# Vocal-visual combinations in wild chimpanzees

**Authors:** Joseph G. Mine[1,2], Claudia Wilke[3], Chiara Zulberti[4], Melika Bejhati[5], Sabine Stoll[1,2], Zarin Machanda[6], Andri Manser[7], Katie. E Slocombe[4] and Simon W. Townsend[1,2,8]

**Affiliations:**
[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] Departments of Anthropology and Biology, Tufts University, Medford, USA
[7] Linguistic Research Infrastructure, University of Zurich, Zurich, Switzerland
[8] Department of Psychology, University of Warwick, Warwick, UK

**Abstract:**

Human communication is strikingly multi-modal, relying on vocal utterances combined with visual gestures, facial expressions and more. Recent efforts to describe multi-modal signal production in our ape relatives have shed important light on the evolutionary trajectory of this core hallmark of human language. However, whilst promising, a systematic quantification of primate signal production which filters out random combinations produced across modalities is currently lacking. Here, through recording the communicative behaviour of wild chimpanzees from the Kibale forest, Uganda we address this issue and generate the first repertoire of non-random combined vocal and visual components. Using collocation analysis, we identify more than 100 vocal-visual combinations which occur more frequently than expected by chance. We also probe how multi-modal production varies in the population, finding no differences between individuals as a function of age, sex or rank. The number of visual components exhibited alongside vocalizations was, however, associated with vocalization type and duration. We demonstrate that chimpanzees produce a vast array of combined vocal and visual components, exhibiting a hitherto underappreciated level of combinatorial complexity. We conclude that a multi-modal approach is crucial to accurately representing the communicative abilities of non-human primates.

## Introduction

Human communication, which is crucial to our daily lives, is an inherently multi-component system [1]. When speaking, humans typically accompany their utterances with gestures, facial expressions and other signals or cues. A smile, for example, or a shrug, may enhance the meaning of an utterance and influence the receiver's interpretation [2]. The combination of vocal utterances with such additional cues, known as extralinguistic cues (ELCs) [3], allows speakers to convey rich and multifaceted meanings and is therefore arguably a cornerstone of the human language faculty [4]. Whether similar multi-modal signals are employed in the communication systems of non-human primates has received growing attention, given the valuable insight such data can provide regarding the evolutionary origins of human communication and language [5,6]. The term "multi-modal" has, however, been used differently in previous communication studies, in some cases denoting multiple signaling channels (e.g. facial expressions vs gestures) [7,8], while in others denoting multiple sensory modalities (e.g. acoustic vs visual modality) [9,10]. Here, we define a multi-modal signal as one that is received in at least two sensory modalities. Previous research in non-human primate communication has shown that apes augment their vocalizations with specific visual gestures, potentially as a way to disambiguate or refine meaning, akin to the function of extralinguistic cues as semantic devices in language [8,11]. 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 [12]. Similarly, in chimpanzees, mothers interacting with infants often combine the "soft hoo" vocalization with the "arm reach" or "present back" gesture, to invite the infant to climb onto their back [13].

To date, the most thorough attempt to document multi-modal signal production in apes has established a repertoire of combinations of existing vocalizations, gestures and facial

63 expressions in chimpanzees [8]. However, since vocalizations may co-occur with other
64 signals or cues simply by chance, differentiating random from non-random multi-modal
65 combinations is a critical step, ultimately providing a more accurate reflection of the multi-
66 modal proclivities of a species. Such a data-driven quantification of the vocal-visual
67 repertoire is currently lacking for any primate [5,6]. We aimed to bridge this gap in
68 understanding through systematically investigating the multi-modal communicative
69 behaviour of wild chimpanzees. As a first step, we build a vocal-visual repertoire by focusing
70 on naturally occurring vocal production and recording the accompanying visual components.
71 Through applying methods borrowed from computational linguistics, namely collocation
72 analysis, we then quantify the non-random nature of identified vocal-visual combinations
73 [14].
74
75 Chimpanzees, like humans, have complex social lives: they reside in groups of ~50-100
76 individuals, forming strong and durable relationships with relatives as well as non-kin [15].
77 Likely as a way to navigate this complex social environment, chimpanzees are also equipped
78 with a rich system of communication comprising signals and cues from both visual and vocal
79 modalities [16-18]. The vocal repertoire consists of approximately 13 different call types
80 [16]. The repertoire is commonly described as graded, meaning that there is acoustic
81 variation within a single category, as well as a degree of overlap in acoustic features also
82 between certain categories. The anatomy of the chimpanzee brain and vocal tract constrains
83 vocal production to a limited range of sounds compared to human vocal production [19,20].
84 By contrast, visual signal production in chimpanzees is highly flexible and the repertoire is
85 vast, comprising at least 9 facial expressions [18] and 66 gesture types [17].
86

Importantly, vocal signals, facial expressions and manual gestures are complemented by an equally broad array of body movements or behaviours, which might be rather described as cues (i.e. behaviours that have not necessarily evolved for a communicative purpose, yet may carry some communicative value) [21,22]. For example, a chimpanzee's body posture (e.g. sitting vs standing), or the orientation of their gaze, which can be towards or away from the recipient, may carry important communicative value for the recipient. As such, we adopted an inclusive, bottom-up approach and considered the combination of vocal signals with both visual signals and behaviours that may act as cues. To this end, we recorded all visible movements, body postures, orientations, behaviours, gestures or facial expressions exhibited by the signaler alongside the vocalization as non-vocal behaviours (NVBs).

In addition to establishing a repertoire of non-random vocal-visual combinations, we aimed to examine the variation underlying NVB production within the population. Previous research has implicated various demographic factors, such as age, sex and rank in driving variation in both gestural and vocal behaviour. For example, females are known to produce a higher rate of call combinations than their male counterparts [23], while highest-ranking males were shown to be the most prolific gesture producers [11]. In line with this existing body of work we therefore also probed how demographic factors influenced the combination of visual components with vocal signals. Given our data-driven and exploratory approach, we formulate no *a priori* predictions regarding patterns of demographic variation. Finally, we probe how NVB production changes in accordance with the characteristics of the call. For example, calls produced while feeding may be associated with different amounts of NVBs compared to calls produced upon encounters with conspecifics. In addition, call duration might affect NVB production as longer calls might be associated with more movements,

111 changes in body posture or gestures. Therefore, we test whether NVB production is
112 influenced by call type and duration.
113
114
115 **Methods**
116
117 *Study site and data collection*
118 The study was conducted on wild chimpanzees from the Kanyawara community in Kibale
119 national park, Uganda [24]. The population consists of ~60 individuals inhabiting a home
120 range of ~15km$^2$. The Kanyawara community has been the object of long-term study since
121 1987 and is entirely habituated. The data used in this study were collected between February-
122 May 2013, and between June 2014 and March 2015 [8]. These data consist in video-audio
123 recordings collected within the chimpanzee home range, between 0800 and 1900 hours. The
124 equipment included a hand-held camcorder (Panasonic HDC-SD90), and an external
125 microphone (Sennheiser MKE 400).
126
127 The individuals observed in this study were 13 females and 14 males, between 10 and 48
128 years of age. Individuals were recorded from a distance of at least 7m while engaged in their
129 natural behaviour. Focal animal sampling was employed [25], involving 15 minutes of
130 continuous video observation of one single animal, with the aim of capturing a clear and
131 complete view of the animal and all its behaviours, including communication. Focal animals
132 were only sampled once a day. Initially focal subjects were chosen on the basis of visibility
133 and ease of pursuit to ensure high-quality recordings. Later in the study period, priority to
134 certain subjects was given in order to homogenize the total focal time across individuals.
135 Thirty-one hours of video data were used in this study.

*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 [16,26]. 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", while "grunt" can refer to "rough grunt" or "general grunt". 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 vocalization 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 1-62 seconds.

In association with each vocal event, between 1-8 NVBs were recorded. NVBs were only annotated during vocal bouts. 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 NVBs included in this list represent an attempt to illustrate the observable variation in NVB 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.

| 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 |

**Table 1.** Full list of NVBs annotated in this study with corresponding behavioural description used to assign NVBs. The term "specific individual" used above refers to the individual who is closest to the signaler.

176 *Data extraction: demographic context of the vocalization*

177 In addition to describing vocal signals and accompanying NVBs, demographic data were
178 annotated for each event. Specifically, identity and sex of the individual were noted and each
179 individual's age in years was calculated based on the long-term data which includes birth
180 dates for all IDs [24]. Next, dominance ranks were calculated using an Elo-rating method
181 [27,28] based on the long-term data on aggressive interactions and submissive pant grunt
182 vocalizations [29]. Rank scores were calculated every 3 months and ranged between 1-24.

183

184 *Inter-observer reliability*

185 To ensure videos were coded reliably, a second independent researcher coded 11% of the
186 events (i.e. 34 events out of 297) and extracted both i) the call type (at least one call for each
187 call type was present in the subset) and ii) non-vocal behaviours (at least one instance of each
188 NVB type was coded in the subset). We calculated a Cohen's kappa value of 0.82 and 0.88
189 for vocalisation type and NVB type respectively, indicating excellent levels of agreement in
190 both cases [30].

191

192 *Collocation analysis*

193 To generate a vocal-visual signal repertoire based on the communicative events observed, we
194 implemented a collocation analysis in R [31]. This method, originating in the field of
195 linguistics and recently adapted to the study of animal communication, estimates the relative
196 attraction between communicative units, based on how frequently they co-occur in the dataset
197 [14]. In this case, the co-occurrence of a particular vocal signal with a specific visual
198 component was examined. For example, if "grunt" + "arm reach" co-occur, collocation
199 analysis compares the frequency of "grunt + arm reach" with the frequency of all other vocal-
200 visual combinations which contain either "grunt" or "arm reach". A multiple distinctive

collocation analysis tests the association between units via one-tailed exact binomial tests on each possible combination, and the log-transformed results provide an estimate of how exclusively units combine with one another. Ultimately, the test indicates whether each combination happens more or less frequently than expected by chance.

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. 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. This threshold of 4 was chosen as 93% of events exhibited between 1-4 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 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.

| 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 |

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

*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 (Kolmogorov-Smirnov test: $P = 0.77$).

## Results

*Vocal-visual repertoire via collocation analysis*

Following collocation analyses, 108 combinations of one vocal signal and between 1-4 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 ($\chi^2_2 = 1.39$, $p = 0.49$), sex ($\chi^2_1 = 1.25$, $p = 0.26$) or rank ($\chi^2_1 = 1.29$, $p = 0.25$) 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 Figure 1.

| | 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 3.** List of 31 single NVBs 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 collocates.

| 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 to |
| climb_feed | approach_get up_sit_walk | approach_extend body | feed_sit | approach_get up_run | scream face | look tow_retract bo |
| feed | gaze alternation | approach_extend body_getup | feed_sit_stand | approach_look tow_run | scream face_stand | look tow_retract bo |
| 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 | | | | | | |

**Table 4.** All combinations of call type and NVBs that were found to co-occur significantly more frequently than expected by chance.

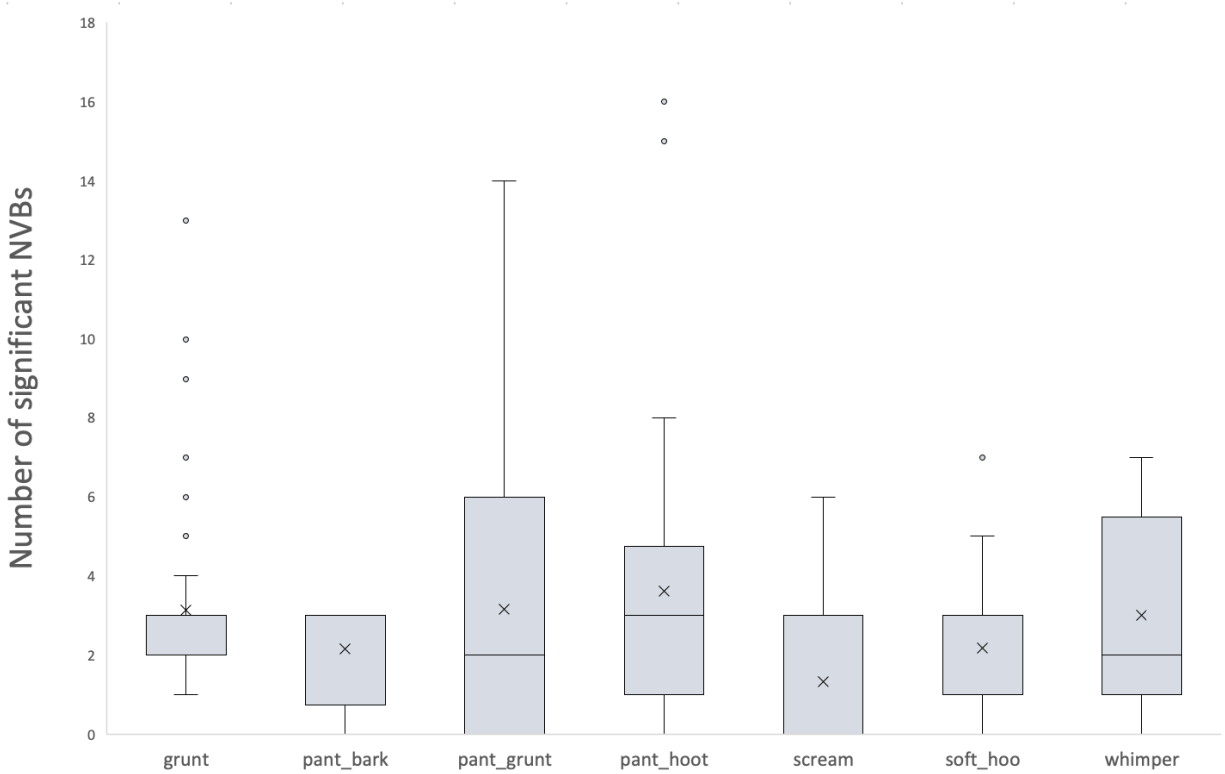


**Figure 1.** Raw data illustrating variation in the number of significant NVBs produced in association with the different call types analysed in this study. Crosses represent means for each call type.

## Discussion

By systematically observing naturally occurring communication events, we show that chimpanzees combine their vocal signals with a wide range of body movements, 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. When a vocalization is produced, the number of accompanying NVBs increases with call duration, but this effect is conditional on call type, such that longer vocalization events are associated with a greater

297 number of NVBs in some call types but not in others. However, the number of NVBs
298 associated with vocal production is not influenced by age, sex, or rank.
299
300 Given the findings of the collocation analysis, it appears that sub-adult and adult chimpanzees
301 have access to a highly diversified repertoire of combined visual and vocal components.
302 Although the constrained vocal repertoire of chimpanzees [19,20] might suggest a limited
303 capacity for information transfer, the productive use of accompanying NVBs instead reveals
304 a high potential for refining the meaning of the limited range of available calls. Indeed, the
305 ~100-strong repertoire of combinations reported in this study highlights the potential for
306 extensive and nuanced information transfer between communicating chimpanzees. A
307 fundamental implication of this investigation is that unimodal approaches to primate
308 communication, which analyze vocal or visual components separately, result in a drastically
309 oversimplified picture of flexibility in signal production. A multi-modal approach is therefore
310 crucial to accurately representing the communicative abilities of non-human primates [5,6],
311 as well as for offering a faithful illustration of real-life communicative exchanges.
312
313 Chimpanzee social life is characterized by a wide variety of interactions, each of which is
314 typically mediated by communication. Thus, it is likely that the diverse repertoire of
315 combined vocal and visual components identified here plays a key role in supporting the
316 demands of a chimpanzee's daily social life [32,33]. It is unknown whether chimpanzee
317 signalers voluntarily combine vocal signals with all of the NVBs reported in this study,
318 nonetheless, chimpanzee receivers may rely on the integration of all the vocal and visual
319 components in order to guide their own adaptive behavioural response [34]. Confirming this
320 hypothesis requires further investigation into how NVBs are perceived by receivers and their
321 potential role in the disambiguation of meaning. Recent developments which combine

insights from linguistics and animal behaviour offer valuable theoretical frameworks and empirical toolkits for addressing the meaning of signal components empirically in nonhumans [35]. One fruitful method involves a systematic analysis of behavioural reactions to signals as a function of signal type [36]. This method 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. A further promising avenue of investigation is to infer which cues are most salient to recipients for meaning disambiguation, using measures of attentional bias. 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 [37].

Our study also investigated the 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. A possible implication of this result is that combinatoriality across modalities may serve a very general function such as that of meaning refinement, which is critical 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 conclusion, our findings reveal a hitherto unappreciated diversity of vocal-visual combinations in the communication system of wild chimpanzees, though 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 appreciation 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 signals represent combinatorial structures, of which vocal and visual components constitute the building-blocks, as opposed to holistic units [38]. 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, suggesting a capacity for complex multi-modal signaling that predates the language faculty and may have played a role in scaffolding language evolution [39-42].

**References**

1. Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639-652.
2. Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, *20*(11), 818-829.
3. Gil, S., Aguert, M., Bigot, L. L., Lacroix, A., & Laval, V. (2014). Children's understanding of others' emotional states: Inferences from extralinguistic or paralinguistic cues?. *International Journal of Behavioral Development*, *38*(6), 539-549.
4. Feldman, R. S., Philippot, P., & Custrini, R. J. (1991). Social competence and nonverbal behavior. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* (pp. 329–350). Cambridge University Press; Editions de la Maison des Sciences de l'Homme.
5. Slocombe, K. E., Waller, B. M., & Liebal, K. (2011). The language void: the need for multimodality in primate communication research. *Animal Behaviour*, *81*(5), 919-924.
6. Liebal, K., Slocombe, K. E., & Waller, B. M. (2022). The language void 10 years on: multimodal primate communication research is still uncommon. *Ethology Ecology & Evolution*, *34*(3), 274-287.
7. Liebal, K., Waller, B. M., Slocombe, K. E., & Burrows, A. M. (2014). *Primate communication: a multimodal approach*. Cambridge University Press.
8. Wilke, C., Kavanagh, E., Donnellan, E., Waller, B. M., Machanda, Z. P., & Slocombe, K. E. (2017). Production of and responses to unimodal and multimodal signals in wild chimpanzees, Pan troglodytes schweinfurthii. *Animal Behaviour*, *123*, 305-316.
9. Fröhlich, M., & van Schaik, C. P. (2018). The function of primate multimodal communication. *Animal Cognition*, *21*, 619-629.

10. Singletary, B., & Tecot, S. (2020). Multimodal pair□bond maintenance: A review of signaling across modalities in pair□bonded nonhuman primates. *American journal of primatology*, *82*(3), e23105.
11. Hobaiter, C., Byrne, R. W., & Zuberbühler, K. (2017). Wild chimpanzees' use of single and combined vocal and gestural signals. *Behavioral Ecology and Sociobiology*, *71*, 1-13.
12. Genty, E., Clay, Z., Hobaiter, C., & Zuberbühler, K. (2014). Multi-modal use of a socially directed call in bonobos. *PloS one*, *9*(1), e84738.
13. Fröhlich, M., Wittig, R. M., & 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. *Animal Cognition*, *19*(3), 483-500.
14. Bosshard, A. B., Leroux, M., Lester, N. A., Bickel, B., Stoll, S., & Townsend, S. W. (2022). From collocations to call-ocations: using linguistic methods to quantify animal call combinations. *Behavioral ecology and sociobiology*, *76*(9), 1-8.
15. Rosati, A. G., Hagberg, L., Enigk, D. K., Otali, E., Emery Thompson, M., Muller, M. N., ... & Machanda, Z. P. (2020). Social selectivity in aging wild chimpanzees. *Science*, *370*(6515), 473-476.
16. Slocombe, K. E., & Zuberbühler, K. (2010). Vocal communication in chimpanzees. *The mind of the chimpanzee: ecological and experimental perspectives*, 192-207.
17. Hobaiter, C., & Byrne, R. W. (2011). The gestural repertoire of the wild chimpanzee. *Animal cognition*, *14*(5), 745-767.
18. Parr, L. A., & Waller, B. M. (2006). Understanding chimpanzee facial expression: insights into the evolution of communication. *Social cognitive and affective neuroscience*, *1*(3), 221-228.
19. Fitch, W. T., De Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science advances*, *2*(12), e1600723.Seyfarth, R. M., & Cheney, D. L. (2010). Primate vocal communication. *Primate neuroethology*, 84-97.
20. Stegmann, U. (Ed.). (2013). *Animal communication theory: information and influence*. Cambridge University Press.
21. Searcy, W. A., & Nowicki, S. (2010). The evolution of animal communication. In *The Evolution of Animal Communication*. Princeton University Press.
22. Roberts, A. I., Roberts, S. G. B., & Vick, S. J. (2014). The repertoire and intentionality of gestural communication in wild chimpanzees. *Animal Cognition*, *17*, 317-336.
23. Leroux, M., Chandia, B., Bosshard, A. B., Zuberbühler, K., & Townsend, S. W. (2022). Call combinations in chimpanzees: a social tool?. *Behavioral Ecology*, *33*(5), 1036-1043.
24. Thompson, M. E., Muller, M. N., Machanda, Z. P., Otali, E., & Wrangham, R. W. (2020). The Kibale Chimpanzee Project: Over thirty years of research, conservation, and change. *Biological conservation*, *252*, 108857.
25. Altmann, J. (1974). Observational study of behavior: sampling methods. *Behaviour*, *49*(3-4), 227-266.
26. Crockford, C., Gruber, T., & Zuberbühler, K. (2018). Chimpanzee quiet hoo variants differ according to context. *Royal Society open science*, *5*(5), 172066.
27. Muller, M. N., Enigk, D. K., Fox, S. A., Lucore, J., Machanda, Z. P., Wrangham, R. W., & Thompson, M. E. (2021). Aggression, glucocorticoids, and the chronic costs of status competition for wild male chimpanzees. *Hormones and behavior*, *130*, 104965.
28. Wilke, C., Lahiff, N. J., Badihi, G., Donnellan, E., Hobaiter, C., Machanda, Z. P., ... & Slocombe, K. E. (2022). Referential gestures are not ubiquitous in wild chimpanzees: alternative functions for exaggerated loud scratch gestures. *Animal Behaviour*, *189*, 23-45.
29. De Vries, H., Stevens, J. M., & Vervaecke, H. (2006). Measuring and testing the steepness of dominance hierarchies. *Animal Behaviour*, *71*(3), 585-592.
30. Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied psychological measurement*, *5*(1), 105-112.
31. Team, R. D. C. (2009). A language and environment for statistical computing. *http://www. R-project. org*.

32. 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*, 390.
33. Freeberg, T. M., Dunbar, R. I., & 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.
34. Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual review of psychology*, *54*(1), 145-173.
35. Berthet, M., Coye, C., Dezecache, G. and Kuhn, J. (2022), Animal linguistics: a primer. Biol Rev. https://doi.org/10.1111/brv.12897
36. Hobaiter, C., & Byrne, R. W. (2014). The meanings of chimpanzee gestures. *Current Biology*, *24*(14), 1596-1600.
37. 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. *Communicative & Integrative Biology*, *10*(2), e1299836.
38. Davila-Ross, M., Jesus, G., Osborne, J., & Bard, K. A. (2015). Chimpanzees (Pan troglodytes) produce the same types of 'laugh faces' when they emit laughter and when they are silent. *PloS one*, *10*(6), e0127337.
39. Seyfarth, R. M., & Cheney, D. L. (2017). Precursors to language: Social cognition and pragmatic inference in primates. *Psychonomic bulletin & review*, *24*(1), 79-84.
40. Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: a promising paradigm whose time has passed. *Evolutionary Anthropology: Issues, News, and Reviews*, *21*(5), 195-205.
41. Seyfarth, R. M., & Cheney, D. L. (2014). The evolution of language from social cognition. *Current opinion in neurobiology*, *28*, 5-9.
42. Seyfarth, R. M., & Cheney, D. L. (2017). *The social origins of language*. Princeton University Press.

**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 thank Piera Filippi for her constructive comments. This project was funded by a Leakey Foundation General Grant to C.W. and the NCCR Evolving Language. 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.

**Competing interests:**
Authors declare that they have no competing interests.

**Data and materials availability:**
All data are available in the supplementary materials.