteau et al., 2011). Importantly, calls may also be directed at infants (Schick et al., 2022). In rhesus macaques, handlers use two distinct call types during infant handling: ‘girneys’, which may attract infant attention specifically, and ‘grunts’, which are produced in many contexts and may simply signal harmless intent (Whitham et al., 2007). Thus, sooty mangabey females may use single calls, bigrams, or iterative sequences, depending on factors like interest in the infant, caller–mother rank difference, kinship, and whether the utterance is directed at the infant, the mother, or both. These findings suggest that utterances in infant-directed affiliations may encode specific information about social negotiation or infant-directed communication (Schick et al., 2022). Further research is needed to clarify these patterns.

Males - contextual usage of vocal sequences

Adult males regularly produced one vocal sequence in our dataset, the ‘shrill_hoo’ bigram, although its components were not observed in isolation. This pattern is consistent with previous findings reporting lower sequence diversity in males compared to females (Sigmundson et al., 2025). While ‘hoo’ calls are rarely produced alone, ‘shrill’ calls occur independently in females. The absence of isolated ‘shrill’ calls in males prevented us from comparing bigrams to components. Instead, we focused on the contextual features of ‘shrill_hoo’, which was produced during animal encounters and in response to heterospecific alarm calls (Figure 5), consistent with its role in signalling threats (León et al., 2023; Range & Fischer, 2004). It also occurred, though less frequently, when callers witnessed conspecific aggression, suggesting a broader communicative function. This raises the possibility of acoustic variants by context, as in chacma baboons (Fischer et al., 2001).

Whole repertoire insights

Contexts of vocal production, such as affiliation, aggression and alarm, overlap with those in many other animals (Seyfarth & Cheney, 2010). Call combinations in sooty mangabeys likely function to expand messaging precision, particularly in socially

benign contexts. Vocal sequences also occurred in alarm contexts and during aggression although the sample size during aggression was too small to assess where sequence structure shifted meaning compared with the component calls. Whilst we clearly demonstrate use of call combinations outside of alarm contexts, the capacity to combine single calls into context-dependent bigrams appears less versatile in sooty mangabeys compared to chimpanzees (Girard-Buttoz et al., 2025) and bonobos (Berthet et al., 2025). Still, whether this reflects a hominid-specific ability to flexibly use compositional-like structures across daily life situations remains unclear, as comparable whole-repertoire analyses are still rare for other monkeys and non-primate species. Broader comparative work across taxa will be crucial to determine the extent to which such combinatorial capacities are shared or represent lineage-specific capacities.

Limitations

Despite careful observation, certain vocal utterances were under-represented in the dataset due to data collection limitations. This was particularly the case during fast-paced interactions, such as aggressive encounters within or between groups, where multiple individuals often vocalised simultaneously, or during alert events when it was not possible to identify the cause of the alert. Other contexts where vocalisations may have been produced, such as in tall trees or at night, when the caller could not be seen, could also not be reliably recorded. In addition, our categorisation of calls into broad types may have overlooked finer acoustic variation within each category. For example, ‘twitters’ may vary acoustically depending on the context of production (Range & Fischer, 2004). If such variation carries context-specific meaning, it could influence the interpretation of utterances, an aspect not addressed in our analyses. We also did not examine the potential roles of adjacent repetitions of vocal elements within call types or the number of call iterations in sequences composed of ‘grunt’ and ‘twitter’, both of which potentially contribute to utterance meaning. Finally, playback experiments are necessary to determine the function of these call combinations, and the meaning conveyed to receivers and third parties.

Conclusion

Our study demonstrates that sooty mangabeys combine singly-used, potentially meaning-bearing calls into sequences. Contextual usage (i.e., feeding vs. infant-directed affiliation) influenced call combinations, order within sequences, and iteration. Specifically, while single calls largely aligned with previously described contextual associations, bigrams, especially 'twitter_grunt', showed a contextual shift in usage, suggesting a form of compositionality shaped by ordering. Iterative sequences further indicate that meaning may be preserved or broadened depending on which call initiates the sequence, with some patterns hinting at the possibility that embedding of bigrams into longer iterated sequences may impact messaging. These findings contribute to the growing body of evidence that nonhuman animals can utilise rule-based vocal combinations to refine their communication during social interactions and expand their messaging beyond what is possible with single calls alone. Crucially, we propose that compositionality in socio-positive contexts might not be restricted to apes but may also be present in monkeys, suggesting that this shared capacity is evolutionarily ancient. However, while sooty mangabeys might exhibit this context-sensitive combinatorial capacity, they do not demonstrate the broad and flexible use of compositional-like structures across daily life seen in *Pan* species. Whether such extensive combinatorial capacities represent a hominid-specific adaptation remains an open question, largely because comparable whole-repertoire analyses are scarce, not only in monkeys but also more generally across non-primate taxa. Thus, more systematic and comparative research across diverse animal taxa is essential to clarify the evolutionary distribution of these communicative abilities and to understand where similar or divergent capacities might lie. Future research should also investigate how recipient attributes, such as caller–receiver kinship and social rank, influence call combination usage, and crucially, whether these vocal sequences convey different meanings to receivers.

References

Altmann, J. (1974). Observational Study of Behavior: Sampling Methods. *Behaviour*, *49*(3–4), 227–266. https://doi.org/10.1163/156853974X00534

Arnold, K., & Zuberbühler, K. (2006). The alarm-calling system of adult male putty-nosed monkeys, Cercopithecus nictitans martini. *Animal Behaviour*, *72*(3), 643–653. https://doi.org/10.1016/j.anbehav.2005.11.017

Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: Compositional or idiomatic expressions? *Brain and Language*, *120*(3), 303–309. https://doi.org/10.1016/j.bandl.2011.10.001

Bauers, K. A. (1993). A Functional Analysis of Staccato Grunt Vocalizations in the Stumptailed Macaque (Macaca arctoides). *Ethology*, *94*(2), 147–161. https://doi.org/10.1111/j.1439-0310.1993.tb00555.x

Bergman, T. J., Beehner, J. C., Painter, M. C., & Gustison, M. L. (2019). The speech-like properties of nonhuman primate vocalizations. *Animal Behaviour*, *151*, 229–237. https://doi.org/10.1016/j.anbehav.2019.02.015

Berthet, M., Coye, C., Dezecache, G., & Kuhn, J. (2023). Animal linguistics: A primer. *Biological Reviews*, *98*(1), 81–98. https://doi.org/10.1111/brv.12897

Berthet, M., Mesbahi, G., Pajot, A., Cäsar, C., Neumann, C., & Zuberbühler, K. (2019). Titi monkeys combine alarm calls to create probabilistic meaning. *Science Advances*, *5*(5), eaav3991. https://doi.org/10.1126/sciadv.aav3991

Berthet, M., Surbeck, M., & Townsend, S. W. (2025). Extensive compositionality in the vocal system of bonobos. *Science*, *388*(6742), 104–108. https://doi.org/10.1126/science.adv1170

Bhat, A. S., Sane, V. A., Seshadri, K. S., & Krishnan, A. (2022). Behavioural context shapes vocal sequences in two anuran species with different repertoire sizes. *Animal Behaviour*, *184*, 111–129. https://doi.org/10.1016/j.anbehav.2021.12.004

Bortolato, T., Friederici, A. D., Girard-Buttoz, C., Wittig, R. M., & Crockford, C. (2023). Chimpanzees show the capacity to communicate about concomitant daily life events. *iScience*, *26*(11), 108090. https://doi.org/10.1016/j.isci.2023.108090

Bouchard, A., & Zuberbühler, K. (2022). An intentional cohesion call in male chimpanzees of Budongo Forest. *Animal Cognition*, *25*(4), 853–866. https://doi.org/10.1007/s10071-022-01597-6

Bouchet, H., Blois-Heulin, C., & Lemasson, A. (2013). Social complexity parallels vocal complexity: A comparison of three non-human primate species. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00390

Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012). Context-related call combinations in female Diana monkeys. *Animal Cognition*, *15*(3), 327–339. https://doi.org/10.1007/s10071-011-0456-8

Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The Syntax and Meaning of Wild Gibbon Songs. *PLoS ONE*, *1*(1), e73. https://doi.org/10.1371/journal.pone.0000073

Clay, Z., & Zuberbühler, K. (2009). Food-associated calling sequences in bonobos. *Animal Behaviour*, *77*(6), 1387–1396. https://doi.org/10.1016/j.anbehav.2009.02.016

Clay, Z., & Zuberbühler, K. (2011). Bonobos Extract Meaning from Call Sequences. *PLoS ONE*, *6*(4), e18786. https://doi.org/10.1371/journal.pone.0018786

Collier, K., Townsend, S. W., & Manser, M. B. (2017). Call concatenation in wild meerkats. *Animal Behaviour*, *134*, 257–269. https://doi.org/10.1016/j.anbehav.2016.12.014

Coye, C., Ouattara, K., Arlet, M. E., Lemasson, A., & Zuberbühler, K. (2018). Flexible use of simple and combined calls in female Campbell's monkeys. *Animal Behaviour*, *141*, 171–181. https://doi.org/10.1016/j.anbehav.2018.05.014

Coye, C., Ouattara, K., Zuberbühler, K., & Lemasson, A. (2015). Suffixation influences receivers' behaviour in non-human primates. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1807), 20150265. https://doi.org/10.1098/rspb.2015.0265

Crockford, C., Gruber, T., & Zuberbühler, K. (2018). Chimpanzee quiet hoo variants differ according to context. *Royal Society Open Science*, *5*(5), 172066. https://doi.org/10.1098/rsos.172066

Cunningham, S., & Magrath, R. D. (2017). Functionally referential alarm calls in noisy miners communicate about predator behaviour. *Animal Behaviour*, *129*, 171–179. https://doi.org/10.1016/j.anbehav.2017.05.021

Darwin, C. (1981). *The descent of man, and selection in relation to sex*. Princeton University Press.

Dunayer, E. S., & Berman, C. M. (2018). Infant Handling Among Primates. *International Journal of Comparative Psychology*, *31*(0). https://doi.org/10.46867/ijcp.2018.31.02.06

Engesser, S., Crane, J. M. S., Savage, J. L., Russell, A. F., & Townsend, S. W. (2015). Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System. *PLOS Biology*, *13*(6), e1002171. https://doi.org/10.1371/journal.pbio.1002171

Engesser, S., Ridley, A. R., & Townsend, S. W. (2016). Meaningful call combinations and compositional processing in the southern pied babbler. *Proceedings of the National Academy of Sciences*, *113*(21), 5976–5981. https://doi.org/10.1073/pnas.1600970113

Engesser, S., Ridley, A. R., & Townsend, S. W. (2017). Element repetition rates encode functionally distinct information in pied babbler 'clucks' and 'purrs.' *Animal Cognition*, *20*(5), 953–960. https://doi.org/10.1007/s10071-017-1114-6

Engesser, S., & Townsend, S. W. (2019). Combinatoriality in the vocal systems of nonhuman animals. *Wiley Interdisciplinary Reviews: Cognitive Science*, e1493. https://doi.org/10.1002/wcs.1493

Faraut, L., Siviter, H., Pesco, F. D., & Fischer, J. (2019). How life in a tolerant society affects the usage of grunts: Evidence from female and male Guinea baboons. *Animal Behaviour*, *153*, 83–93. https://doi.org/10.1016/j.anbehav.2019.05.003

Fichtel, C., & Kappeler, P. M. (2002). Anti-predator behavior of group-living Malagasy primates: Mixed evidence for a referential alarm call system. *Behavioral Ecology and Sociobiology*, *51*(3), 262–275. https://doi.org/10.1007/s00265-001-0436-0

Fischer, J., Hammerschmidt, K., Cheney, D. L., & Seyfarth, R. M. (2001). Acoustic Features of Female Chacma Baboon Barks. *Ethology*, *107*(1), 33–54. https://doi.org/10.1111/j.1439-0310.2001.00630.x

Fitch, W. T. (2010). *The evolution of language* (Cambridge University Press).

Freeberg, T. M. (2008). Complexity in the chick-a-dee call of carolina chickadees (Poecile carolinesis): Associations of context and signlater behavior to call structure. *The Auk*, *125*(4), 896–907. https://doi.org/10.1525/auk.2008.07025

Freeberg, T. M., Dunbar, R. I. M., & Ord, T. J. (2012). Social complexity as a proximate and ultimate factor in communicative complexity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1597), 1785–1801. https://doi.org/10.1098/rstb.2011.0213

Fruteau, C., van de Waal, E., van Damme, E., & Noë, R. (2011). Infant access and handling in sooty mangabeys and vervet monkeys. *Animal Behaviour*, *81*(1), 153–161. https://doi.org/10.1016/j.anbehav.2010.09.028

Gabry, J., Češnovar, A., Johnson, A., & Bronder, S. (2023). *cmdstanr: R Interface to "CmdStan"* [Computer software]. https://mc-stan.org/cmdstanr

Gallot, Q., Depriester, C., Moran, S., & Zuberbühler, K. (2024). A non-human primate combinatorial system for long-distance communication. *iScience*, *27*(11). https://doi.org/10.1016/j.isci.2024.111172

Gill, S. A., & 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. *Behavioral Ecology and Sociobiology*, *56*(1), 71–80. https://doi.org/10.1007/s00265-003-0736-7

Girard-Buttoz, C., Bortolato, T., Laporte, M., Grampp, M., Zuberbühler, K., Wittig, R. M., & Crockford, C. (2022). Population-specific call order in chimpanzee greeting vocal sequences. *iScience*, 104851. https://doi.org/10.1016/j.isci.2022.104851

Girard-Buttoz, C., Neumann, C., Bortolato, T., Zaccarella, E., Friederici, A. D., Wittig, R. M., & Crockford, C. (2025). Versatile use of chimpanzee call combinations promotes meaning expansion. *Science Advances*, *11*(19), eadq2879. https://doi.org/10.1126/sciadv.adq2879

Girard-Buttoz, C., Zaccarella, E., Bortolato, T., Friederici, A. D., Wittig, R. M., & Crockford, C. (2022). Chimpanzees produce diverse vocal sequences with ordered and recombinatorial properties. *Communications Biology*, *5*(1), 410. https://doi.org/10.1038/s42003-022-03350-8

Grampp, M., Samuni, L., Girard-Buttoz, C., León, J., Zuberbühler, K., Tkaczynski, P., Wittig, R. M., & Crockford, C. (2023). Social uncertainty promotes signal complexity during approaches in wild chimpanzees (*Pan troglodytes verus*) and mangabeys (*Cercocebus atys atys*). *Royal Society Open Science*, *10*(11), 231073. https://doi.org/10.1098/rsos.231073

Grieves, L. A., Logue, D. M., & Quinn, J. S. (2014). Joint-nesting smooth-billed anis, Crotophaga ani, use a functionally referential alarm call system. *Animal Behaviour*, *89*, 215–221. https://doi.org/10.1016/j.anbehav.2014.01.008

Hauser, M. D. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Hedwig, D., & Kohlberg, A. (2024). Call combination in African forest elephants Loxodonta cyclotis. *PLOS ONE*, *19*(3), e0299656. https://doi.org/10.1371/journal.pone.0299656

Jansen, D. A., Cant, M. A., & Manser, M. B. (2012). Segmental concatenation of individual signatures and context cues in banded mongoose (Mungos mungo) close calls. *BMC Biology*, *10*(1), 97. https://doi.org/10.1186/1741-7007-10-97

K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology. (2024). *Raven Pro: Interactive Sound Analysis Software* (Version 1.6.5) [Computer software]. The Cornell Lab of Ornithology. https://www.ravensoundsoftware.com/

Kirchhof, J., & Hammerschmidt, K. (2006). Functionally Referential Alarm Calls in Tamarins (Saguinus fuscicollis and Saguinus mystax)—Evidence from Playback Experiments. *Ethology*, *112*(4), 346–354. https://doi.org/10.1111/j.1439-0310.2006.01165.x

Le Floch, A., Girard-Buttoz, C., Azaiez, T. S., Bande, N., Wittig, R. M., Moran, S., Zuberbühler, K., & Crockford, C. (2025). *Rule-based sequences in sooty mangabey vocal communication*. BioRxiv. https://doi.org/10.1101/2025.07.07.663462

León, J., Thiriau, C., Bodin, C., Crockford, C., & Zuberbühler, K. (2022). Acquisition of predator knowledge from alarm calls via one-trial social learning in monkeys. *iScience*, *25*(9), 104853. https://doi.org/10.1016/j.isci.2022.104853

León, J., Thiriau, C., Crockford, C., & Zuberbühler, K. (2023). Comprehension of own and other species' alarm calls in sooty mangabey vocal development. *Behavioral Ecology and Sociobiology*, *77*(5), 56. https://doi.org/10.1007/s00265-023-03318-6

Leroux, M., Bosshard, A. B., Chandia, B., Manser, A., Zuberbühler, K., & Townsend, S. W. (2021). Chimpanzees combine pant hoots with food calls into larger structures. *Animal Behaviour*, *179*, 41–50. https://doi.org/10.1016/j.anbehav.2021.06.026

Leroux, M., Chandia, B., Bosshard, A. B., Zuberbühler, K., & Townsend, S. W. (2022). Call combinations in chimpanzees: A social tool? *Behavioral Ecology*, arac074. https://doi.org/10.1093/beheco/arac074

Leroux, M., Schel, A. M., Wilke, C., Chandia, B., Zuberbuhler, K., Slocombe, K. E., & Townsend, S. W. (2023). Call combinations and compositional processing in wild chimpanzees. *Nature Communications*, *14*, 2225. https://doi.org/10.1038/s41467-023-37816-y

Manser, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 11.

McComb, K., & Semple, S. (2005). Coevolution of vocal communication and sociality in primates. *Biology Letters*, *1*(4), 381–385. https://doi.org/10.1098/rsbl.2005.0366

McGraw, W. S., Zuberbuhler, K., & Noë, R. (2007). *Monkeys of the Taï forest: An African primate community* (Vol. 51). Cambridge University Press.

Mielke, A., Crockford, C., & Wittig, R. M. (2019). Snake alarm calls as a public good in sooty mangabeys. *Animal Behaviour*, *158*, 201–209. https://doi.org/10.1016/j.anbehav.2019.10.001

Moran, S., & Bickel, B. (2020). Rejoinder to Huijbregts's: Biting into Evolution of Language. *Journal of Language Evolution*, *5*(2), 184–187. https://doi.org/10.1093/jole/lzaa005

Nellissen, L., Fuh, T., Zuberbühler, K., & Masi, S. (2024). Vocal consensus building for collective departures in wild western gorillas. *Proceedings of the Royal Society B: Biological Sciences*, *291*(2033), 20240597. https://doi.org/10.1098/rspb.2024.0597

Neumann, C., & Zuberbühler, K. (2016). Vocal correlates of individual sooty mangabey travel speed and direction. *PeerJ*, *4*, e2298. https://doi.org/10.7717/peerj.2298

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009a). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, *106*(51), 22026–22031. https://doi.org/10.1073/pnas.0908118106

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009b). Campbell's Monkeys Use Affixation to Alter Call Meaning. *PLoS ONE*, *4*(11), e7808. https://doi.org/10.1371/journal.pone.0007808

Price, T., Arnold, K., Zuberbühler, K., & Semple, S. (2009). Pyow but not hack calls of the male putty-nosed monkey (Cercopithcus nictitans) convey information about caller identity. *Behaviour*, *146*(7), 871–888. https://doi.org/10.1163/156853908X396610

Quintero, F., Touitou, S., Magris, M., & Zuberbühler, K. (2022a). An Audience Effect in Sooty Mangabey Alarm Calling. *Frontiers in Psychology*, *13*, 816744. https://doi.org/10.3389/fpsyg.2022.816744

Quintero, F., Touitou, S., Magris, M., & Zuberbühler, K. (2022b). The Evolution of Food Calls: Vocal Behaviour of Sooty Mangabeys in the Presence of Food. *Frontiers in Psychology*, *13*, 897318. https://doi.org/10.3389/fpsyg.2022.897318

R Core Team. (2021). *R: A language and environment for statistical computing* (Version 2021.09.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Range, F., & Fischer, J. (2004). Vocal Repertoire of Sooty Mangabeys (Cercocebus torquatus atys) in the Tai National Park. *Ethology*, *110*(4), 301–321. https://doi.org/10.1111/j.1439-0310.2004.00973.x

Range, F., & Noë, R. (2002). Familiarity and dominance relations among female sooty mangabeys in the Taï National Park: Relations Among Female Sooty Mangabeys. *American Journal of Primatology*, *56*(3), 137–153. https://doi.org/10.1002/ajp.1070

Rauber, R., Kranstauber, B., & Manser, M. B. (2020). Call order within vocal sequences of meerkats contains temporary contextual and individual information. *BMC Biology*, *18*(1), 119. https://doi.org/10.1186/s12915-020-00847-8

Salis, A., Léna, J.-P., & Lengagne, T. (2022). Which acoustic parameters modify the great tit's response to conspecific combinatorial mobbing calls? *Behavioral Ecology and Sociobiology*, *76*(4), 46. https://doi.org/10.1007/s00265-022-03157-x

Schick, J., Fryns, C., Wegdell, F., Laporte, M., Zuberbühler, K., van Schaik, C. P., Townsend, S. W., & Stoll, S. (2022). The function and evolution of child-directed communication. *PLOS Biology*, *20*(5), e3001630. https://doi.org/10.1371/journal.pbio.3001630

Schlenker, P., Chemla, E., Arnold, K., & Zuberbühler, K. (2016). Pyow-hack revisited: Two analyses of Putty-nosed monkey alarm calls. *Lingua*, *171*, 1–23. https://doi.org/10.1016/j.lingua.2015.10.002

Schlenker, P., Chemla, E., Schel, A. M., Fuller, J., Gautier, J.-P., Kuhn, J., Veselinović, D., Arnold, K., Cäsar, C., Keenan, S., Lemasson, A., Ouattara, K., Ryder, R., & Zuberbühler, K. (2016). Formal monkey linguistics. *Theoretical Linguistics*, *42*(1–2), 1–90. https://doi.org/10.1515/tl-2016-0001

Seyfarth, R. M., & Cheney, D. L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain and Language*, *115*(1), 92–100. https://doi.org/10.1016/j.bandl.2009.10.003

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, *210*(4471), 801–803.

Sharma, P., Gero, S., Payne, R., Gruber, D. F., Rus, D., Torralba, A., & Andreas, J. (2024). Contextual and combinatorial structure in sperm whale vocalisations. *Nature Communications*, *15*(1), 3617. https://doi.org/10.1038/s41467-024-47221-8

Sigmundson, R., Girard-Buttoz, C., Le Floch, A., Azaiez, T. S., McElreath, R., Zuberbühler, K., Wittig, R. M., & Crockford, C. (2025). Vocal sequence diversity and length remain stable across ontogeny in a catarrhine monkey (Cercocebus atys). *Communications Biology*, *8*(1), 465. https://doi.org/10.1038/s42003-025-07922-2

Silk, J. B., Kaldor, E., & Boyd, R. (2000). Cheap talk when interests conflict. *Animal Behaviour*, *59*(2), 423–432. https://doi.org/10.1006/anbe.1999.1312

Silk, J. B., Rendall, D., Cheney, D. L., & Seyfarth, R. M. (2003). Natal Attraction in Adult Female Baboons (Papio cynocephalus ursinus) in the Moremi Reserve, Botswana. *Ethology*, *109*(8), 627–644. https://doi.org/10.1046/j.1439-0310.2003.00907.x

Silk, J. B., Roberts, E. R., Städele, V., & Strum, S. C. (2018). To grunt or not to grunt: Factors governing call production in female olive baboons, Papio anubis. *PLOS ONE*, *13*(11), e0204601. https://doi.org/10.1371/journal.pone.0204601

Silk, J. B., Seyfarth, R. M., & Cheney, D. L. (2016). Strategic Use of Affiliative Vocalizations by Wild Female Baboons. *PLOS ONE*, *11*(10), e0163978. https://doi.org/10.1371/journal.pone.0163978

Stan Development Team. (2023). *Stan Modeling Language Users Guide and Reference Manual* (Version 2.33.1) [Computer software]. https://mc-stan.org

Stephan, C., & Zuberbühler, K. (2008). Predation increases acoustic complexity in primate alarm calls. *Biology Letters*, *4*(6), 641–644. https://doi.org/10.1098/rsbl.2008.0488

Struhsaker, T. T. (1967). Auditory communication among vervet monkeys (Cercopithecus aetthiops). *Social Communication among Primates*.

Suzuki, T. N. (2012). Referential mobbing calls elicit different predator-searching behaviours in Japanese great tits. *Animal Behaviour*, *84*(1), 53–57. https://doi.org/10.1016/j.anbehav.2012.03.030

Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nature Communications*, *7*(1), 10986. 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

Torti, V., Gamba, M., Rabemananjara, Z. H., & Giacoma, C. (2013). The songs of the indris (Mammalia: Primates: Indridae): contextual variation in the long-distance calls of a lemur. *Italian Journal of Zoology*, *80*(4), 596–607. https://doi.org/10.1080/11250003.2013.845261

Townsend, S. W., Engesser, S., Stoll, S., Zuberbühler, K., & Bickel, B. (2018). Compositionality in animals and humans. *PLOS Biology*, *16*(8), e2006425. https://doi.org/10.1371/journal.pbio.2006425

Whitham, J. C., Gerald, M. S., & Maestripieri, D. (2007). Intended Receivers and Functional Significance of Grunt and Girney Vocalizations in Free-Ranging Female Rhesus Macaques. *Ethology*, *113*(9), 862–874. https://doi.org/10.1111/j.1439-0310.2007.01381.x

Wich, S. A., Koski, S., de Vries, H., & van Schaik, C. P. (2003). Individual and Contextual Variation in Thomas Langur Male Loud Calls. *Ethology*, *109*(1), 1–13. https://doi.org/10.1046/j.1439-0310.2003.00837.x

Wittig, R. M. (2018). Taï Chimpanzees. In J. Vonk & T. Shackelford (Eds.), *Encyclopedia of Animal Cognition and Behavior* (pp. 1–7). Springer International Publishing. https://doi.org/10.1007/978-3-319-47829-6_1564-1

Zuberbühler, K. (2001). Predator-specific alarm calls in Campbell's monkeys, Cercopithecus campbelli. *Behavioral Ecology and Sociobiology*, *50*(5), 414–422. https://doi.org/10.1007/s002650100383

Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, *63*(2), 293–299. https://doi.org/10.1006/anbe.2001.1914

Zuberbühler, K., Noë, R., & Seyfarth, R. M. (1997). Diana monkey long-distance calls: Messages for conspecifics and predators. *Animal Behaviour*, *53*(3), 589–604. https://doi.org/10.1006/anbe.1996.0334

Zuberbühler, K., & Wittig, R. M. (2011). Field experiments with non-human primates: A tutorial. In J. M. Setchell & D. J. Curtis (Eds.), *Field and Laboratory Methods in Primatology* (2nd ed., pp. 207–224). Cambridge University Press. https://doi.org/10.1017/CBO9780511921643.013

Supplementary material

Table S1. Ethogram of contextual features surrounding vocal production. Based on the ethogram from Range & Fischer (2002), this table lists all features and their associated definitions. Contextual features frequently associated with vocal production in the literature are marked with an asterisk (*); see references below the table. The caller's role (actor or receiver) is indicated for each type of social interaction, except in cases where the caller was a bystander to aggression (four observations), involved in a supplant (three observations), or engaged in copulation (which always occurred between a mature male and female). Whenever the caller's role is specified, the interaction partner is considered part of the context.

Contextual feature type	**Code**	**Definition**
Stable activity	FE*	Feeding. Can be either: -Eating: individual sits on the ground or in a tree and puts food in their mouth continuously. -Foraging: individual moves slowly forward while visually scanning the forest floor and the lower parts of vegetation, occasionally putting food in their mouth.
	MO*	Moving: individual walks steadily forward without visually scanning the forest floor and the lower parts of vegetation in search of food.
	RE	Resting: individual rests or sleeps while sitting or lying down and can occasionally empty their cheek pouches and chew some food.
Transitional activity	FM*	Transition from feeding to moving.
	FR	Transition from feeding to resting.

	MF	Transition from moving to feeding.
	MR	Transition from moving to resting.
	RF	Transition from resting to feeding.
	RM*	Transition from resting to moving.
Social interaction	GM/MG*	Give grooming/Receive grooming. Continuous grooming interaction between individuals, excluding other affiliative behaviours. Recorded as a bout.
	GA/RA*	Give approach/receive approach. Approach: one individual moves towards another (the receiver), closing the distance until they are within one meter of each other, and positions their body towards the receiver.
	GG/RG/BG*	Give aggression/receive aggression/bystander of an aggression. Aggression: comprises hard contacts (i.e., biting, gripping, slapping), arm wave, chase (i.e., running after another individual), charge (i.e., changing speed while coming at another one while staring at them to displace them), display (stares at another individual while showing their teeth/tongue and/or head bobbing).
	GI/RI*	Give affiliation/receive affiliation. Affiliation: comprises embrace (i.e., one individual puts one or two arms around another individual), infant handling/touching (when the infant is carried

		by the mother or less than one meter from her), start grooming (i.e., first grooming contact initiated toward another individual), genital inspection, hugging using the tail (one individual wraps their tail around another one), mouth sniffing, mounting (i.e., copulation-like interaction but without penetration), play, genital presenting (i.e., one individual shows their genitals to another one by placing their body so the genitals are in front of the other's face).
	GZ	Gaze: one individual standing or sitting still and looking in the direction of another individual without engaging in any other social behaviour. In sooty mangabeys, gaze is difficult to assess in the wild due to frequent and rapid head movements. Nevertheless, we systematically assessed gaze whenever a vocalisation was recorded. We only included conspicuous gazes—those where the individual's head remained still or moved slowly to follow the gaze target, and where the recipient or object of the gaze could be clearly identified within approximately 5 metres. If gaze occurred alongside other social behaviours, we classified the context based on those behaviours, as gaze in such cases was considered part of a broader interaction rather than a distinct behavioural event. These strict criteria likely contributed to the limited number of gaze observations in our dataset.
	SU	Supplant: an individual (the actor) approaches another individual (the receiver) occupying a resource site (i.e., a 1-metre circumference area containing food around the receiver) without

		displaying aggression. The receiver is either sitting and eating or standing and foraging. The receiver immediately leaves the site upon the actor's approach, either at a normal speed, quickly, or even fleeing. The actor then takes the receiver's place at the site to eat or forage.
Environmental changes	AE*	Animal encounter: encounter with another animal (i.e., eagle, chimpanzee, leopard, Gabon/rhinoceros viper, hippo, duiker, squirrel).
	IT	Intergroup encounter: encounter with a neighbouring group of sooty mangabeys.
	MK*	Monkey alarm: alarm call given by another monkey species (i.e., black-and-white colobus, Campbell's monkey, Diana monkey, lesser spot-nosed monkey).
	TR	Tree: falling branch or tree.
Other interactions	CO*	Copulation. For copulation, we did not assign a partner as contextual feature , as these interactions always involved a mature male and a mature female, and we had no rationale to expect differences in vocal production based on the age class (subadult vs. adult).
	MA	Male adult arrival: arrival of an adult male from the group at the location of the caller.

References:

Grampp, M., Samuni, L., Girard-Buttoz, C., León, J., Zuberbühler, K., Tkaczynski, P., ... & Crockford, C. (2023). Social uncertainty promotes signal complexity during approaches in wild chimpanzees (Pan troglodytes verus) and mangabeys (Cercocebus atys atys). *Royal Society Open Science*, *10*(11), 231073.

León, J., Thiriau, C., Crockford, C., & Zuberbühler, K. (2023). Comprehension of own and other species' alarm calls in sooty mangabey vocal development. *Behavioral Ecology and Sociobiology*, *77*(5), 56.

León, J., Quintero, F., & Zuberbühler, K. (2023). Acquisition of predator knowledge in sooty mangabeys. *Animal Behaviour*, *205*, 1-14.

Mielke, A., Crockford, C., & Wittig, R. M. (2019). Snake alarm calls as a public good in sooty mangabeys. *Animal Behaviour*, *158*, 201-209.

Neumann, C., & Zuberbühler, K. (2016). Vocal correlates of individual sooty mangabey travel speed and direction. *PeerJ*, *4*, e2298.

Quintero, F., Touitou, S., Magris, M., & Zuberbühler, K. (2022). The evolution of food calls: vocal behaviour of sooty mangabeys in the presence of food. *Frontiers in Psychology*, *13*, 897318.

Quintero, F., Touitou, S., Magris, M., & Zuberbühler, K. (2023). Audience Effects in Sooty Mangabey Agonistic Behaviour.

Range, F., & Fischer, J. (2004). Vocal repertoire of sooty mangabeys (Cercocebus torquatus atys) in the Taï National Park. *Ethology*, *110*(4), 301-321.

Table S2. Social interaction partner with age-class classification.

Age Class	Code	Age/biological marker
Infant	IN	<1 years old, more than 1m from the mother; for affiliation and aggression

		Biological marker: mostly carried by mother, or close to mother.
Infant	IZ	<1 years old, more than 1m from the mother; as being the receiver of a gaze.
Juvenile	JU	Between >= 1 and < 4 years old for females Between >= 1 and < 4 years old for males Biological marker: not carried by mother, independent from mother.
Subadult female	SF	Between >= 4 and < 5 years old Biological marker: first appereance of sexual swelling.
Subadult male	SM	Biological marker: -min the size of an adult female. -tests begin to appear and/or canines begin to develop.
Adult female not carrying an infant	AF	> 5 years old Biological marker:

		-full size. -first parturition.
Adult female carrying/close to her infant (i.e., <1m) at the time of data collection for aggression and approaches (not affiliations)	MI	
Infant being carried/being closed to (<1m) by/to their mother and being the receiver of an affiliation	XI	
Adult female carrying/close to her infant (i.e., <1m) at the time of data collection being the receiver of an affiliation (not the infant)	XM	
Adult female carrying/close to her infant (i.e., <1m) at the time of data collection and both her and her infant are being the receivers of an affiliation (i.e., behaviours of the actor are given to both)	XX	
Adult male	AM	Biological marker: -tests completely out and canines fully developed.

Table S3. Inter-observer tests on context classification.

We conducted two sets of inter-observer reliability tests on context classification: one during the 2022 field session between ALF and TA, and another during the 2023 field session between ALF and NB. We used an unweighted Cohen's kappa test and considered a value of at least 0.7 (i.e., substantial agreement) as satisfactory.

Gaze direction was recorded infrequently due to our strict and conservative criteria for assessing it (see Table S1, 'Gaze' in the ethogram).

We did not conduct a reliability test for mother–infant distance between ALF and TSA because data collection began in June, after infants were no longer being carried or staying close to their mothers.

Grooming was included under social interaction behaviours in the reliability test because it was initially not classified as an activity. We reclassified it post hoc to unify all long-duration behaviours and to distinguish them from brief affiliative events.

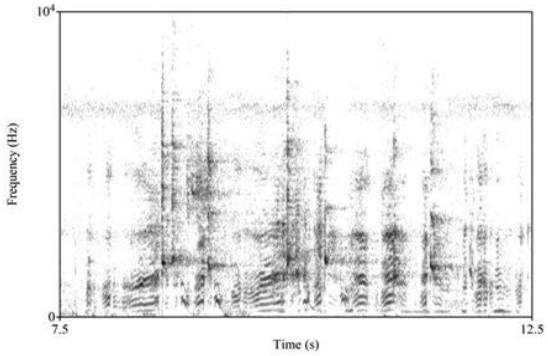
Category	**ALF vs TSA**	**ALF vs NB**
Activities	$\kappa = 0.72$, $z = 6.17$, $p < 0.001$, $N = 50$ observations	$\kappa = 0.87$, $z = 7$, $p < 0.001$, $N = 64$ observations
Social interaction behaviours	$\kappa = 0.81$, $z = 12.4$, $p < 0.001$, $N = 23$ observations	$\kappa = 0.74$, $z = 20.1$, $p < 0.001$, $N = 63$ observations
Caller role during social interaction (actor/receiver)	$\kappa = 1$, $z = 7.88$, $p < 0.001$, $N = 18$ observations	$\kappa = 1$, $z = 6.08$, $p < 0.001$, $N = 37$ observations
Gaze direction	$\kappa = 1$, $z = 2.65$, $p < 0.01$, $N = 7$ observations	$\kappa = 0.73$, $z = 2.62$, $p < 0.01$, $N = 8$ observations
Distance between a mother and their infant	-	$\kappa = 0.81$, $z = 5.02$, $p < 0.001$, $N = 38$ observations

Document S1. Call types.

Range and Fischer (2004) have thoroughly described the complete vocal repertoire of sooty mangabeys, and readers seeking more detail should refer to their paper. For convenience, we provide a condensed version here, including all the call types examined in our study.

Growl

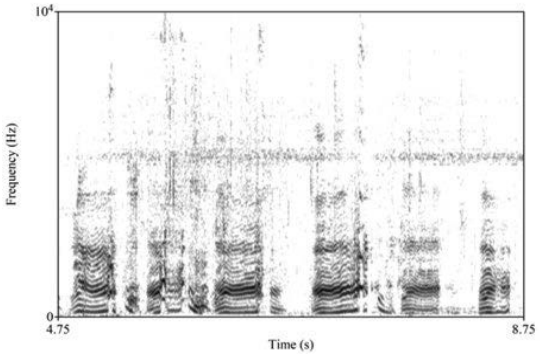
Growls are characterised by a low fundamental frequency band with many visible overtones, and they vary in both frequency and duration, often comprising a combination of shorter and prolonged elements.



Spectrographic examples of growls. y axis represents the frequency in Hz and x axis the time in s.

Grumble

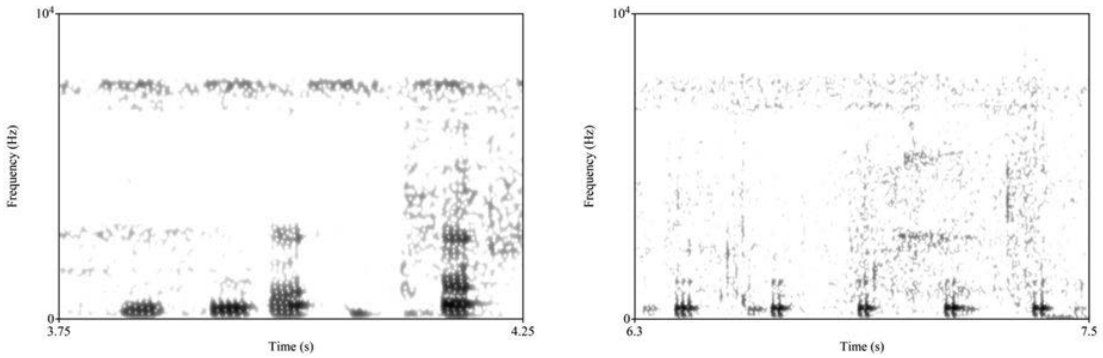
Grumbles are longer than growls, characterised by a low fundamental frequency and a rich harmonic structure with up to 12 overtones.



Spectrographic examples of grumbles interspersed with growls (shorter vocalisations). y axis represents the frequency in Hz and x axis the time in s.

Grunt

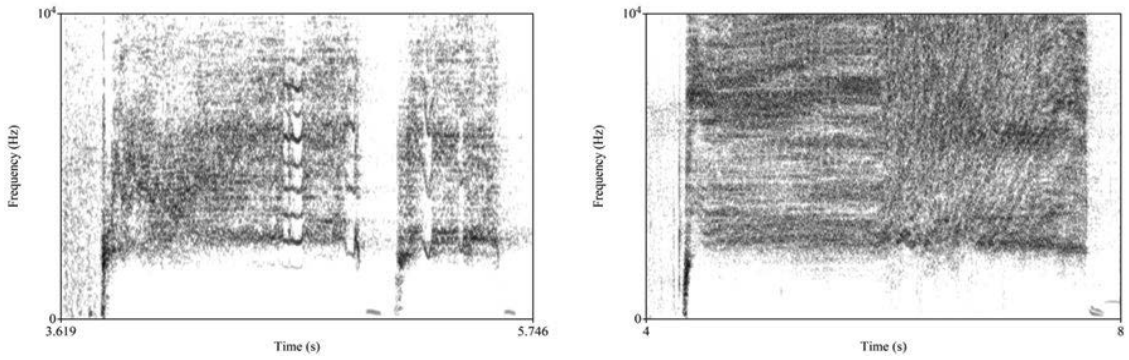
Grunts are low-frequency vocalisations, consistently short in duration (less than 200 milliseconds), and produced across a wide range of contexts. Adult females and males diverge in grunt production (see Range and Fischer).



Spectrograms of grunt calls, from an adult female (left) and an adult male (right). The y axis represents the frequency in Hz and x axis the time in s.

Scream

Screams are harsh calls characterised by wide-band noise extending up to 15 kHz, often lasting for several seconds.

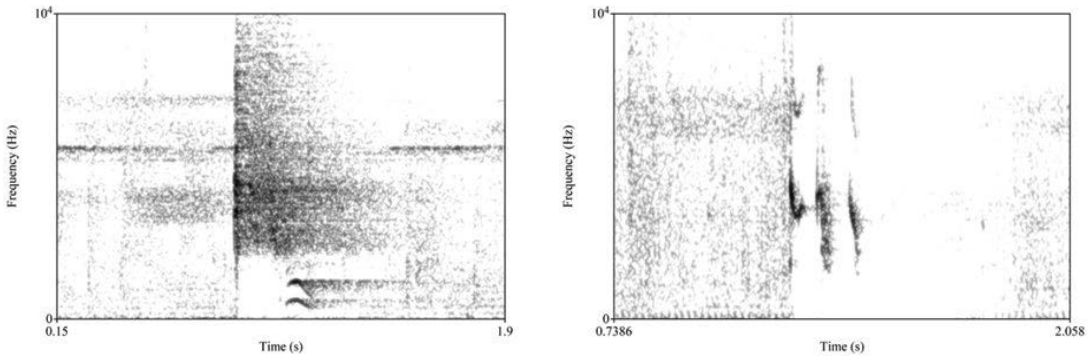


Spectrographic examples of screams. The y axis represents the frequency in Hz and the x axis is time in seconds (s).

Shrill and hoo

Referred to as the alarm call in Range and Fischer, 2004). Shrill calls consist of several high-frequency elements, sometimes displaying a harmonic structure, though this is

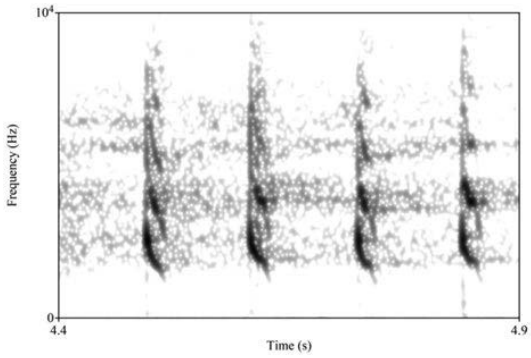
less common, and are frequently combined with a hoo call, which features a low fundamental frequency and a rich, readily visible harmonic structure.



Spectrograms of shrill calls combined with a hoo call (left) and shrill call produced singly (right). The y axis represents the frequency in Hz and the x axis is time in s.

Twitter

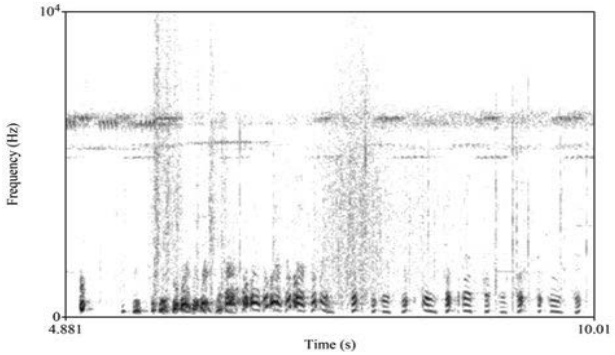
Twitters are short, high-frequency calls with a fundamental frequency around 1 kHz and many visible overtones extending up to 20 kHz. Adult males do not produce them.



Spectrographic example of twitter. The y axis represents the frequency in Hz and the x axis is time in s.

Vibrato

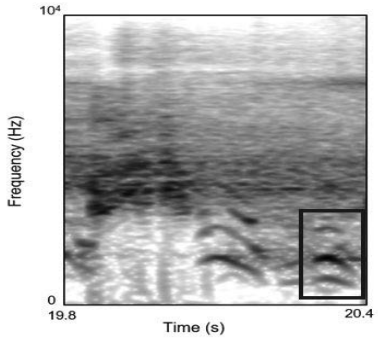
Referred to as the copulation call in Range and Fischer, 2004. Complex call used exclusively by sexually mature females. It consists of many variable elements with a low fundamental frequency and a rich harmonic structure, with individual elements alternating between pitch shifts upwards and downwards, creating the characteristic "vibrato."



Spectrographic example of vibrato. The y axis represents the frequency in Hz and the x axis is time in s.

Wau

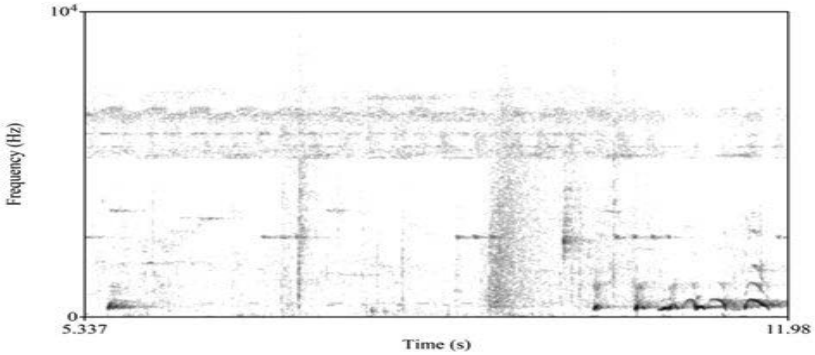
Short, low-frequency call with an ascending-descending pitch pattern and several visible harmonics.



Spectrographic example of wau (in the black box) preceded by shrill and hoo. The y axis represents the frequency in Hz and the x axis is time in s.

Whoop-gobble

Loud, low-frequency complex call that can be heard over long distances and produced by adult males. It begins with a single low-frequency element, followed by a pause that may last several seconds, before concluding with a series of repetitive, low-frequency elements. Although the pause between the first and final elements typically violates our usual rule for defining a call type, we treated it as a special case due to its consistent production in this form.



Spectrographic example of whoop-gobble. The y axis represents the frequency in Hz and the x axis is time in s.

Reference

Range, F., & Fischer, J. (2004). Vocal repertoire of sooty mangabeys (Cercocebus torquatus atys) in the Taï National Park. Ethology, 110(4), 301-321.

Table S4. Adult male vocal utterances sampled from audio recordings.

The table below presents the recorded vocal utterance types, along with sample sizes of adult males, for which we could define the context of production. ^{n} indicates that for each call, any number of repetitions of vocal elements is possible (from zero upwards).

Utterance type	**N utterances (total = 249)**	**N individuals (total = 14)**
grunt^{n}	178	14
growl^{n}	34	7
shrill^{n}_hoo^{n}	19	3
shrill^{n}_hoo^{n}_shrill^{n}_hoo^{n}	5	3
growl^{n}_scream^{n}	2	2

shrill^{n}_hoo^{n}_growl^{n}	2	1
growl^{n}_grumble^{n}_growl^{n}_grunt^{n}	1	1
growl^{n}_scream^{n}_growl^{n}	1	1
growl^{n}_scream^{n}_growl^{n}_scream^{n}_growl^{n}	1	1
growl^{n}_scream^{n}_hoo^{n}_growl^{n}	1	1
growl^{n}_shrill^{n}_hoo^{n}	1	1
growl^{n}_shrill^{n}_hoo^{n}_growl^{n}	1	1
scream^{n}_hoo^{n}_grumble^{n}_growl^{n}_grumble^{n}_growl^{n}	1	1
shrill^{n}_hoo^{n}_wau^{n}	1	1
whoopgobble^{n}	1	1

Table S5. Adult female vocal sequences of grunts and twitters sampled from audio recordings.

The table below presents the raw sequences, composed of ‘grunt’ and ‘twitter’ calls, with sample sizes of adult females for which we could define the context of production. ^{n} indicates that for each call, any number of repetitions of vocal elements is possible (from zero upwards).

Utterance type	**N utterances**	**N individuals**
grunt^{n}_twitter^{n}	76	26
grunt^{n}_twitter^{n}_grunt^{n}	38	14

gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ	3	2
gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ	15	7
gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ	2	2
gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ	5	3
gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ	2	2
twitterⁿ_gruntⁿ	43	17
twitterⁿ_gruntⁿ_twitterⁿ	8	4
twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ	9	6
twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ	2	2
twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ	2	2
twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ	1	1
twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ	4	4
twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ	1	1
twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ	1	1
twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ	1	1

twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ_gruntⁿ_twitterⁿ	1	1

Table S6. Adult female vocal utterances.

The table below presents the recorded vocal utterance types, along with sample sizes of adult females for which we could define the context of production. ⁿ indicates that for each call, any number of repetitions of vocal elements is possible (from zero upwards).

Utterance type	**N utterances (total = 1591)**	**N individuals (total = 41)**
gruntⁿ	890	36
growlⁿ	60	26
gruntⁿ_twitterⁿ	76	26
twitterⁿ	119	29
screamⁿ	43	20
gruntⁿ_twitterⁿX	65	18
twitterⁿ_gruntⁿ	43	17
shrillⁿ	26	15
vibratoⁿ	107	12
twitterⁿ_gruntⁿX	30	11

growlⁿ_screamⁿ	9	6
screamⁿ_growlⁿ	4	4
shrillⁿ_hooⁿ	9	3
growlⁿ_grumbleⁿ_growlⁿ	2	2
growlⁿ_screamⁿ_growlⁿ	2	2
screamⁿ_growlⁿ_screamⁿ	2	2
grumbleⁿ_growlⁿ	1	1
growlⁿ_hooⁿ_growlⁿ_grumbleⁿ_growlⁿ	1	1
growlⁿ_hooⁿ_growlⁿ_hooⁿ_growlⁿ	1	1
growlⁿ_hooⁿ_growlⁿ_hooⁿ_growlⁿ_hooⁿ_growlⁿ	1	1
growlⁿ_hooⁿ_growlⁿ_hooⁿ_growlⁿ_hooⁿ_growlⁿ_hooⁿ_growlⁿ_hooⁿ_growlⁿ_hooⁿ_grumbleⁿ_hooⁿ_grumbleⁿ_growlⁿ_grumbleⁿ_hooⁿ_growlⁿ_hooⁿ_grumbleⁿ_growlⁿ_hooⁿ_growlⁿ_hooⁿ_grumbleⁿ_hooⁿ_growlⁿ_grumbleⁿ_growlⁿ_hooⁿ_growlⁿ_grumbleⁿ_growlⁿ_hooⁿ_growlⁿ_hooⁿ_growlⁿ	1	1
growlⁿ_screamⁿ_growlⁿ_grumbleⁿ	1	1
growlⁿ_screamⁿ_growlⁿ_screamⁿ_hooⁿ_growlⁿ_screamⁿ_hooⁿ	1	1
screamⁿ_hooⁿ	1	1
screamⁿ_hooⁿ_screamⁿ	1	1

scream^{n}_hoo^{n}_scream^{n}_hoo^{n}_scream^{n}	1	1
scream^{n}_twitter^{n}_scream^{n}_growl^{n}	1	1
shrill^{n}_hoo^{n}_shrill^{n}	1	1
twitter^{n}_growl^{n}	1	1
twitter^{n}_growl^{n}_shrill^{n}_twitter^{n}_growl^{n}	1	1

Table S7. **Recipient of affiliation in adult female vocal utterance composed of single and combined 'grunts' and 'twitter'.**

^{n}indicates that for each call, any number of repetitions of vocal elements is possible (from zero upwards).

Utterance type	**Affiliation recipient**	**N utterances**	**N individuals**
grunt^{n}	Adult Female (no infant)	12	10
grunt^{n}	Adult Male	1	1
grunt^{n}	Infant (far from mother)	4	4
grunt^{n}	Juvenile	2	2
grunt^{n}	Subadult Male	1	1
grunt^{n}	Mother with infant: Affiliation to infant only	90	21
grunt^{n}	Mother with infant: Affiliation to mother only	9	5
grunt^{n}	Mother with infant: Affiliation to both	64	21
grunt^{n}_twitter^{n}	Infant (far from mother)	1	1

grunt^{n}_twitter^{n}	Mother with infant: Affiliation to infant only	15	8
grunt^{n}_twitter^{n}	Mother with infant: Affiliation to mother only	1	1
grunt^{n}_twitter^{n}	Mother with infant: Affiliation to both	5	5
grunt^{n}_twitter^{n}X	Adult Female (no infant)	1	1
grunt^{n}_twitter^{n}X	Mother with infant: Affiliation to infant only	19	11
grunt^{n}_twitter^{n}X	Mother with infant: Affiliation to both	9	5
twitter^{n}	Infant (far from mother)	2	2
twitter^{n}	Mother with infant: Affiliation to infant only	13	8
twitter^{n}	Mother with infant: Affiliation to mother only	1	1
twitter^{n}	Mother with infant: Affiliation to both	11	7
twitter^{n}_grunt^{n}	Infant (far from mother)	1	1
twitter^{n}_grunt^{n}	Mother with infant: Affiliation to infant only	18	9
twitter^{n}_grunt^{n}	Mother with infant: Affiliation to mother only	2	2
twitter^{n}_grunt^{n}	Mother with infant: Affiliation to both	4	3
twitter^{n}_grunt^{n}X	Mother with infant: Affiliation to infant only	15	6
twitter^{n}_grunt^{n}X	Mother with infant: Affiliation to both	7	5

Chapter 3: Sooty mangabey alarm system reflects phylogenetic history and environmental pressures

Abstract

How animals use alarm calls offers insight into how they convey external events to others. Many species communicate about danger, often distinguishing between terrestrial and aerial predators. Two basic systems have been identified: information is conveyed either through acoustically distinct or graded call types or through calls combined according to specific rules. This diversity raises questions about the evolution of alarm systems, such as the extent to which they reflect phylogenetic constraints and ecological factors, including predator diversity and visibility. To explore this, we studied the alarm system of sooty mangabeys, a terrestrial forest-dwelling monkey species. We compared it to both sympatric monkey species, most of which use predator-specific combinations, and more distantly located but closely related Papionini, which mostly rely on a single graded call. We conducted predator model experiments simulating eagles, leopards, and vipers, and analysed the initial responses of the detecting individual. While sooty mangabeys showed distinct behaviours towards each predator, evidence for predator-specific vocal utterances was limited. They consistently used a single call type, ‘shrill’, across contexts, sometimes appending a ‘hoo’ affixation. This structure may convey additional information, such as threat type, urgency or visibility. We conclude that the sooty mangabey alarm system reflects both inherited traits and adaptations to ecological challenges.

Introduction

Alarm calls have long been central to the study of meaning in animal communication as they can encode information about threats using specific call types or call sequences. Early work proposed that some single call types function to refer to external events (Macedonia & Evans, 2010; Seyfarth et al., 1980; Zuberbühler, 2003). Indeed, several species produce acoustically distinct calls in response to specific predator types, typically distinguishing between aerial predators (e.g., raptors) and general alarms, and in some cases producing separate calls for ground predators (e.g., felids) (Dezecache & Berthet, 2018). This has been documented across diverse taxa, including primates (e.g., vervets (Seyfarth et al., 1980; Struhsaker, 1967), Campbell's (Zuberbühler, 2001), Diana, putty-nosed (Zuberbühler, 2000), blue monkeys (Murphy et al., 2013), lemurs (Fichtel & Kappeler, 2002)), mongooses (e.g., dwarf mongooses (Collier et al., 2020) and meerkats (Manser, 2001)), squirrels (e.g., red squirrels (Greene & Meagher, 1998)), and birds (e.g., chestnut-crowned babblers (Crane et al., 2016), Japanese great tits (Suzuki, 2012), noisy miners (Cunningham & Magrath, 2017), and pale-winged trumpeters (Seddon et al., 2002)). In some species, alarm calls closely resemble calls used in social contexts (e.g., vervet monkeys (Price et al., 2015), lemurs (Fichtel & Kappeler, 2002)) or are given across a wide range of threat types (e.g., lesser spot-nosed monkeys (Le Floch et al., 2021), woolly monkeys (Casamitjana, 2002), Belding's ground squirrels (Leger et al., 2010; Robinson, 2010), domestic fowls (Marler et al., 1987)), indicating that listeners may also rely on contextual cues to interpret their meaning.

Beyond individual calls, some species also combine calls into sequences which appear to convey meaning as they produce different responses in listeners, a capacity that has been likened to a rudimentary form of syntax (Leroux & Townsend, 2020). For instance, Campbell's monkeys modify predator-specific calls by adding suffixes that generalise or shift meaning (e.g., 'krak' for leopard, 'krak-oo' for general alert (Coye et al., 2015; Ouattara, Lemasson, & Zuberbühler, 2009)). Putty-nosed monkeys produce predator- and disturbance-specific calls, but only a sequence of both prompts group movement, interpreted as a possible idiomatic expression (Arnold & Zuberbühler, 2012; Price et al., 2009; Schlenker, Chemla, Arnold, et al., 2016). Titi monkeys vary the proportions of their calls within sequences to signal both predator type and location (Berthet et al., 2019), while colobus monkeys use similar vocal elements for leopards and eagles but distinguish threats through call series structure and timing (Schel et al.,

2009). This diversity in alarm vocal systems, ranging from distinct call types to flexible combinations and context-dependent usage, raises questions about the extent to which these systems reflect inherited phylogenetic constraints and adaptations to specific ecological challenges.

To address these questions, we investigated the alarm call system of sooty mangabeys (*Cercocebus atys*), a forest-dwelling monkey species from the Taï Forest (Côte d'Ivoire). We compared their alarm call system to those of several other monkey species: some that inhabit the same forest but are not closely related (from the genera *Cercopithecus* and *Colobus*), and others that belong to the tribe Papionini but live in different habitats. Among the previously investigated species that share the Taï Forest, almost all produce specific vocal utterances in response to eagles, in contrast to other threats, such as leopards or general disturbances (Table 1). This is achieved through either distinct call types or call combinations, with the exception of lesser spot-nosed monkeys, which produce a general alarm call (Le Floch et al., 2021).

Regarding closely related species, predator-specific alarm calls or sequences appear to be largely absent, although this may partly reflect a lack of dedicated research. Studies have shown that mandrills, the closest species whose vocalisations have been described, vary their alarm call rates in response to leopard and eagle models, but there is no report of potential differences in their acoustic structure depending on the type of threat (Kudo, 1987; Yorzinski & Vehrencamp, 2008). Black mangabeys, the next closest relatives to sooty mangabeys after mandrills with studied vocalisations, use the same ‘staccato-bark’ for both crowned eagles and human hunters (Horn, 1987). In macaques, which are more distantly related to sooty mangabeys than mandrills and black mangabeys, bonnet macaques vary their alarm calls in acoustic structure depending more on perceived threat level than on predator type (Coss et al., 2007). Barbary macaques have been reported to produce distinct ‘shrill bark’ calls in response to predators and general disturbances (Fischer et al., 2010). While these calls differ acoustically in response to snakes, humans, and dogs, playback experiments have failed to confirm predator-specific reactions (Fischer & Hammerschmidt, 2001). Despite extensive investigation of their vocal communication, baboons show very limited evidence of predator-specific alarm calls (Hammerschmidt & Fischer, 2019). Instead, they appear to rely on a graded vocal system, where acoustic variants differ across contexts (Byrne, 1981; Fischer et al., 2002; Fischer, Hammerschmidt, et al., 2001; Fischer, Metz, et al., 2001). Mandrills and geladas

appear to use call types similar to those of baboons (Kudo, 1987), although to date, we have not found studies that investigate the context-specificity of their alarm calls.

Table 1. Alarm call systems of sympatric and closely related monkey species to sooty mangabeys. We report here on monkey species whose alarm call systems have been investigated and that either inhabit the Taï forest in Ivory Coast (grey rows) or belong to the tribe *Papionini* (white rows). For each, we indicate which information they can encode using vocal utterances and their alarm system, whether they are sympatric with sooty mangabeys and provide their estimated divergence time from sooty mangabeys (i.e., time since their last common ancestor) (Paddock et al., 2024; Pozzi et al., 2014). We also indicate their habitat type and body weight for discussion (Registry-Migration.Gbif.Org, 2023; Rowe, 1996).

Monkey Species	Habitat Type	Body weight (kg)	Last common ancestor (Ma)	Information encoded	Alarm call system	Source
Diana monkeys	Closed forest	♀ ~5.4 ♂ ~5	~14	Eagle vs leopard	Distinct call types	(Zuberbühler, 2000)
Campbell's monkeys	Closed forest	♀ ~2.5 ♂ ~4.3	~14	a. Eagle vs leopard b. Predator vs general disturbance c. Urgency	a. Distinct call types b. Call combinations c. Call rate	(Coye et al., 2015; Lemasson et al., 2010; Ouattara, Lemasson, & Zuberbühler, 2009; Ouattara, Lemasson, & Zuberbuhler, 2009; Zuberbühler, 2001, 2002)

Lesser spot-nosed monkeys	Closed forest	♀ ~3 ♂ ~3.8	~14	Not specific	One call type	(Le Floch et al., 2021)
Putty-nosed monkeys	Closed forest	♀ ~4.1 ♂ ~6.4	~14	Eagle vs terrestrial disturbance	Distinct call types	(Arnold & Zuberbühler, 2006)
Olive colobus	Closed forest	♀ ~4.2 ♂ ~4.7	~21	Eagle vs leopard	Call combinations	(Gallot et al., 2024)
Black-and-white colobus	Closed forest	♀ ~8.3 ♂ ~9.9	~21	Eagle vs leopard	Distinct series lengths of the same call type	(Schel et al., 2009)
Chacma baboons	Open habitats	♀ ~16.8 ♂ ~20.4	~12.5	Mammalian carnivores vs crocodiles	Same call with acoustic variants	(Fischer, Hammerschmidt, et al., 2001)
Guinea baboons	Closed forest and open habitats	♀ ~12.5 ♂ ~21.5	~12.5	Threat detection vs predator movement	Distinct series lengths of the same call type	(Byrne, 1981)
Barbary macaques	Closed forest and	♀ ~10.5 ♂ ~16	~11	Snakes vs humans vs dogs (not confirmed	Same call with acoustic variants	(Fischer et al., 2010; Fischer &

	open habitats			via playback)		Hammerschmidt, 2001)
Bonnet macaques	Closed forest and open habitats	♀ ~4.1 ♂ ~6.5	~11	Urgency	Same call with acoustic variants	(Coss et al., 2007)
Black mangabeys	Closed forest	♀ ~15.3 ♂ ~21	~9	Eagle vs human	One call type	(Horn, 1987)
Mandrills	Closed forest	♀ ~11.5 ♂ ~26.9	~5	Eagle vs leopard	Same call with different rates	(Yorzinski & Vehrencamp, 2008)

If certain features of the sooty mangabey alarm system are inherited through phylogeny, we might expect a single call type to be used across predator contexts, possibly modulated acoustically depending on the threat, without the use of distinct call types or structured combinations. Conversely, if we observe the use of distinct calls or call sequences depending on predator type, this could point to adaptations shaped by environmental pressures, such as predator diversity or habitat structure. Given the substantial overlap in body weight between sooty mangabey females (5.6–7.0kg) and other Taï Forest monkeys, and the fact that even larger males (9.5–11.4kg) overlap in size with black-and-white colobus, we considered it unlikely that body size, and thus size-related vulnerability (i.e. larger monkeys being more targeted by predators (Zuberbühler & Jenny, 2002), would account for predator-specific vocal utterances (Table 1). Sooty mangabeys are thought to produce distinct alarm calls in response to vipers, leopards, and eagles (León, Thiriau, et al., 2023; Mielke et al., 2019; Quintero et al., 2022a; Range & Fischer, 2004), yet no study to date has systematically analysed, described, or compared their acoustic structure and whether they can be combined in sequences.

In this study, we conducted predator model experiments using these three predator types and analysed the initial behavioural and vocal responses of the first individual to detect the model. Our goals were twofold: (1) to confirm model ecological validity by comparing behavioural responses to those described during natural encounters and (2) to investigate differences in vocal responses across predator types. In particular, we focused on vocal utterance acoustic differences in relation to predator type (whether different call types were produced based on their acoustic structure), composition (whether these calls were combined into sequences) and rate, with these variables selected based on the alarm systems described in Table 1.

Methods

To investigate sooty mangabey alarm calls, we conducted model presentations simulating three of their main natural threats: eagles, leopards, and large vipers (Gaboon or Rhinoceros vipers). Sooty mangabeys frequently encounter these predators and alert others to their presence, although the snakes do not prey on monkeys but may pose a threat if accidentally stepped on (León, Thiriau, et al., 2023; McGraw et al., 2007; Mielke et al., 2019; Quintero et al., 2022a). See Document S1 for details on the models.

Data collection

Study sites and subjects

Data were collected by ALF, NB, and TSA on two habituated groups of sooty mangabeys in Taï National Park (Ivory Coast) (McGraw et al., 2007; Wittig, 2018) across two field seasons: March–July 2022 and February–October 2023. During the 2022 season, both the Taï Monkey Project (TMP) and the Taï Chimpanzee Project (TCP) groups consisted of similar sizes, each comprising approximately 60 individuals. In 2023, data collection focused solely on the TCP group, which had grown to around 80 individuals, following a substantial influx of migrating males and a high number of births. Experiments targeted adult and subadult males and females across the two habituated groups.

Experimental set-tup

Two types of trial set-ups were employed. In the first scenario, the model was already positioned by a field assistant in the direction of the approaching group. Here, two observers (ALF and either NB or TSA) waited until the group advanced naturally, and opportunistically recorded the behavioural and vocal responses of the first individual to spot the model. In the second scenario, which aimed to balance the sample across age and sex classes, specific individuals were targeted so we could expose the same individuals to different predator types. ALF followed these individuals until they reached the periphery of the group, at which point the model, either leopard or eagle, was presented directly to them by NB, who remained concealed behind a tree at a minimum distance of seven metres. For the viper model, it was strategically placed on the ground along the individual's walking path, but the placement was concealed from view to ensure the model was not seen during its positioning. In both scenarios, the presentation procedure was identical. For leopard and eagle, upon a signal from ALF, NB or the field assistant placed the model either on the ground (leopard) or at an elevated position (~2m, eagle), using a camouflage net to conceal their arms. Once ALF was certain the focal individual had seen the model, she signalled to the person behind the tree to remove it, conceal it in a bag, and remain hidden until the trial was complete. In two leopard trials, instead of presenting a model, a field assistant wore a cape made of artificial leopard-patterned fabric and emerged from behind a tree walking on all fours, before quickly retreating from view upon receiving the signal to end the trial. For the viper model, it was only removed from the ground when no individual was close enough to see the removal (approximately 15 metres).

To avoid habituation and pseudo-replication, we used two different eagle models, three different leopard models (outside of the cape) and three different viper models (see Document S1 for model details). The viper models were repainted after each trial, inspired by the patterns of real gaboon or rhinoceros vipers encountered in the forest. Additionally, models of the leopard and eagle were hidden as soon as one individual spotted them, with the aim of minimising the likelihood that any tested individual saw the same model more than once, although repeated exposures could not be entirely ruled out. A minimum of ten days was observed between experiments using the same model. Before conducting experiments, and every ten days, we conducted dummy runs (i.e., simulating an experiment without the model) to first try the setup without the risk of failing an experiment and to limit habituation to the experimental setup. Lastly,

after each experiment, GPS coordinates were noted to ensure that experiments were not conducted in the same location.

All vocal and behavioural responses of the focal individual (i.e., the first individual to spot the model) were recorded using a handheld camera (Panasonic models HV-V727/HC-V700/ HC-V700M) equipped with a short directional microphone (Sennheiser MKE 400). Vocal notes were added to the video to describe subtle or partially obscured behaviours (e.g., an individual raising its tail while on a branch, up in the canopy), to indicate the identity of the individual producing a call and to specify the individual's position in the tree relative to the ground when applicable. The distance between the model and the focal individual was also noted.

This work was conducted under the research permit for the Taï Chimpanzee Project (Wittig/006/MESRS/DRI) issued by the Ivorian Ministry of Higher Education and Scientific Research and the Ivorian Office of Parks and Reserves, which granted access to the Taï National Park. The use of wild animals in this research adheres to the guidelines set forth by the Association for the Study of Animal Behaviour and the International Union for Conservation of Nature.

In addition, we included five viper experiments with vocal responses, conducted by AM with the TCP group in 2015, using a similar setup (Mielke et al., 2019). These experiments involved different individuals from those tested in 2022 and 2023 and were used to increase the overall sample size, particularly for inter-rater reliability assessments.

Data coding

Video

ALF coded the behaviours of the focal individuals from the video recordings using Adobe Premiere Pro (version 22.5.0), annotating the onset of each behaviour with time markers (in seconds) and the offset for locomotive behaviours (i.e., retreat, approach model, or move vertically in the trees). All behaviours are based on previous studies (León et al., 2022; León, Quintero, et al., 2023; León, Thiriau, et al., 2023; Mielke et al., 2019) and detailed in the ethogram (Table 2). For each behaviour onset and offset, the individual's position in the tree was also noted using three categories: (1) less than

50cm from the ground (e.g., forest floor, log, root, or extremely low branches), (2) between 0.5-3 m from the ground and (3) more than 3 m from the ground.

Table 2. Ethogram of focal individual behaviours recorded from videos upon predator model detection.

Behaviour	**Definition**
Approach model	Movement in the direction of the model.
Bipedal stance	Standing upright on two feet, either on the ground or on a branch, with arms along the body or holding onto a support.
Retreat	Sudden, rapid movement to increase distance from the model (including flight and startle reactions).
Move down / up	Descend or ascend within the canopy at normal speed (not retreating). Pauses under 5 seconds are considered part of the same movement.
Tail raise	Tail lift above the body in a pronounced arch, forming a handle-like curve over the back.
Yawn	The individual opens their mouth widely to expose their canines, without vocalising.

To confirm the reliability of the coding, TSA independently recoded $N = 18$ videos (55% of all coded trials, see Table S2), including $N = 5$ eagle, $N = 5$ leopard, and $N = 8$ viper trials, in which she had not been involved in the data collection. TSA coded the onset of each behaviour listed in the ethogram, except for “approach model” and “yawn”, which were too infrequent or not always clearly identifiable on video; the latter required supplementary information from vocal notes. These behaviours were not used in our statistical analyses. Inter-rater reliability between ALF and TSA was assessed using Cohen’s Kappa, resulting in a substantial agreement ($\kappa = 0.82$, $z = 6.12$, $p < 0.001$) (Gamer et al., 2019; McHugh, 2012).

Audio

ALF used Raven Pro version 1.6.5 (K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology, 2024) to code each recording and extract 11 acoustic parameters (see Table S1 for selection criteria and definitions). She visually inspected the spectrograms (FFT size: 1024, Hann window, hop size: 256 samples) while simultaneously listening to the recordings to manually annotate each vocal element by marking its onset and offset, as well as its quality (good or bad, depending on interference and background noise). We defined a vocal element as a continuous sound with a distinct beginning and end. Following the established vocal repertoire of sooty mangabeys described by Range and Fischer (Range & Fischer, 2004), we first identified three element types: 'shrill' (corresponding to the initial high-frequency elements of the alarm call), 'hoo' (the optional final low-frequency element of the alarm call), and 'growl'. We observed that most 'shrill' elements were very short (median = 0.06 s, SD = 0.03 s, *N* = 386), with 98% being shorter than 0.1 s. However, eight 'shrills' elements produced by males were significantly longer (median = 0.21 s, SD = 0.04 s, range = 0.17–0.28 s). We classified these as a distinct variant, 'shrill2', to capture this male-specific variation. A full description of the element types, including spectrogram illustrations, is available in the supplementary materials (Document S2).

After coding each element, we examined the intervals between consecutive vocal elements to define both calls and combinations. The majority of these intervals (72%, n = 307) were shorter than 0.1 seconds (median = 0.03 s, SD = 2.25 s), indicating that elements typically occur in rapid succession. The high variability in interval duration further justified the need for a threshold. Based on this distribution, we defined a call as any pair of consecutive elements of the same type separated by less than 0.1 seconds, and a combination as any pair of consecutive calls of different types (e.g., 'shrill' and 'hoo') occurring within 0.1 seconds of each other. This approach contrasts with previous studies on sooty mangabey vocal sequences (Sigmundson et al., 2025) but allowed for a more fine-grained analysis of alarm calls. Vocal delivery rates are particularly relevant in alarm calls, as they can indicate different levels of danger (Lemasson et al., 2010; Yorzinski & Vehrencamp, 2008). Using a larger interval (e.g., 1 or 2 seconds) as a threshold might obscure this specific aspect of alarm call dynamics. We also defined a vocal utterance as either a single call type or a combination of distinct call types.

Data analysis

Behaviours

We ran four Bayesian regression models to assess the influence of predator type (eagle/leopard/viper) on initial behavioural responses of the first individual to detect the model. Due to the high frequency of retreat behaviours across all conditions (85.2% of trials), we first focused our analysis on this response, while reporting all observed behaviours (see Tables 3 and S7). Two models focused on retreat behaviours: one modelled the duration of retreat and the other the probability of retreating by climbing a tree. The retreat duration data contained a small number of zero values (4 out of 27 trials), corresponding to individuals who did not retreat. For this reason, we used a hurdle Gamma regression model, which accounts for this structure by separately modelling the probability of retreat (Bernoulli component) and the duration of retreat for those who did retreat (Gamma component) (Min & Agresti, 2005). A second model used a Bernoulli regression to assess the likelihood of retreating up a tree (N = 16 trials with retreat up and N = 11 without retreat up). We then examined the two next most frequently observed behaviours: bipedal stance (33.3% of trials; N = 8 trials with bipedal stance and N = 19 trials without) and tail raise (40.7%; N = 11 trials with tail raise, N = 16 trials without). For these binary outcomes, we used two additional Bernoulli regression models to evaluate how predator type influenced their likelihood.

For all models, we included sex (male vs. female) and age class (adult vs. subadult) as fixed effects. These factors were considered relevant because subadults and females are smaller in body size than adult males and may be more vulnerable to predation, leading to potentially different behaviours. Moreover, adult males have been observed to mob predators, which may also lead to distinct behavioural responses (McGraw et al., 2007). Individual identity was not included as a random effect due to the unbalanced and sparse nature of the data: 12 individuals were tested under only one predator condition, three under two conditions, and three under all three conditions. Including individual identity in such a design risks convergence issues and unreliable estimates, especially when estimating between-individual variance, which generally requires a sufficient number of observations per level and reasonably balanced sampling across conditions.

All models were fitted using four Markov chains. Each chain ran for 4,000 iterations, with 2,000 warm-up iterations per chain for the retreat duration, bipedal stance, and

tail raise models, resulting in 12,000 post-warm-up samples per model. For the model assessing the likelihood of retreating up a tree, each chain included only 2,000 iterations with 1,000 warm-up iterations (8,000 post-warm-up samples), as initial model runs showed reliable convergence with fewer iterations. In contrast, we increased the number of iterations for the retreat duration, bipedal stance, and tail raise models to ensure adequate convergence and effective sample sizes. The adapt delta parameter was kept as default for the models bipedal and tail raise and was set to 0.99 for the models retreat duration and likelihood of retreat up a tree to ensure adequate exploration of the posterior distribution. We used the default flat priors for all parameters. Convergence was assessed using the potential scale reduction factor (R-hat), with values of 1 indicating good convergence, and model fit was evaluated using posterior predictive checks (see Table S5).

Vocal responses

To analyse vocal responses, we considered trials in which the focal individual began vocalising within 30 seconds of model detection (N = 23 trials with vocalisations produced within 30 seconds of model detection and N = 2 trials with vocalisations produced after). We then included every vocalisation that started within a 30-second window following the onset of the first vocal element.

Acoustic analyses

We first tested for highly correlated acoustic parameters using a Pearson correlation test (coefficient > 0.8). As a result, we retained ten out of 11 parameters for the acoustic analyses (see Table S1 for details on the parameters used). We then ran three separate permuted discriminant function analyses (Mundry & Sommer, 2007), a method widely adopted in acoustic studies of animal vocal repertoires (Collier et al., 2020; Keenan et al., 2020; Leroux et al., 2021). For this, we included only vocal elements flagged as having good acoustic quality (i.e., a good signal-to-noise ratio and no overlap with other sounds). In pDFA-1, we first tested our set of acoustic parameters to determine if they could discriminate between ‘shrill’ and ‘hoo’ elements, with element type as the test variable. This initial test run helped assess the relevance of these parameters for subsequent analyses. In pDFA-2, we examined whether ‘shrill’

elements could be assigned to specific conditions (eagle/leopard/viper), with condition as test variable. Similarly, in pDFA-3, we investigated whether 'hoo' elements could be classified between eagle and leopard conditions, again using condition as the test variable, but excluding 'hoo' elements produced in the viper condition due to limited sample size (N = 6 'hoo' elements).

We used for our three pDFAs the *pDFA.incomplete* function. Unlike conventional DFAs, pDFA accounts for non-independence of data due to repeated measurements from the same individuals by including caller identity as a control factor. However, standard pDFA requires that each individual contributes data to all levels of the test factor (e.g., all call types or contexts), an assumption that is not met in our dataset, since some individuals produced only a subset of vocal element types or were recorded in only some contexts. The pDFA incomplete method is specifically designed to handle such incomplete datasets by allowing inclusion of individuals with partial data across test factor levels, thus maximising sample size while controlling for repeated measures. We set the parameter *balance.contr.fac.contrib* to balance the inclusion of individuals according to our data: for the first analysis discriminating between 'shrill' and 'hoo' elements (pDFA-1), we set this to "mode" to include the largest subset of individuals producing both call types (10 out of 16). For the subsequent analyses classifying 'shrill' and 'hoo' elements by predator context (pDFA-2 and 3), where only two individuals had observations for all three contexts, we set it to "no" to include all available data. This approach enabled us to assess acoustic discrimination while accounting for the unbalanced and incomplete nature of our dataset.

Utterance composition and rate of production

To better understand utterance composition, we examined both the number of 'shrill' elements per 'shrill' call and the probability of a 'hoo' call occurring in combination. The number of 'shrill' elements varied substantially across conditions, sexes, and age classes (see Table S3), which prevented model fitting; we therefore reported these results descriptively. However, we could analyse the probability of combining a 'hoo' using a Bayesian Bernoulli regression model with a binary outcome (whether or not a 'hoo' was combined; N = 96 utterances with 'hoo', N = 74 without) and predator type (eagle/leopard/viper) as the main explanatory variable. Additionally, given the importance of early vocal responses in signalling urgency (Narbona Sabaté et al.,

2022; Schlenker, Chemla, Schel, et al., 2016), we ran a complementary analysis restricted to vocal utterances produced within five seconds of the first vocal element onset, to assess immediate vocal behaviour in response to threats (*N* = 63 utterances with 'hoo', *N* = 18 without). This two-tiered analytical approach allowed us to capture both overall and immediate vocal responses to predator presence in our dataset.

We then ran two additional models to analyse the rate of vocal utterance production depending on predator type (predictor variable), one using the 30-second window and one using the 5-second window after the onset of the first vocal element. Utterance rate was calculated as the number of vocal utterances produced within each time window, divided by the duration of that window, resulting in continuous, strictly positive values. Given the continuous nature of the outcome and the skewed distribution, we used a Bayesian Gamma regression model with a log link function. For both models, we accounted for the fixed effects of sex (male vs. female) and age class (subadult vs. adult) but did not include the identity of the subjects (those who first saw the model per trial) as a random effect due to the relatively small number of observations per individual (11 individuals tested with one condition, three individuals tested with two conditions and two individuals tested with three conditions).

All models were fitted with 4 Markov chains, each with 4,000 iterations with 2,000 warm-up iterations per chain, resulting in 8,000 post-warm-up samples. We used the default flat priors for all parameters. Convergence was assessed using the potential scale reduction factor (R-hat), with values of 1 indicating good convergence and model fit was evaluated using posterior predictive checks (see Table S6).

All analyses were conducted using R version 4.3.3 with the *brms* package version 2.20.4, the *emmeans* package version 1.10.0 (Bürkner, 2015; Lenth, 2017; R Core Team, 2021) and for the pDFA, an R function designed by Roger Mundry based on the *lda* function from the *MASS* package version 7.5-53 (Mundry & Sommer, 2007; Venables & Ripley, 2002).

Results

Out of the *N* = 42 trials conducted, we discarded: *N* = 1 trial where Diana monkeys spotted the model and gave alarm calls first, *N* = 2 trials where the first individual detecting the model was a juvenile and *N* = 6 trials where we could not identify which

individual saw the model first. This led to *N* = 33 trials initially selected. After video coding, *N* = 27 trials remained where focal individual behaviours were clearly observable for 30 seconds following model detection (see Table S2 for details). Our analyses ultimately included *N* = 8 eagle trials, *N* = 9 leopard trials, and *N* = 10 viper trials.

Behaviours

Retreat was observed in 85.2% of trials (*N* = 23), making it the most frequent behaviour across all conditions. This was followed by tail raises (40.7%, *N* = 11 trials) and bipedal stance (33.3%, *N* = 8 trials), both almost exclusively in eagle and viper conditions, with only one individual raising their tail in the leopard condition. Other behaviours, such as approaching the model, moving up or down in the trees (without retreat) and yawning, were less common (see Table 3 for a summary of behaviours and Table S7 for details on behaviour succession across individuals and conditions).

Table 3. Behaviours recorded during the first 30 seconds following model detection by the first individual spotting the model. Numbers indicate the number of individuals for which a given behaviour was recorded (with the total number of occurrences in brackets).

Condition	**Approach model**	**Bipedal stance**	**Retreat**	**Move down (normal speed)**	**Move up (normal speed)**	**Tail raise**	**Yawn**
Eagle (N=8)	1 (1)	4 (5)	5 (5)	3 (3)	1 (1)	6 (10)	2 (3)
Leopard (N=9)	0 (0)	0 (0)	9 (9)	0 (0)	1 (1)	1 (1)	0 (0)
Viper (N=10)	3 (3)	4 (4)	9 (9)	1 (1)	5 (5)	4 (5)	0 (0)
N total behaviour	4 (4)	8 (9)	23 (23)	4 (4)	7 (7)	11 (16)	2 (3)
% of trials featuring each behaviour	14.8%	33.3%	85.2%	14.8%	25.9%	40.7%	7.4%

Retreat duration

In the analysis of retreat duration, the mean duration was shortest in the viper condition (mean posterior distribution = 1.60, 95%-CI [0.95–2.88]), followed by the eagle condition (mean posterior distribution = 3.34, 95%-CI [1.77–6.95]), and the longest duration was observed in the leopard condition (mean posterior distribution = 7.34, 95%-CI [4.25–13.53]) (Figure 1a). The pairwise comparisons showed that retreat duration in the leopard condition was significantly longer than in the eagle (estimate = -0.80, 95%-CI [-1.40–-0.06]) and viper (estimate = 1.53, 95%-CI [0.91–2.12]) conditions. Moreover, the retreat duration was also significantly different between the eagle and the viper condition (estimate = 0.73, 95%-CI [0.08–1.46]).

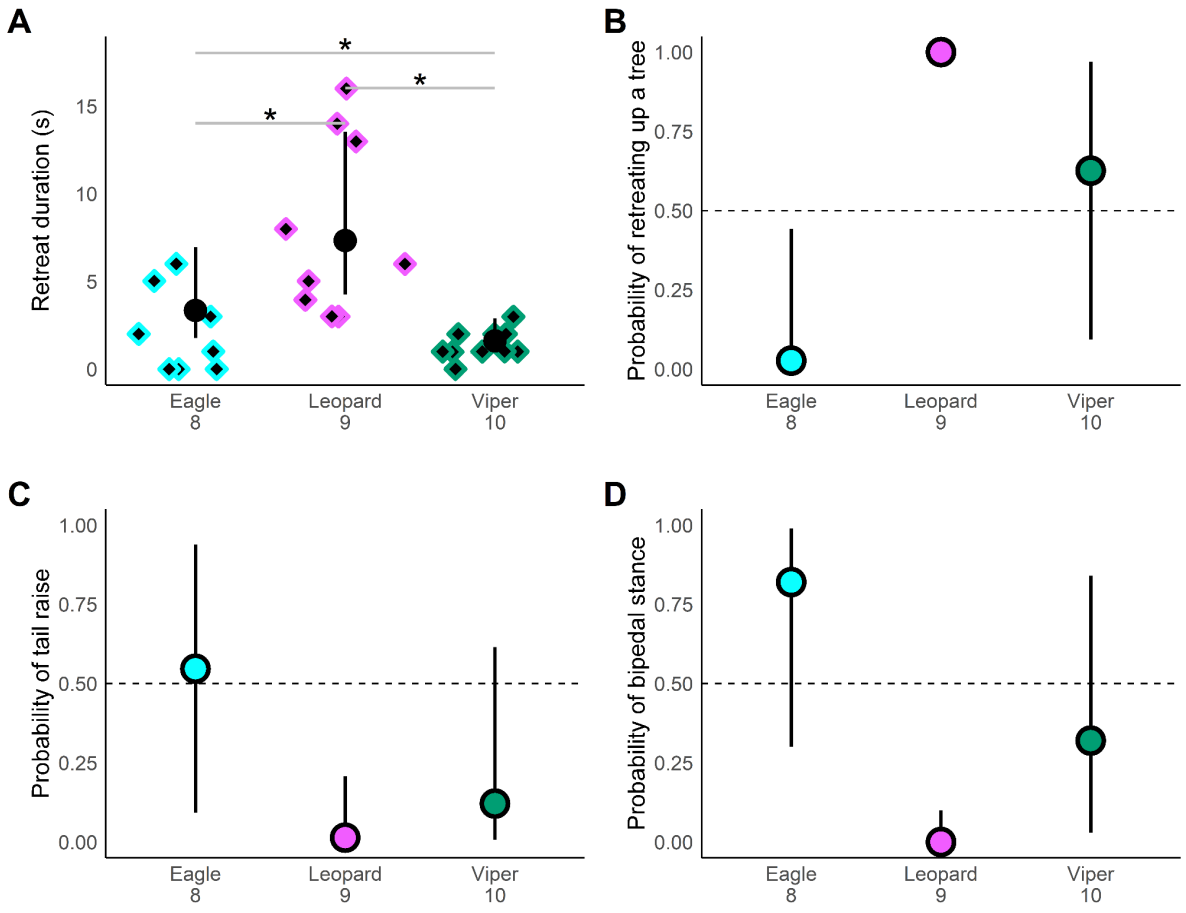


Figure 1. Most common initial reactions to the different predator types during a period of 30 s after model detection. **(a)** Probability of retreating up a tree. **(b)** Retreat duration. **(c)** Probability of bipedal stance. **(d)** Probability of tail raise. Indicated below predator types are the number of individuals (one individual = one trial). The bars represent the credible interval at 95% and the point the mean posterior distribution of the model.

Probability of retreating up a tree

We found strong support that the probability of retreating up a tree was high in the leopard condition (mean probability and 95% CI = 1, 95%-CI[0.99–1.00]), with almost all individuals (8/9) ending up more than three metres from the ground (Figure 1b and Table 4). We also found consistent support for the eagle condition for not retreating up a tree (mean probability and 95% CI = 0.03, 95%-CI[0.0002–0.44]). Amongst the individuals that did retreat up, one ended up below three metres and two above three metres (Figure 1b and Table 4). In contrast, the viper condition did not show a significant preference for either retreating up a tree or not (mean probability and 95%

CI = 0.63, 95%-CI[0.09–0.97]). Whenever individuals did retreat up (*N* = 4), they always ended up less than three metres from the ground (Figure 1b and Table 4).

Table 4. Retreat direction and final canopy position. At the start of the experiment, individuals were initially positioned either on the ground (N = 17 trials) or between 50 cm and 3 m above the ground (N = 10 trials). As a result, not all individuals could retreat downwards, and the retreat direction was constrained by the starting position. The table shows the number of individuals retreating in each direction (down, straight, or up) per predator condition. In grey, we indicate the final canopy position for individuals who retreated upward: between 0.05 and 3 m or above 3 m.

Condition (N trials with retreat/N total trials)	**Retreat down**	**Retreat straight**	**Retreat up**	**0.05–3 m**	**>3 m**
Eagle (5/8)	2	0	3	1	2
Leopard (9/9)	0	0	9	1	8
Viper (9/10)	0	5	4	4	0

Probability of tail raise

The probability of raising the tail was very low in the leopard condition (mean = 0.01; 95% CI: [0.00, 0.21]), and higher, though variable, in the eagle (mean = 0.55; 95% CI: [0.09, 0.94]) and viper (mean = 0.12; 95% CI: [0.01, 0.62]) conditions (Figure 2c).

Probability of bipedal stance

The probability of bipedal stance was very low in the leopard condition (mean = 0.00; 95% CI: [0.00, 0.10]), but higher and more variable in the eagle (mean = 0.82; 95% CI: [0.30, 0.99]) and viper (mean = 0.32; 95% CI: [0.03, 0.84]) conditions (Figure 1d). See Table S5 for details on model summaries.

Vocal responses

Of the *N* = 27 trials, *N* = 23 contained vocal responses from the focal individual that started during the first 30 seconds following model detection. Of the four remaining trials not used in the analyses of vocal responses, in *N* = 2 viper trials individuals did not call after detecting the models and over *N* = 1 eagle trial and *N* = 1 leopard trial, the same individual, an adult male, waited at least 2 minutes after model detection before calling.

We recorded a total of *N* = 378 'shrill' elements from *N* = 16 individuals, *N* = 8 'shrill2' elements (i.e., long 'shrill' variants produced by males) from *N* = 4 individuals, *N* = 75 'hoo' elements from *N* = 10 individuals and *N* = 9 'growl' elements from *N* = 2 individuals.

At the utterance level (i.e., after grouping the same vocal elements within call types), 'shrill' calls could be produced either singly (*N* = 95 single 'shrill' calls from *N* = 16 individuals across 23 trials) or in combination with 'hoo' calls (*N* = 63 'shrill_hoo' from *N* = 10 individuals) (Table 5). In contrast, 'hoo' calls were almost never produced alone (one occurrence only), and in combination after 'shrill' or 'shrill2' calls (see Table S4 for details on utterance type per condition and individual age and sex).

Table 5. Utterance type recorded during predator model presentation experiments. Note that utterances can be either a single call or a combination.

Utterance type	**N utterance**	**N individual**
shrill	95	16
shrill_hoo	63	10
shrill_hoo_growl	1	1
shrill_hoo_shrill_hoo	1	1
shrill2_hoo	5	2
shrill2_hoo_growl	1	1

shrill2_shrill_hoo	1	1
hoo	1	1
growl	1	1
growl_shrill2_shrill_hoo_growl	1	1

Acoustic analyses

pDFA-1 confirmed that our set of acoustic parameters could reliability classify sooty mangabey element types, with 'shrill' and 'hoo' elements (i.e., vocal sounds with a clear beginning and end constituting each call) being significantly set apart (*N* = 374 'shrills' elements from *N* = 16 individuals, *N* = 75 'hoos' elements, from *N* = 10 individuals, p = 0.001, correctly cross-classified = 95%, expected = 55%) (Figure 3).

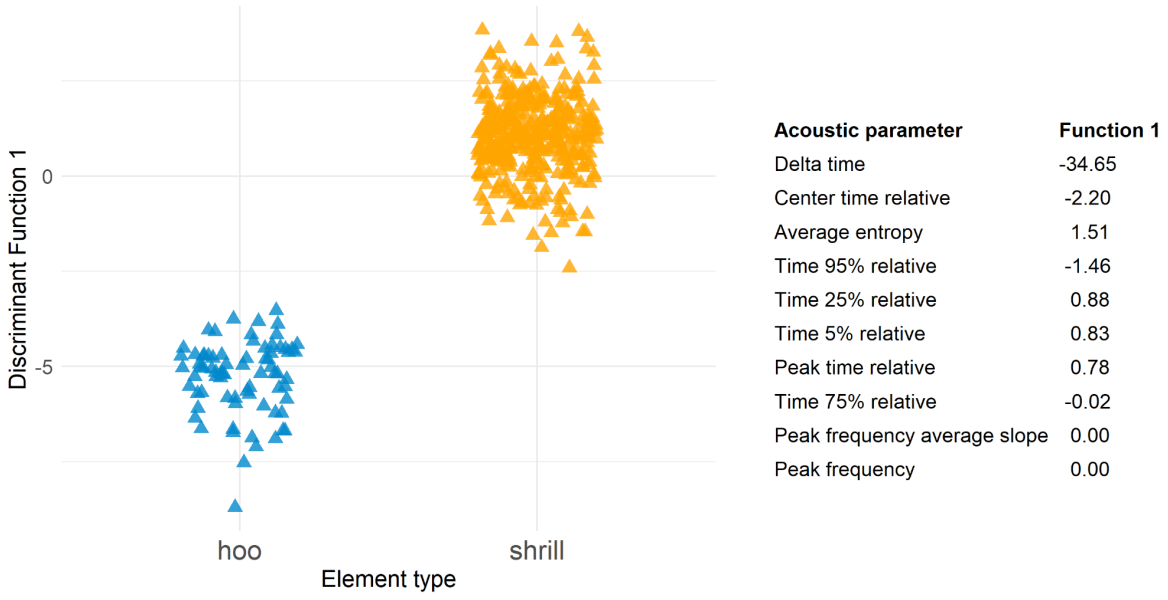
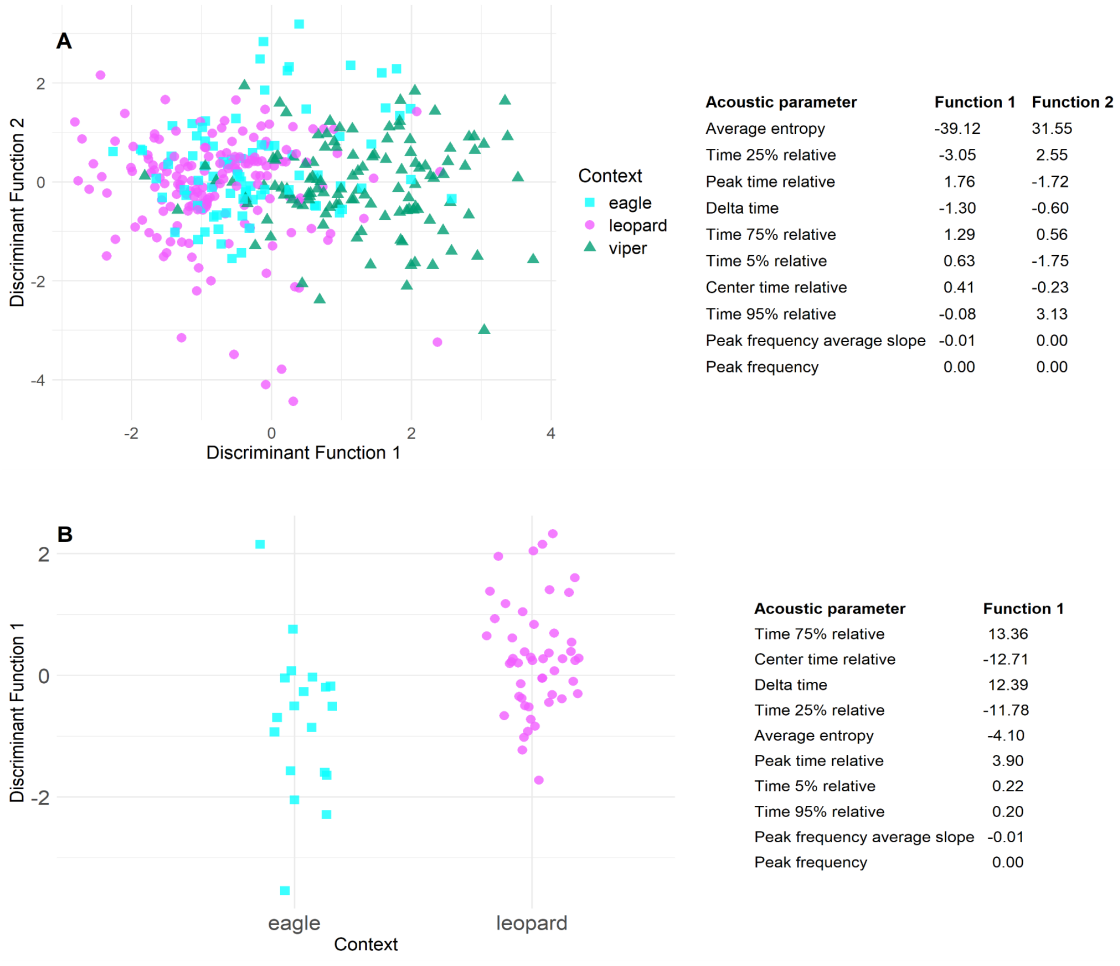


Figure 3. 'Shrill' and 'hoo' call elements plotted along the first discriminant function from a linear discriminant analysis (LDA) of acoustic parameters used in the pDFA-1. This plot visualises how the call types are separated based on their acoustic features. The table on the right displays the contribution of each acoustic parameter to the discriminant function.

We found no statistically significant differences in the acoustic structure of 'shrill' elements across predator types, though the classification accuracy exceeded chance levels by 6%, suggesting a weak trend (pDFA2: *N* = 82 'shrills' from *N* = 7 individuals for eagle; *N* = 172 'shrills' from *N* = 8 individuals for leopard; *N* = 120 'shrills' from *N* = 8 individuals for viper; 65% correctly classified vs. 59% expected, p = 0.06) (Figure 4a).

Regarding 'hoo' elements, we did not find evidence for acoustic variations between the eagle and the leopard conditions (pDFA3: *N* = 19 'hoos' from *N* = 6 individuals for eagle; *N* = 49 'hoos' from *N* = 8 individuals for leopard; p = 0.67, correctly cross-classified = 50%, expected = 55%) (Figure 4b).



Acoustic parameter	Function 1	Function 2
Average entropy	-39.12	31.55
Time 25% relative	-3.05	2.55
Peak time relative	1.76	-1.72
Delta time	-1.30	-0.60
Time 75% relative	1.29	0.56
Time 5% relative	0.63	-1.75
Center time relative	0.41	-0.23
Time 95% relative	-0.08	3.13
Peak frequency average slope	-0.01	0.00
Peak frequency	0.00	0.00

Acoustic parameter	Function 1
Time 75% relative	13.36
Center time relative	-12.71
Delta time	12.39
Time 25% relative	-11.78
Average entropy	-4.10
Peak time relative	3.90
Time 5% relative	0.22
Time 95% relative	0.20
Peak frequency average slope	-0.01
Peak frequency	0.00

Figure 4. (a) 'shrill' elements plotted along the first two discriminant functions from a linear discriminant analysis (LDA) of acoustic parameters used in pDFA-2; **(b)** 'hoo' elements plotted along the first discriminant function from an LDA of acoustic parameters used in pDFA-3. The table on the right displays the contribution of each acoustic parameter to the discriminant functions.

Number of ‘shrill’ elements per ‘shrill’ calls

The number of ‘shrill’ elements within ‘shrill’ calls varied greatly across condition, sex and age class (Table 6 and Table S3 for details). However, we noticed that in the viper condition, ‘shrill’ calls tended to include either one or two elements compared to the eagle and leopard conditions, where ‘shrill’ calls appeared to carry two or more elements.

Table 6. Number of ‘shrill’ elements per ‘shrill’ calls and condition. “3-6” indicates between three and six.

Condition	N elements per call	N calls	N individuals
Eagle	2	13	4
	3–6	16	4
Leopard	1	2	2
	2	21	6
	3–6	39	8
Viper	1	22	4
	2	41	7
	3–6	6	1

Probability of producing ‘hoo’ calls in combination – first 30 seconds

For eagle, the mean probability of producing utterances without a ‘hoo’ call was 0.59 (95% CI [0.36–0.79]), while the probability of producing a ‘hoo’-combined utterance was 0.41 (95% CI [0.21–0.64]) (Figures 5 and 7a). In the leopard condition, the probability of producing ‘hoo’-combined utterances appeared higher (mean = 0.69, 95% CI [0.47–0.85]) than that of utterances without a ‘hoo’ (mean = 0.31, 95% CI [0.15–0.53]), although the credible interval slightly overlapped 0.5, suggesting this pattern should be interpreted with caution (Figures 5 and 7a). In contrast, in the viper condition, utterances were clearly less likely to include a ‘hoo’ (mean = 0.08, 95% CI [0.03–0.21]) than to exclude it (mean = 0.92, 95% CI [0.79–0.97]) (Figures 5 and 7a).

Our model also indicated that males were more likely than females to produce 'hoo'-combined utterances (Estimate = 1.14, 95% CI [0.14–2.17]) (Table S6).

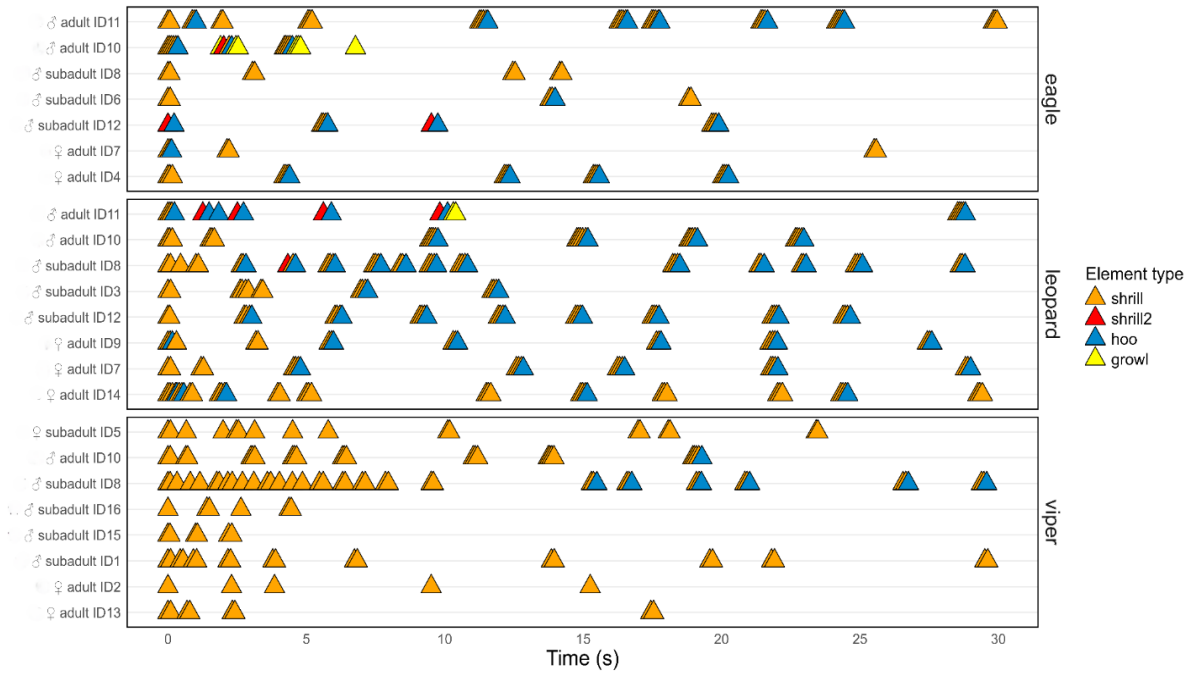


Figure 5. Vocal responses per focal individual and condition over 30 seconds following the onset of the first vocal element. Each line represents the vocal response produced during a single trial corresponding to a specific focal individual. Indicated to the left of the y-axis are the sex, age class and identity label of each individual.

Probability of producing 'hoo' calls in combination – first 5 seconds

In the eagle condition, the likelihood of producing utterances without a 'hoo' call was 0.56 (95% CI [0.22–0.86]), and those including a 'hoo' combined were at 0.44 (95% CI [0.14–0.78]), reflecting a similar pattern to what was seen across the whole 30-second period (Figures 6 and 7b). For the leopard condition, probabilities for utterances without a 'hoo' (mean = 0.51, 95% CI [0.19–0.82]) and with a 'hoo' combined (mean = 0.49, 95% CI [0.19–0.82]) were roughly balanced, which differs from the 30-second data where 'hoo'-combined utterances were somewhat more frequent (Figures 6 and 7b). In the viper condition, the production of utterances without a 'hoo' call was nearly certain (mean = 1.00, 95% CI [0.93–1.00]), and 'hoo'-combined utterances were almost completely absent (mean = 0.00, 95% CI [0.00–0.07]), consistent with the 30-second

results (Figures 6 and 7b). Unlike the findings from the longer time window, no sex differences were observed in the first 5 seconds (Table S6).

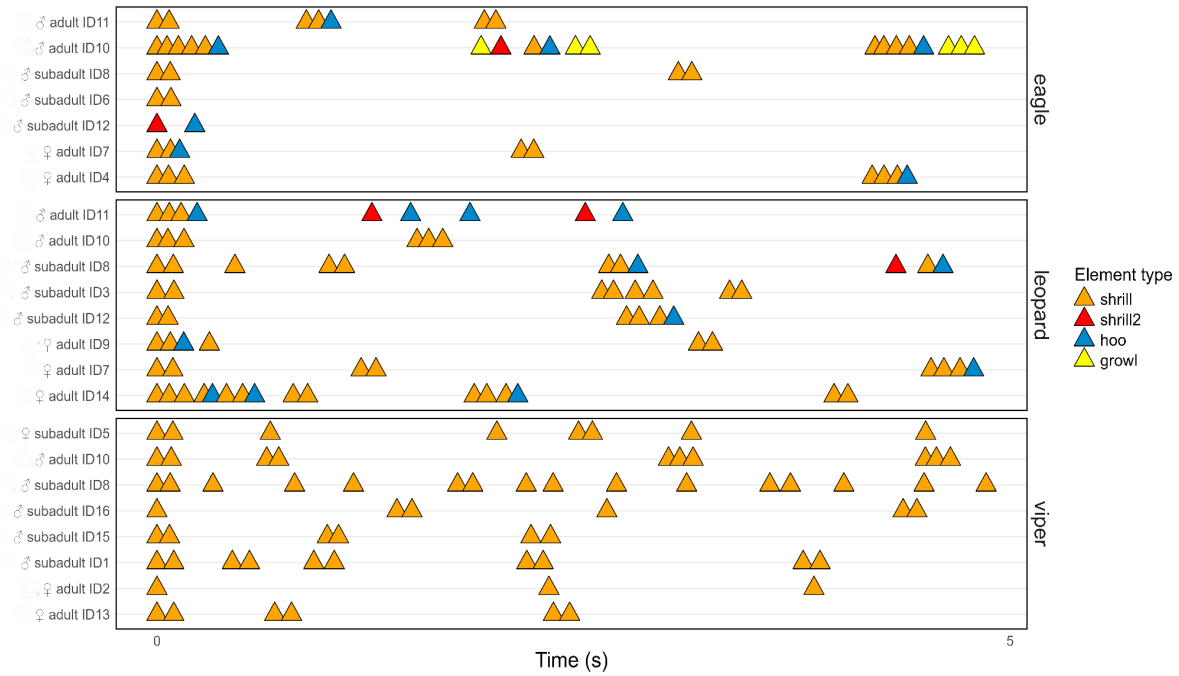


Figure 6. Vocal responses per focal individual and condition over 5 seconds following the onset of the first vocal element. Each line represents the vocal response produced during a single trial corresponding to a specific focal individual. Indicated to the left of the y-axis are the sex, age class and identity label of each individual.

Rate of vocal utterance production – 30 seconds

The rate of utterance production, defined as the number of vocal utterances per second (calculated over 30 seconds following the production of the first vocal element), was lower in the eagle condition (mean posterior distribution = 0.14, 95% CI [0.08–0.25]), and the leopard and viper conditions had similar rates (leopard: mean posterior distribution = 0.25, 95% CI [0.16–0.44]; viper: mean posterior distribution = 0.25, 95% CI [0.15–0.46]) (Figure. 7c). Pairwise comparisons indicated a difference between the eagle and leopard conditions (estimate = -0.61, 95% CI [-1.22, -0.03]), a marginal difference between eagle and viper (estimate = -0.58, 95% CI [-1.20, 0.07]), and no difference between leopard and viper (estimate = 0.03, 95% CI [-0.58, 0.61]).

Rate of vocal utterance production – 5 seconds

The rate of utterance production over the first five seconds also showed that eagle condition had the lowest rate (mean posterior distribution = 0.14, 95% CI [0.08–0.25]), while the leopard and viper conditions showed similar rates (leopard: mean posterior distribution = 0.26, 95% CI [0.17–0.44]; viper: mean posterior distribution = 0.25, 95% CI [0.15–0.46]) (Figure 7d). Pairwise comparisons showed a difference between the eagle and leopard conditions (estimate = -0.61, 95% CI [-1.22, -0.03]), some indication of a difference between eagle and viper (estimate = -0.58, 95% CI [-1.20, 0.07]), and no detectable difference between the leopard and viper conditions (estimate = 0.03, 95% CI [-0.58, 0.61]). See Table S6 for full model summaries.

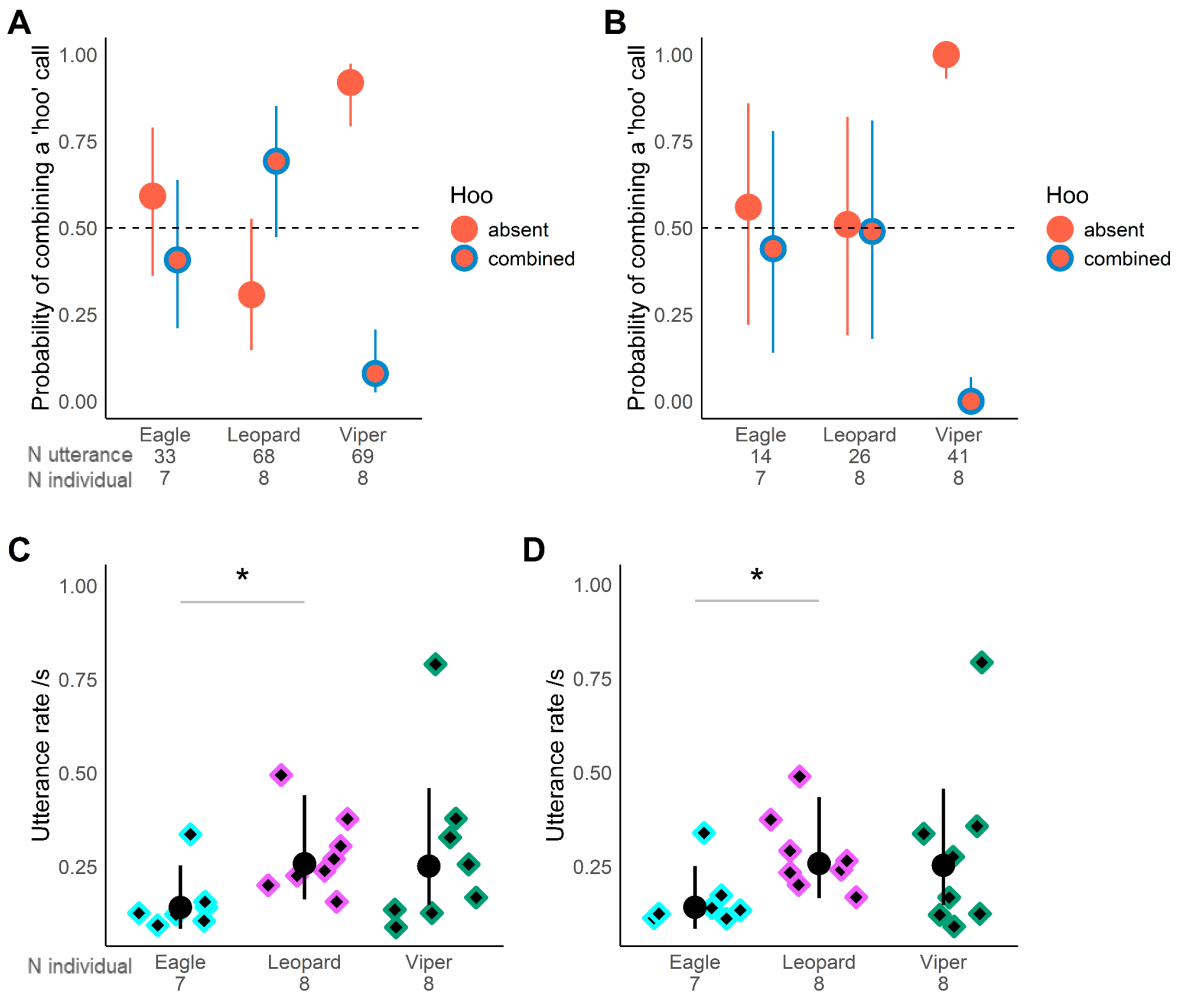


Figure 7. Vocal utterance production depending on predator type. **(a)** Probability of producing a vocal utterance with 'hoo' calls combined or absent over 30 seconds following the onset of vocal response. **(b)** Probability of producing a vocal utterance

with ‘hoo’ calls combined or absent over 5 seconds following the onset of vocal response. **(c)** Vocal utterance rate over 30 seconds following the onset of vocal response. **(d)** Vocal utterance rate over 5 seconds following the onset of vocal response.

Discussion

In our study, we investigated the initial behavioural and vocal responses of sooty mangabeys to three common predatory threats: eagles, leopards and vipers. Following experimental presentations of predator models, individuals showed predator-specific behaviours: viper models typically triggered startle responses with brief retreats on the ground or a few metres up in tress, eagle models often led to stationary responses or short retreats and rarely tree climbing, while leopard models consistently elicited rapid climbing high into the canopy, with responses lasting longer than in the other conditions. Evidence for predator-specific vocal behaviour was more mixed. Sooty mangabeys primarily produced ‘shrill’ calls across conditions. While we did not find any clear patterns regarding acoustic structure or number of elements, we were able to identify an affixation rule: subjects were more likely to affix ‘hoo’ calls after ‘shrill’ calls in response to eagle and leopard models than to vipers. In addition, utterance rate differed across conditions, with higher rates observed in response to leopards than to eagles.

Behavioural responses

Upon detecting viper models, sooty mangabeys typically displayed a startle reaction (i.e., brief retreats lasting less than 2 seconds) in most trials (90%; Table 3 and Figure 1a). These occurred either on the ground or in the lower canopy, below three metres (Table 4). Bipedal stances and tail raises were each observed in nearly half of the trials (40%; Table 3). In eagle trials, bipedal stances were observed in half of the trials (50%), often accompanied by visual scanning of the model positioned approximately two metres above ground. Tail raises occurred in most trials (75%; Table 3). When present, retreats were typically brief and not directed into the trees (Figure 1a and b, and Table 4). In response to leopard models, individuals consistently retreated into the canopy, usually to heights above three metres. These retreats lasted longer than those

observed in the other conditions (Figure 1a and b and Table 4). Bipedal stances and tail raises were rare in this condition (Figure 1c and d and Table 3).

The frequent occurrence of bipedal stances and tail raises in both viper and eagle trials may reflect the use of visual inspection or signalling behaviours. Bipedal stances likely allow individuals to scan their environment (Enstam & Isbell, 2002) while tail raises might serve as social signalling. Observations from newly formed captive groups report that resident adult males sometimes raised their tails in an arched position and yawned at newly introduced individuals (Bernstein, 1971), a behaviour we also observed twice by adult males in response to eagle models in our study, suggesting that these may serve a display function. Alternatively, tail raises could function to communicate information to conspecifics through a non-vocal channel, similar to the use of tail positioning described in vervet monkeys across different social contexts (Bernstein et al., 1978). Bipedal stances and tail raises were extremely rare in the initial responses of sooty mangabeys to leopards, suggesting that inspecting or displaying in response to a leopard is not an appropriate survival strategy.

The behaviour that occurred across all three conditions and showed predator-specific variation was retreat. As noted, viper encounters typically triggered startle retreats, eagle encounters involved either no retreat or short ones, with a low likelihood of climbing into the trees, and leopard encounters consistently elicited retreats into the canopy most often to heights above three metres, that lasted longer than those observed in the other two conditions (Figure 1a and b and Table 4). Overall, these results align with previous findings on sooty mangabey predator-specific anti-predator behaviours, documented through both experimental tests and natural observations (León, Quintero, et al., 2023; León, Thiriau, et al., 2023; Mielke et al., 2019; Quintero et al., 2022a; Range & Fischer, 2004), and support the validity of our predator models, which elicited ecologically appropriate and specific responses.

Vocal responses

Sooty mangabeys produced a vocal response within 30 seconds of detecting the predator model in most of the selected trials. This included 88% of eagle trials and 89% of leopard trials; in both cases, the only individual who did not respond promptly was the same adult male, who vocalised more than two minutes after detection. In contrast,

during viper trials, 80% of individuals responded vocally, while no vocalisations were recorded in the remaining trials (see Table S7).

Across conditions, the most frequent call types were ‘shrill’ and ‘hoo’, forming the core of the alarm repertoire. Males produced two additional call types, ‘shrill2’ and ‘growl’, exclusively during eagle and leopard conditions. ‘Shrill2’ calls were never used in isolation but occurred in combinations, contrary to ‘growls’, which were produced both alone and combined (Tables 5 and S4). Given that ‘growls’ are typically associated with aggression (Quintero et al., 2022b; Range & Fischer, 2004), their use here may reflect a sex-specific mobbing strategy.

Focusing on the ‘shrill’ and ‘hoo’ elements, our acoustic parameters reliably distinguished between the two call types. However, we found no consistent acoustic differences in ‘shrill’ variants across predator contexts, although a weak trend suggested some variation (Figure 4a). Similarly, ‘hoo’ elements failed to be differentiated between the eagle and leopard conditions. Analysis of the number of elements per ‘shrill’ call was not possible due to high variability related to age class and sex (Table S3). However, single-element ‘shrill’ calls were observed predominantly in the viper condition, in contrast to eagle and leopard trials, where they were rare (Table 6). This pattern, characterised by the use of a single call type with little to no evidence for acoustic variation and variation in element number across predator types, broadly aligns with alarm systems reported in other Papionini species (see Table 1).

Regarding the combination with ‘hoo’ calls, we found that appending a ‘hoo’, usually after a ‘shrill’ or a ‘shrill2’, was much more likely in the eagle and leopard conditions compared to the viper condition (Figures 5, 6, 7a and b). Interestingly, ‘hoo’ calls were almost exclusively produced in combination (except for one single ‘hoo’ produced by an adult male in response to a leopard model). Although calls restricted to sequences, like ‘hoo’, are widespread across species (Bhat et al., 2022; Bouchet et al., 2010; Candiotti et al., 2012; Freeberg, 2008; Girard-Buttoz et al., 2022; Luís et al., 2019; Riesch et al., 2008; Zhang et al., 2019), one particular case of combination appears especially comparable. Campbell’s monkeys also use a similar short, low-frequency ‘hoo’ affix-like call, indicating a less urgent threat, by appending it to their leopard alarm ‘krak’ forming ‘krak_hoo’ to signal a more general ground disturbance, and to their eagle alarm ‘hok’ forming ‘hok_hoo’ to indicate a general aerial disturbance (Coye et al., 2015; Ouattara, Lemasson, & Zuberbühler, 2009). In sooty mangabeys, however, ‘hoo’ calls appear to be produced in response to eagles and leopards, but not vipers

even though the former poses a much greater danger. Vipers, in contrast, typically do not prey on sooty mangabeys; the main risk they present is the possibility of being accidentally stepped on (Foerster, 2008; Penner et al., 2008). Thus, compared to Campbell's monkeys, sooty mangabeys might use the 'hoo' affixation to signal a more serious threat, like a hunting predator.

The fact that sooty mangabeys also frequently produce single 'shrill' calls in response to eagles and leopards, in addition to vipers, raises questions about their function (Figures 5, 6, 7a and b). One possibility is that single 'shrill' calls are used to signal the detection of a danger. An alternative, non-mutually exclusive explanation is that 'hoo' calls might be given when individuals no longer have visual contact with the predator. While single 'shrill' calls could indicate "there is something dangerous here, and I have eyes on it," 'shrill_hoo' combinations might signal "watch out, there is a serious threat nearby, but I do not have visual contact." This interpretation would be consistent with the two cases in which individuals produced 'shrill_hoo' utterances towards the viper models. In both instances, the callers startled until reaching a distance of more than one metre from the snake and may have been unable to locate it afterwards, given how well-concealed vipers are. Similar systems have been described in Guinea baboons, signalling the detection of a threat using single 'barks' and predator movements with double-phased 'barks' (Byrne, 1981), and suggested in black-and-white colobus 'snort' calls that could signal the location of a predator (Schel et al., 2009).

The only reliable indicator that we found that discriminates between an eagle and a leopard response is the call rate (Figure 7c and d), which is similar to what has been found in Campbell's monkeys to indicate urgency and in closely related mandrill responses to leopard and eagle models (Lemasson et al., 2010; Yorzinski & Vehrencamp, 2008). However, this might be more related to the perceived level of threat rather than indicating its type. Sooty mangabeys are more terrestrial than other Taï Forest monkeys (McGraw et al., 2007), which may shape how different predators are perceived as threats. Although crowned eagles account for a greater proportion of observed mangabey deaths (13 % versus 6 % for leopards) (Shultz et al., 2004), a perched eagle, once detected, requires a suitable launch position and a clear line of sight to attack. By contrast, a leopard positioned 15 metres away on the forest floor may represent a more immediate and unpredictable danger for a ground-dwelling species. In such cases, the acoustic structure of sooty mangabey 'shrill' calls and their

combinations could not convey information about the type of predator, and receivers would have to rely on contextual cues to assess the situation. For instance, seeing an individual producing single ‘shrill’ calls while standing bipedally and looking at the ground likely indicates a viper. Observing an individual on the ground producing single ‘shrill’ calls or ‘shrill_hoo’ combinations at a low rate while scanning the surroundings, possibly with a bipedal posture, would suggest the presence of an eagle. Seeing an individual running high in the trees while producing ‘shrill_hoo’ combinations at a high rate would point to a leopard.

Limitations

Overall, the lack of predator-specific calls in sooty mangabeys challenges previous assumptions (León, Thiriau, et al., 2023; Range & Fischer, 2004). To confirm that ‘shrill’ calls are not about predator specificity but instead reflect predator detection, location, and/or perceived danger levels, future research should examine caller behaviours at the time of vocalisation (e.g., whether the caller has visual contact with the threat). This is particularly important for interpreting the function of ‘hoo’ affixation, as our study did not analyse caller behaviour at the moment of vocal production. Furthermore, while our regression models revealed condition-dependent differences in behaviour and vocal production, they were based on a relatively small and uneven number of observations per individual. Although we controlled for sex and age class in all models, the limited data prevented us from including individual identity as a random effect. As a result, we interpret our findings as reflecting general patterns, while acknowledging potential individual variation that could not be formally modelled. Playback experiments would also be necessary to determine whether receivers can extract predator-specific information from ‘shrill’ elements produced in response to different threats, by testing whether receivers respond as if the corresponding predator were present. This will be the focus of a forthcoming study.

Conclusion

Sooty mangabeys have an alarm system with a limited capacity to convey the type of predator, which contrasts with sympatric monkey species in the Taï Forest. Instead, it closely resembles alarm systems of the Papionini tribe, characterised by acoustic

variations of one alarm call type (Table 1). We also found a call affixation mechanism that is likely to operate as a semantic modifier, similar to what has been described in sympatric Campbell's monkeys (Ouattara, Lemasson, & Zuberbühler, 2009). Additionally, call rate varied across conditions, with higher rates observed in response to leopards compared to eagles, suggesting that rate may encode information about threat level. This raises the question of what factors drove the structure of the sooty mangabey alarm system. All Papionini, including sooty mangabeys, are primarily terrestrial, while all non-Papionini in the Taï Forest are strictly arboreal (Table 1). From a ground-based perspective, predators are unlikely to approach from below, and dense vegetation can obstruct monkey abilities to perceive the direction of an attack. In such conditions, selection may favour a generalised alarm call that can be modified (e.g., through affixation) to indicate whether the caller has visual contact with the threat or to signal the level of urgency, rather than relying on distinct calls for different predator types, usually reflecting different direction of attack. Notably, Papionini species, including baboons and macaques, have adapted to both open and closed habitats. Among those restricted to closed forests, such as mandrills and mangabeys, this may have favoured the evolution of combinatorial patterns, though this remains to be further explored.

References

Arnold, K., & Zuberbühler, K. (2006). The alarm-calling system of adult male putty-nosed monkeys, Cercopithecus nictitans martini. *Animal Behaviour*, *72*(3), 643–653. https://doi.org/10.1016/j.anbehav.2005.11.017

Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: Compositional or idiomatic expressions? *Brain and Language*, *120*(3), 303–309. https://doi.org/10.1016/j.bandl.2011.10.001

Bernstein, I. S. (1971). The influence of introductory techniques on the formation of captive mangabey groups. *Primates*, *12*(1), 33–44. https://doi.org/10.1007/BF01730380

Bernstein, P. L., Smith, W. J., Krensky, A., & Rosene, K. (1978). Tail Positions of *Cercopithecus aethiops*. *Zeitschrift Für Tierpsychologie*, *46*(3), 268–278. https://doi.org/10.1111/j.1439-0310.1978.tb01449.x

Berthet, M., Mesbahi, G., Pajot, A., Cäsar, C., Neumann, C., & Zuberbühler, K. (2019). Titi monkeys combine alarm calls to create probabilistic meaning. *Science Advances*, *5*(5), eaav3991. https://doi.org/10.1126/sciadv.aav3991

Bhat, A. S., Sane, V. A., Seshadri, K. S., & Krishnan, A. (2022). Behavioural context shapes vocal sequences in two anuran species with different repertoire sizes. *Animal Behaviour*, *184*, 111–129. https://doi.org/10.1016/j.anbehav.2021.12.004

Bouchet, H., Pellier, A.-S., Blois-Heulin, C., & Lemasson, A. (2010). Sex differences in the vocal repertoire of adult red-capped mangabeys (Cercocebus torquatus): A multi-level acoustic analysis. *American Journal of Primatology*, *72*(4), 360–375. https://doi.org/10.1002/ajp.20791

Bürkner, P.-C. (2015). *brms: Bayesian Regression Models using "Stan"* (p. 2.22.0) [Dataset]. https://doi.org/10.32614/CRAN.package.brms

Byrne, R. W. (1981). Distance Vocalisations of Guinea Baboons (Papio Papio) in Senegal: An Analysis of Function. *Behaviour*, *78*(3–4), 283–312. https://doi.org/10.1163/156853981X00365

Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012). Context-related call combinations in female Diana monkeys. *Animal Cognition*, *15*(3), 327–339. https://doi.org/10.1007/s10071-011-0456-8

Casamitjana, J. (2002). The vocal repertoire of the woolly monkey (Lagothrix lagotricha). *Bioacoustics*, *13*(1), 1–19. https://doi.org/10.1080/09524622.2002.9753483

Collier, K., Radford, A. N., Stoll, S., Watson, S. K., Manser, M. B., Bickel, B., & Townsend, S. W. (2020). Dwarf mongoose alarm calls: Investigating a complex non-human animal call. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1935), 20192514. https://doi.org/10.1098/rspb.2019.2514

Coss, R. G., McCowan, B., & Ramakrishnan, U. (2007). Threat-Related Acoustical Differences in Alarm Calls by Wild Bonnet Macaques (*Macaca radiata*) Elicited by Python and Leopard Models. *Ethology*, *113*(4), 352–367. https://doi.org/10.1111/j.1439-0310.2007.01336.x

Coye, C., Ouattara, K., Zuberbühler, K., & Lemasson, A. (2015). Suffixation influences receivers' behaviour in non-human primates. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1807), 20150265. https://doi.org/10.1098/rspb.2015.0265

Crane, J. M. S., Savage, J. L., & Russell, A. F. (2016). Diversity and function of vocalisations in the cooperatively breeding Chestnut-crowned Babbler. *Emu - Austral Ornithology*, *116*(3), 241–253. https://doi.org/10.1071/MU15048

Cunningham, S., & Magrath, R. D. (2017). Functionally referential alarm calls in noisy miners communicate about predator behaviour. *Animal Behaviour*, *129*, 171–179. https://doi.org/10.1016/j.anbehav.2017.05.021

Dezecache, G., & Berthet, M. (2018). Working hypotheses on the meaning of general alarm calls. *Animal Behaviour*, *142*, 113–118. https://doi.org/10.1016/j.anbehav.2018.06.008

Enstam, K. L., & Isbell, L. A. (2002). Comparison of responses to alarm calls by patas (*Erythrocebus patas*) and vervet (*Cercopithecus aethiops*) monkeys in relation to habitat structure. *American Journal of Physical Anthropology*, *119*(1), 3–14. https://doi.org/10.1002/ajpa.10104

Fichtel, C., & Kappeler, P. M. (2002). Anti-predator behavior of group-living Malagasy primates: Mixed evidence for a referential alarm call system. *Behavioral Ecology and Sociobiology*, *51*(3), 262–275. https://doi.org/10.1007/s00265-001-0436-0

Fischer, J., & Hammerschmidt, K. (2001). Functional referents and acoustic similarity revisited: The case of Barbary macaque alarm calls. *Animal Cognition*, *4*(1), 29–35. https://doi.org/10.1007/s100710100093

Fischer, J., Hammerschmidt, K., Cheney, D. L., & Seyfarth, R. M. (2001). Acoustic Features of Female Chacma Baboon Barks. *Ethology*, *107*(1), 33–54. https://doi.org/10.1111/j.1439-0310.2001.00630.x

Fischer, J., Hammerschmidt, K., Cheney, D. L., & Seyfarth, R. M. (2002). Acoustic features of male baboon loud calls: Influences of context, age, and individuality. *The Journal of the Acoustical Society of America*, *111*(3), 1465–1474. https://doi.org/10.1121/1.1433807

Fischer, J., Hammerschmidt, K., & Todt, D. (2010). Factors Affecting Acoustic Variation in Barbary-macaque (Macaca sylvanus) Disturbance Calls. *Ethology*, *101*(1), 51–66. https://doi.org/10.1111/j.1439-0310.1995.tb00345.x

Fischer, J., Metz, M., Cheney, D. L., & Seyfarth, R. M. (2001). Baboon responses to graded bark variants. *Animal Behaviour*, *61*(5), 925–931. https://doi.org/10.1006/anbe.2000.1687

Foerster, S. (2008). Two incidents of venomous snakebite on juvenile blue and Sykes monkeys (Cercopithecus mitis stuhlmanni and C. m. Albogularis). *Primates*, *49*(4), 300–303. https://doi.org/10.1007/s10329-008-0098-x

Freeberg, T. M. (2008). Complexity in the chick-a-dee call of carolina chickadees (Poecile carolinesis): Associations of context and signlater behavior to call structure. *The Auk*, *125*(4), 896–907. https://doi.org/10.1525/auk.2008.07025

Gallot, Q., Depriester, C., Moran, S., & Zuberbühler, K. (2024). A non-human primate combinatorial system for long-distance communication. *iScience*, *27*(11). https://doi.org/10.1016/j.isci.2024.111172

Gamer, M., Lemon, J., & Ian Fellows Puspendra Singh. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement* (Version R package version 0.84.1) [Computer software]. https://CRAN.R-project.org/package=irr

Girard-Buttoz, C., Zaccarella, E., Bortolato, T., Friederici, A. D., Wittig, R. M., & Crockford, C. (2022). Chimpanzees produce diverse vocal sequences with ordered and recombinatorial properties. *Communications Biology*, *5*(1), 410. https://doi.org/10.1038/s42003-022-03350-8

Greene, E., & Meagher, T. (1998). Red squirrels,Tamiasciurus hudsonicus, produce predator-class specific alarm calls. *Animal Behaviour*, *55*(3), 511–518. https://doi.org/10.1006/anbe.1997.0620

Hammerschmidt, K., & Fischer, J. (2019). Baboon vocal repertoires and the evolution of primate vocal diversity. *Journal of Human Evolution*, *126*, 1–13. https://doi.org/10.1016/j.jhevol.2018.10.010

Horn, A. D. (1987). The socioecology of the black mangabey (*Cercocebus aterrimus*) near Lake Tumba, Zaire. *American Journal of Primatology*, *12*(2), 165–180. https://doi.org/10.1002/ajp.1350120204

K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology. (2024). *Raven Pro: Interactive Sound Analysis Software* (Version 1.6.5) [Computer software]. The Cornell Lab of Ornithology. https://www.ravensoundsoftware.com/

Keenan, S., Mathevon, N., Stevens, J. M. G., Nicolè, F., Zuberbühler, K., Guéry, J.-P., & Levréro, F. (2020). The reliability of individual vocal signature varies across the bonobo's graded repertoire. *Animal Behaviour*, *169*, 9–21. https://doi.org/10.1016/j.anbehav.2020.08.024

Kudo, H. (1987). The study of vocal communication of wild mandrills in Cameroon in relation to their social structure. *Primates*, *28*(3), 289–308. https://doi.org/10.1007/BF02381013

Le Floch, A., Bouchard, A., Gallot, Q., & Zuberbühler, K. (2021). Lesser spot-nosed monkeys coordinate alarm call production with associated Campbell's monkeys. *Behavioral Ecology and Sociobiology*, *75*(8), 112. https://doi.org/10.1007/s00265-021-03053-w

Leger, D. W., Owings, D. H., & Gelfand, D. L. (2010). Single-note Vocalizations of California Ground Squirrels: Graded Signals and Situation-specificity of Predator and Socially Evoked Calls. *Zeitschrift Für Tierpsychologie*, *52*(3), 227–246. https://doi.org/10.1111/j.1439-0310.1980.tb00714.x

Lemasson, A., Ouattara, K., Bouchet, H., & Zuberbühler, K. (2010). Speed of call delivery is related to context and caller identity in Campbell's monkey males. *Naturwissenschaften*, *97*(11), 1023–1027. https://doi.org/10.1007/s00114-010-0715-6

Lenth, R. V. (2017). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (p. 1.10.5) [Dataset]. https://doi.org/10.32614/CRAN.package.emmeans

León, J., Quintero, F., & Zuberbühler, K. (2023). Acquisition of predator knowledge in sooty mangabeys. *Animal Behaviour*, *205*, 1–14. https://doi.org/10.1016/j.anbehav.2023.08.012

León, J., Thiriau, C., Bodin, C., Crockford, C., & Zuberbühler, K. (2022). Acquisition of predator knowledge from alarm calls via one-trial social learning in monkeys. *iScience*, *25*(9), 104853. https://doi.org/10.1016/j.isci.2022.104853

León, J., Thiriau, C., Crockford, C., & Zuberbühler, K. (2023). Comprehension of own and other species' alarm calls in sooty mangabey vocal development. *Behavioral Ecology and Sociobiology*, *77*(5), 56. https://doi.org/10.1007/s00265-023-03318-6

Leroux, M., Bosshard, A. B., Chandia, B., Manser, A., Zuberbühler, K., & Townsend, S. W. (2021). Chimpanzees combine pant hoots with food calls into larger structures. *Animal Behaviour*, *179*, 41–50. https://doi.org/10.1016/j.anbehav.2021.06.026

Leroux, M., & Townsend, S. W. (2020). Call combinations in great apes and the evolution of syntax. *Animal Behavior and Cognition*, *7*(2), 131–139. https://doi.org/10.26451/abc.07.02.07.2020

Luís, A. R., Alves, I. S., Sobreira, F. V., Couchinho, M. N., & Dos Santos, M. E. (2019). Brays and bits: Information theory applied to acoustic communication sequences of bottlenose dolphins. *Bioacoustics*, *28*(3), 286–296. https://doi.org/10.1080/09524622.2018.1443285

Macedonia, J. M., & Evans, C. S. (2010). Essay on Contemporary Issues in Ethology: Variation among Mammalian Alarm Call Systems and the Problem of Meaning in Animal Signals. *Ethology*, *93*(3), 177–197. https://doi.org/10.1111/j.1439-0310.1993.tb00988.x

Manser, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 11.

Marler, P., Pickert, R., & Gyger, M. (1987). Semantics of an Avian Alarm Call System: The Male Domestic Fowl, Gallus Domesticus. *Behaviour*, *102*(1–2), 15–39. https://doi.org/10.1163/156853986X00027

McGraw, W. S., Zuberbuhler, K., & Noë, R. (2007). *Monkeys of the Taï forest: An African primate community* (Vol. 51). Cambridge University Press.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 276–282. https://doi.org/10.11613/BM.2012.031

Mielke, A., Crockford, C., & Wittig, R. M. (2019). Snake alarm calls as a public good in sooty mangabeys. *Animal Behaviour*, *158*, 201–209. https://doi.org/10.1016/j.anbehav.2019.10.001

Min, Y., & Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling*, *5*(1), 1–19. https://doi.org/10.1191/1471082X05st084oa

Mundry, R., & Sommer, C. (2007). Discriminant function analysis with nonindependent data: Consequences and an alternative. *Animal Behaviour*, *74*(4), 965–976. https://doi.org/10.1016/j.anbehav.2006.12.028

Murphy, D., Lea, S. E. G., & Zuberbühler, K. (2013). Male blue monkey alarm calls encode predator type and distance. *Animal Behaviour*, *85*(1), 119–125. https://doi.org/10.1016/j.anbehav.2012.10.015

Narbona Sabaté, L., Mesbahi, G., Dezecache, G., Cäsar, C., Zuberbühler, K., & Berthet, M. (2022). Animal linguistics in the making: The Urgency Principle and titi monkeys' alarm system. *Ethology Ecology & Evolution*, *34*(3), 378–394. https://doi.org/10.1080/03949370.2021.2015452

Ouattara, K., Lemasson, A., & Zuberbuhler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, *106*(51), 22026–22031. https://doi.org/10.1073/pnas.0908118106

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's Monkeys Use Affixation to Alter Call Meaning. *PLoS ONE*, *4*(11), e7808. https://doi.org/10.1371/journal.pone.0007808

Paddock, C. L., Silva, M. J. F. D., McCabe, G. M., Fernández, D., McGraw, W. S., & Bruford, M. W. (2024). *On the phylogenetic history of the Sanje mangabey (Cercocebus sanjei)*. In Review. https://doi.org/10.21203/rs.3.rs-5484999/v1

Penner, J., Fruteau, C., Range, F., & Rödel, M.-O. (2008). Finding a Needle in a Haystack: New Methods of Locating and Working with Rhinoceros Vipers (Bitis rhinoceros). *Herpetological Review*, *39*(3), 310–314.

Pozzi, L., Hodgson, J. A., Burrell, A. S., Sterner, K. N., Raaum, R. L., & Disotell, T. R. (2014). Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution*, *75*, 165–183. https://doi.org/10.1016/j.ympev.2014.02.023

Price, T., Arnold, K., Zuberbühler, K., & Semple, S. (2009). Pyow but not hack calls of the male putty-nosed monkey (Cercopithcus nictitans) convey information about caller identity. *Behaviour*, *146*(7), 871–888. https://doi.org/10.1163/156853908X396610

Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., & Fischer, J. (2015). Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Scientific Reports*, *5*(1), 13220. https://doi.org/10.1038/srep13220

Quintero, F., Touitou, S., Magris, M., & Zuberbühler, K. (2022a). An Audience Effect in Sooty Mangabey Alarm Calling. *Frontiers in Psychology*, *13*, 816744. https://doi.org/10.3389/fpsyg.2022.816744

Quintero, F., Touitou, S., Magris, M., & Zuberbühler, K. (2022b). *Audience Effects in Sooty Mangabey Agonistic Behaviour* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-1836808/v1

R Core Team. (2021). *R: A language and environment for statistical computing* (Version 2021.09.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Range, F., & Fischer, J. (2004). Vocal Repertoire of Sooty Mangabeys (Cercocebus torquatus atys) in the Tai National Park. *Ethology*, *110*(4), 301–321. https://doi.org/10.1111/j.1439-0310.2004.00973.x

Registry-Migration.Gbif.Org. (2023). *GBIF Backbone Taxonomy* [Dataset]. GBIF Secretariat. https://doi.org/10.15468/39OMEI

Riesch, R., Ford, J. K. B., & Thomsen, F. (2008). Whistle sequences in wild killer whales (*Orcinus orca*). *The Journal of the Acoustical Society of America*, *124*(3), 1822–1829. https://doi.org/10.1121/1.2956467

Robinson, S. R. (2010). Alarm Communication in Belding's Ground Squirrels. *Zeitschrift Für Tierpsychologie*, *56*(2), 150–168. https://doi.org/10.1111/j.1439-0310.1981.tb01293.x

Rowe, N. (1996). *The pictorial guide to the living primates*. Pogonias Press.

Schel, A. M., Tranquilli, S., & Zuberbühler, K. (2009). The alarm call system of two species of black-and-white colobus monkeys (Colobus polykomos and Colobus guereza). *Journal of Comparative Psychology*, *123*(2), 136–150. https://doi.org/10.1037/a0014280

Schlenker, P., Chemla, E., Arnold, K., & Zuberbühler, K. (2016). Pyow-hack revisited: Two analyses of Putty-nosed monkey alarm calls. *Lingua*, *171*, 1–23. https://doi.org/10.1016/j.lingua.2015.10.002

Schlenker, P., Chemla, E., Schel, A. M., Fuller, J., Gautier, J.-P., Kuhn, J., Veselinović, D., Arnold, K., Cäsar, C., Keenan, S., Lemasson, A., Ouattara, K., Ryder, R., & Zuberbühler, K. (2016). Formal monkey linguistics. *Theoretical Linguistics*, *42*(1–2), 1–90. https://doi.org/10.1515/tl-2016-0001

Seddon, N., Tobias, J., & Alvarez, A. (2002). Vocal communication in the pale-winged trumpeter (Psophia leucoptera): Repertoire, context and functional reference. *Behaviour*, *139*(10), 1331–1359. https://doi.org/10.1163/156853902321104190

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, *210*(4471), 801–803.

Shultz, S., Noë, R., McGraw, W. S., & Dunbar, R. I. M. (2004). A community–level evaluation of the impact of prey behavioural and ecological characteristics on predator diet composition. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1540), 725–732. https://doi.org/10.1098/rspb.2003.2626

Sigmundson, R., Girard-Buttoz, C., Le Floch, A., Azaiez, T. S., McElreath, R., Zuberbühler, K., Wittig, R. M., & Crockford, C. (2025). Vocal sequence diversity and length remain stable across ontogeny in a catarrhine monkey (Cercocebus atys). *Communications Biology*, *8*(1), 465. https://doi.org/10.1038/s42003-025-07922-2

Struhsaker, T. T. (1967). Auditory communication among vervet monkeys (Cercopithecus aetthiops). *Social Communication among Primates*.

Suzuki, T. N. (2012). Referential mobbing calls elicit different predator-searching behaviours in Japanese great tits. *Animal Behaviour*, *84*(1), 53–57. https://doi.org/10.1016/j.anbehav.2012.03.030

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (4. ed., [Nachdr.]). Springer.

Wittig, R. M. (2018). Taï Chimpanzees. In J. Vonk & T. Shackelford (Eds.), *Encyclopedia of Animal Cognition and Behavior* (pp. 1–7). Springer International Publishing. https://doi.org/10.1007/978-3-319-47829-6_1564-1

Yorzinski, J. L., & Vehrencamp, S. L. (2008). Preliminary report: Antipredator behaviors of mandrills. *Primate Report*, *75*, 11–18.

Zhang, K., Liu, T., Liu, M., Li, A., Xiao, Y., Metzner, W., & Liu, Y. (2019). Comparing context-dependent call sequences employing machine learning methods: An indication of syntactic structure of greater horseshoe bats. *Journal of Experimental Biology*, jeb.214072. https://doi.org/10.1242/jeb.214072

Zuberbühler, K. (2000). Referential labelling in Diana monkeys. *Animal Behaviour*, *59*(5), 917–927. https://doi.org/10.1006/anbe.1999.1317

Zuberbühler, K. (2001). Predator-specific alarm calls in Campbell's monkeys, Cercopithecus campbelli. *Behavioral Ecology and Sociobiology*, *50*(5), 414–422. https://doi.org/10.1007/s002650100383

Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, *63*(2), 293–299. https://doi.org/10.1006/anbe.2001.1914

Zuberbühler, K. (2003). Natural semanticity in wild primates. In *Animal Complexity. Intelligence, culture, and individualized societies.* (Tyack PL: Harvard University Press, pp. 362–367).

Zuberbühler, K., & Jenny, D. (2002). Leopard predation and primate evolution. *Journal of Human Evolution*, *43*(6), 873–886. https://doi.org/10.1006/jhev.2002.0605

Supplementary material

Document S1. Models of predators

a The three life-sized viper models, painted to replicate patterns of the Gaboon viper *(Bitis gabonica)* or the rhinoceros viper *(Bitis nasicornis).* Models were kindly provided by the Taï Chimpanzee Project.

b The two eagle models replicating the African crowned eagle *(Stephanoaetus coronatus).* 3D printed in plastic fabric at the Comparative Cognition Lab (University of Neuchâtel) by Auriane Le Floch and Quentin Gallot. Modelled by Alwin Durez. Height: 50 cm; Length (from tip of tail to tip of beak measured along the ground): 61.5 cm.

c The leopard cape and three leopard models. From left to right, the 4 leopard models: (1) cape made in leopard fabric (approximately 1.5mx2m), (2) back part of a leopard stuffed toy (approximately 50 cm length without the tail, 38 cm width and 40 cm high), (3) front part of the same leopard stuffed toy as in model 2 (approximately 50 cm length, 38 cm width and 40 cm high at the withers) and (4) full leopard stuffed toy (80 cm length without the tail and approximately 30 cm high at the withers).

d NB is placing the leopard model in the vicinity of sooty mangabeys from behind a tree.

Document S2. Call types

Shrill and hoo (referred to as alarm call in Range and Fischer, 2004)

See below for spectrographic representations by age class and sex.

-‘Shrill’ calls consist of several high-frequency elements, occasionally displaying a harmonic structure, although this is relatively uncommon.

-‘Shrill2’ calls consist of one prolonged ‘shrill’ that extends along the same frequency band. In contrast to ‘shrill’ calls, individual elements cannot be distinguished.

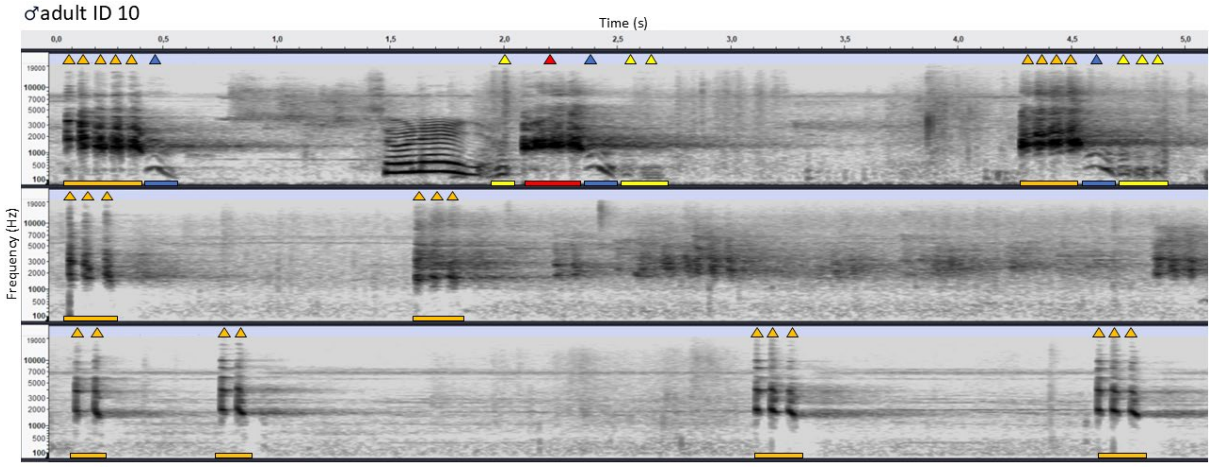
-‘Hoo’ calls features a low fundamental frequency and a rich, readily visible harmonic structure.

Growl

See below for spectrographic representations in adult males.

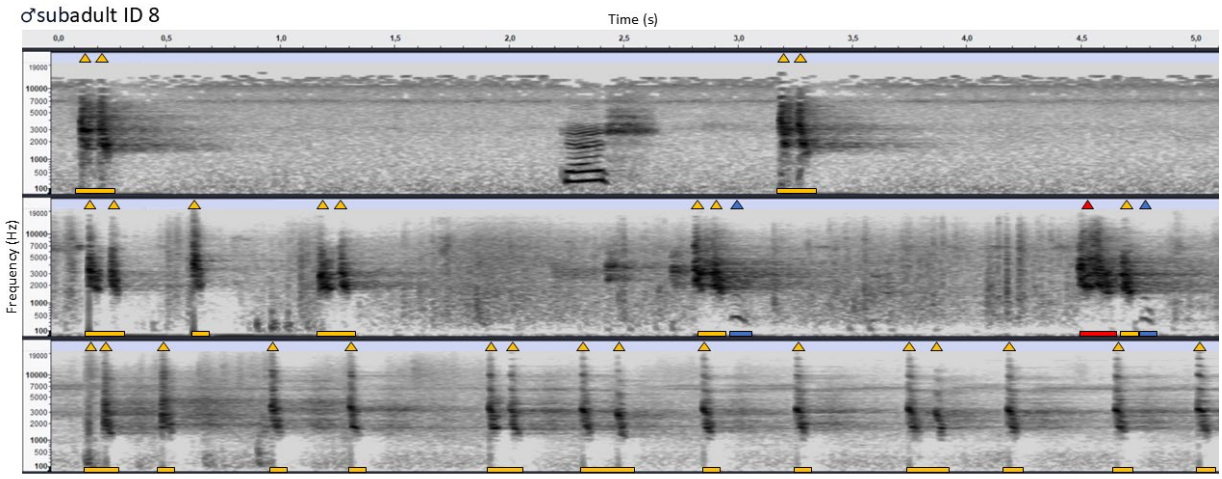
Growls are characterised by a low fundamental frequency band with many visible overtones, and they vary in both frequency and duration, often comprising a combination of shorter and prolonged elements.

Adult male vocal utterances



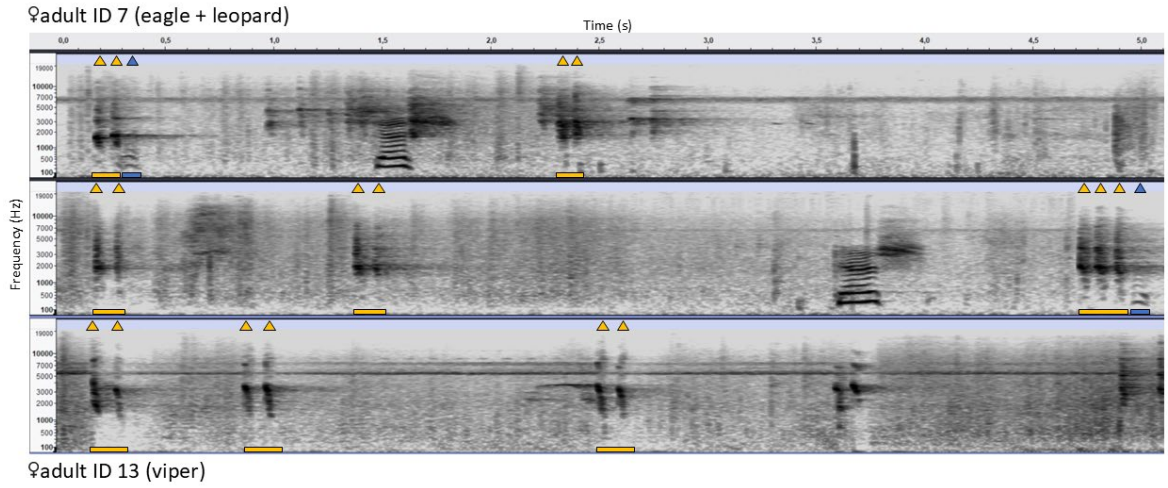
Spectrographic examples of the first 5 seconds of vocal response of the same adult male to eagle (top), leopard (middle) and viper (bottom) models. Each triangle (above the spectrogram) indicates one vocal element, and each bar (below the spectrogram) a call. Colour code: orange = ‘shrill’; blue = ‘hoo’; red = ‘shrill2’; yellow = ‘growl’. The y-axis represents the frequency in Hz, and the x-axis the time in seconds.

Subadult male vocal utterances



Spectrographic examples of the first 5 seconds of vocal response of the same subadult male to eagle (top), leopard (middle) and viper (bottom) models. Each triangle (above the spectrogram) indicates one vocal element, and each bar (below the spectrogram) a call. Colour code: orange = ‘shrill’; blue = ‘hoo’; red = ‘shrill2’. The y-axis represents the frequency in Hz, and the x-axis represents the time in seconds.

Adult female vocal utterances



Spectrographic examples of the first 5 seconds of vocal response of the same adult female to eagle (top) and leopard (middle) models and a different adult female to the viper (bottom) model. Each triangle (above the spectrogram) indicates one vocal element, and each bar (below the spectrogram) a call. Colour code: orange = ‘shrill’; blue = ‘hoo’. The y-axis represents the frequency in Hz, and the x-axis the time in seconds.

Table S1. Acoustic parameters extracted from Raven Pro

Because some parameters can be influenced by manual selection — particularly those related to frequency range, which can be affected by the distance between the vocalising individual and the recording device — we focused exclusively on parameters that are less sensitive to variability in frequency visibility on the spectrogram. Specifically, we selected parameters that primarily describe temporal features and frequency measures unaffected by spectrogram range settings. We provide short definitions for each extracted parameter here. For more detailed definitions, please refer to the Raven Pro manual (reference K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology. Raven Pro: Interactive Sound Analysis Software. The Cornell Lab of Ornithology (2024)).

Parameter	**Definition**	**Unit**	**pDFA shrill vs hoo**	**pDFA shrill vs condition**	**pDFA hoo vs condition**
Aggregate Entropy	The total entropy across the entire selected element, reflecting the overall signal randomness and uncertainty.	Bits	No	No	No
Average Entropy	The mean entropy per unit of time within the selected element, representing the average level of randomness or uncertainty.	Bits	Yes	Yes	Yes
Center Time Relative	Temporal midpoint of energy distribution within the selected	Relative (0-1)	Yes	Yes	Yes

	element, relative to element duration.				
Delta Time	Total duration of the selected element, in seconds.	Seconds (s)	Yes	Yes	Yes
Peak Frequency	Peak frequency of the selected element, corresponding to the frequency of maximum energy.	Hertz (Hz)	Yes	Yes	Yes
Peak Frequency Average Slope	Average slope (rate of change) of peak frequency contour over time, measured in Hz/ms.	Hz/ms	Yes	Yes	Yes
Peak Time Relative	Time point of the peak frequency within the selected element, relative to element duration.	Relative (0-1)	Yes	Yes	Yes
Time 25% Relative	Relative time at which 25% of the energy is accumulated within the element.	Relative (0-1)	Yes	Yes	Yes
Time 5% Relative	Relative time at which 5% of the energy is accumulated within the element.	Relative (0-1)	Yes	No	Yes

Time 75% Relative	Relative time at which 75% of the energy is accumulated within the element.	Relative (0-1)	Yes	Yes	Yes
Time 95% Relative	Relative time at which 95% of the energy is accumulated within the element.	Relative (0-1)	Yes	Yes	Yes

Table S2. Sample size of experimental trials

N trials conducted = number of trials conducted (including five viper trials retrieved from AM (Mielke et al., 2019)). N trials coded = number of trials for which we initially coded the videos. N trials analysed = number of trials used in our analyses.

Condition	N trials conducted	N trials coded	N trials analysed
Eagle	11	10	8
Leopard	11	9	9
Viper	19	14	10

We discarded one eagle, one leopard, and four viper trials because it was not possible to determine which individual first detected the model. We also discarded one leopard because Diana monkeys spotted the model before sooty mangabeys and alarm called. We removed two vipers because the first individual spotting the model was a juvenile. Following video coding, we further excluded two eagle and four viper trials in which focal individual behaviours could not be clearly observed for 30 seconds after model detection. This 30-second window was our criterion for systematically comparing the initial response of the individual detecting the model.

Table S3. Number of ‘shrill’ elements per ‘shrill’ call, condition, age and sex

Condition	Age-sex category	N elements per call	N call	N individual
Eagle	Adult Female	2	3	1
	Adult Female	3	5	1
	Adult Male	2	3	1
	Adult Male	3	3	1
	Adult Male	4	5	2
	Adult Male	5	1	1
	Subadult Male	2	7	2
	Subadult Male	3	2	1
Leopard	Adult Female	1	1	1
	Adult Female	2	11	3
	Adult Female	3	13	3
	Adult Female	6	1	1
	Adult Male	3	3	2
	Adult Male	4	4	2
	Adult Male	5	1	1
	Subadult Male	1	1	1

	Subadult Male	2	10	3
	Subadult Male	3	16	3
	Subadult Male	4	1	1
Viper	Adult Female	1	5	1
	Adult Female	2	4	1
	Adult Male	2	2	1
	Adult Male	3	4	1
	Adult Male	4	2	1
	Subadult Female	1	5	1
	Subadult Female	2	6	1
	Subadult Male	1	12	2
	Subadult Male	2	29	4

Table S4. Utterance type per condition, age and sex

Condition	**Sex**	**Age Class**	**Utterance type**	**N utterance**	**N individual**
Eagle	f	adult	shrill	3	2
Eagle	f	adult	shrill_hoo	5	2
Eagle	m	adult	growl	1	1

Eagle	m	adult	growl_shrill2_shrill_hoo_growl	1	1
Eagle	m	adult	shrill	4	1
Eagle	m	adult	shrill_hoo	7	2
Eagle	m	subadult	shrill	6	2
Eagle	m	subadult	shrill2_hoo	2	1
Eagle	m	subadult	shrill_hoo	3	2
Leopard	f	adult	shrill	11	3
Leopard	f	adult	shrill_hoo	14	3
Leopard	f	adult	shrill_hoo_shrill_hoo	1	1
Leopard	m	adult	hoo	1	1
Leopard	m	adult	shrill	2	1
Leopard	m	adult	shrill2_hoo	3	1
Leopard	m	adult	shrill2_hoo_growl	1	1
Leopard	m	adult	shrill_hoo	6	2
Leopard	m	subadult	shrill	7	3
Leopard	m	subadult	shrill2_shrill_hoo	1	1
Leopard	m	subadult	shrill_hoo	21	3
Viper	f	adult	shrill	9	2

Viper	f	subadult	shrill	11	1
Viper	m	adult	shrill	7	1
Viper	m	adult	shrill_hoo	1	1
Viper	m	subadult	shrill	35	4
Viper	m	subadult	shrill_hoo	6	1

Table S5. Behavioural responses to predator model presentations

Bayesian regression model results for behavioural responses to predator model presentations (retreat duration, probability of fleeing up a tree, bipedal stance, and tail raise). Posterior mean estimates with standard errors (Error), 95% credible intervals (CI), and effective sample sizes (ESS) are shown.
† Shape parameter of the Gamma distribution.
‡ Hurdle component indicating probability of zero-inflation.

Model: Retreat duration (hurdle Gamma)						
Parameter	**Estimate**	**Error**	**Lower 95% CI**	**Upper 95% CI**	**Bulk ESS**	**Tail ESS**
Intercept	1.41	0.33	0.80	2.12	4869	4508
context_experiment: leopard	0.79	0.34	0.09	1.45	4532	4087
context_experiment: viper	-0.74	0.35	-1.45	-0.07	4595	3799
age class: subadult	-0.37	0.31	-0.99	0.24	4239	4571

sex: male	0.01	0.32	-0.63	0.64	4378	4523
shape†	3.14	0.97	1.53	5.34	4580	4236
hurdle‡	0.17	0.07	0.06	0.33	6183	4497
Model: Probability of retreating up a tree (Bernoulli)						
Parameter	**Estimate**	**Error**	**Lower 95% CI**	**Upper 95% CI**	**Bulk ESS**	**Tail ESS**
Intercept	-3.61	2.07	-8.35	-0.23	6536	5370
context_experiment: leopard	13.78	8.61	4.31	35.75	1947	1416
context_experiment: viper	0.51	1.44	-2.27	3.45	6240	6385
age class: subadult	3.02	1.91	-0.22	7.38	5482	5210
sex: male	0.50	1.94	-3.22	4.49	5650	5394
Model: Probability of bipedal stance (Bernoulli)						
Parameter	**Estimate**	**Error**	**Lower 95% CI**	**Upper 95% CI**	**Bulk ESS**	**Tail ESS**
Intercept	1.52	1.32	-0.86	4.31	4931	4840
context_experiment: leopard	-10.59	7.86	-31.23	-2.22	1761	1208

context_experiment: viper	-0.77	1.26	-3.34	1.63	4266	3997
age class: subadult	-0.22	1.35	-2.82	2.47	4119	4008
sex: male	-1.77	1.45	-4.78	0.94	3983	4062
Model: Probability of tail raise (Bernoulli)						
Parameter	**Estimate**	**Error**	**Lower 95% CI**	**Upper 95% CI**	**Bulk ESS**	**Tail ESS**
Intercept	0.19	1.23	-2.28	2.72	7493	5137
context_experiment: leopard	-4.28	1.68	-7.95	-1.34	5224	4608
context_experiment: viper	-1.98	1.36	-4.87	0.47	5846	4894
age class: subadult	0.37	1.33	-2.29	3.02	5541	5444
sex: male	1.64	1.41	-0.97	4.57	6068	5452

Table S6. Vocal responses to predator model presentations

Bayesian regression model summaries for vocal responses to predator model presentations. Posterior mean estimates with standard errors (Error), 95% credible intervals (CI), and effective sample sizes (ESS) are shown.

† Shape parameter of the Gamma distribution.

Model: Probability of producing 'hoo' calls in the first 30 seconds						
Parameter	**Estimate**	**Error**	**Lower 95% CI**	**Upper 95% CI**	**Bulk ESS**	**Tail ESS**
Intercept	-0.37	0.49	-1.33	0.57	6400	6531
context_experiment: leopard	0.81	0.48	-0.11	1.75	5325	6047
context_experiment: viper	-2.44	0.58	-3.62	-1.34	5267	5436
sex: male	1.14	0.52	0.14	2.17	6035	6407
age class: subadult	-0.46	0.50	-1.43	0.51	5899	5541
Model: Probability of producing 'hoo' calls in the first 5 seconds						
Parameter	**Estimate**	**Error**	**Lower 95% CI**	**Upper 95% CI**	**Bulk ESS**	**Tail ESS**
Intercept	-0.25	0.77	-1.79	1.26	4798	4658
context_experiment: leopard	-0.05	0.75	-1.54	1.42	4222	3674

context_experiment: viper*	-9.56	9.13	-31.90	-2.54	792	459
sex: male	1.07	0.88	-0.65	2.83	3711	4472
age class: subadult	-1.75	0.90	-3.56	0.02	3830	3885
Model: Rate of vocal utterance production in the first 30 seconds						
Parameter	**Estimate**	**Error**	**Lower 95% CI**	**Upper 95% CI**	**Bulk ESS**	**Tail ESS**
Intercept	-1.95	0.28	-2.47	-1.37	6315	5035
context_experiment: leopard	0.61	0.30	0.01	1.21	5491	5330
context_experiment: viper	0.58	0.32	-0.06	1.22	4776	4686
age class: subadult	0.08	0.29	-0.50	0.64	5187	4832
sex: male	0.15	0.28	-0.43	0.71	5813	5365
shape†	3.52	1.10	1.73	5.97	6364	5749
Model: Rate of vocal utterance production in the first 5 seconds						
Parameter	**Estimate**	**Error**	**Lower 95% CI**	**Upper 95% CI**	**Bulk ESS**	**Tail ESS**
Intercept	-1.95	0.28	-2.47	-1.38	7101	5903

context_experiment: leopard	0.60	0.29	0.01	1.17	6789	5906
context_experiment: viper	0.58	0.31	-0.04	1.20	6005	5447
age class: subadult	0.08	0.29	-0.49	0.65	6369	5468
sex: male	0.14	0.28	-0.42	0.71	6534	5916
shape†	3.50	1.09	1.72	5.97	5987	6010

Note: *No 'hoo' calls were observed in the first 5 seconds following viper model presentations. This resulted in complete separation for that condition, meaning the absence of variation prevented a reliable estimation of the effect. While the Bayesian model still returned an estimate, it should be interpreted with caution due to the high uncertainty and lack of data for that condition.*

Table S7. Succession of behaviours per trial in the first 30s following model detection

Individual	**Vocal response**	**Behaviour**	**Condition**
♂ adult ID18	2min45 after model detection	Retreat down_Tail raise_Yawn_Yawn	**eagle**
♂ adult ID11	Yes	Bipedal stance_Retreat up	
♂ adult ID10	Yes	Tail raise_Approach model_Tail raise_Move up_Tail raises_Move down	
♂ subadult ID8	Yes	Tail raise_Bipedal stance_Tail raise	
♂ adult ID6	Yes	Retreat up_Move down_Tail raise	
♂ adult ID12	Yes	Retreat up_Tail raise_Yawn_Move down	
♀ adult ID7	Yes	Bipedal stance_Retreat down	

♀ adult ID4	Yes	Bipedal stance_Tail raise_Tail raise_Bipedal stance	
♂ adult ID18	2min after model detection	Retreat up	**leopard**
♂ adult ID11	Yes	Retreat up	
♂ adult ID10	Yes	Retreat up	
♂ subadult ID8	Yes	Retreat up_Tail raise_Move up	
♂ subadult ID3	Yes	Retreat up	
♂ subadult ID12	Yes	Retreat up	
♀ adult ID9	Yes	Retreat up	
♀ adult ID7	Yes	Retreat up	
♀ adult ID14	Yes	Retreat up	
♀ adult ID17	No	Retreat straight_Approach model	**viper**
♂ subadult ID12	No	Retreat up_Move up	
♀ subadult ID5	Yes	Retreat straight_Bipedal stance_Move up	
♂ adult ID10	Yes	Retreat straight_Tail raise_Bipedal stance_Tail raise_Approach model	
♂ subadult ID8	Yes	Retreat up_Tail raise	
♂ subadult ID16	Yes	Retreat straight_Tail raise	

♂ subadult ID15	Yes	Retreat up_Move down_Bipedal stance_Move up	
♂ subadult ID1	Yes	Approach model	
♀ adult ID2	Yes	Retreat up_Move up	
♀ adult ID13	Yes	Retreat straight_Bipedal stance_Move up_Tail raise	

General Discussion

Summary of the thesis and main findings

With this thesis, I aimed to contribute to our understanding of how structured vocal sequences and potentially meaningful combinatorial patterns evolved in non-human animals. A key challenge in the study of language evolution is to identify which features of human language, such as the use of rules for combining signals or the compositional structure of utterances, are mirrored by resembling features in other species. While many studies have focused on isolated combinations in a handful of species, comprehensive analyses of combinatorial abilities have so far been conducted primarily in apes, especially chimpanzees and bonobos. In contrast, comprehensive investigations of vocal repertoire remain scarce, making it generally unclear how flexible, structured, and functionally relevant vocal sequences are across the full vocal repertoire of a given species.

To address this, the thesis focused on two central questions: (1) What mechanisms are used to generate meaning through call combinations, such as combining distinct calls, altering their order, or the reoccurrence of specific call types? and (2) How widespread is this use of combination for meaning across the vocal repertoire? I investigated these questions in sooty mangabeys (*Cercocebus atys*), a forest-dwelling monkey species, by collecting observational and experimental data on two wild groups habituated to human presence in Taï National Park, Ivory Coast, over a total of 15 months. The research is presented in three empirical chapters that address complementary goals.

In Chapter 1, I systematically characterised the structural organisation of all vocal sequences. This analysis revealed that a subset of combinations occurred frequently and followed consistent structural patterns. Within these, I identified fixed call ordering, the iteration of specific call types, and constrained positional arrangements. Additionally, I observed structural regularities suggesting a limited form of hierarchical organisation in some multi-call sequences. The chapter also revealed sex differences in sequence production, with certain combinations appearing more frequently in one sex than the other. This suggests that sex-specific social roles may influence the use of vocal sequences. In addition to structured combinations, I also found a number of sequences that occurred rarely and did not appear to follow consistent rules.

In Chapter 2, I investigated the contextual use of vocal sequences across the full vocal repertoire. While the analysis included all call types, the bigram composed of ‘grunt’

and ‘twitter’, exclusively produced by females, emerged as the most frequently produced combination with clear and sufficient contextual information to examine in detail. This chapter showed that call combinations were used in context-specific ways, and that this specificity was associated with structural properties including ordering and iteration. The results supported the idea that sequence structure contributed to context-specificity and that these effects were consistent with compositionality, where the meaning of the sequence was derived from the meanings of its individual calls. Recipient attributes such as age and sex also appeared to influence the type of vocal utterance produced, with differences observed depending on whether the affiliative behaviour was directed toward, for example, an infant carried by its mother or an adult female without an infant.

In Chapter 3, I examined whether sooty mangabeys produce distinct alarm call types or call combinations in response to different predator classes, such as eagles, leopards, and vipers, similar to findings reported in other monkey species. The results showed that sooty mangabeys do not have predator-specific alarm calls. Instead, they use a general alarm call, ‘shrill’, which can be produced alone, and a second call, ‘hoo’, which was never observed independently but occurred only in combination following ‘shrill’. This consistent positioning and dependence suggest an affixation-like structure. ‘Shrill’ was predominantly produced during viper encounters, likely reflecting a generalised warning to a low-level threat. In contrast, the combination ‘shrill_hoo’ occurred more frequently in response to leopards and eagles, both highly mobile and actively hunting predators. This structure may encode the presence of a more dangerous threat. These results indicate that while sooty mangabeys do not produce acoustically distinct alarm calls for each predator type, they do structure call combinations in a way that may modulate signal specificity to convey threat-related information.

Overall, I show that sooty mangabeys rely on four main types of sequences which impact contextual usage: (1) order-sensitive bigrams, specifically ‘grunt_twitter’ and ‘twitter_grunt’, each used in differing socio-positive contexts such as affiliation and feeding; (2) order-sensitive iterative sequences (e.g., ‘grunt_twitter_grunt’, ‘twitter_grunt_twitter’), also used in differing socio-positive contexts; (3) affixation, where ‘hoo’ is added to ‘shrill’ in some alarm contexts; and (4) a set of structurally variable sequences that do not follow rule-based patterns, most likely produced in aggressive contexts based on the call types they contain (‘scream’, ‘hoo’, ‘grumble’,

and ‘growl’). While this last category was not analysed in detail, it is worth mentioning here to provide a more complete overview of the range of sequence structures observed in sooty mangabeys. Importantly, these sequences were not used uniformly between sexes: some, such as grunt–twitter combinations, were only produced by females, while others, like ‘shrill_hoo’, were more frequently observed in males. Together, these findings highlight the link between vocal sequence structure and meaning, and how different combinatorial mechanisms are associated with usage in distinct social or ecological contexts. In the following sections, I discuss how these findings contribute to our understanding of vocal communication in non-human animals and the evolution of syntactic-like systems.

Sex differences in sooty mangabey vocal sequence production

Adult female and male sooty mangabeys differed in their vocal repertoires, with some call types used only by one sex, and, partly as a result, also differed in how they used call sequences.

‘Shrill_hoo’ combinations were more frequently recorded in males during observational data collection (but not during predator experiments), while only females produced sequences composed of ‘twitter’ and ‘grunt’, as adult males do not produce twitters. This calls for further explanation, which can be gained by comparing these patterns to those in other primate species that differ in social structure, dispersal patterns, or group composition. Family-living and pair-bonded monkeys such as gibbons and marmosets typically exhibit minimal sex differences in vocal behaviour, likely due to shared social roles (Snowdon, 2017). In contrast, in multi-male, multi-female species, social dynamics appear to shape vocal production. For instance, in rhesus macaques, females vocalise more frequently than males and also have a higher rate of social bonding among themselves (Greeno & Semple, 2009). In red-capped mangabeys, females produce more calls and sequences, and at a higher rate than males, likely because their social role includes mediation of intragroup social relationships (Bouchet et al., 2010). In harem groups with a single adult male and several females and offspring, such as in many Cercopithecus species, only males produce distinctive loud calls (e.g., putty-nosed monkeys (Arnold & Zuberbühler, 2006), Diana monkeys (Zuberbühler, 2000), Campbell’s monkeys (Zuberbühler, 2001)), likely due to their central role in territory and predator defence (Zuberbühler, 2004). Sooty mangabeys

show sex differences not only in call types but also in the production of call sequences. Males, who disperse from their natal group (McGraw et al., 2007), produced alarm calls such as ‘shrill_hoo’ more frequently than females, potentially due to their peripheral position in the group, which may make them more exposed to predators. In contrast, females, who remain in the group throughout their lives and occupy central roles within the social structure (Range & Noë, 2002), are the only individuals to produce grunt and twitter sequences in socio-positive contexts and during daily activities such as foraging. Sooty mangabeys thus appear to follow patterns observed in other multi-male, multi-female monkey species, and even in harem systems, where vocal production, including call combination, reflects sex-specific social roles. This contrasts with chimpanzees and bonobos, where both sexes share the full vocal sequence repertoire (Berthet et al., 2025; Boesch & Crockford, 2005; Girard-Buttoz et al., 2025), suggesting that socio-ecological factors may promote sex-specific call combinations. However, confirming this would require further investigation.

Comparison of sequence mechanisms and meaning with other species

Order-sensitive bigram

In sooty mangabeys, the ‘twitter_grunt’ bigram provides a potential example of both compositional and order-sensitive sequences. Both component calls (i.e., ‘twitter’ and ‘grunt’) can occur independently and are primarily associated with feeding contexts and, to a lesser extent, affiliative interactions involving infants. When combined in the order ‘grunt_twitter’, the bigram remains broadly associated with these same contexts. In contrast, when the order is reversed to form ‘twitter_grunt’, the bigram shows a pronounced contextual shift and narrowing: it is used almost exclusively during affiliative interactions directed at infants. This shift suggests that while the individual calls have overlapping meanings, combining them in a specific order results in a more precise or clarified message. This pattern indicates a compositional process, where the meaning of the sequence is derived from the meanings of its parts, and also demonstrates that call order plays a critical role in shaping interpretation.

Comparable examples of compositional structure, without regard to call order, have been documented in several species. In chimpanzees, ‘alarm-huu’ and ‘waa-bark’ calls, each associated with surprise and recruitment, respectively, are combined during snake encounters which elicits stronger responses than either call alone, suggesting

that listeners integrate the meanings of both (Leroux et al., 2023). In pied babblers, ‘alert’ and ‘recruitment’ calls are combined into mobbing sequences produced specifically in response to terrestrial predators, and playback experiments show enhanced vigilance and approach behaviour when the combination is heard (Engesser et al., 2015). In African forest elephants, combinations of identity-signalling ‘rumbles’ and high-intensity ‘broadband’ calls are used in a broader range of contexts than either call alone, particularly in high-stakes competitive or separation situations (Hedwig & Kohlberg, 2024). In each of these species, call combinations appear to generate new or enriched meanings by integrating the functions of the component calls.

However, order-sensitive compositionality has only been demonstrated in a few species, including chimpanzees and Japanese great tits. In great tits, the sequence ‘ABC–D’ triggers a dual response, danger scanning followed by approach, while the reversed sequence ‘D–ABC’ fails to elicit the same reaction, despite both components carrying meaning when used independently (Suzuki et al., 2016, 2017). In chimpanzees, bigrams such as ‘hoo_grunt’ and ‘grunt_hoo’ differ in their contextual deployment. Importantly, both component calls are also produced alone: ‘hoo’ calls occur in resting and travel contexts, and ‘grunts’ in feeding and greeting contexts. When combined as ‘hoo_grunt’, the bigram is used during resting at a feeding site, whereas the reversed order, ‘grunt_hoo’, is more frequent in travel and social fusion events (Girard-Buttoz et al., 2025). This shows that the same calls, each associated with distinct meanings, can generate different meanings when combined depending on their order. In sooty mangabeys, a similar mechanism appears in the contrast between ‘grunt_twitter’ and ‘twitter_grunt’: while both calls occur independently in feeding and affiliative contexts, only ‘twitter_grunt’ shows a marked narrowing to infant-directed affiliative interactions. These findings suggest that in all three species, meaning is derived not only from the combination of distinct vocal elements but also from the specific order of the calls in the sequence in which they are produced.

Order-sensitive iterative sequences

In sooty mangabeys, iterative sequences composed of ‘grunt’ and ‘twitter’ provided further evidence of structured vocal production. These sequences were most commonly used in infant-directed affiliative contexts, but the order in which calls occurred also appeared to influence their contextual use. Sequences beginning with

‘twitter’, such as ‘twitter_grunt_twitter’, were used almost exclusively in affiliative interactions directed at infants, mirroring the pattern observed in the ‘twitter_grunt’ bigram. In contrast, sequences beginning with ‘grunt’, such as ‘grunt_twitter_grunt’, showed a broader contextual profile, occurring in both infant-directed affiliation and feeding situations, though still more frequently in the former. This contrast suggested that while iterative sequences were generally associated with affiliation, the type of the initial call modulated their specificity, with ‘twitter’-initiated sequences being more narrowly infant-focused and ‘grunt’-initiated ones more flexible in use.

Comparable uses of recurrence of call types within sequences have been described in a few other species, although the mechanism of iteration has not been analysed as such. In white-handed gibbons, sequences used in predator encounters (‘predator songs’) contain a higher number of certain note types, particularly ‘hoo’ and ‘sharp wow’, compared to duet songs, suggesting that specific notes are iterated to signal threat-related information (Andrieu et al., 2020; Clarke et al., 2006). In bonobos, vocal sequences associated with food contexts often include multiple calls that are not fixed in order but can involve the reoccurrence of particular calls across the sequence, depending on food preference (Clay & Zuberbühler, 2009, 2011). These examples point to the possibility that the reiteration of calls within a sequence may contribute to generating meaning.

To my knowledge, however, no previous study has demonstrated that iterative sequences are order-sensitive depending on the context of production. It is also possible that such sequences reflect the embedding of previously meaningful bigrams and single calls, rather than an effect of overall sequence length or call count. This interpretation would suggest that meaning is not distributed uniformly across longer sequences but emerges from the internal structure of embedded units, a hypothesis that would need to be explicitly tested in future research.

Affixation

In sooty mangabeys, one sequence structure consistently observed during alarm contexts resembles affixation. Individuals produced the call ‘shrill’ both alone and in combination with a second call, ‘hoo’, which was never observed independently. The combination ‘shrill_hoo’ occurred more frequently in response to leopards and eagles, whereas single ‘shrill’ calls were more commonly produced in response to vipers. While

we also recorded a few instances of ‘hoo’ appended to other call types typically given during aggressive interactions, these were not frequent enough to analyse. Thus, only the ‘shrill_hoo’ combination could be systematically linked to specific alarm contexts. These findings suggest that sooty mangabeys may use affixation-like structures to communicate threat-related information, with the presence of ‘hoo’ potentially modulating the meaning of the initial ‘shrill’ call.

Comparable systems have been described in monkeys and mongooses: in male Campbell’s monkeys, the ‘krak’ alarm call is produced in response to leopards, while the affixed version ‘krak_oo’ signals a more general terrestrial disturbance (Ouattara, Lemasson, & Zuberbühler, 2009). Playback experiments with sympatric Diana monkeys showed that listeners responded more strongly to the unmodified form, suggesting they attended to the affix to adjust their interpretation (Coye et al., 2015). A similar structure has been described for eagle alarms (‘hok’ and ‘hok_oo’) (Ouattara, Lemasson, & Zuberbühler, 2009). In female Campbell’s monkeys, affixation appears in social contexts: contact calls can occur alone or in combination with a high-pitched arched unit, with combined forms being more conspicuous and carrying identity-related information (Coye et al., 2018). Banded mongooses also use affixation-like structures in their close calls: a constant initial element conveys individual identity, while a variable second element signals behavioural context (Jansen et al., 2012). These systems suggest that affixation-like structures may serve different functions across species, including signalling threat specificity or social cues.

Among these examples, the ‘shrill_hoo’ combination in sooty mangabeys most closely resembles the structure observed in male Campbell’s monkeys, where affixation systematically modifies the meaning of alarm calls. The fact that ‘shrill_hoo’ and ‘shrill’ were produced in different predator contexts raises the possibility that this structure is also compositional in function.

Other mechanisms

In addition to the structured and contextually specific sequences described above, sooty mangabeys also produced other bigrams that did not appear to follow consistent ordering patterns. Among these were the bigrams ‘growl_scream’ and ‘scream_growl’. Sooty mangabey vocal sequence production also included a large proportion of unique

sequences, defined by the specific order of call types (e.g., A_B, A_B_A, A_C, D_F, etc.).

In contrast to the ‘twitter_grunt’ combinations described earlier, these bigrams and unique sequences were most frequently composed of call types typically produced during aggressive interactions (ref vocal repertoire). Among these, ‘hoo’, ‘grumble’ and ‘wau’ were not recorded alone and may also function as affixes.

Although we did not analyse the context of production for these sequences due to limited sample size, it is possible that such sequences are produced during rapid changes in social events, for instance, during escalating conflicts or when new individuals join an ongoing altercation. In these situations, sequences may result from closely timed call production reflecting dynamic shifts in behaviour. They could offer a vocal read-out of fast-evolving social interactions, allowing non-visible group members to track unfolding events in real time. Conversely, as with the ‘grunt_twitter’ and ‘twitter_grunt’ bigrams, it remains possible that bigrams such as ‘growl_scream’ and ‘scream_growl’ are compositional, but since we did not analyse their context of production, no conclusions can be drawn at this stage.

Linguistic perspectives

The sequence structures identified in sooty mangabeys, particularly the order-sensitive bigrams, iterative sequences, and affixation-like combinations, raise the question of whether meaning in this system is generated through compositional mechanisms. From a linguistic perspective, Schlenker and colleagues (Schlenker et al., 2023, 2024) have proposed the notion of minimal compositionality to describe animal combinations where the meaning of a sequence appears to depend on the meanings of its component calls and their order, without assuming the presence of complex syntactic rules as in human language. For example, in the case of ‘twitter_grunt’, both calls are used independently and associated with broader affiliative or feeding contexts, yet when combined in this order, the sequence is used almost exclusively in infant-directed affiliation, suggesting that the meaning of the sequence may be influenced by its structure. However, Schlenker and colleagues have cautioned against interpreting such findings as evidence of compositional syntax too readily. They argue that some apparent structure in animal communication may instead be explained by pragmatic inference or learned contextual associations, rather than by the application of structural

rules. In this view, animals may not compute meaning through a rule-based system but rather rely on patterns of usage and contextual cues, which have been termed animal implicatures. This alternative explanation could also apply to sooty mangabeys: the context-specific use of ‘twitter_grunt’ might reflect regularities in social interactions, rather than a compositional process in the linguistic sense. These perspectives highlight the importance of cautious interpretation and the need for targeted experimental testing to distinguish between structurally encoded meaning and inference based on co-occurrence and context. While sooty mangabey sequences are consistent with the predictions of minimal compositionality, further work is needed to assess how meaning is generated and processed by receivers.

Evolutionary drivers of syntax

The evolution of communicative systems capable of conveying a wide range of meanings is likely the result of combined social and ecological pressures. Species that need to manage complex social dynamics have been suggested to produce a greater diversity of vocal utterances dedicated to this function, an idea captured by the social complexity hypothesis, which proposes that more complex social environments select for more elaborate communication systems (Bouchet et al., 2013; Freeberg, 2006; Freeberg et al., 2012; Grampp et al., 2023; Humphrey, 1976; McComb & Semple, 2005). Supporting this, call combinations are common across daily life in chimpanzees (Leroux et al., 2022), more diverse in meerkats during social interactions compared to other contexts (Collier et al., 2017), and more frequent in larger social groups of magpies than in smaller ones (Walsh et al., 2024). Sooty mangabeys, with their frequent use of call combinations in affiliative contexts (and possibly also during aggression, although this was not analysed) fit this pattern well.

Another suggested selection pressure for complex signalling is predation (Stephan & Zuberbühler, 2008). Signalling in alarm contexts, whether to convey information about predator type, behaviour, perceived threat level, or to recruit others for mobbing, has been shown to involve specific call combinations, including in titi monkeys (Berthet et al., 2018), olive colobus monkeys (Gallot et al., 2024), Campbell’s monkeys (Ouattara, Lemasson, & Zuberbühler, 2009; Ouattara, Lemasson, & Zuberbuhler, 2009), meerkats (Rauber et al., 2020), Japanese great tits (Suzuki et al., 2016), and chimpanzees (Leroux et al., 2023). In sooty mangabeys, alarm contexts also frequently

triggered combinations, particularly in response to high-threat predators such as leopards and eagles, rather than less immediate dangers like vipers. These patterns support the idea that the need to convey diverse meanings in response to specific threats may favour the use of vocal sequences.

Low-visibility habitats such as rainforests have also been suggested to favour the use of particular vocal signals to maintain group cohesion (Coye et al., 2018; Marler, 1975). In gorillas, for instance, both the calling rate and acoustic structure of close-distance calls vary with visibility, with higher-pitched units produced more often when individuals are out of sight or dispersed (Hedwig et al., 2015). However, I am not aware of any study showing that call combinations specifically serve this function.

Taken together, these patterns suggest that social and ecological pressures have likely contributed to the evolution of call combinations in sooty mangabeys.

General conclusion

In this thesis, I used a whole-repertoire approach combining field observations and experiments to investigate the vocal sequence system of sooty mangabeys. I showed that they rely on rule-based sequences in two daily contexts, particularly in affiliative interactions directed at infants and in alarm contexts involving external threats such as leopards, eagles and vipers. These include order-sensitive bigrams, iterative sequences and affixation-like structures. They also produce sequences without apparent rules, specifically during aggressive interactions, possibly reflecting rapid shifts in social dynamics during conflicts. Sequence use differed between sexes, suggesting that social roles shape production. While playback experiments were not conducted and remain essential to confirm whether sequences are meaningful to receivers, the structural consistency and contextual use of these sequences suggest that they likely contribute to meaning generation. Certain contexts were also under-sampled, including those that are more difficult for observers to follow, such as events occurring high in the trees or during high-arousal situations like intense aggressive encounters within or between groups. Still, by analysing the full vocal repertoire, this thesis shows that within a single monkey species, multiple mechanisms of rule-based sequences (ordering, iteration and affixation) are used in ways that likely contribute to meaning generation. This provides new insights into how vocal sequences can expand messaging beyond what can be offered by single calls alone. These patterns likely evolved in response to social and environmental pressures: the need to manage social relationships and to respond effectively to diverse external threats may have favoured the emergence of rule-based sequences in specific contexts. Across species, demonstrations of compositional-like vocal structures outside of alarm contexts, here in affiliative contexts, are rare except in great apes. Whether this reflects the limits of animal capacities or rather the limit of reporting remains to be assessed. Compared to apes, which are more closely related to humans, sooty mangabeys rely on a less versatile combinatorial system with use of combinations in fewer domains of daily life, which may reflect differences in communicative needs or cognitive capacities. Together, this thesis offers new insights into the evolution of compositional abilities that may have ultimately contributed to the emergence of human language.

References

Andrieu, J., Penny, S. G., Bouchet, H., Malaivijitnond, S., Reichard, U. H., & Zuberbühler, K. (2020). White-handed gibbons discriminate context-specific song compositions. *PeerJ*, *8*, e9477. https://doi.org/10.7717/peerj.9477

Arnold, K., & Zuberbühler, K. (2006). The alarm-calling system of adult male putty-nosed monkeys, Cercopithecus nictitans martini. *Animal Behaviour*, *72*(3), 643–653. https://doi.org/10.1016/j.anbehav.2005.11.017

Berthet, M., Neumann, C., Mesbahi, G., Cäsar, C., & Zuberbühler, K. (2018). Contextual encoding in titi monkey alarm call sequences. *Behavioral Ecology and Sociobiology*, *72*(1), 8. https://doi.org/10.1007/s00265-017-2424-z

Berthet, M., Surbeck, M., & Townsend, S. W. (2025). Extensive compositionality in the vocal system of bonobos. *Science*, *388*(6742), 104–108. https://doi.org/10.1126/science.adv1170

Boesch, C., & Crockford, C. (2005). Call combinations in wild chimpanzees. *Behaviour*, *142*(4), 397–421. https://doi.org/10.1163/1568539054012047

Bouchet, H., Blois-Heulin, C., & Lemasson, A. (2013). Social complexity parallels vocal complexity: A comparison of three non-human primate species. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00390

Bouchet, H., Pellier, A.-S., Blois-Heulin, C., & Lemasson, A. (2010). Sex differences in the vocal repertoire of adult red-capped mangabeys (Cercocebus torquatus): A multi-level acoustic analysis. *American Journal of Primatology*, *72*(4), 360–375. https://doi.org/10.1002/ajp.20791

Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The Syntax and Meaning of Wild Gibbon Songs. *PLoS ONE*, *1*(1), e73. https://doi.org/10.1371/journal.pone.0000073

Clay, Z., & Zuberbühler, K. (2009). Food-associated calling sequences in bonobos. *Animal Behaviour*, *77*(6), 1387–1396. https://doi.org/10.1016/j.anbehav.2009.02.016

Clay, Z., & Zuberbühler, K. (2011). Bonobos Extract Meaning from Call Sequences. *PLoS ONE*, *6*(4), e18786. https://doi.org/10.1371/journal.pone.0018786

Collier, K., Townsend, S. W., & Manser, M. B. (2017). Call concatenation in wild meerkats. *Animal Behaviour*, *134*, 257–269. https://doi.org/10.1016/j.anbehav.2016.12.014

Coye, C., Ouattara, K., Arlet, M. E., Lemasson, A., & Zuberbühler, K. (2018). Flexible use of simple and combined calls in female Campbell's monkeys. *Animal Behaviour*, *141*, 171–181. https://doi.org/10.1016/j.anbehav.2018.05.014

Coye, C., Ouattara, K., Zuberbühler, K., & Lemasson, A. (2015). Suffixation influences receivers' behaviour in non-human primates. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1807), 20150265. https://doi.org/10.1098/rspb.2015.0265

Engesser, S., Crane, J. M. S., Savage, J. L., Russell, A. F., & Townsend, S. W. (2015). Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System. *PLOS Biology*, *13*(6), e1002171. https://doi.org/10.1371/journal.pbio.1002171

Freeberg, T. M. (2006). Social Complexity Can Drive Vocal Complexity: Group Size Influences Vocal Information in Carolina Chickadees. *Psychological Science*, *17*(7), 557–561. https://doi.org/10.1111/j.1467-9280.2006.01743.x

Freeberg, T. M., Dunbar, R. I. M., & Ord, T. J. (2012). Social complexity as a proximate and ultimate factor in communicative complexity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1597), 1785–1801. https://doi.org/10.1098/rstb.2011.0213

Gallot, Q., Depriester, C., Moran, S., & Zuberbühler, K. (2024). A non-human primate combinatorial system for long-distance communication. *iScience*, *27*(11). https://doi.org/10.1016/j.isci.2024.111172

Girard-Buttoz, C., Neumann, C., Bortolato, T., Zaccarella, E., Friederici, A. D., Wittig, R. M., & Crockford, C. (2025). Versatile use of chimpanzee call combinations promotes meaning expansion. *Science Advances*, *11*(19), eadq2879. https://doi.org/10.1126/sciadv.adq2879

Grampp, M., Samuni, L., Girard-Buttoz, C., León, J., Zuberbühler, K., Tkaczynski, P., Wittig, R. M., & Crockford, C. (2023). Social uncertainty promotes signal complexity during approaches in wild chimpanzees (*Pan troglodytes verus*) and mangabeys (*Cercocebus atys atys*). *Royal Society Open Science*, *10*(11), 231073. https://doi.org/10.1098/rsos.231073

Greeno, N. C., & Semple, S. (2009). Sex differences in vocal communication among adult rhesus macaques. *Evolution and Human Behavior*, *30*(2), 141–145. https://doi.org/10.1016/j.evolhumbehav.2008.09.002

Hedwig, D., & Kohlberg, A. (2024). Call combination in African forest elephants Loxodonta cyclotis. *PLOS ONE*, *19*(3), e0299656. https://doi.org/10.1371/journal.pone.0299656

Hedwig, D., Mundry, R., Robbins, M. M., & Boesch, C. (2015). Audience effects, but not environmental influences, explain variation in gorilla close distance vocalizations—A test of the acoustic adaptation hypothesis. *American Journal of Primatology*, *77*(12), 1239–1252. https://doi.org/10.1002/ajp.22462

Humphrey, N. (1976). The social function of intellect. In *Growing points in ethology* (Bateson PPG, Hinde RA, pp. 303–317). Cambridge University Press.

Jansen, D. A., Cant, M. A., & Manser, M. B. (2012). Segmental concatenation of individual signatures and context cues in banded mongoose (Mungos mungo) close calls. *BMC Biology*, *10*(1), 97. https://doi.org/10.1186/1741-7007-10-97

Leroux, M., Chandia, B., Bosshard, A. B., Zuberbühler, K., & Townsend, S. W. (2022). Call combinations in chimpanzees: A social tool? *Behavioral Ecology*, arac074. https://doi.org/10.1093/beheco/arac074

Leroux, M., Schel, A. M., Wilke, C., Chandia, B., Zuberbuhler, K., Slocombe, K. E., & Townsend, S. W. (2023). Call combinations and compositional processing in wild chimpanzees. *Nature Communications*, *14*, 2225. https://doi.org/10.1038/s41467-023-37816-y

Marler, P. (1975). On the origin of speech from animal sounds. *The Role of Speech in Language*, *11*(37).

McComb, K., & Semple, S. (2005). Coevolution of vocal communication and sociality in primates. *Biology Letters*, *1*(4), 381–385. https://doi.org/10.1098/rsbl.2005.0366

McGraw, W. S., Zuberbuhler, K., & Noë, R. (2007). *Monkeys of the Taï forest: An African primate community* (Vol. 51). Cambridge University Press.

Ouattara, K., Lemasson, A., & Zuberbuhler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, *106*(51), 22026–22031. https://doi.org/10.1073/pnas.0908118106

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's Monkeys Use Affixation to Alter Call Meaning. *PLoS ONE*, *4*(11), e7808. https://doi.org/10.1371/journal.pone.0007808

Range, F., & Noë, R. (2002). Familiarity and dominance relations among female sooty mangabeys in the Taï National Park: Relations Among Female Sooty Mangabeys. *American Journal of Primatology*, *56*(3), 137–153. https://doi.org/10.1002/ajp.1070

Rauber, R., Kranstauber, B., & Manser, M. B. (2020). Call order within vocal sequences of meerkats contains temporary contextual and individual information. *BMC Biology*, *18*(1), 119. https://doi.org/10.1186/s12915-020-00847-8

Schlenker, P., Coye, C., Leroux, M., & Chemla, E. (2023). The ABC-D of animal linguistics: Are syntax and compositionality for real? *Biological Reviews*, *98*(4), 1142–1159. https://doi.org/10.1111/brv.12944

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*, brv.13068. https://doi.org/10.1111/brv.13068

Snowdon, C. T. (2017). Vocal Communication in Family-Living and Pair-Bonded Primates. In R. M. Quam, M. A. Ramsier, R. R. Fay, & A. N. Popper (Eds.), *Primate Hearing and Communication* (pp. 141–174). Springer International Publishing. https://doi.org/10.1007/978-3-319-59478-1_6

Stephan, C., & Zuberbühler, K. (2008). Predation increases acoustic complexity in primate alarm calls. *Biology Letters*, *4*(6), 641–644. https://doi.org/10.1098/rsbl.2008.0488

Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nature Communications*, *7*(1), 10986. 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

Walsh, S. L., Townsend, S. W., Engesser, S., & Ridley, A. R. (2024). Call combination production is linked to the social environment in Western Australian magpies (

Gymnorhina tibicen dorsalis). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *379*(1905), 20230198. https://doi.org/10.1098/rstb.2023.0198

Zuberbühler, K. (2000). Referential labelling in Diana monkeys. *Animal Behaviour*, *59*(5), 917–927. https://doi.org/10.1006/anbe.1999.1317

Zuberbühler, K. (2001). Predator-specific alarm calls in Campbell's monkeys, Cercopithecus campbelli. *Behavioral Ecology and Sociobiology*, *50*(5), 414–422. https://doi.org/10.1007/s002650100383

Zuberbühler, K. (2004). Effects of Natural and Sexual Selection on the Evolution of Guenon Loud Calls. In M. E. Glenn & M. Cords (Eds.), *The Guenons: Diversity and Adaptation in African Monkeys* (pp. 289–306). Kluwer Academic Publishers. https://doi.org/10.1007/0-306-48417-X_20