Article:

Mine, Joseph G., Wilke, Claudia, Zulberti, Chiara et al. (7 more authors) (2024) Vocal-visual combinations in wild chimpanzees. Behavioral Ecology and Sociobiology. 108. ISSN 1432-0762

https://doi.org/10.1007/s00265-024-03523-x

Behavioral Ecology and Sociobiology (2024) 78:108
https://doi.org/10.1007/s00265-024-03523-x

ORIGINAL ARTICLE

Vocal-visual combinations in wild chimpanzees

Joseph G. Mine[1,2] · Claudia Wilke[3] · Chiara Zulberti[4] · Melika Behjati[5,6] · Alexandra B. Bosshard[1,2] · Sabine Stoll[1,2] · Zarin P. Machanda[7] · Andri Manser[8] · Katie E. Slocombe[4] · Simon W. Townsend[1,2,9]

Received: 6 October 2023 / Revised: 26 September 2024 / Accepted: 2 October 2024

Abstract

Living organisms throughout the animal kingdom habitually communicate with multi-modal signals that use multiple sensory channels. Such composite signals vary in their communicative function, as well as the extent to which they are recombined freely. Humans typically display complex forms of multi-modal communication, yet the evolution of this capacity remains unknown. One of our two closest living relatives, chimpanzees, also produce multi-modal combinations and therefore may offer a valuable window into the evolutionary roots of human communication. However, a currently neglected step in describing multi-modal systems is to disentangle non-random combinations from those that occur simply by chance. Here we aimed to provide a systematic quantification of communicative behaviour in our closest living relatives, describing non-random combinations produced across auditory and visual modalities. Through recording the behaviour of wild chimpanzees from the Kibale forest, Uganda we generated the first repertoire of non-random combined vocal and visual components. Using collocation analysis, we identified more than 100 vocal-visual combinations which occurred more frequently than expected by chance. We also probed how multi-modal production varied in the population, finding no differences in the number of visual components produced with vocalisations as a function of age, sex or rank. As expected, chimpanzees produced more visual components alongside vocalizations during longer vocalization bouts, however, this was only the case for some vocalization types, not others. We demonstrate that chimpanzees produce a vast array of combined vocal and visual components, exhibiting a hitherto underappreciated level of multi-modal complexity.

Significance

In humans and non-humans, acoustic communicative signals are typically accompanied by visual information. Such “multi-modal communication” has been argued to function for increasing redundancy as well as for creating new meaning. However, a currently neglected step when describing multi-modal systems and their functions is to disentangle non-random combinations from those that occur simply by chance. These data are essential to providing a faithful illustration of a species’ multi-modal communicative behaviour. Through recording the behaviour of wild chimpanzees from the Kibale forest, Uganda we aimed to bridge this gap in understanding and generated the first repertoire of non-random combined vocal and visual components in animals. Our data suggest chimpanzees combine many components flexibly and these results have important implications for our understanding of the complexity of multi-modal communication already existing in the last common ancestor of humans and chimpanzees.

Keywords Animal communication · Chimpanzees · Multimodal communication · Language evolution

---

Communicated by D. P. Watts.

Extended author information available on the last page of the article

Introduction

In many species, individuals rely on multiple sensory channels for communication, which can also be deployed simultaneously, resulting in what is commonly referred to as "multi-modal communication". A recent upsurge in interest in multi-modal communication paved the way for the realization that studying communicative channels such as vocalizations and visual displays in a vacuum fails to capture the full complexity of a communicative system, and since then, multi-modal communication has been described in groups as diverse as insects, anurans, birds, reptiles and mammals (Higham and Hebets 2013). For example, in ants (*Aphaenogaster albisetosus*) and spiders (*Cupiennius salei*), signals involved in recruitment and mate attraction may involve a pheromone release in conjunction with a vibrational element (Rovner and Barth 1981; Hölldobler 1999), while in African elephants (*Loxodonta africana)*, "greeting" behaviours may comprise a visual and a tactile component (Allen et al. 2021). Whether multi-modal signals are employed in the communication systems of non-human primates has received growing attention, given the valuable insight such data can also provide regarding the evolutionary origins of human communication and language (Slocombe et al. 2011; Liebal et al. 2022). The term "multi-modal" has, however, been used differently in previous primate communication studies, in some cases denoting multiple signaling channels (e.g. facial expressions vs gestures; Liebal et al. 2014; Wilke et al. 2017), while in others denoting multiple sensory modalities (e.g. acoustic vs. visual modality; Fröhlich and van Schaik 2018; Singletary and Tecot 2020). Here, we define a multi-modal signal as one that is received in at least two sensory modalities (Partan and Marler 2005; Fröhlich et al. 2019).

Previous research on multi-modal communication in non-human primates has shown that acoustic and visual signal components are often tightly linked (Liebal et al. 2014; Fröhlich et al. 2019; Aychet et al. 2021). For example, in monkeys, receivers are adept at recognizing the correspondence between facial expressions and vocalizations (Ghazanfar 2013) and integrating this composite information for crucial tasks such as individual recognition (Habbershon et al. 2023). Furthermore, great apes are known to augment their vocalizations with specific visual gestures. Vocalisations combined with a gesture are more likely to elicit a response from receivers compared to unimodal vocal signals (Wilke et al. 2017), and combining vocal and gestural signals may also serve to disambiguate or refine meaning, akin to the function of so-called 'extralinguistic cues' as semantic devices in language (Holle and Gunter 2007; Hobaiter et al. 2017; Wilke et al. 2017). For example, in bonobos, the "contest hoot" vocalization can be combined with a threatening "stomp" gesture during agonistic challenges, or with a playful "wrist shake" in friendly play (Genty et al. 2014). In chimpanzees, mothers interacting with infants often combine the "soft hoo" vocalization, which is given in multiple contexts, with an "arm reach" or "present back" gesture, when they want to invite the infant to climb onto their back (Fröhlich et al. 2016).

These examples and more have additionally provided valuable insight into the proximate and ultimate mechanisms underlying multi-modal communication. At the proximate level, multi-modal signals can be described as either "fixed", whereby the distinct components which make up the signal are necessarily emitted together (e.g. the croaking of a frog is necessarily accompanied by the visually salient inflation of the throat sac), or "free", when signal components can also occur separately, such as the independently produced visual bowing and vocal cooing in male pigeon courtship displays (Partan et al. 2005). Furthermore, distinct components can be combined into a multi-modal signal for different functions (Partan and Marler 1999; Fröhlich and van Schaik 2018). Combinations can serve to increase the reliability of message content ("redundancy hypothesis") or the likelihood of signal transmission across a variety of environmental conditions ("efficacy-backup hypothesis" Hebets and Papaj 2005). Combinations may also elicit enhanced responses compared to those observed when either component occurs alone, such as the increased anti-predator response observed in squirrels when tail flagging and alarm calling behaviour are combined (Partan et al. 2010). In addition, non-redundant multi-modal signalling may provoke an entirely new response in receivers, which has been termed "emergence" (Partan and Marler 2005). This function is well-attested in humans and is arguably partly responsible for the enormous generativity of our communication system (Scherer 2013).

Regarding the ultimate function of multi-modal communication, it may be that more sophisticated multi-modal signaling is associated with increased social demands, in line with previous work indicating an evolutionary relationship between communicative complexity and social complexity (Freeberg et al. 2012). Complexity in multi-modal communication could be manifested via abundant multi-modal combination types, many signal components within a combination or a diversity of functions of multi-modal signals and indeed evidence from chimpanzees suggests that signal combinations represent an adaptation for navigating challenging social situations (Leroux et al. 2022). Further, signaling behaviour in chimpanzees is known to vary based on demographic factors (e.g. age, rank; Tomasello et al. 1994; Hobaiter et al. 2017), denoting a high degree of flexibility that can be tailored to individual social needs. However, investigating either the functional mechanisms or the evolutionary drivers of multi-modal combinations requires first a comprehensive and systematic documentation of multimodal repertoires (Aychet et al. 2021). Focusing on non-human species with an extensive vocal and visual repertoire and an attested ability to produce "free" multi-modal signals would be particularly promising. Chimpanzees, one of our two closest living relatives, represent precisely such a system.

Chimpanzees, like humans, have complex social lives: they reside in groups of ~20–200 individuals, forming strong and durable relationships with relatives as well as non-kin (Rosati et al. 2020). Likely as a way to navigate this complex social environment, chimpanzees are also equipped with a rich system of communication comprising signals in several modalities (Parr and Waller 2006; Slocombe and Zuberbühler 2010; Hobaiter and Byrne 2011). In line with a wealth of literature addressing the notion of communicative signals in non-humans, we use the term "signal" to mean a vocal or visual behaviour that has evolved specifically for the purpose of communication, as opposed to behaviours designed to fulfil other mechanical or social functions (Seyfarth and Cheney 2003; Genty et al. 2009; Hobaiter and Byrne 2011; Laidre and Johnstone 2013; Grund et al. 2024). The chimpanzee's repertoire of vocal signals consists of approximately 13 different call types (Slocombe and Zuberbühler 2010). The repertoire is commonly described as graded (Slocombe and Zuberbühler 2010), meaning that there is acoustic variation within a single category, as well as a degree of overlap in acoustic features also between certain categories. Chimpanzee vocal production is constrained to a limited range of sounds compared to human vocal production (Seyfarth and Cheney 2010; Fitch et al. 2016). By contrast, visual signal production in chimpanzees is highly flexible and the repertoire comprises at least 66 visual gesture types (Hobaiter and Byrne 2011) and 9 facial expressions (Parr and Waller 2006). Although some gestures may also be received in the tactile modality (e.g. slap, poke) and important information relating to identity and reproductive state may be received in the olfactory modality (Liebal and Oña 2018), we focus here on the core modalities which have been most thoroughly investigated and which are empirically most tractable, namely vocal and visual communication (Liebal et al. 2014). Previous work has shown that both wild and captive chimpanzees combine vocal and gestural signals (Taglialatela et al. 2015; Hobaiter et al. 2017) and a repertoire of combinations of vocalizations, gestures and facial expressions has been described (Wilke et al. 2017). However, since vocalizations may co-occur with other visual components simply by chance, differentiating random from non-random multi-modal combinations is paramount.

Importantly, in chimpanzees as well as humans, vocal signals, facial expressions and manual gestures are also complemented by an equally broad array of bodily behaviours, such as posture (e.g. sitting vs. standing), head and body orientation, (towards vs. away from the recipient), or movements and actions (e.g. walking, running, playing, eating). In comparison to signals, these behaviours might be rather described as cues, i.e. behaviours that have not evolved for a communicative purpose, and serve direct mechanical functions other than communication, yet may nonetheless carry communicative value (Searcy and Nowicki 2010; Roberts et al. 2014). In humans, such non-verbal cues are known to play an important role in augmenting the meaning of an utterance, constituting a vital component of the communicative interaction (Müller et al. 2013). For example, body posture can inform the receiver regarding the believability of what is being said (Müller et al. 2013). However, these cues are rarely, if ever considered in discussions of primate multi-modal communication. Given the importance of such cues in human communication and in light of the evidence that primates integrate a wide range of information sources when responding to a signal (Seyfarth et al. 2017), we regarded all visual information offered by the signaler during vocal production as integral to a comprehensive analysis of chimpanzee multi-modal communication.

In this study, we systematically investigated the multimodal communicative behaviour of wild chimpanzees. Firstly, taking a bottom-up, inductive approach we built a vocal-visual repertoire by focusing on all non-vocal behaviour (NVB) offered by the signaler during vocal production (body postures, orientations, behaviours, gestures or facial expressions). Secondly, we quantified the non-random nature of identified vocal-visual combinations using collocation analysis, a method borrowed from computational linguistics (Bosshard et al. 2022). The use of quantitative models and theoretical concepts from linguistics has become increasingly advocated by primatologists studying communication. For example, previous work has highlighted the validity of applying conversation-analytic tools, along with numerous linguistic laws, to describe gestural and vocal communicative behaviour in apes (e.g. Fröhlich 2017; Heesen et al. 2019; Watson et al. 2020; Safryghin et al. 2022). Here, building on work from Leroux et al. (2022), we investigated chimpanzee signals using collocation analysis. This method quantifies whether signal combinations are non-random, defined here as the co-occurrence of signal components with a frequency that is statistically above chance. In contrast to observational data and field experiments that focus on receiver responses to gauge if a single combination produced is communicatively relevant (e.g. Hobaiter et al. 2017; Wilke et al. 2017; Leroux et al. 2023), collocation analysis provides an overview of combinatoriality across the whole repertoire and therefore can help in identifying candidate, potentially meaningful combinations in a species. In line with related approaches such as network analyses (see Aychet et al. 2021), our study therefore represents an additional novel data-driven quantification of non-random multi-modal combinations, illustrating which vocalizations and non-vocal behaviours are systematically associated at frequencies above chance in a non-human primate.

In addition to establishing a repertoire of non-random vocal-visual combinations, we aimed to further examine variation in multi-modal combinations within the population. Previous research has highlighted substantial inter-individual flexibility in primate signal production across demographic classes (Tomasello et al. 1994; Luef and Pika 2019) and social contexts (Clark 1993; Bouchard and Zuberbühler 2022) and as a function of audience presence (Schel et al. 2013; Hobaiter et al. 2017). In particular, these studies have implicated demographic factors such as age, sex and rank in driving variation in both gestural and vocal behaviour. For example, female chimpanzees are known to produce a higher rate of call combinations than their male counterparts (Leroux et al. 2022), while highest-ranking males were shown to be the most prolific gesture producers (Hobaiter et al. 2017). However, in contrast, Wilke et al. (2017) found no effects of age, rank or sex on the rate of multi-modal signal production in chimpanzees. Given the mixed evidence for demographic factors influencing the rate of unimodal and multi-modal signal production, we explored whether demographic factors affected the number of visual components individuals combined with vocal signals. We also examined low-level explanations (i.e. call duration and call type) for variation in number of NVBs produced alongside a call. Longer vocalization bouts might provide more opportunity for movements, changes in body posture or gestures to be produced. Finally, as call types produced in evolutionarily urgent events such as agonistic encounters may be associated with more NVBs to reduce ambiguity, we tested whether NVB production was influenced by call type.

Methods

Study site and data collection

The study was conducted on wild chimpanzees from the Kanyawara community in Kibale national park, Uganda (Thompson et al. 2020). The population consists of ~60 individuals inhabiting a home range of ~15km^2. The Kanyawara community has been the object of long-term study since 1987 and is entirely habituated. The data used in this study were collected between February-May 2013, and between June 2014 and March 2015 (Wilke et al. 2017). These data consist of video-audio recordings collected within the chimpanzee home range, between 0800 and 1900 h. The equipment included a hand-held camcorder (Panasonic HDC-SD90), and an external microphone (Sennheiser MKE 400).

The individuals observed in this study were 13 females and 14 males, between 10 and 48 years of age. Individuals were recorded from a distance of at least 7 m while engaged in their natural behaviour. Focal animal sampling was employed (Altmann 1974), involving 15 min of continuous video observation of one single animal, with the aim of capturing a clear and complete view of the animal and all its behaviours, including communication. Focal animals were only sampled once a day. Initially focal subjects were chosen on the basis of visibility and ease of pursuit to ensure high-quality recordings. Later in the study period, priority to certain subjects was given in order to homogenize the total focal time across individuals. Thirty-one hours of video/audio data were used in this study. It was not possible to record data blind because our study involved focal animals in the field.

Data extraction: the vocal-visual combinations

Subsequent data extraction was carried out on the video/audio recordings using Noldus Observer XT 10 events logging software (http://www.noldus.com/animal-behaviour-research). The annotation of video/audio footage was centered around events of vocal production ($N=297$). For each of these events, the researcher coded information on both the vocal as well as the visual components of signal production.

Vocalizations were classified according to the call types described in existing chimpanzee repertoires and specific empirical studies (Slocombe and Zuberbühler 2010; Crockford et al. 2018). Of the ~13 call types described in the repertoires, this study focused on the seven most commonly produced: grunt, soft hoo, pant bark, pant grunt, pant hoot, scream and whimper. The minimum number of occurrences necessary for a call to be included in the analyses was 5. In the case of the calls "grunt" and "soft hoo", the existing literature describes different call subtypes, whereby "soft hoo" can be divided into "travel hoo", "rest hoo" and "alarm hoo" (Crockford et al. 2018), while "grunt" can refer to "rough grunt" or "general grunt" (Slocombe and Zuberbühler 2010). Here however, all respective subtypes were lumped into the broad categories of "soft hoo" and "grunt". Rough grunts and general grunts were collapsed given that our sample only included low-frequency rough grunts, which are acoustically similar to general grunts. High-pitched rough grunts and rare call types did not occur in the available video-audio footage with sufficient frequency to be included in this study. Additional call types that were not observed at least 5 times and therefore not included in the study were the following: bark, waa bark, pant, cough, wraa, laughter, squeak. The number of events observed for the seven call types included ranged from 5 to 98. Chimpanzee vocalizations are often produced in bouts. A bout was defined as a sequence of the same call type with pauses shorter than 10s between the individual acoustic elements. A bout was considered terminated when followed by 10s of silence or by the production of a different call type. Bouts constituted single data points. The duration of vocal bouts ranged between 1 and 62 s.

In association with each vocal event, between 1 and 8 non-vocal behaviours (NVBs) were recorded. NVBs were only annotated if the behaviour occurred during the vocalization or vocal bout. A total of 31 different NVB types were recorded in this study. Table 1 provides the full list of NVBs annotated in this study, as well as a description of the behavioural criteria used to assign each NVB type. The list of NVBs illustrated in Table 1 includes behaviours such as feed, groom and rest. Whilst previous research has used the concurrent behaviour of the vocalizing individual to determine the broader behavioural context in which communication is produced (e.g. Wilke et al. 2017), these behaviours also potentially provide conspecific receivers with valuable visual information in the same manner as a body posture such as "sit", or a gaze orientation such as "look towards". Therefore, we include such behaviours as NVBs. NVBs included in this list represent an attempt to illustrate the observable variation in visual behaviour, and the level of granularity takes into account the risks of an over-representation of NVBs, general feasibility in coding, and complying with inter-observer reliability. Additional measures taken to maximally standardize the annotation procedure can be found in the ESM.

Data extraction: demographic context of the vocalization

In addition to describing vocal signals and accompanying NVBs, demographic data were annotated for each event. Specifically, identity and sex of the individual were noted and each individual's age in years was calculated based on the long-term data which includes birth dates for all IDs (Thompson et al. 2020). Next, dominance ranks were calculated using an Elo-rating method (Muller et al. 2021; Wilke et al. 2022) based on the long-term data on aggressive interactions and submissive pant grunt vocalizations (de Vries et al. 2006). Rank scores were calculated every 3 months and ranged between 1 and 24.

Table 1 Full list of NVBs (non-vocal behaviour) annotated in this study with corresponding behavioural description used to assign NVBs. The term "specific individual" refers to the individual who is closest to the signaler

NVB name	NVB description
rest	signaler is lying down or in resting position with chest or back touching the ground
sit	signaler sits with bottom touching ground, chest or back are not touching ground
get_up	signaler transitions from lying or sitting position to standing or walking
stand	signaler is in erect quadrupedal position without movement
walk	signaler moves quadrupedally by more than 1 meter
run	quadrupedal movement that occurs at a faster pace than normal walking, often gallop-like appearance with both feet in the air at once
climb	signaler moves up, down or along the trunk or branch of a tree
look_towards	head orientation is shifted toward specific individual by at least 90 degrees resulting in specific individual being in line of sight of signaler
look_away	head orientation shifted away from specific individual by at least 90 degrees
gaze_upwards	head orientation is shifted towards the canopy/sky
gaze_alternation	head orientation changes 3 or more times by approximately 90 degrees
turn_body_towards	body orientation changed by at least 90 degrees in direction of specific individual
turn_body_away	body orientation is shifted away from specific individual by at least 90 degrees
extend_body_towards	signaler moves chest, back or bottom toward a specific individual but legs do not usually move
retract_body	signaler's body axis connecting hips to head either changes angle or moves away from specific individual
crouch_down	signaler brings bottom, body or shoulders close to the ground
present_back	signaler orients back and bottom toward a specific individual by at least 90 degrees
arm_reach	arm is fully or partially extended towards a specific individual with or without contact
arm_wave	arm performs repetitive back and forth or side to side motion
scratch_self	fingers perform loud scratching gesture against any body surface
approach	signaler moves in direction of specific individual with 45 degree accuracy on either side
embrace	arms or legs are wrapped around a specific individual with degree of surface body contact consisting in at least hand/foot + forelimb
chase	signaler runs or climbs quickly after a specific individual in aggressive manner
hit	hand or foot is moved aggressively with the intent to make contact with body part of another individual
grab_branch	tree branch is grabbed and shaken or dragged along the floor while running or displaying
slap_ground	hands or feet are brought violently against the ground to produce a smacking noise, sometimes repeatedly
feed	signaler grabs food items and places in mouth, or chews food items already in mouth
groom	signaler probes own hair or that of other individual and extracting small particles, using one or both hands
play	signaler interacts with another individual via non-aggressive grabbing, biting, chasing, climbing, tickling
relaxed_open_mouth_face	open mouth with intermediate separation between upper and lower jaw, while engaged in play
scream_face	wide open mouth with maximum separation between upper and lower jaw, lip corners pulled up, teeth bared

Inter-observer reliability

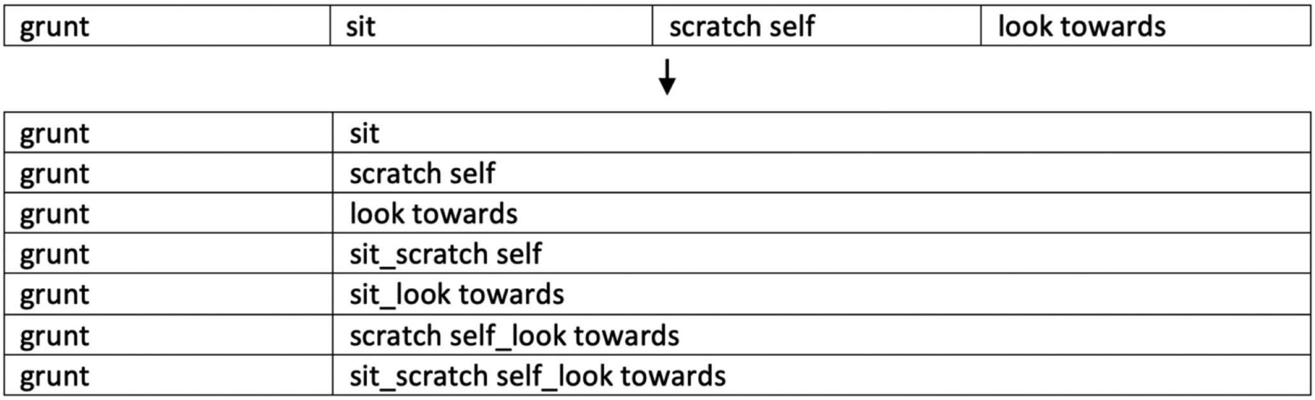
To ensure videos were coded reliably, a second independent researcher coded 11% of the events (i.e. 34 events out of 297) and extracted both (i) the call type (at least one call for each call type was present in the subset) and (ii) non-vocal behaviours (at least one instance of each NVB type was coded in the subset). We calculated a Cohen's kappa value of 0.82 and 0.88 for vocalisation type and NVB type respectively, indicating excellent levels of agreement in both cases (Fleiss 1981).

Collocation analysis

To generate a vocal-visual combination repertoire based on the communicative events observed, we implemented a collocation analysis in R (R Development Core Team 2009). This method, originating in the field of linguistics and recently adapted to the study of animal communication, estimates the relative attraction between communicative units, based on how frequently they co-occur in the dataset (Bosshard et al. 2022). We specifically implemented a Multiple Distinctive Collocation Analysis (MDCA; Bosshard et al. 2022) given that this method has been shown to account for datasets that are small and skewed (Gries and Stefanowitsch 2004; Hilpert 2006; Gries 2014). This feature was especially relevant in our analysis given that certain call types occurred infrequently (pant bark $N=7$ and whimper $N=5$), which was therefore explicitly accounted for via implementing the MDCA method. Using MDCA, we examined the co-occurrence of a particular vocal signal with a specific visual component. For example, if "grunt" + "arm reach" co-occur, collocation analysis compares the frequency of "grunt + arm reach" with the frequency of all other vocal-visual combinations which contain either "grunt" or "arm reach". In multiple distinctive collocation analysis, the association between units is calculated using binomial probability values on each possible two-unit combination. Specifically, the binomial probability mass function renders probability values for each unit with each other unit, and the log-transformed results provide a numerical estimate of how exclusively units combine with one another. Values above 1.3 represent co-occurrence at above-chance level, while values below -1.3 represent significant repulsion between collocates.

A feature of the communicative events included in this dataset is that one vocal signal commonly co-occurs with more than one NVB simultaneously. For example, a "grunt" vocalization may co-occur with a "sit" posture, a "scratch self" gesture and a "look towards" movement within the same event. Consequently, our analysis aimed to investigate not only the above-chance occurrence of vocalizations and NVBs individually, but also the association between a given call and multiple NVBs at once. Therefore, a modified collocation analysis was designed to test the association between one call and up to four concomitant NVBs. A threshold of 4 was chosen as 93% of events exhibited between 1 and 4 NVBs. In order to test associations between vocalizations and NVBs at all levels of combination, each event where >1 NVB occurred was entered into the dataset multiple times; first with each NVB individually, and then with all possible combinations of two, three and four NVBs given the NVBs present in that event. When such combinations were entered into the data table, this was done while maintaining the two-column structural requirement of collocation analyses. As shown in Table 2, the event "grunt + sit + scratch self + look towards" generated 7 separate rows of data to account for each unique combination within the communicative event. Given that what is important in MDCA is the frequency of specific combinations and not the frequency at which any given element appears within the sample, the inclusion of several entries per event does not lead to an inflation of collocation strengths. Instead, the inclusion of multiple entries per event results in a greater extent of recombination,

Table 2 Illustration of procedure for entering each communication event into a suitable dataset for implementing the multiple-NVBs collocation analysis



grunt	sit	scratch self	look towards

↓

grunt	sit
grunt	scratch self
grunt	look towards
grunt	sit_scratch self
grunt	sit_look towards
grunt	scratch self_look towards
grunt	sit_scratch self_look towards

which has been shown to subsequently decrease collocation strength (Bosshard et al. 2022). As such, we regard our MDCA results to be a conservative estimate of the combinations of vocal signals alongside non-vocal behaviours.

Statistical analyses: demographic and call-related drivers of NVB production

To examine variation in the number of NVBs produced alongside vocalizations as a function of demographic variation and call characteristics (i.e. call type and call duration), we performed a generalized linear mixed model (GLMM) with a negative binomial error structure and log link function using the glmmTMB function, glmmTMB package in R. We modeled the number of NVBs produced per event as a numerical integer response variable. As demographic predictors, we fitted age (years) as a second-order polynomial, sex as a binary categorical variable (M/F) and rank as a numerical integer. As call-related predictors, we fitted call type as a 7-level categorical variable, and duration of call bout (seconds) as a numerical predictor. Given that the effect of call type and duration may not be independent, an interaction term was fitted between these predictors. Individual identity was fitted as a random factor to account for multiple events from single individuals.

We first compared the full model including all predictors and random effects with a null model which was identical in structure minus the predictors, for which we report a likelihood ratio test (chi-squared statistic and p-value). We ascertained the relative contribution of each variable to the model by comparing the full model to a reduced model lacking each individual predictor in turn. We then report chi-squared values of likelihood ratio tests regarding the effect of each individual predictor, as well as p-values using a 95% significance threshold.

Model assumptions were checked using the DHARMa package in R. The model was not found to exhibit overdispersion (nonparametric dispersion test $P = 0.74$), no outliers were detected ($P = 0.4$) and visual inspection of the Q-Q plots confirmed normality (K-S test: $P = 0.77$).

Results

Vocal-visual repertoire via collocation analysis

Following collocation analyses, 108 combinations of one vocal signal and between 1 and 4 NVBs were found to co-occur significantly more frequently than expected by chance (all p-values < 0.05). The number of significant combinations varied between call types: for example, four combinations were documented for the "pant bark" call, six for the "scream", 11 for the "whimper", 16 for the "soft hoo", 22 for the "pant grunt", 24 for the "pant hoot" and 25 combinations for the "grunt" call. Of the 31 NVB types present in the raw data, 21 featured in significant combinations with vocal signals. Eighteen out of these 21 NVB types (i.e. 86%) were recombined productively across multiple call types. The full set of significant combinations which constitute the vocal-visual repertoire is presented in Tables 3 and 4.

Demographic and call-related drivers of NVB production

Our GLMM analysis indicated that the full model, including all predictors, explained significantly more variation in the response variable compared to a null model ($\chi^2_{16} = 38.96$, $p = 0.001$). Likelihood ratio tests revealed that there was no significant main effect of age ($\chi^2_2 = 1.39$, $p = 0.497$), sex ($\chi^2_1 = 1.25$, $p = 0.262$) or rank ($\chi^2_1 = 1.29$, $p = 0.254$) on the number of NVBs produced per vocalization. However, there was a significant interaction between call type and duration ($\chi^2_6 = 19.68$, $p = 0.003$), such that the effect of duration on the number of NVBs differed between call types. Longer call duration was associated with more NVBs in "pant grunt", "pant hoot" and "soft hoo" calls, while no such effect was observed in the other call types. Overall, the "pant grunt" call was produced in association with the most NVBs while the "scream" was associated with the fewest, as shown in Fig. 1.

Discussion

By systematically observing naturally occurring communication events, and through implementing frequency-based analyses, we show that chimpanzees combine their vocal signals in non-random ways with a wide range of body movements, behaviours, postures, gestures and facial expressions, collectively referred to here as non-vocal behaviours (NVBs). More than 100 such combinations of vocal and visual components occur more frequently than expected by chance, indicating a strikingly diverse repertoire of vocal-visual combinations. Some NVBs are used productively across multiple call types, yet each call type is associated with its own set of single and combined NVBs. Further, variation in the number of NVBs produced per vocalization is not predicted by demographic variables such as age, sex or rank.

It has not escaped our attention that within the ~100 combinations found to occur most frequently, approximately two thirds involve visual behaviours that need not necessarily be interpreted as signals, i.e. evolved for the purpose of communication, but rather as visual cues (see ESM for

Table 3 List of 31 single NVBs (non-vocal behaviour) and 7 call types included in this analysis. Colour codes denote strength of attraction/repulsion between NVBs and each call type: darkest green = strongest attraction, darkest red = strongest repulsion. All values above 1.3 represent co-occurrence at above-chance level with 95% confidence interval, while values below – 1.3 represent significant repulsion between components

	grunt	pant bark	pant grunt	pant hoot	scream	soft hoo	whimper
approach	-1.4	-0.447	4.166	0.464	-1.401	-1.288	0.612
arm reach	-0.222	-0.149	-0.384	0.47	0.241	0.591	-0.012
arm wave	-0.074	-0.05	0.593	-0.06	-0.124	-0.043	-0.004
chase	-0.074	-0.05	-0.128	-0.06	0.606	-0.043	-0.004
climb	1.639	-0.747	-0.396	-0.967	0.41	0.302	-0.114
crouch	-0.37	0.361	0.961	-0.3	-0.195	-0.215	-0.02
embrace	0.306	-0.198	-0.137	-0.24	0.168	0.486	-0.016
extend body towards	-0.591	0.676	1.524	-0.48	-0.428	-0.344	-0.032
feed	6.839	-1.885	-3.721	-1.457	-4.697	6.955	-0.154
gaze alternation	-0.503	-0.785	-0.442	2.611	-0.918	-0.321	2.611
gaze upwards	2.478	-0.178	-1.409	-0.659	-1.36	1.822	-0.045
get up	-0.435	-0.978	0.439	2.058	-2.339	1.021	0.565
grab branch	-0.222	-0.149	-0.384	2.669	-0.371	-0.129	-0.012
groom	7.08	-1.687	-3.253	-1.258	-4.202	5.202	-0.138
hit	-0.148	-0.099	-0.256	-0.12	1.212	-0.086	-0.008
look away	1.143	-0.295	-0.654	0.226	-1.08	0.791	0.884
look towards	4.517	-0.851	-1.121	-0.706	-3.454	3.086	0.286
play	0.54	-0.099	0.351	-0.12	-0.247	-0.086	-0.008
relaxed open mouth face	0.805	-0.05	-0.128	-0.06	-0.124	-0.043	-0.004
present back	0.52	-0.347	1.121	0.208	-0.865	-0.301	-0.028
rest	4.008	-0.595	-1.537	-0.276	-1.483	1.016	-0.049
retract body	-1.035	1.896	0.303	-0.839	0.599	-0.601	-0.057
run	-1.331	0.227	-1.45	4.687	-0.277	-0.773	0.81
scratch	1.978	-0.635	-1.064	0.671	-3.09	2.164	-0.101
scream face	-0.37	-0.248	-0.641	-0.3	3.031	-0.215	-0.02
sit	8.889	-6.517	-7.357	-0.27	-17.692	25.456	1.054
slap ground	-0.148	-0.099	-0.256	0.618	0.363	-0.086	-0.008
stand	2.01	-0.492	-0.661	-0.269	-1.119	1.02	-0.17
turn body away	0.319	0.53	-0.293	-0.599	0.66	-0.429	-0.041
turn body towards	2.638	-0.843	-0.819	-0.473	-0.234	-0.288	0.833
walk	1.557	-2.945	-0.339	2.535	-5.24	2.733	-0.397

Table 4 All combinations of vocal and visual components that were found to co-occur more frequently than expected by chance

grunt	pant hoot	pant grunt	soft hoo	whimper	scream	pant bark
climb	approach_get up_sit	approach	feed	approach_get up_look tow_run	look towards_scream face	extend body_look tow_retract body
climb_feed	approach_get up_sit_walk	approach_extend body	feed_sit	approach_get up_run	scream face	look tow_retract body
feed	gaze alternation	approach_extend body_getup	feed_sit_stand	approach_look tow_run	scream face_stand	look tow_retract body_sit
feed_sit	gaze alternation_get up	approach_extend body_get up_walk	feed_stand	approach_run	scream face_turn body tow	retract body
feed_walk	gaze alt_get up_sit	approach_extend body_walk	gaze upwards	gaze alternation	stand_turn body towards	
gaze upwards	gaze alt_get up_sit_walk	approach_get up	gaze upwards_sit	gaze alternation_sit	stand_turn body tow_walk	
gaze upwards_sit	gaze alt_get up_walk	approach_get up_walk	groom	gaze alt_sit_turn body towards		
groom	gaze alternation_scratch	approach_look towards_walk	groom_sit	gaze alternation_turn body tow		
groom_sit	gaze alt_scratch_sit	approach_present back	look towards	get up_look towards_run		
look away_stand_walk	gaze alternation_sit	approach_present back_walk	look tow_scratch self	get up_run		
look towards	gaze alternation_sit_walk	approach_walk	look tow_scratch_sit	look towards_run		
look towards_rest	get up	climb_gaze alternation_stand	look towards_sit			
look towards_sit	get up_sit	climb_gaze alternation_stand_walk	scratch self			
look tow_sit_turn body towards	get up_sit_walk	climb_gaze alternation_walk	scratch self_sit			
look tow_sit_turn body tow_walk	get up_walk	climb_stand	sit			
look towards_turn body towards	grab branch	climb_stand_walk	walk			
rest	grab branch_run	climb_walk				
scratch self	grab branch_run_walk	extend body				
scratch self_sit	grab branch_walk	extend body_get up				
sit	run	extend body_get up_walk				
sit_turn body towards	run_walk	extend body_walk				
sit_turn body towards_walk	scratch self_walk	present back_walk				
stand	sit_walk					
turn body towards	walk					
walk						

a separation of combinations containing visual signals and those containing only visual cues). Whilst signals and cues are traditionally considered separately, we argue in favor of casting the net wide and taking both together, asserting that the traditional stark dichotomy between signals and cues may benefit from being represented more as a graded continuum. For instance, in our dataset, bodily movements such as standing up, retracting or extending the body, whilst superficially cue-like, could be deployed for communication purposes, signaling willingness to initiate travel, or submission. Thus, we maintain that not only signals, but all visual information offered by the vocalizer constitutes an integral component of the multi-modal communicative interaction. Ultimately, carefully designed experiments using captive animals and video playbacks are needed to understand the relative importance of cues or signals being combined with vocalizations (Waller et al. 2016; Whitehouse et al. 2016).

Given the findings of the collocation analysis, it appears that sub-adult and adult chimpanzees have access to a

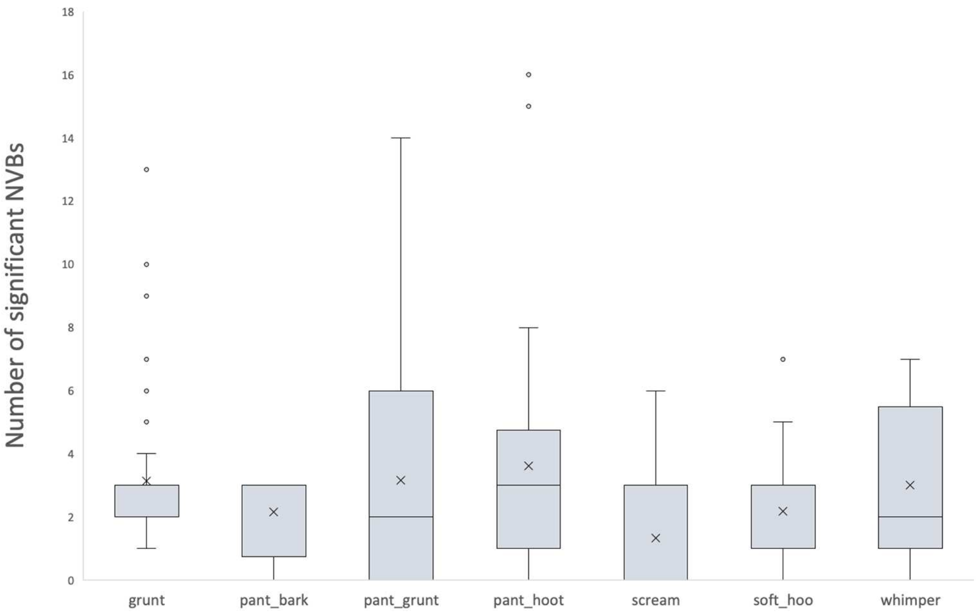


Fig. 1 Raw data illustrating variation in the number of NVBs (non-vocal behaviours) produced at significantly above chance frequencies with the different call types analysed in this study. Boxes indicate the inter quartile range (IQR), the central line depicts the median and the whiskers extend to the maximum and minimum values excluding outliers. Circles represent outliers, and crosses represent means for each call type

highly diversified repertoire of combined visual and vocal components. Although the constrained vocal repertoire of chimpanzees (Seyfarth and Cheney 2010) might suggest a limited capacity for information transfer, the free and productive use of accompanying NVBs instead reveals a high potential for refining the meaning of the limited range of available calls. This finding was enabled by the application of methods developed in the field of linguistics, further adding to the growing body of work which highlights the importance of inter-disciplinary integration between linguistic concepts and biological research (Semple et al. 2022).

A fundamental implication of this investigation is that unimodal approaches to primate communication, which analyze vocal or visual components separately, result in a drastically oversimplified picture of flexibility in signal production. A multi-modal approach is therefore crucial to accurately represent the communicative abilities and real-life communicative exchanges of non-human primates (Slocombe et al. 2011; Liebal et al. 2022), as well as other species outside the primate taxon. For instance, many non-primate examples of communication traditionally regarded as unimodal could be revisited using the broad inductive approach outlined here as a means to identify non-random cues that may enhance receiver reception of signals or otherwise modify signal meaning.

Chimpanzee social life is characterized by a wide variety of interactions, each of which is typically mediated by communication. Thus, it is likely that the suite of combined vocal and visual components identified here plays a key role in supporting the demands of a chimpanzee's daily social life (Freeberg et al. 2012; Bouchet et al. 2013). Now that we have identified a suite of systematically re-occurring vocal-visual combinations, further investigation may shed light on the communicative function of combining vocal and visual components, especially with regard to the function of "emergence" whereby new meaning is created (Partan and Marler 2005). Importantly, we have identified potentially meaningful vocal-visual combinations using a data-driven, bottom-up approach rather than by selecting combinations a priori. Consequently, we may now begin to appreciate the full diversity of possible nuances that can be generated by combining a given vocalization with many different non-vocal behaviours, as in the case of the "grunt", which was associated with the longest suite of NVBs in our study. Vocalizations described as "grunt" are common across many primate species (Cheney and Seyfarth 1982; Gamba and Giacoma 2007; Salmi et al. 2013; Katsu 2016; Silk et al. 2018) and are involved in a variety of interaction types including reconciliation (Cheney et al. 1995), travel initiation (Sperber et al. 2017) and more (Rendall et al. 1999; Whitham et al. 2007). Within chimpanzees, "grunt" is a call type that can be used in many different contexts (Slocombe and Zuberbühler 2010). Therefore, this vocalization represents an appropriate model for a signal whose meaning may be disambiguated by integrating the signaler's body posture, gaze, movements and gestures. For example, our results show that "grunt" is commonly combined with "sit" or "rest", but also with "stand", "walk" or "climb". Such substantial variation in body postures is likely to correlate with distinct behavioural contexts of emission and thus engender contrasting outcomes on behalf of receivers. Similarly, if we compare the "grunt + look towards + sit + turn body towards" combination reported in our collocation results with the alternative "grunt + look away + stand + walk", it is reasonable to assume that the respective chimpanzee receivers in these circumstances might exhibit divergent behavioural responses. As such, simply by inspecting the variation in NVBs associated with each call type, we may already begin to infer the communicative importance of combining vocal and visual elements in chimpanzees.

It is unknown whether chimpanzee signalers voluntarily combine vocal signals with all of the NVBs reported in this study. Nonetheless, chimpanzee receivers likely rely on the integration of all the vocal and visual information available in order to guide their own adaptive behavioural response (Seyfarth and Cheney 2003). Confirming this hypothesis requires further investigation into how NVBs are perceived by receivers and their potential role in the disambiguation of meaning. One fruitful method to address meaning empirically involves a systematic analysis of behavioural reactions to signals as a function of signal type (Hobaiter and Byrne 2014), showing, for example, that a chimpanzee grunt vocalization combined with a gesture is more likely to exhibit a response than if the grunt is produced on its own (Wilke et al. 2017). Furthermore, recent developments which combine insights from linguistics and animal behaviour offer new theoretical frameworks and empirical toolkits for addressing the meaning of signal components in nonhumans (Berthet et al. 2022). Berthet and colleagues illustrate how semantics and pragmatics, key concepts in linguistics which relate to meaning interpretation, can be adapted to primatological research to achieve a more detailed picture of signal meaning in primates. This valuable approach could be applied to the wide range of vocalization and NVB combinations highlighted in this study, offering critical insights into the meaning of chimpanzee vocal-visual combinations (see also Wheeler and Fischer 2012). A complementary avenue of investigation is to infer which cues are most salient to recipients for meaning disambiguation, using measures of attentional bias during video playbacks. The application of eye-tracking technology in captive great apes, for example, has enjoyed a recent surge of advances, bringing this goal confidently within reach (Kano et al. 2017).

Our study also investigated variation in the number of NVBs produced per vocalization as a function of individual demographic attributes such as age, sex and rank. However, males and females did not differ in the number of NVBs produced, nor was the observed variation explained by age or rank. These results broadly support Wilke et al.'s (2017) finding that demographic factors did not influence rate of multimodal signal production but are in contrast to Leroux et al.'s (2022) finding that females and low-ranking chimpanzees produce more vocal combinations. A possible implication of this result is that unlike vocal sequences comprising two acoustic items, combining components from different modalities may predominantly serve a more general function such as that of redundancy or efficacy-backup (Hebets and Papaj 2005), which is crucial irrespective of demographic status. Replicating this work in other communities of chimpanzees would prove useful for establishing the universality of this finding. Indeed, it remains possible that a population which experiences different ecological or social pressures, may display more pronounced demographic patterns in NVB production than those observed here. In contrast to demographic factors, we found a significant effect of call duration on number of NVBs produced, but this effect was conditional on call type, such that longer vocalization events were associated with a greater number of NVBs in some call types but not in others. Interestingly, a positive effect of duration was observed most strongly in pant hoot, soft hoo and pant grunt vocalisations, not in screams or pant barks which are associated with more evolutionarily urgent events, such as agonistic encounters or threatening interactions with dominant individuals. Given the urgency of these contexts, it is possible that the NVBs associated with screams and pant barks tend to be produced within the first few seconds of the vocalization bout, while NVB production in more relaxed and nuanced contexts can be more temporally protracted.

In conclusion, our findings reveal a hitherto unappreciated diversity of vocal-visual combinations in the communication of wild chimpanzees. Follow-up behavioural observations and experimental work are key to unpacking the function and meaning of such combinations. Nonetheless, the extent and variety of non-random vocal-visual combinations described here broadens our understanding of the potential combinatorial information available to receivers in our closest-living relative. Furthermore, ~90% of the visual components of communicative exchanges observed in this study were shown to be produced in association with multiple call types. In line with previous work, this is suggestive that multi-modal events represent combinatorial structures, of which vocal and visual components constitute the building-blocks (Davila-Ross et al. 2015). By virtue of our phylogenetic proximity to chimpanzees, the range of vocal-visual combinations presented here also informs our understanding of the communicative behaviour of our hominin ancestors, corroborating convergent evidence that multi-modal communication has an ancient history in our primate lineage (Evans et al. 2005; Kulahci et al. 2014) and supporting the notion that complex multi-modal signaling may have played a role in scaffolding language evolution (Wheeler and Fischer 2012; Seyfarth and Cheney 2014, 2017).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00265-024-03523-x.

Acknowledgements We are grateful to the directors of Kibale Chimpanzee Project for permitting and supporting us to carry out this research on the Kanyawara community of chimpanzees. We are also thankful to the KCP field manager Emily Otali and the KCP field assistants, Dan Akaruhanga, Seezi Atwijuze, Sunday John, Richard Karamagi, James Kyomuhendo, Francis Mugurusi, Solomon Musana and Wilberforce Tweheyo, for their valuable assistance and support in the field. We appreciate the permission of the Uganda National Council for Science and Technology, the President's Office and the Uganda Wildlife Authority for us to carry out this study in Uganda. We thank Piera Filippi for her constructive comments. We thank Edoardo Properzi for assistance with data processing. We thank the reviewers for insightful input which has contributed to improving the final version of this manuscript.

Funding This work was supported by the Swiss National Science Foundation (PP00P3_198912) to SWT, by a European Research Council Consolidator Grant (724608) to KES, by National Science Foundation Grants (NSF 0849380 and NSF 1355014) to ZPM and by the NCCR Evolving Language (SNSF Agreement #51NF40_180888). Open access funding provided by University of Zurich

Data availability All data are available in the supplementary materials.

Declarations

Ethical approval This study complied with the ASAB/ABS guidelines for the use of animals in research; ethical approval was granted by the Biology Ethics Committee (University of York). The Ugandan Wildlife Authority and the Ugandan National Council for Science and Technology granted consent to carry out the data collection in Uganda.

Conflict of interests Authors declare that they have no competing interests.

References

Allen CR, Croft DP, Testard C, Brent LJ (2021) Function of trunk-mediated greeting behaviours between male African elephants: insights from choice of partners. Animals 11:2718. https://doi.org/10.3390/ani11092718

Altmann J (1974) Observational study of behavior: sampling methods. Behaviour 49:227–267. https://doi.org/10.1163/156853974x00534

Aychet J, Blois-Heulin C, Lemasson A (2021) Sequential and network analyses to describe multiple signal use in captive mangabeys. Anim Behav 182:203–226. https://doi.org/10.1016/j.anbehav.2021.09.005

Berthet M, Coye C, Dezecache G, Kuhn J (2023) Animal linguistics: a primer. Biol Rev 98:81–98. https://doi.org/10.1111/brv.12897

Bosshard AB, Leroux M, Lester NA, Bickel B, Stoll S, Townsend SW (2022) From collocations to call-ocations: using linguistic methods to quantify animal call combinations. Behav Ecol Sociobiol 76:122. https://doi.org/10.1007/s00265-022-03224-3

Bouchard A, Zuberbühler K (2022) Male chimpanzees communicate to mediate competition and cooperation during feeding. Anim Behav 186:41–55. https://doi.org/10.1016/j.anbehav.2022.01.009

Bouchet H, Blois-Heulin C, Lemasson A (2013) Social complexity parallels vocal complexity: a comparison of three non-human primate species. Front Psychol 4:390. https://doi.org/10.3389/fpsyg.2013.00390

Cheney DL, Seyfarth RM (1982) How vervet monkeys perceive their grunts: field playback experiments. Anim Behav 30:739–751. https://doi.org/10.1016/S0003-3472(82)80146-2

Cheney DL, Seyfarth RM, Silk JB (1995) The role of grunts in reconciling opponents and facilitating interactions among adult female baboons. Anim Behav 50:249–257. https://doi.org/10.1006/anbe.1995.0237

Clark AP (1993) Rank differences in the production of vocalizations by wild chimpanzees as a function of social context. Am J Primatol 31:159–179. https://doi.org/10.1002/ajp.1350310302

Crockford C, Gruber T, Zuberbühler K (2018) Chimpanzee quiet hoo variants differ according to context. R Soc Open Sci 5:172066. https://doi.org/10.1098/rsos.172066

Davila-Ross M, Jesus G, Osborne J, Bard KA (2015) Chimpanzees (*Pan troglodytes*) produce the same types of 'laugh faces' when they emit laughter and when they are silent. PLoS ONE 10:e0127337. https://doi.org/10.1371/journal.pone.0127337

de Vries H, Stevens JM, Vervaecke H (2006) Measuring and testing the steepness of dominance hierarchies. Anim Behav 71:585–592. https://doi.org/10.1016/j.anbehav.2005.05.015

Evans TA, Howell S, Westergaard GC (2005) Auditory-visual cross-modal perception of communicative stimuli in tufted capuchin monkeys (*Cebus apella*). J Exp Psychol Anim B 31:399–406. https://doi.org/10.1037/0097-7403.31.4.399

Fitch WT, de Boer B, Mathur N, Ghazanfar AA (2016) Monkey vocal tracts are speech-ready. Sci Adv 2:e1600723. https://doi.org/10.1126/sciadv.1600723

Fleiss JL (1981) Balanced incomplete block designs for inter-rater reliability studies. Appl Psychol Meas 5:105–112. https://doi.org/10.1177/014662168100500115

Freeberg TM, Dunbar RI, Ord TJ (2012) Social complexity as a proximate and ultimate factor in communicative complexity. Phil Trans R Soc B 367:1785–1801. https://doi.org/10.1098/rstb.2011.0213

Fröhlich M (2017) Taking turns across channels: conversation-analytic tools in animal communication. Neurosci Biobehav Rev 80:201–209. https://doi.org/10.1016/j.neubiorev.2017.05.005

Fröhlich M, van Schaik CP (2018) The function of primate multimodal communication. Anim Cogn 21:619–629. https://doi.org/10.1007/s10071-018-1197-8

Fröhlich M, Wittig RM, Pika S (2016) Should I stay or should I go? Initiation of joint travel in mother–infant dyads of two chimpanzee communities in the wild. Anim Cogn 19:483–500. https://doi.org/10.1007/s10071-015-0948-z

Fröhlich M, Sievers C, Townsend SW, Gruber T, van Schaik CP (2019) Multimodal communication and language origins: integrating gestures and vocalizations. Biol Rev 94:1809–1829. https://doi.org/10.1111/brv.12535

Gamba M, Giacoma C (2007) Quantitative acoustic analysis of the vocal repertoire of the crowned lemur. Ethol Ecol Evol 19:323–343. https://doi.org/10.1080/08927014.2007.9522555

Genty E, Breuer T, Hobaiter C, Byrne RW (2009) Gestural communication of the gorilla (*Gorilla gorilla*): repertoire, intentionality and possible origins. Anim Cogn 12:527–546. https://doi.org/10.1007/s10071-009-0213-4

Genty E, Clay Z, Hobaiter C, Zuberbühler K (2014) Multi-modal use of a socially directed call in bonobos. PLoS ONE 9:e84738. https://doi.org/10.1371/journal.pone.0084738

Ghazanfar AA (2013) Multisensory vocal communication in primates and the evolution of rhythmic speech. Behav Ecol Sociobiol 67:1441–1448. https://doi.org/10.1007/s00265-013-1491-z

Gries ST Coll.analysis 3.5. A script for R to compute perform collostructional analyses., Script R (2014) https://www.stgries.info/teaching/groningen/index.html

Gries ST, Stefanowitsch A (2004) Extending collostructional analysis: a corpus-based perspective onalternations'. Int J Corpus Linguist 9:97–129. https://doi.org/10.1075/ijcl.9.1.06gri

Grund C, Badihi G, Graham KE, Safryghin A, Hobaiter C (2024) GesturalOrigins: a bottom-up framework for establishing systematic gesture data across ape species. Behav Res Methods 56:986–1001. https://doi.org/10.3758/s13428-023-02082-9

Habbershon HM, Ahmed SZ, Cohen YE (2023) Rhesus macaques recognize unique multimodal face-voice relations of familiar individuals and not of unfamiliar ones. Brain Behav Evol 81:219–225. https://doi.org/10.1159/000351203

Hebets EA, Papaj DR (2005) Complex signal function: developing a framework of testable hypotheses. Behav Ecol Sociobiol 57:197–214. https://doi.org/10.1007/s00265-004-0865-7

Heesen R, Hobaiter C, Ferrer-i-Cancho R, Semple S (2019) Linguistic laws in chimpanzee gestural communication. Proc R Soc B 286:20182900. https://doi.org/10.1098/rspb.2018.2900

Higham JP, Hebets EA (2013) An introduction to multimodal communication. Behav Ecol Sociobiol 67:1381–1388. https://doi.org/10.1007/s00265-013-1590-x

Hilpert M (2006) Distinctive collexeme analysis and diachrony. Corpus Linguist Ling Theor 2:243–256. https://doi.org/10.1515/CLLT.2006.012

Hobaiter C, Byrne RW (2011) The gestural repertoire of the wild chimpanzee. Anim Cogn 14:745–767. https://doi.org/10.1007/s10071-011-0409-2

Hobaiter C, Byrne RW (2014) The meanings of chimpanzee gestures. Curr Biol 24:1596–1600. https://doi.org/10.1016/j.cub.2014.05.066

Hobaiter C, Byrne RW, Zuberbühler K (2017) Wild chimpanzees' use of single and combined vocal and gestural signals. Behav Ecol Sociobiol 71:96. https://doi.org/10.1007/s00265-017-2325-1

Hölldobler B (1999) Multimodal signals in ant communication. J Comp Physiol 184:129–141. https://doi.org/10.1007/s003590050313

Holle H, Gunter TC (2007) The role of iconic gestures in speech disambiguation: ERP evidence. J Cogn Neurosci 19:1175–1192. https://doi.org/10.1162/jocn.2007.19.7.1175

Kano F, Krupenye C, Hirata S, Call J (2017) Eye tracking uncovered great apes' ability to anticipate that other individuals will act according to false beliefs. Commun Integr Biol 10:e1299836. https://doi.org/10.1080/19420889.2017.1299836

Katsu N, Yamada K, Nakamichi M (2016) Function of grunts, girneys and coo calls of Japanese macaques (*Macaca fuscata*) in relation to call usage, age and dominance relationships. Behaviour 153:125–142. https://doi.org/10.1163/1568539X-00003330

Kulahci IG, Drea CM, Rubenstein DI, Ghazanfar AA (2014) Individual recognition through olfactory–auditory matching in Lemurs. Proc R Soc B 281:20140071. https://doi.org/10.1098/rspb.2014.0071

Laidre ME, Johnstone RA (2013) Animal signals. Curr Biol 23:R829–R833. https://doi.org/10.1016/j.cub.2013.07.070

Leroux M, Chandia B, Bosshard AB, Zuberbühler K, Townsend SW (2022) Call combinations in chimpanzees: a social tool? Behav Ecol 33:1036–1043. https://doi.org/10.1093/beheco/arac074

Leroux M, Schel AM, Wilke C, Chandia B, Zuberbühler K, Slocombe KE, Townsend SW (2023) Call combinations and compositional processing in wild chimpanzees. Nat Commun 14:2225. https://doi.org/10.1038/s41467-023-37816-y

Liebal K, Oña L (2018) Different approaches to meaning in primate gestural and vocal communication. Front Psychol 9:478. https://doi.org/10.3389/fpsyg.2018.00478

Liebal K, Waller BM, Slocombe KE, Burrows AM (2014) Primate communication: a multimodal approach. Cambridge University Press, Cambridge

Liebal K, Slocombe KE, Waller BM (2022) The language void 10 years on: multimodal primate communication research is still uncommon. Ethol Ecol Evol 34:274–287. https://doi.org/10.1080/03949370.2021.2015453

Luef EM, Pika S (2019) Social relationships and greetings in wild chimpanzees (*Pan troglodytes*): use of signal combinations. Primates 60:507–515. https://doi.org/10.1007/s10329-019-00758-5

Muller MN, Enigk DK, Fox SA, Lucore J, Machanda ZP, Wrangham RW, Emery-Thomson M (2021) Aggression, glucocorticoids, and the chronic costs of status competition for wild male chimpanzees. Horm Behav 130:104965. https://doi.org/10.1016/j.yhbeh.2021.104965

Müller C, Cienki A, Fricke E, Ladewig S, McNeill D, Teßendorf S (eds) (2013) Body-language-communication: an international handbook on multimodality in human interaction, vol 1. De Gruyter Mouton, Berlin, pp 131–232

Parr LA, Waller BM (2006) Understanding chimpanzee facial expression: insights into the evolution of communication. Soc Cogn Affect Neurosci 1:221–228. https://doi.org/10.1093/scan/nsl031

Partan S, Marler P (1999) Communication goes multimodal. Science 283:1272–1273. https://doi.org/10.1126/science.283.5406.1272

Partan SR, Marler P (2005) Issues in the classification of multimodal communication signals. Am Nat 166:231–245. https://doi.org/10.1086/431246

Partan S, Yelda S, Price V, Shimizu T (2005) Female pigeons, *Columba livia*, respond to multisensory audio/video playbacks of male courtship behaviour. Anim Behav 70:957–966. https://doi.org/10.1016/j.anbehav.2005.03.002

Partan SR, Fulmer AG, Gounard MAM, Redmond JE (2010) Multimodal alarm behavior in urban and rural gray squirrels studied by means of observation and a mechanical robot. Curr Zool 56:313–326. https://doi.org/10.1093/czoolo/56.3.313

R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org

Rendall D, Seyfarth RM, Cheney DL, Owren MJ (1999) The meaning and function of grunt variants in baboons. Anim Behav 57:583–592. https://doi.org/10.1006/anbe.1998.1031

Roberts AI, Roberts SGB, Vick SJ (2014) The repertoire and intentionality of gestural communication in wild chimpanzees. Anim Cogn 17:317–336. https://doi.org/10.1007/s10071-013-0664-5

Rosati AG, Hagberg L, Enigk DK, Otali E, Emery Thompson M, Muller MN, Wrangham RW, Machanda ZP (2020) Social selectivity in aging wild chimpanzees. Science 370:473–476. https://doi.org/10.1126/science.aaz9129

Rovner JS, Barth FG (1981) Vibratory communication through living plants by a tropical wandering spider. Science 214:464–466. https://doi.org/10.1126/science.214.4519.464

Safryghin A, Cross C, Fallon B, Heesen R, Ferrer-i-Cancho R, Hobaiter C (2022) Variable expression of linguistic laws in ape gesture: a case study from chimpanzee sexual solicitation. R Soc Open Sci 9:220849. https://doi.org/10.1098/rsos.220849

Salmi R, Hammerschmidt K, Doran-Sheehy DM (2013) Western gorilla vocal repertoire and contextual use of vocalizations. Ethology 119:831–847. https://doi.org/10.1111/eth.12122

Schel AM, Machanda Z, Townsend SW, Zuberbühler K, Slocombe KE (2013) Chimpanzee food calls are directed at specific individuals. Anim Behav 86:955–965. https://doi.org/10.1016/j.anbehav.2013.08.013

Scherer KR (2013) The functions of nonverbal signs in conversation. In: St Clair RN, Giles H (eds) The social and psychological contexts of language. Psychology, New York, pp 281–282

Searcy WA, Nowicki S (2010) The evolution of animal communication: reliability and deception in signaling systems. Princeton University Press, Princeton

Semple S, Ferrer-i-Cancho R, Gustison ML (2022) Linguistic laws in biology. Trends Ecol Evol 37:53–66. https://doi.org/10.1016/j.tree.2021.08.012

Seyfarth RM, Cheney DL (2003) Signalers and receivers in animal communication. Annu Rev Psychol 54:145–173. https://doi.org/10.1146/annurev.psych.54.101601.145121

Seyfarth RM, Cheney DL (2010) Primate vocal communication. In: Platt ML, Ghazanfar AA (eds) Primate neuroethology. Oxford University Press, Oxford, pp 84–97. https://doi.org/10.1093/acprof:oso/9780195326598.003.0005

Seyfarth RM, Cheney DL (2014) The evolution of language from social cognition. Curr Opin Neurobiol 28:5–9. https://doi.org/10.1016/j.conb.2014.04.003

Seyfarth RM, Cheney DL (2017) Precursors to language: social cognition and pragmatic inference in primates. Psychon Bull Rev 24:79–84. https://doi.org/10.3758/s13423-016-1059-9

Silk JB, Roberts ER, Städele V, Strum SC (2018) To grunt or not to grunt: factors governing call production in female olive baboons, *Papio anubis*. PLoS ONE 13:e0204601. https://doi.org/10.1371/journal.pone.0204601

Singletary B, Tecot S (2020) Multimodal pair-bond maintenance: a review of signaling across modalities in pair-bonded nonhuman primates. Am J Primatol 82:e23105. https://doi.org/10.1002/ajp.23105

Slocombe KE, Zuberbühler K (2010) Vocal communication in chimpanzees. In: Lonsdorf EV, Ross S, Matsuzawa T (eds) The mind of the Chimpanzee: ecological and experimental perspectives. University of Chicago Press, Chicago, pp 192–207. https://doi.org/10.7208/9780226492810-018

Slocombe KE, Waller BM, Liebal K (2011) The language void: the need for multimodality in primate communication research. Anim Behav 81:919–924. https://doi.org/10.1016/j.anbehav.2011.02.002

Sperber AL, Werner LM, Kappeler PM, Fichtel C (2017) Grunt to go—vocal coordination of group movements in redfronted Lemurs. Ethology 123:894–905. https://doi.org/10.1111/eth.12663

Taglialatela JP, Russell JL, Pope SM, Morton T, Bogart S, Reamer LA, Schapiro SJ, Hopkins WD (2015) Multimodal communication in chimpanzees. Am J Primatol 77:1143–1148. https://doi.org/10.1002/ajp.22449

Thompson ME, Muller MN, Machanda ZP, Otali E, Wrangham RW (2020) The Kibale Chimpanzee Project: over thirty years of research, conservation, and change. Biol Conserv 252:108857. https://doi.org/10.1016/j.biocon.2020.108857

Tomasello M, Call J, Nagell K, Olguin R, Carpenter M (1994) The learning and use of gestural signals by young chimpanzees: a trans-generational study. Primates 35:137–154. https://doi.org/10.1007/BF02382050

Waller BM, Whitehouse J, Micheletta J (2016) Macaques can predict social outcomes from facial expressions. Anim Cogn 19:1031–1036. https://doi.org/10.1007/s10071-016-0992-3

Watson SK, Heesen R, Hedwig D, Robbins MM, Townsend SW (2020) An exploration of Menzerath's law in wild mountain gorilla vocal sequences. Biol Lett 16:20200380. https://doi.org/10.1098/rsbl.2020.0380

Wheeler BC, Fischer J (2012) Functionally referential signals: a promising paradigm whose time has passed. Evol Anthropol 21:195–205. https://doi.org/10.1002/evan.21319

Whitehouse J, Micheletta J, Kaminski J, Waller BM (2016) Macaques attend to scratching in others. Anim Behav 122:169–175. https://doi.org/10.1016/j.anbehav.2016.10.020

Whitham JC, Gerald MS, Maestripieri D (2007) Intended receivers and functional significance of grunt and girney vocalizations in free-ranging female rhesus macaques. Ethology 113:862–874. https://doi.org/10.1111/j.1439-0310.2007.01381.x

Wilke C, Kavanagh E, Donnellan E, Waller BM, Machanda ZP, Slocombe KE (2017) Production of and responses to unimodal and multimodal signals in wild chimpanzees, *Pan troglodytes schweinfurthii*. Anim Behav 123:305–316. https://doi.org/10.1016/j.anbehav.2016.10.024

Wilke C, Lahiff NJ, Badihi G et al (2022) Referential gestures are not ubiquitous in wild chimpanzees: alternative functions for exaggerated loud scratch gestures. Anim Behav 189:23–45. https://doi.org/10.1016/j.anbehav.2022.04.007

Authors and Affiliations

Joseph G. Mine[1,2] · Claudia Wilke[3] · Chiara Zulberti[4] · Melika Behjati[5,6] · Alexandra B. Bosshard[1,2] · Sabine Stoll[1,2] · Zarin P. Machanda[7] · Andri Manser[8] · Katie E. Slocombe[4] · Simon W. Townsend[1,2,9]

✉ Joseph G. Mine
joseph.mine@uzh.ch

[1] Department of Comparative Language Science, University of Zurich, Zurich, Switzerland

[2] Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Zurich, Switzerland

[3] Department of Psychology, University of York, York, UK

[4] Institute for the Study of Human Biology and Primate Cognition, University of Leipzig, Lepizig, Germany

[5] Natural Language Understanding Group, Idiap Research Institute, Martigny, Switzerland

[6] École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[7] Departments of Anthropology and Biology, Tufts University, Medford, USA

[8] Linguistic Research Infrastructure, University of Zurich, Zurich, Switzerland

[9] Department of Psychology, University of Warwick, Warwick, UK