nature ecology & evolution

Article

https://doi.org/10.1038/s41559-024-02420-w

African elephants address one another with individually specific name-like calls

Received: 24 October 2023

Accepted: 22 April 2024

[redacted]

Check for updates

Michael A. Pardo[1] ✉, Kurt Fristrup[2], David S. Lolchuragi[3], Joyce Poole[4], Petter Granli[4], Cynthia Moss[5], Iain Douglas-Hamilton[3] & George Wittemyer[1,3]

Personal names are a universal feature of human language, yet few analogues exist in other species. While dolphins and parrots address conspecifics by imitating the calls of the addressee, human names are not imitations of the sounds typically made by the named individual. Labelling objects or individuals without relying on imitation of the sounds made by the referent radically expands the expressive power of language. Thus, if non-imitative name analogues were found in other species, this could have important implications for our understanding of language evolution. Here we present evidence that wild African elephants address one another with individually specific calls, probably without relying on imitation of the receiver. We used machine learning to demonstrate that the receiver of a call could be predicted from the call's acoustic structure, regardless of how similar the call was to the receiver's vocalizations. Moreover, elephants differentially responded to playbacks of calls originally addressed to them relative to calls addressed to a different individual. Our findings offer evidence for individual addressing of conspecifics in elephants. They further suggest that, unlike other non-human animals, elephants probably do not rely on imitation of the receiver's calls to address one another.

Q1 Q2 Q3 Q4 Q5 Q6 A hallmark of spoken human language is the use of vocal labels: learned sounds that refer to an object or individual (the 'referent')[1]. Many species produce functionally referential calls for food and predators[2,3], but the production of these calls is typically innate[4]. Learned vocal labels expand the expressive scope of communication by making it possible to establish labels for new referents. Thus, they increase the sophistication of cooperative behaviour and are central to humans' ability to articulate symbolic thought[5]. Personal names are a type of vocal label that refers to another individual. Names must involve vocal learning, as an individual cannot be born knowing the names for all its future social affiliates. Thus, non-human analogues of personal names are highly relevant to understanding the evolution of language and complex cognition.

Most human words, including names, are arbitrary: they are not imitations of sounds typically made by the referent or tied to its physical properties[6]. Arbitrariness is crucial to language because it enables communication about referents that do not make any imitable sound. However, clear evidence for arbitrary names in other species is lacking. Bottlenose dolphins (*Tursiops truncatus*) and orange-fronted parakeets (*Eupsittula canicularis*) address individual conspecifics by imitating the receiver's 'signature' call, a sound that is most commonly produced by the receiver to broadcast their identity[7,8]. While considered arbitrary when used for self-identification[9], it may be argued that copied signature calls used to address the call's owner are iconic (non-arbitrary) labels since they are an imitation of a sound most often produced by the individual to whom the call refers. Non-imitative learned vocal labelling may be more cognitively demanding than imitative labelling, as it requires individuals to make an abstract connection between a sound and referent. Evidence that arbitrary vocal labelling is not unique to humans would expand the breadth of models for the evolution of language and cognition.

---

[1]Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO, USA. [2]Department of Electronic and Computer Engineering, Colorado State University, Fort Collins, CO, USA. [3]Save The Elephants, Nairobi, Kenya. [4]ElephantVoices, Sandefjord, Norway. [5]Amboseli Elephant Research Project, Nairobi, Kenya. ✉e-mail: map385@cornell.edu

Elephants are among the few mammals capable of mimicking novel sounds, although the function of this vocal learning ability is unknown[10,11]. The most common elephant call type is the rumble, a harmonically rich, low-frequency sound that is individually distinct[12,13] and distinguishable[14] and is produced across most behavioural contexts[15]. Contact rumbles are long-distance calls produced when the caller is out of sight and more than ~50 m from one or more social affiliates and attempting to reinitiate contact. Greeting rumbles are affiliative calls produced when one individual approaches another to within touching distance[15]. Caregiving rumbles are affiliative calls produced by an adult or adolescent female while suckling, comforting or rousing a calf[15].

In this Article, we analysed contact, greeting and caregiving rumbles from female–offspring groups of wild African savannah elephants (*Loxodonta africana*) to assess whether they contain individual vocal labels. We investigated (1) if elephants address conspecifics using receiver-specific vocal labels, (2) if the labels are imitative of the receiver's calls or arbitrary, (3) if different callers share the same label for the same receiver and (4) if playbacks to the assumed receiver elicit behavioural responses indicating label recognition (Table 1).

For contact calls, we defined the receiver as the only adult member of the family group separated (>50 m) from the caller or the only individual who responded to the call by vocalizing or approaching. For greeting calls, the receiver was the individual who approached or was approached by the caller. For caregiving calls, the receiver was the calf being suckled, comforted or roused by the caller. We excluded calls with uncertain or multiple recipients. Given the complexity of elephant vocalizations, it was not clear what acoustic features were optimal for capturing the relevant variation in the calls. Thus, we ran models separately for two different sets of features measured on each call (spectral and cepstral; Extended Data Fig. 1 and Extended Data Table 1). The results reported in the text and figures are for the spectral features (see tables for cepstral results, which were similar).

Table 1 | Hypotheses and predictions tested in this study and whether they were supported

Hypotheses	Predictions	Supported?
1. Elephants vocally label individual conspecifics	1. Receiver ID can be predicted from call structure 2. Calls with same caller and same receiver will be more similar than calls with same caller and different receivers, while controlling for caller–receiver relationship type 3. Elephants will respond more strongly to playback of call originally addressed to them than to playback of call from same caller originally addressed to another individual	1. Yes 2. Yes 3. Yes
2. Vocal labels are arbitrary (not imitative of receiver's calls)	1. Receiver can be predicted from call structure regardless of whether calls are convergent or divergent from receiver's calls relative to other calls by the same caller 2. Calls from caller A to receiver B will be no more similar to receiver B's calls than calls from caller A to other receivers are to receiver B's calls	1. Yes 2. Yes
3. Different callers use same label for same receiver	1. Calls with different callers and same receiver will be more similar than calls with different callers and different receivers 2. Receiver ID can be predicted from call structure independently of caller ID	1. Yes 2. No

Calls were specific to individual receivers

We ran a random forest[16] with sevenfold cross-validation to predict the receiver of each rumble as a function of the acoustic features. Call structure varied with the identity of the targeted receiver (Extended Data Figs. 2 and 3) as expected if elephants vocally label other individuals. Our model correctly identified the receiver for 27.5% of calls analysed, Q7 a significantly greater proportion than achieved by models with randomly permuted acoustic features (permutation test, mean ± standard deviation (s.d.) accuracy for 10,000 permuted models: 8.0 ± 0.66% correct, one-tailed $P < 0.0001$) (Fig. 1 and Extended Data Table 2). This indicated that receivers of calls could be correctly identified from call structure statistically significantly better than chance (Table 1, hypothesis 1, prediction 1).

As caller ID and receiver ID were partially aliased in our dataset (Supplementary Table 1), the random forest could theoretically use acoustic cues to caller ID[15] to predict receiver ID, even if the calls did not contain any vocal label. To assess this possibility, we compared the mean similarity of pairs of calls with the same caller and receiver to the mean similarity of pairs of calls with the same caller and different receivers, using proximity scores derived from the random forest as a metric of call similarity[17]. If the random forest relied entirely on cues to caller ID to predict receiver ID, there should be no difference in proximity score between 'same caller/same receiver' pairs and 'same caller/different receiver' pairs. To control for the possibility that calls were specific to the type of relationship between the caller and receiver rather than to individual receivers, we categorized social relationship on the basis of relatedness and age (Extended Data Table 3) and only considered pairs of calls with the same type of relationship between caller and receiver. Calls with the same caller and receiver were significantly more similar (higher proximity scores) than calls with the same caller and different receivers, even after controlling for social relationship, behavioural context and recording date (rank-transformed linear model, $n = 1{,}105$ call pairs with same receiver, 179 pairs with different receivers, $\chi^2_1 = 13.0$, $P = 0.0003$, partial $\eta^2 = 0.063$) (Fig. 1 and Extended Data Table 4). This indicates that rumbles contain information specific to the individual receiver, not merely to the caller or to the type of relationship between the caller and receiver (Table 1, hypothesis 1, prediction 2).

Vocal labels more likely in certain contexts and age classes

For 87.4% of calls, receiver ID was predicted consistently correctly or consistently incorrectly across >95% of random forest iterations. We used logistic regression to assess factors influencing the probability of correct classification. Contact ($n = 138$, 42.0% correct) and caregiving rumbles ($n = 62$, 46.8% correct) were more likely to be correctly classified than greeting rumbles ($n = 127$, 3.9% correct) (care/contact: $P = 0.264$, odds ratio 6.4; care/greeting: $P = 0.014$, odds ratio 48.9; contact/greeting: $P = 0.047$, odds ratio 7.6) (Extended Data Table 5). Calls from adult females ($n = 274$, 32.8% correct) were more likely to be predicted correctly than calls from juveniles ($n = 53$, 3.8% correct) ($\chi^2_1 = 6.5$, $P = 0.011$, odds ratio 0.067). Calls that occurred later in the bout were more likely to be predicted correctly ($\chi^2_1 = 3.8$, $P = 0.0498$, odds ratio 2.8), as were calls addressed to receivers with more total calls in our dataset ($\chi^2_1 = 7.6$, $P = 0.006$, odds ratio 1.4).

No evidence for imitation of receiver in vocal labels

Elephants are not known to produce discrete 'signature' calls like dolphins and parrots; instead, the caller specificity of elephant rumbles is probably a product of voice characteristics[12,13]. If elephants address individual receivers by imitating the receiver's voice, they should sound more like the receiver when addressing her than when addressing other individuals. Among the calls for which we had recordings of the receiver and recordings of the caller addressing other individuals ($n = 236$), 59.7% were divergent from the receiver's calls; that is, less similar to the receiver's calls than typical for that caller. The random forest's prediction accuracy was significantly better than baseline expectations for both convergent and divergent calls (Table 1, hypothesis 2, prediction 1) (permutation test; convergent calls: 20.1% correct, permuted models

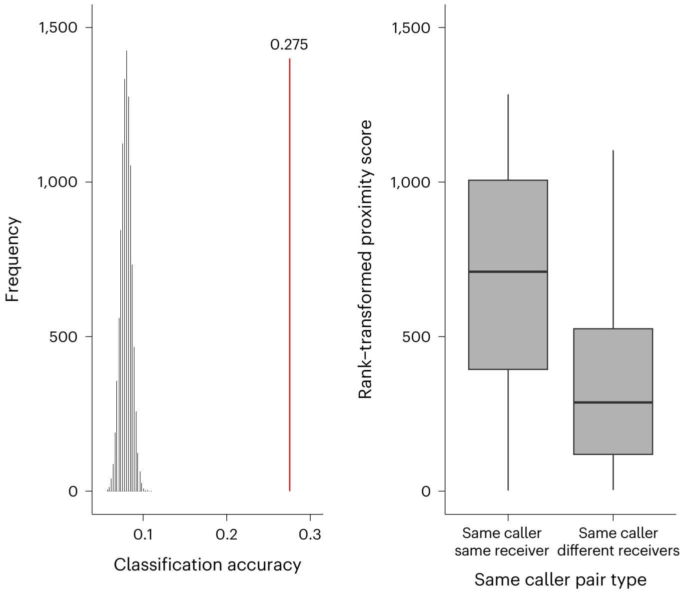


Fig. 1 | Evidence that calls are specific to individual receivers within a caller. Left: the classification accuracy of random forest predicting receiver ID from acoustic features (red line) was significantly higher than the classification accuracies of 10,000 models predicting receiver ID from randomized acoustic features (black histogram) ($n = 437$ calls, permutation test, one-tailed $P = 0.0000$). Cross-validation folds were stratified so that the model was trained and tested on the same combinations of caller and receiver; thus, the classification accuracy represents the receiver specificity of calls within a caller. Right: calls with the same caller and same receiver were significantly more similar (higher proximity score) than calls with the same caller and different receivers who had the same type of relationship to the caller ($n = 1{,}105$ call pairs with same receiver, 179 pairs with different receivers, ANOVA on ranks, $\chi^2 = 13.0$, d.f. 1, two-tailed $P = 0.0003$, partial $\eta^2 = 0.063$). Boxplot centre lines, medians; box limits, 25th and 75th quantiles; whiskers, 1.5× interquartile range.

mean ± s.d. accuracy of 7.7 ± 1.3%, $n = 95$ calls, one-tailed $P < 0.0001$; divergent calls: 32.6% correct, permuted models mean ± s.d. accuracy of 17.9 ± 1.6%, $n = 141$ calls, one-tailed $P < 0.0001$) (Fig. 2 and Extended Data Table 2).

Proximity scores for pairs of calls in which the receiver of one call made the other call were marginally higher than for pairs in which this was not the case, but this was not statistically significant (rank-transformed linear model, $n = 943$ call pairs where receiver of one call made the other call, 1,553 pairs where this was not the case, $\chi^2_1 = 3.7$, $P = 0.056$, partial $\eta^2 = 0.001$) (Fig. 2 and Extended Data Table 6). This suggests that calls addressed to a given receiver were no more convergent with the receiver's calls than with calls from other individuals (Table 1, hypothesis 2, prediction 2). Collectively, the evidence suggests that vocal labelling in elephants probably does not rely on imitation of the receiver's calls. However, a definitive conclusion about the role of imitation will require exhaustively sampling the vocal repertoire of each caller.

Mixed evidence for shared labels across callers

In humans and bottlenose dolphins, different callers generally use the same label for a given receiver. To determine if elephants do the same, we further examined call proximity scores. Calls from different callers to the same receiver were significantly more similar than calls from different callers to different receivers (Table 1, hypothesis 3, prediction 1) (rank-transformed linear model, $n = 693$ call pairs with same receiver, 7,522 pairs with different receivers, $\chi^2_1 = 10.7$, two-tailed $P = 0.001$, partial $\eta^2 = 0.004$) (Fig. 3 and Extended Data Table 7). This suggests that there was some vocal convergence among different callers addressing the same receiver.

We then ran a random forest structured to predict receiver ID from different callers than the model was trained on ($n = 437$ calls) (Table 1, hypothesis 3, prediction 2). This model correctly classified 1.1% of calls, no better than the corresponding models with randomly permuted acoustic features (permutation test, mean ± s.d. accuracy of permuted models 1.4 ± 0.33% correct, one-tailed $P = 0.896$) (Fig. 3 and Extended Data Table 2). Therefore, the random forest was not able to predict receiver ID independently of caller ID, suggesting convergence across callers was weak.

Playback confirms receiver recognition of vocal labels

To determine if elephants perceive and respond to the vocal labels in calls addressed to them (Table 1, hypothesis 1, prediction 3), we compared reactions of 17 wild elephants to playback of a call that was originally addressed to them (test) relative to playback of a call from the same caller that was originally addressed to a different individual (control). By using test and control stimuli from the same caller, we controlled for the possibility of the caller's relationship to the subject influencing the results. To control for the possibility that calls were specific to the type of relationship between the caller and receiver rather than to the individual receiver, we included the type of relationship between the caller and the original receiver as a factor in the analysis. Further supporting the existence of vocal labels, subjects approached the speaker more quickly (Cox regression, $\chi^2 = 6.8$, $P = 0.009$, hazards ratio 8.77), vocalized more quickly (Cox regression, $\chi^2 = 7.9$, $P = 0.005$, hazards ratio 7.45) and produced more vocalizations (Poisson regression, $\chi^2 = 6.7$, $P = 0.009$, rate ratio 2.41) in response to test playbacks than control playbacks (Fig. 4 and Table 2). In trials where an approach or vocalization occurred, the mean ± s.d. latency to the first approach or vocalization was 99.7 ± 161.4 s.

Discussion and conclusions

Very few species are known to address conspecifics with vocal labels. Our discovery of individual vocal labels in a species that diverged from both the primate and cetacean lineages ~90–100 million years ago provides an important opportunity to study the convergent evolution of unusually sophisticated communication[18]. Moreover, where evidence for vocal labels has been found in non-human species, they are either clearly imitative[7,8] or of unknown structure[19–21]. Our data suggest that elephants may label conspecifics without relying on imitation of the receiver's calls, a phenomenon previously known to occur only in human language. If further research supports the absence of receiver imitation in elephant vocal labels, then investigating the social context, acoustic structure and ontogeny of vocal labels in elephants may shed light on why elephants and humans developed non-imitative vocal labels in contrast to other species known to vocally label conspecifics. Our results also have significant implications for elephant cognition, as inventing or learning sounds to address one another suggests the capacity for some degree of symbolic thought.

The existence of individual vocal labelling in elephants is supported by multiple lines of evidence that exclude simpler alternative explanations. Receiver ID could be predicted from call structure significantly better than chance. Moreover, analysis of random forest proximity scores showed that calls from the same caller to the same receiver were significantly more similar than calls from the same caller to two different receivers who had the same type of relationship with the caller. This ruled out the alternative explanations that call structure predicted receiver ID because of the correlation between caller ID and receiver ID in our dataset or that call structure reflected only the type of relationship between caller and receiver and not the individual identity of the receiver. We also controlled for behavioural context and recording date in the proximity score analysis, ensuring that receiver specificity was not an artefact of context-related cues or autocorrelation among calls from the same day. The results did not change when

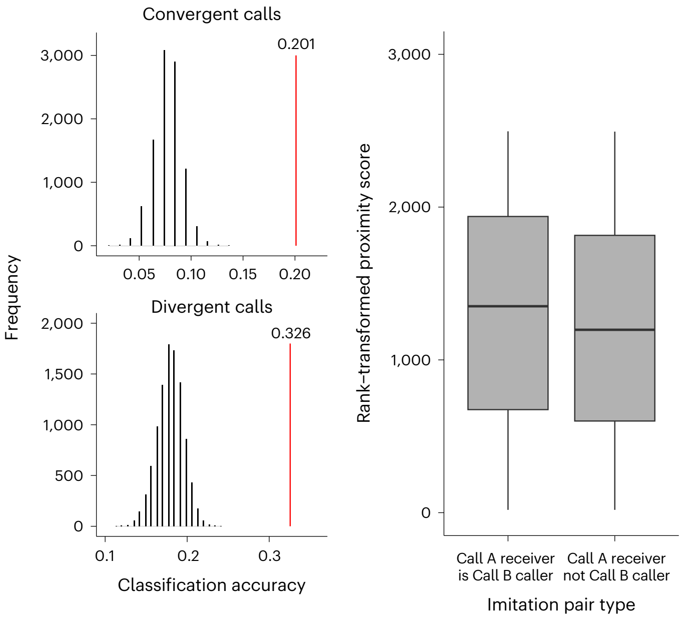


Fig. 2 | Evidence that vocal labelling probably did not rely on imitation of the receiver's calls. Random forest predicted receiver ID significantly better than models with randomly permuted features both among calls that were identified as convergent to receiver's calls (top left) ($n = 95$ calls, permutation test, one-tailed $P = 0.0000$) and divergent from receiver's calls (bottom left) ($n = 141$ calls, permutation test, one-tailed $P = 0.0000$). The red lines represent classification accuracy of original random forest model, and the black histograms represent the distribution of classification accuracies of null models with randomized acoustic features. Right: pairs of calls in which the receiver of one call made the other call did not differ significantly in mean proximity score from pairs of calls in which the receiver of one call did not make the other call ($n = 943$ call pairs where receiver of one call made the other call, 1,553 pairs where this was not the case, ANOVA on ranks, $\chi^2 = 3.7$, d.f. 1, $P = 0.056$, partial $\eta^2 = 0.001$). Boxplot centre lines, medians; box limits, 25th and 75th quantiles; whiskers, 1.5× interquartile range.

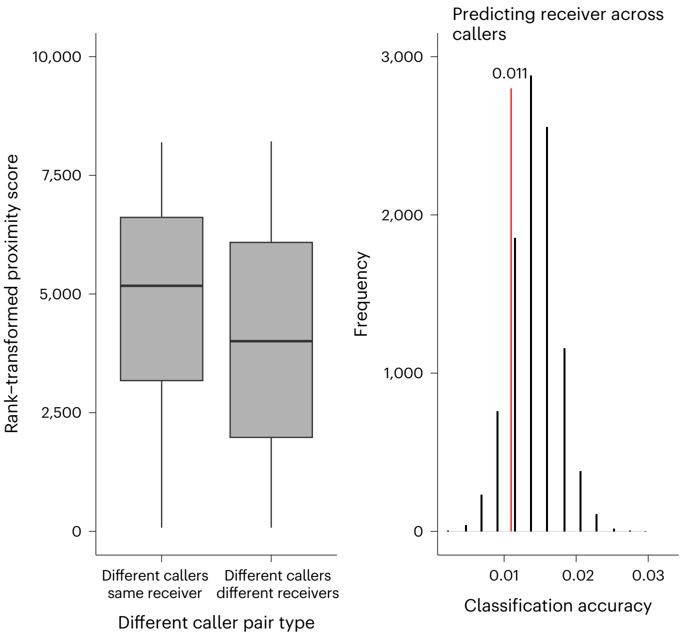


Fig. 3 | Mixed evidence that different callers use similar labels for the same receiver. Left: pairs of calls with different callers and the same receiver were significantly more similar (higher proximity score) than pairs of calls with different callers and different receivers, indicating some convergence among callers addressing the same receiver ($n = 693$ call pairs with same receiver, 7,522 pairs with different receivers, ANOVA on ranks, $\chi^2 = 10.7$, d.f. 1, two-tailed $P = 0.001$, partial $\eta^2 = 0.004$). Boxplot centre lines, medians; box limits, 25th and 75th quantiles; whiskers, 1.5× interquartile range. Right: classification accuracy (red line) of random forest designed to predict receiver ID from acoustic features independently of caller ID (all calls with the same caller and receiver allocated to the same cross-validation fold) was not significantly different from classification accuracies of models with randomized acoustic features (black histogram), indicating that receiver ID could not be predicted independently of caller ID ($n = 437$ calls, permutation test, one-tailed $P = 0.896$). The fact that elephant calls contain multiple messages and are structurally highly complex may account for the model's poor generalization of receiver ID across callers.

two individuals that accounted for a disproportionate number of calls in the dataset (M6 and M6.99) were excluded, indicating that our results were not driven by a few highly influential individuals (Supplementary Information). Most importantly, elephants responded more strongly to playback of calls addressed to them than to playback of calls from the same caller addressed to a different receiver, indicating that the calls contained receiver-specific information that was salient to the elephants. The difference in response to test and control trials was often pronounced. For example, subject R26 vocalized eight times and approached the speaker in response to the test playback but vocalized only once and did not approach the speaker in response to the control playback. Only one subject exhibited an unambiguously stronger response to the control playback than to the test playback. These results are particularly notable in that we could not be certain that all playback stimuli contained vocal labels.

The social behaviour and ecology of elephants create an environment in which individual vocal labelling may be particularly advantageous. Elephants maintain lifelong differentiated social bonds with many individuals, and due to their fission–fusion social dynamics are often separated from their closely bonded social partners[22,23]. In contact calls, where the caller and receiver are separated, vocal labels probably allow elephants to attract the attention of a specific distant receiver. In close-distance calls such as greeting and caregiving rumbles, vocal labels may help strengthen social bonds, similar to the way in which humans experience a positive affective response and increased willingness to cooperate when someone remembers their name[24].

Our random forest model correctly predicted receiver ID for slightly over a quarter of calls (albeit significantly better than random), suggesting that vocal labels may not be necessary in all or even most contexts. Indeed, both humans and bottlenose dolphins only use individual vocal labels (that is, names or imitated signature whistles) in a small percentage of utterances[25]. We found that receiver ID was more likely to be correctly predicted for contact and caregiving rumbles than for greeting rumbles, which suggests that vocal labels may be used more in the former two contexts. Vocally identifying the intended receiver seems especially likely to be beneficial in contact calls, where the caller and receiver are out of visual and tactile contact. It is somewhat surprising, however, that caregiving rumbles were more likely to be correctly classified than greeting rumbles, as both are close-distance affiliative calls. Perhaps labels are included in caregiving rumbles to help calves learn the labels with which others address them or because hearing the label is comforting for calves. Calls made by adult females were also more likely to be correctly classified than calls made by juveniles. This suggests that adult females may use vocal labels more than calves, possibly because the behaviour takes years to develop.

Elephant rumbles are highly complex and simultaneously encode multiple messages, including but not limited to caller identity, age, sex, emotional state and behavioural context[12,15,26,27]. The top acoustic features for predicting receiver ID were not those that explained the most variation in the calls (Supplementary Discussion), suggesting that vocal labels account for only a small fraction of the total variation in rumbles. This appears to contrast with human names, in which the vocal label accounts for most of the acoustic variation in the signal, even

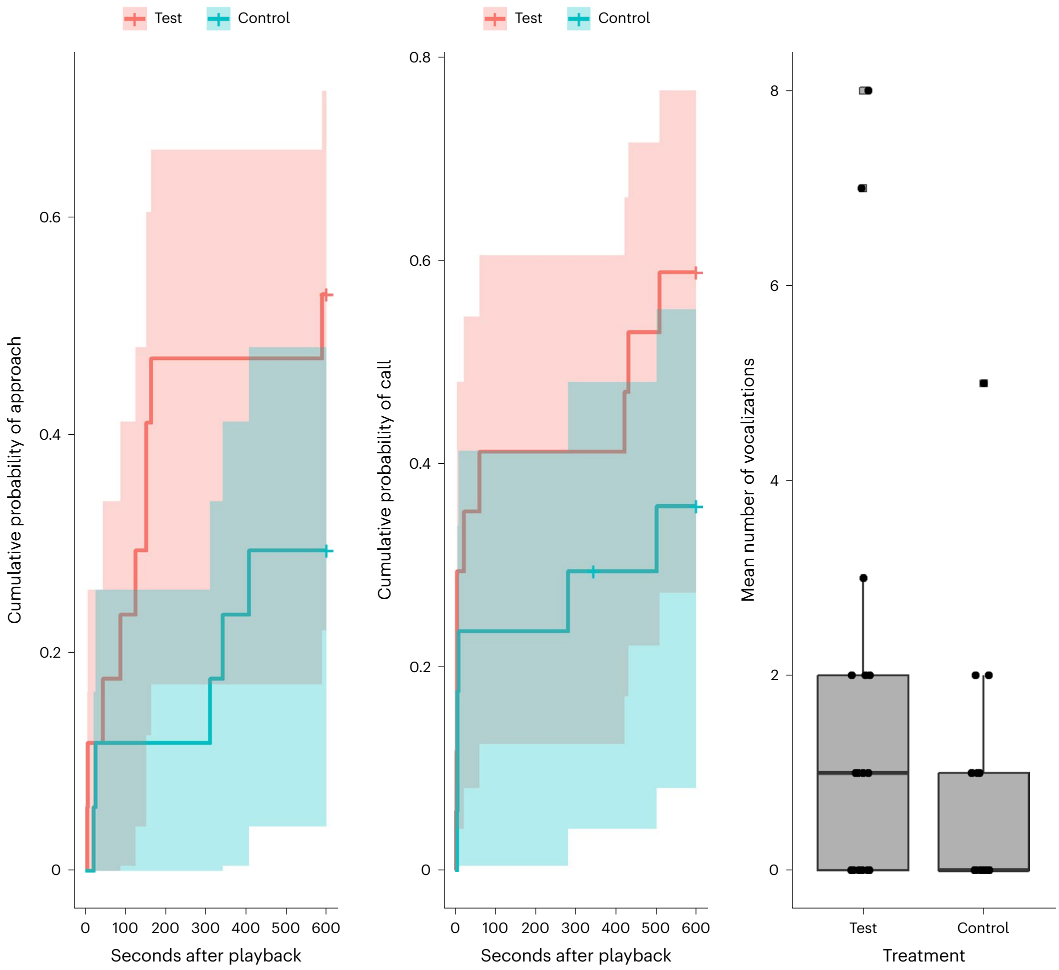


Fig. 4 | Response to playbacks of test stimuli (calls originally addressed to the subject) versus control stimuli (calls from the same caller originally addressed to a different individual). Left: subjects approached the speaker more quickly ($n = 17$ individuals, Cox regression, $\chi^2 = 6.8$, d.f. 1, two-tailed $P = 0.009$, hazards ratio 8.77) in response to test playbacks than controls. Centre: subjects vocalized more quickly in response to test playbacks than controls ($n = 17$ individuals, Cox regression, $\chi^2 = 7.9$, d.f. 1, two-tailed $P = 0.005$, hazards ratio 7.45). Right: subjects produced more vocalizations in response to test playbacks than controls ($n = 17$ individuals, Poisson GLM, $\chi^2 = 6.7$, d.f. 1, two-tailed $P = 0.009$, hazards ratio 2.41). The shaded areas in the left and centre panels represent 95% confidence intervals around survival curves. Boxplot centre line, median; box limits, 25th and 75th quantiles; whiskers, 1.5× interquartile range; grey boxes, location of outliers; black circles, all individual data points. The median and the 25th quantile of the control box are both 0. No corrections were done for multiple comparisons as the analyses presented in this figure were three distinct models with different response variables.

Q8

though information such as the identity, age, sex and emotional state of the speaker is also encoded in the speaker's voice characteristics[28]. Whereas human language conveys complex messages via sequential encoding of information, elephants may rely more on simultaneous encoding, packing more information into a single vocalization than humans typically do.

The richness in the information content of elephant vocalizations makes it difficult to identify the specific acoustic parameters that encode receiver ID, although the variable importance scores from the random forest suggest possible candidate features (Supplementary Discussion). Unlike dolphin and parrot signature calls[20,25,29], elephant vocal labels cannot be discerned by visual inspection of the spectrogram and are probably encoded by a complex and subtle interaction among many acoustic parameters. As a result, we employed machine learning in this analysis, but innovative approaches in signal processing may be necessary to isolate the aspects of rumbles encoding vocal labels.

We found mixed support for the hypothesis that different callers use the same label to address the same receiver. While the random forest failed to predict receiver ID independently of caller ID, analysis of proximity scores indicated at least some convergence among different callers addressing the same receiver. It is possible that all callers within a family group use the same label for the same receiver and the poor performance of the random forest was due to limitations of our data. The dense information content and high variability of rumbles coupled with the small number of calls per receiver in our dataset may have prevented the random forest from learning cues to receiver ID that generalized across callers. Moreover, as the acoustic features we extracted were based on the mel frequency scale, which was inspired by human vocal tract models[30], it is possible that they provided peripheral measures of the principal modes of label encoding. Acoustic features more closely tailored to the properties of the elephant vocal tract might result in a higher classification accuracy for receiver ID.

Alternatively, it is possible that callers only partially share labels for a given receiver. Such a system would greatly increase the number of labels that elephants need to understand, although partial overlap in the labels addressed to a given receiver could mitigate the difficulty of this task. Nonetheless, partial convergence among labels might be favoured if it is easier for receivers to learn to respond to multiple labels than it is for callers to learn to produce the exact same label for each receiver. This seems possible, as modifying the structure of calls based on auditory experience (vocal production learning) requires

Table 2 | Results for type III analyses of deviance on playback experiment models

Response variable (model type)	Subject ID (s.d.)	Treatment (d.f. 1)	Relationship of caller to original receiver (d.f. 4)	Distance (d.f. 1)	dBC (d.f. 1)	Other adults (d.f. 1)	Speaker location (d.f. 1)	Cumulative playback exposure (d.f. 1)
Latency to approach (Cox)	3.43	χ^2=6.8, P=**0.009,** RR 8.77	χ^2=1.7, P=0.80	χ^2=2.4, P=0.12, RR 0.79	χ^2=0.65, P=0.42, RR 1.38	χ^2=0.41, P=0.52, RR 3.13	χ^2=0.59, P=0.44, RR 4.62	χ^2=0.11, P=0.73, RR 0.88
Latency to vocalize (Cox)	2.84	χ^2=7.9, P=**0.005,** RR 7.45	χ^2=6.4, P=0.17	χ^2=0.97, P=0.32, RR 0.87	χ^2=0.02, P=0.90, RR 0.96	χ^2=0.64, P=0.42, RR 3.25	χ^2=0.20, P=0.66, RR 2.02	χ^2=0.10, P=0.75, RR 0.91
Number of calls (Poisson)	–	χ^2=6.2, P=**0.013,** RR 2.41	χ^2=19.9, P=**0.0005**	χ^2=0.32, P=0.57, RR 0.98	χ^2=0.48, P=0.49, RR 1.09	χ^2=0.72, P=0.40, RR 1.54	χ^2=0.13, P=0.72, RR 0.84	χ^2=0.01, P=0.91, RR 0.99
Latency to vigilance (Cox)	0.02	χ^2=3.1, P=0.08, RR 2.07	χ^2=10.1, P=**0.038**	χ^2=1.8, P=0.18, RR 0.93	χ^2=1.9, P=0.16, RR 0.84	χ^2=5.5, P=**0.019,** RR 4.24	χ^2=0.55, P=0.46, RR 0.64	χ^2=0.02, P=0.88, RR 0.99
Vigilance duration after–before (linear)	9.95	χ^2=0.06, P=0.81, β=1.70	χ^2=2.1, P=0.72	χ^2=4.0, P=**0.045,** β=−1.98	χ^2=0.02, P=0.89, β=−0.30	χ^2=0.43, P=0.51, β=7.58	χ^2=0.33, P=0.56, β=6.68	χ^2=0.83, P=0.36, β=−1.73

Subject ID was included as a random effect in all models except the Poisson regression for number of calls, because it had variance of 0 for this model. Values in the 'Subject ID' column represent the square root of the variance explained by that random effect. Significant P values in bold. Latency to vigilance exhibited a non-significant trend towards faster onset of vigilance in response to test playbacks. In addition to the d.f., χ^2 statistic and two-tailed P value from the analysis of deviance, this table includes the hazard or rate ratios (RR) for the Cox and Poisson models and the estimated slope parameters (β) for the linear model. Ratios and slopes are not shown for relationship of caller to original receiver, as this covariate had more than two levels.

more specialized neural circuitry than modifying the context in which calls are produced (usage learning)[31]. Spectacled parrotlets (*Forpus conspicillatus*) and budgerigars (*Melopsittacus undulatus*) reportedly address individual conspecifics with vocal labels that are not shared across callers[19,20], although this could reflect imperfect imitation of the receiver's calls rather than discrete 'nicknames'[32]. Further work to identify how vocal labels are encoded in elephant calls is necessary to determine to what degree different callers use the same label for the same receiver. Isolating the labels for individual elephants will allow investigation of questions such as whether elephants understand the labels used by third parties or even refer to third parties in their absence.

Both African and Asian elephants have a demonstrated capacity for vocal mimicry in captivity, but no study has documented a function of this ability in the wild[10,11]. Depending on whether callers share labels for the same receiver, vocal labelling in elephants could rely on either vocal production learning or vocal innovation combined with usage learning. However, given the evidence for partial convergence among callers, it seems likely that production learning is involved. Dolphins and parrots, which show evidence for individual vocal addressing via imitation of the receiver, are adept vocal learners. Another vocal learner, the Egyptian fruit bat (*Rousettus aegyptiacus*), produces calls that are specific to individual receivers and may be vocal labels as well, although it is currently unknown if the bats perceive this information[21]. Humans, dolphins, parrots, bats and elephants all form long-term social bonds and live in groups with a high degree of fission–fusion dynamics[22,32–35]. A mechanism to direct communication to individual conspecifics could be especially beneficial for animals that frequently separate and rejoin with bonded social partners. This raises the possibility that social selection pressures creating a need to address individual conspecifics may have led to multiple independent origins of vocal production learning, a precursor for language.

The use of learned arbitrary labels is part of what gives human language its uniquely broad range of expression[6]. Our results suggesting possible use of arbitrary vocal labels in elephants provide an opportunity to investigate the selection pressures that may have led to the evolution of this rare ability in two divergent lineages. Moreover, these findings raise intriguing questions about the complexity of elephant social cognition, considering the potential relevance of symbolic communication to their social decision-making.

Methods

Field recording

We collected audio recordings of wild female–calf groups in Amboseli National Park, Kenya in 1986–1990 and 1997–2006 and Samburu and Buffalo Springs National Reserves (hereafter, Samburu), Kenya in November 2019 to March 2020 and June 2021 to April 2022. Both populations have been continuously monitored for decades, and all individuals can be individually identified by external ear morphology[22,36]. We recorded calls from a vehicle during daylight hours with all-occurrence sampling[37] using an Earthworks QTC1 microphone (4 Hz to 40 kHz ± 1 dB) with a Nagra IV-SJ reel-to-reel tape recorder or an HHB PDR1000 DAT recorder in Amboseli, and an Earthworks QTC40 microphone (3 Hz to 40 kHz ± 1 dB) with a Sound Devices MixPre3 or MixPre3-II digital recorder in Samburu. Recordings were recorded at a 48 kHz sampling rate with 16 bits of amplitude resolution and stored at 2 kHz in Amboseli and recorded and stored at 44.1 kHz with 24 or 32 bits of amplitude resolution in Samburu.

When possible, we recorded for each call the identity of the caller, the behavioural context and the identity of the receiver (criteria for identifying receiver defined in Main). The caller was identified using behavioural and contextual cues, such as an open mouth, flapping ears or being the only individual of the right age class in the immediate vicinity (calls made by young calves are audibly shorter and higher pitched than adult calls)[15]. Behavioural observations were recorded by a single observer at each field site (M.A.P. in Samburu, J.P. in Amboseli). Since the observations at each field site were conducted without accompanying video in most cases, there was no way to calculate inter-observer reliability.

Scoring behavioural context

For this study, we only used rumbles produced in the contexts of 'contact calling', 'greeting' and 'caregiving', as these are the contexts in which vocal labelling seems most likely to be beneficial[15]. We did not include rumbles from other behavioural contexts as these typically either involve multiple simultaneous receivers (for example, 'let's go' rumbles) or occur in contexts where vocal labelling is less likely to be necessary (for example, 'begging', 'protest', 'oestrus' and 'musth' rumbles)[15]. Nonetheless, there was a great deal of variation in the precise social context surrounding the production of each call and the age and internal state of the callers. As elephant rumbles vary with behavioural context, age and the emotional state of the caller[12,15,27], this contextual heterogeneity of the recordings probably added substantial noise to the data.

Following published methodology[15], we defined contact rumbles as calls produced by or addressed to an individual who was separated from the receiver by >~50 m and apparently attempting to reinitiate contact. Our category of 'greeting' rumbles encompasses two different categories distinguished by Poole[15]: 'little-greeting' and 'greeting'. Both call types are produced when one individual approaches another in an affiliative manner, but Poole's 'greeting rumbles' are produced after a greater period of separation than 'little-greeting rumbles', are more likely to involve a face-to-face approach and typically involve greater emotive behaviour such as temporal gland streaming and pirouetting to stand in parallel[15]. The context of 'caregiving' in our study is primarily synonymous with 'coo rumbles' described by Poole[15], which are rumbles produced by adult or adolescent females to a calf when gently touching or suckling the calf or in an apparent attempt to reassure a calf who exhibited distress (for example, being pushed by another elephant, being separated from its mother and so on). We also included in this category two calls from adult females attempting to rouse a calf who was sleeping when the group began to move off.

Scoring certainty of caller ID, behavioural context and receiver ID

In Samburu, we recorded the certainty with which we knew caller ID, behavioural context and receiver ID as 1 over the number of possible alternatives[38]. For example, in cases where we thought the call was plausibly addressed to a single individual but there were two possible candidates for who the receiver was, we designated one of the two individuals as the putative receiver and assigned the certainty of receiver ID a value of 0.5. In Amboseli, certainty of caller ID and behavioural context were scored as 'certain', 'fairly confident', 'educated guess' or 'no idea'. The certainty of receiver ID was not systematically recorded in Amboseli, but sometimes the field notes specified that the receiver ID was uncertain.

Call selection

For all analyses in this paper, we only used rumbles with the highest possible certainty for receiver ID (that is, certainty of 1 for Samburu calls, no notes indicating uncertain receiver ID for Amboseli calls). We also required rumbles to have the first two formants clearly visible in the spectrogram with no significant overlap with other calls or loud sounds in the same frequency range. This dataset consisted of 469 calls, 101 unique callers and 117 unique receivers, with 1–36 (median 2) calls per caller, 1–40 (median 2) calls per receiver, 1–7 (median 2) receivers per caller and 1–7 (median 1) callers per receiver (Supplementary Table 1).

There were 32 calls for which the receiver ID was certain but the caller ID was not. We used these calls in the random forest model that was used to generate the proximity score matrix and the conditional inference forest used to calculate variable importance scores for predicting receiver ID, as caller ID was irrelevant to these models. However, for all other analyses, including the linear mixed models with proximity score as a response variable, we only used calls where the caller ID was known for certain.

Q10 For analyses that examined behavioural context (LMMs, logistic regression), we required the certainty of behavioural context to be 1 in Samburu or 'certain' in Amboseli. For analyses that did not explicitly include behavioural context, we also included calls with uncertain contexts as long as the only possible options were contact, greeting or caregiving.

Call segmentation

In Amboseli, we wrote down the elapsed time on the recorder and contextual information for each call heard in the field; in Samburu, we recorded verbal annotations onto a second channel of the recorder in real time using a Martel Stenomask, which isolated the sound of the observer's voice from the Earthworks microphone[38]. We manually drew a selection box around the spectrogram of each call in Raven Pro 1.5 (Cornell Lab of Ornithology, Ithaca, NY), with a buffer of approximately 1 s on either side of the call (Samburu (44.1 kHz sampling rate): Hann window, 50% overlap, window 11,878 samples, DFT 16,384 samples; Amboseli (2 kHz sampling rate): Hann window, 50% overlap, window 312 samples, DFT 512 samples). Q11 This automatically generated a selection table in .txt format with the file name and start and end times of each selection box, to which we added caller ID, receiver, ID, behavioural context and the certainty of each. We performed all further acoustic and statistical analyses in R version 4.1.3 (ref. 39).

To determine the precise onset and offset of each call, we low-pass filtered the calls (Butterworth filter, order 5, cut-off 490 Hz), downsampled them to 2,000 Hz if not already at that sampling rate, applied a high-pass filter (Butterworth filter, order 10, cut-off 30 Hz) and normalized them to 70% of max amplitude and 16 bits of amplitude resolution using the packages seewave[40] and tuneR[41]. We then used the function segment() in the package soundgen[42] to detect the onset and offset of each call based on the amplitude envelope. We verified the automatically detected start and end time for each call by visual inspection of the amplitude envelope and spectrogram and manually adjusted the times when necessary.

Acoustic measurements

We trimmed the original unfiltered sound clips to the automatically detected start and end times, low-pass filtered the clips (Butterworth filter, order 5, cut-off 800 Hz), downsampled them to 2,000 Hz if not already at that sampling rate, applied a high-pass filter (Butterworth filter, order 2, cut-off 4 Hz) and finally normalized them to 70% of the max amplitude and 16 bits of amplitude resolution. For each call, we measured the smoothed Hilbert amplitude envelope (moving average window, window length 350 ms, overlap 90%) and two alternative sets of features: normalized mel spectrogram and mel-frequency cepstral coefficients (MFCCs).

A mel spectrogram is similar to a traditional spectrogram (raster plot with time on the x axis, frequency on the y axis, and amplitude indicated by pixel darkness) but with frequency transformed to the logarithmic mel scale[30]. While the mel scale was designed to approximate human hearing sensitivity, most other mammals, including elephants, perceive frequency on a similar logarithmic scale[43]. We calculated a mel spectrogram for each call using the audspec() function of the tuneR package (26 mel-frequency bands between 0 Hz and 500 Hz, 350 ms Hamming window, 90% overlap). We then normalized the mel spectrogram by dividing the energy value in each cell of the spectrogram by its column sum so that the energies would be a proportion of the total energy in each time window, and logit-transformed these proportional energies so the values would not be limited between 0 and 1. We also calculated delta and delta–delta values for each mel spectral band, with delta values being the differences between successive energy values in the mel spectral band (that is, the change in energy over time within a mel spectral band) and delta–delta values being the differences between successive delta values (that is, the acceleration of energy over time within a mel spectral band) (Extended Data Fig. 1). We saved the vector of energies in each mel spectral band and their corresponding delta and delta–delta values as acoustic contours for further processing. While mel spectral bands have not previously been used as acoustic features for analysing elephant calls, they describe more of the variation in the call than commonly used features such as fundamental frequency and formants, while remaining easily interpretable.

We also calculated MFCCs for each call, which are less interpretable than mel spectral bands but have been previously used successfully to classify elephant vocalizations[13,27,44]. MFCCs are calculated by applying a discrete cosine transform to each time window of a mel spectrogram, with the coefficients of the discrete cosine transform being the cepstral coefficients[45]. Each cepstral coefficient can be thought of as representing the degree of modulation of the spectrum at a different period, with lower numbered coefficients representing slower periods of modulation. Since MFCCs are calculated for each time window of the mel spectrogram, the output is a vector of values for each cepstral coefficient. We calculated MFCCs using the melfcc() function in the tuneR package, with a time window of 350 ms with 90% overlap, 40 mel-frequency bands between 0 Hz and 500 Hz, and a pre-emphasis filter with a cut-off frequency of 10 Hz, and kept the first 12 coefficients (12 vectors per call) for further processing. We also calculated delta and delta–delta values for the first 12 cepstral coefficient contours.

Extraction of derived features from acoustic contours

We rescaled the acoustic contours by arranging them in a matrix with each contour in a separate row, and then subtracting the column median from each value and dividing the result by the column mean average deviation. We decorrelated the contours with robust principal components analysis in the rpca package in R, which separates the data into a low-rank matrix of robust principal components without outliers, and a sparse matrix containing the outlier values ($\lambda = 0.00996$)[46]. Robust principal component analysis (PCA) has the advantage over standard PCA of being more resilient to noisy data. We extracted four measurements from the sparse matrix to use for statistical analysis: median, robust skewness and two measures of spread: minimum extent and equivalent statistical extent. We also calculated the means of the first *n* low-rank principal components required to explain 99.9% of the variation (74 for spectral features, 12 for cepstral features).

We used multi-taper spectral estimation[47] to derive the frequency spectra of the low-rank principal components that explained 99.9% of the variation (treating each principal component as if it were a waveform) and calculated an *F* ratio for each point in each spectrum, testing the null hypothesis that the spectral value in question could have been derived from a random waveform. We calculated the mean of the *F* ratios at each point across the aligned spectra and selected the four largest peaks in the series of mean *F* ratios. We sorted these peaks in order of increasing frequency and calculated the frequency and magnitude of each peak.

We calculated the same metrics on spectra that were weighted according to the proportion of variation that was explained by the principal component from which the spectrum was derived. We multiplied the *F* ratios in each of the spectra by the proportion of variation in the data explained by the principal component in question, summed the weighted *F* ratios at each point in the aligned spectra and then calculated the frequencies and magnitudes of the four largest peaks in the summed *F* ratios, sorted in order of increasing frequency. The final acoustic features used in our models are summarized in Extended Data Table 1. We ran all subsequent statistical analyses separately for the spectral and cepstral acoustic features.

Statistical analysis of acoustic data

Unless otherwise specified, all statistical tests were two-tailed and all measurements were taken from distinct samples. The significance level was set to 0.05 for all tests. We used partial η^2 as a measure of effect size for linear models, calculated according to the formula partial $\eta^2 = \frac{\mathrm{SSE_C} - \mathrm{SSE_A}}{\mathrm{SSE_C}}$, where $\mathrm{SSE_A}$ is the sum of the variances for all the error terms (random effects and residual error) in the full model and $\mathrm{SSE_C}$ is the sum of the variances for all the error terms in the same model minus the fixed effect of interest[48]. For all regression models, we calculated *P* values for the fixed effects using type III analysis of deviance.

Are calls specific to individual receivers (hypothesis 1)? We ran a sevenfold cross-validated random forest model in the R package ranger[49] to predict the identity of the receiver of each call (receiver ID) as a function of the acoustic features (Table 1, hypothesis 1, prediction 1). We stratified the cross-validation folds by caller ID and receiver ID to ensure as even a distribution as possible of all caller–receiver dyads across all folds. Thus, if calls contain acoustic cues to receiver ID, this model was expected to predict receiver ID better than chance regardless of whether the label for a given receiver is shared across callers (Table 1, hypothesis 1, prediction 1). We only used calls where caller ID was known for certain ($n = 437$ calls). The model used 500 trees, 6 variables per node, 60% of observations per tree, a minimum node size of 1 and no maximum tree depth. To increase the stability of the model's classification accuracy, we ran the model 2,000 times and used the mean classification accuracy across the 2,000 runs. To determine if the model predicted receiver ID better than expected by chance, we ran the model 10,000 times with the acoustic features randomly permuted and compared the classification accuracy of the original model (averaged across 2,000 runs) with the null distribution of classification accuracies generated by the 10,000 models with randomized acoustic features (one-tailed permutation test).

To disentangle the effects of caller ID and receiver ID on call structure, we compared the mean pairwise similarities between pairs of calls with the same caller and receiver and pairs with the same caller and different receivers (same caller pair type). As a metric of call similarity, we extracted a proximity score for each pairwise combination of calls from a random forest trained to predict receiver ID as a function of the acoustic features on the full dataset (469 training observations, 8,000 trees, other hyperparameters same as above). The proximity score for a given pair of calls was the proportion of trees in which both calls were classified in the same terminal node, corrected for the size of each node and represented the degree of similarity between the two calls in terms of the acoustic features most relevant to predicting receiver ID[17]. If calls are specific to individual receivers within a given caller, then pairs of calls with the same caller and same receiver should be more similar (have higher proximity scores) than pairs of calls with the same caller and different receivers (Table 1, hypothesis 1, prediction 2).

Previous work has shown that elephants vary the structure of their rumbles when interacting with dominant versus subordinate conspecifics[12]. To rule out the possibility that calls were specific to the type of relationship between caller and receiver rather than to individual receivers per se, we restricted the analysis of same caller pair type to pairs of calls that had the same type of relationship between caller and receiver. We defined the caller–receiver relationship using 12 categories based on sex, family group membership, relative age and mother–offspring relationship, reflecting the fact that dominance in elephants is primarily determined by age[50,51] and that mother–calf bonds are the strongest social bonds in elephants[22,52] (Extended Data Table 2). As calls from different behavioural contexts differ in acoustic structure[15], we categorized each pair of calls according to whether the two calls had the same or different behavioural contexts ('same context') and included this variable as a factor in the analysis. We also included a binary factor indicating whether the two calls were recorded on the same date, as exploratory analyses indicated that calls recorded on the same date were more similar than calls recorded on different dates. We only used calls in this model for which the caller ID and behavioural context were known for certain.

The proximity scores were highly skewed to the right, so we rank-transformed them and ran a linear mixed model with rank-transformed proximity score as the response variable and same caller pair type, same context and same date as fixed effects. To account for the fact that there were multiple call pairs with the same combination of callers and receivers, we included 'pair ID' (a unique identifier for each caller–receiver–caller–receiver combination) as a random effect. We excluded pair IDs with only one observation as it was not possible to estimate within-class variability for these pair IDs (final $n = 1{,}284$ call pairs).

Which calls are most likely to contain vocal labels? Vocal labels might be more likely to occur in certain behavioural contexts than others. Similarly, callers may only use a vocal label in some of the calls within a bout, as it would be redundant to include the same information in all the calls. To assess whether behavioural context or position within a bout influenced the likelihood of a call containing a vocal label, we calculated the proportion of the 2,000 iterations of the random forest in which the receiver ID was correctly predicted for each call (probability of correct classification). We designated calls that were correctly predicted in ≥95% of iterations as 'correct' and calls that were correctly predicted in ≤5% of iterations as 'incorrect' and excluded all calls that did not meet this criteria, as well as all calls with uncertain caller ID or behavioural context, and receivers that occurred only once after applying the previous criteria ($n = 327$). Then, we ran a mixed-effects logistic regression with prediction outcome (1 or 0) as the response, receiver ID as a random effect, and behavioural context, caller age class, position within the bout and the total number of calls addressed to the receiver in question as fixed effects. The latter effect was included because receivers with more calls in our dataset were expected to be predicted with greater accuracy, as there were more training opportunities for the random forest to learn them. Caller age class was defined as juvenile (<10 years old for females, not yet dispersed from natal group for males) or adult (>10 years old for females). There were no adult male callers in our dataset. We defined a bout as calls produced by the same caller within the same sound file with no more than 30 s between successive calls.

Are vocal labels based on imitation of the receiver's calls (hypothesis 2)? To assess whether imitation of the receiver's calls was necessary for vocal labelling, we examined the calls in the dataset for which we had at least one recording of the receiver's calls and at least one recording of the caller addressing someone other than the receiver ($n = 236$ calls). For each of these calls, we calculated its mean proximity score to all the calls made by the receiver (mean proximity to targeted receiver). We also calculated the mean proximity score between the same caller and receiver when the caller was addressing other individuals (mean proximity when targeting others). Calls in which the mean proximity to targeted receiver was greater than the mean proximity when targeting others were classified as 'convergent' ($n = 95$) and divergent otherwise ($n = 141$). We then examined the proportion of convergent and divergent calls that were classified correctly by the random forest model with receiver ID and the acoustic features as input variables, and cross-validation folds stratified by caller ID and receiver ID. If vocal labelling relies on imitation of the receiver's calls, we expected only the convergent calls to be classified correctly more often than by the null model, but if imitation is not necessary for vocal labelling, we expected both convergent and divergent calls to be classified correctly more often than by the null model (Table 1, hypothesis 2, prediction 1).

If elephants imitate the calls of the receiver that they are addressing, then callers should sound more like a given conspecific when they are addressing her than when they are addressing someone else (Table 1, hypothesis 2, prediction 2). To assess whether this was the case, we classified each pair of calls into one of two types (hereafter, 'imitation pair type'): pairs in which the receiver of one call was the caller of the other call, and pairs in which this was not the case. We separately classified each call pair according to whether the two calls had the same relationship between caller and receiver (hereafter, 'same relationship'). We also created a categorical variable caller dyad ID, which was an identifier for each unique combination of callers that composed a call pair. We ran a linear mixed model with rank-transformed proximity score as the response variable, imitation pair type, same relationship, same context and same date as fixed effects, and caller dyad ID and pair ID as random effects. By including caller dyad ID as a random effect, we assessed the effect of imitation pair type within a given pair of callers, that is, whether calls from caller A to receiver B were more similar to the receiver B's calls than calls from caller A addressed to other receivers were to receiver B's calls. We excluded pairs of calls with the same caller or receiver, uncertain caller ID or behavioural context for either call, that were recorded from different family groups, for which caller dyad ID did not occur with both levels of imitation pair type, or for which pair ID occurred only once ($n = 2,360$ call pairs). Pairs of calls from different family groups were excluded because they comprised a small percentage of pairs where the receiver of one call was the caller of the other, and because it is possible that different families have different vocal signatures, which would influence call similarity.

Do different callers use the same label for the same receiver (hypothesis 3)? If different callers use similar labels for the same receiver, then pairs of calls with different callers and the same receivers should be more similar than pairs of calls with different callers and different receivers (Table 1, hypothesis 3, prediction 1). To test whether this was the case, we ran another linear mixed model with rank-transformed proximity score as the response variable, different caller pair type (different callers/same receiver or different callers/same receiver), same relationship and same context as fixed effects, and pair ID as a random effect. As before, we excluded calls with uncertain caller ID or behavioural context, pairs of calls recorded from different family groups, and levels of pair ID that occurred only once ($n = 8,215$ call pairs).

To determine if receiver ID could be predicted independently of caller ID, which would be possible only if callers use similar labels for a given receiver, (Table 1, hypothesis 3, prediction 2), we ran another sevenfold cross-validated random forest model to predict receiver ID as a function of the acoustic features but partitioned the cross-validation folds such that all calls with the same caller and receiver were always allocated to the same fold (observations and hyperparameters same as first model). We averaged the classification accuracy of the model across 2,000 runs and compared this value with the distribution of classification accuracies generated by 10,000 iterations of the same model with the acoustic features randomly permuted (one-tailed permutation test).

Checking model assumptions. For all rank-transformed linear mixed models, we checked the assumption of normality by visually examining histograms of the residuals. We checked the assumption of equal variances by visually examining boxplots of all groups. The residuals for all models exhibited only minor deviations from normality, with the absolute values of skewness and excess kurtosis being less than 1 for all models. As linear models have been shown to be robust even to severe deviations from normality with skewness as high as 2 and excess kurtosis as high as 6 (a normal distribution has a skewness of 0 and excess kurtosis of 0)[53], we deemed the choice of model appropriate. Boxplots indicated similar variances across groups.

How are labels encoded in calls? To investigate which acoustic features encode receiver ID and caller ID, we extracted variable importance scores (Supplementary Table 2) from a conditional inference random forest model in the R package 'party'[54] trained on the full dataset to predict the response variable in question (receiver ID or caller ID) as a function of the acoustic features (469 training observations for receiver ID, 437 for caller ID; 1,000 trees; all other hyperparameters same as other random forests). We used a conditional inference forest because, unlike traditional random forest, it is not biased towards correlated variables[54]. We only calculated variable importance scores for the spectral features, as cepstral coefficients are difficult to interpret intuitively. To assess the relative importance of the original acoustic contours, we weighted the loadings of the acoustic contours on each principal component by the variable importance score of the mean of the principal component in question and then calculated the sum of the absolute values of these weighted loadings for each acoustic contour (Supplementary Table 3). Acoustic contours with a higher sum of the absolute values of the weighted loadings were deemed more important. This weighting process only considered the means of low-rank principal components.

Playback experimental design

To determine if elephants respond more strongly to calls addressed to them (Table 1, hypothesis 1, prediction 3), we played back rumbles with known adult (>10-year-old) female callers and known receivers to 17 elephants (15 adult females, one 9-year-old female, one 9–10-year-old male) in the Samburu study area. Fourteen subjects received one 'test' playback of a call that was originally addressed to them and one 'control' playback of a call from the same caller that was originally addressed to another individual. One subject received two sets of test and control playbacks from two different callers, one received only a test playback, and one received only a control playback (Supplementary Table 4). Most stimuli functioned as the test stimulus for one subject and the control stimulus for another, but no stimulus was used as the same experimental condition for more than one subject. The order of presentation was balanced across subjects, and we waited at least 7 days (mean ± s.d., 29.5 ± 27.1 days) between successive playbacks to the same subject.

Playback stimuli

Playback stimuli were recorded in Samburu and Buffalo Springs between January 2020 and March 2022 from adult female callers. In all but two cases, the playback stimuli were contact calls. In one case we used a loud greeting call (similar in original amplitude to a typical contact call but produced at a much closer distance), and in one case we used a call that was produced in a similar context to contact calls (caller and receiver >100 m apart and out of sight of each other) but was lower in original amplitude than a typical contact call and was part of a lengthy antiphonal exchange between two individuals and, therefore, was probably a 'cadenced rumble'[15]. These non-contact calls were used to complete a pair of test and control stimuli because we were unable to obtain contact calls to two different receivers from the same caller. Three playback stimuli were elicited by another playback, and we assumed that the individual whose call was broadcast from the speaker was the intended receiver of the call that was produced in response to that playback. We identified the receiver of natural calls as the only adult member of the family group who was separated from the caller during the call or the only individual who responded to the call. In one case, there were two adult females separated from the caller, and we assumed the receiver was the older of the two females who was in the lead and who rejoined the caller first (Extended Data Table 7). We note that there was no mechanism to ensure the playback stimulus contained a vocal label, and it is possible not all stimuli were labelled. We prepared all playback stimuli in Audacity 3.0.2. Each stimulus consisted of a single rumble preceded by one second of background noise with a fade-in and followed by 1 s of background noise with a fade-out. In three cases, we applied a high-pass (5 Hz cut-off, 6 dB roll-off) or low-pass filter (1,000 Hz cut-off, 6 dB roll-off) to remove excessive noise.

Playback system and volume

We played back all stimuli as .wav files (uncompressed audio) from an iPhone SE (Apple) attached to a QLXD1 wireless bodypack transmitter (Shure) transmitting to a custom-built loudspeaker (Bag End Loudspeakers). The cord connecting the playback device to the wireless transmitter had to be replaced three times over the course of the experiment, each time changing the output level of the speaker. Thus, depending on which cord was in use, we normalized the stimuli to −24, −22.5 or −18 dB in Audacity 3.0.2 to ensure a functionally equivalent normalization level across all trials.

The speaker's frequency response was flat from 10 Hz to 500 Hz up to a given maximum output level (maximum output 89 dB sound pressure level (SPL) at 10 Hz, 101 dB SPL at 20 Hz and 113 dB SPL at 40 Hz). If the signal exceeded the maximum output at a given frequency, the speaker automatically reduced the level of the frequencies in question to avoid damage. Reported amplitudes for natural contact calls range from 94 to 115 dB SPL (extrapolated value at 1 m from source)[15,55]. We did not have access to an SPL meter with a flat frequency response at low frequencies, but our playback stimuli ranged from 96.2 to 104.3 dBC at 1 m measured with a Protmex PT6708 sound level meter (Protech International Group Co.) or 93.4 to 102.9 dB SPL at 1 m measured with the SoundMeter 10.5.8 iPhone application (Faber Acoustical). Mean measured volume did not differ between test and control stimuli (dBC: t-test, $t_{32.0} = 0.03$, $P = 0.97$; dB SPL: t-test, $t_{32.0} = 0.15$, $P = 0.88$). Q12

Playback trial protocol

We placed the speaker 40.2–59.0 m from the subject (mean 49.1 ± 4.2 m), either on the ground in front of a tree or shrub and covered by camouflage netting or on the edge of the rear seat of a Toyota double cab Landcruiser facing the door with all four doors and windows and both roof hatches open. Rerecordings at 50 m revealed no obvious difference between sounds played with the speaker on the ground or inside the vehicle. We conducted playbacks only when the original caller and 'alternate receiver' (the other subject receiving playbacks from the same caller) were >180 m from and out of sight of the subject (>270 m from the alternate receiver if she had not yet received all her playbacks). When the original caller's location was known (19/34 trials) the speaker was placed in approximately the same direction relative to the subject as the original caller. In the remaining trials, the caller could not be located after searching a ~300 m radius around the subject. Trials were redone after at least 7 days if the speaker malfunctioned, the subject moved her head out of sight right before the playback started or we discovered after the playback that the speaker was not in the correct location relative to the subject and the original caller. During each trial, we filmed the subject from inside the vehicle for at least 1 min before the playback, then played the stimulus once and continued filming for at least another 10 min. We also recorded audio with an Earthworks QTC40 microphone and Sound Devices MixPre3-II recorder. The observers were blind to the playback condition (test or control) until all trials were complete, and all videos and audio recordings were scored.

Statistical analysis of playback data

From the video and audio recordings of each playback trial, we measured the subject's latency to approach the speaker, latency to vocalize, number of calls produced within 10 min following the playback, latency to vigilance and change in vigilance duration in the minute following the playback compared with the minute preceding the playback. Latencies were defined as the time from the start of the playback until the behaviour of interest occurred and were censored when the subject moved out of sight or at 10 min, whichever came first. Vigilance was defined as lifting head above shoulder level, moving head from side to side, holding ears away from body without flapping, or lifting trunk while sniffing towards speaker[56]. We ran a separate model for each response variable with subject ID as a random effect and treatment and the following covariates/factors as fixed effects: caller–original receiver relationship (relationship between the caller and the original receiver of the call; Extended Data Table 3), distance (distance in metres between the speaker and the subject), dBC (amplitude of the playback stimulus in dBC at 1 m), other adults (whether other adults were within 50 m of subject during playback), speaker location (whether speaker was on ground or in vehicle) and cumulative playback exposure (cumulative number of playbacks to which subject was exposed at distance of 300 m or less, including trials that were redone and playbacks to other subjects). We used Cox proportional hazards regression in the coxme package[57] for the latency variables, a generalized linear model with a Poisson error distribution in the lme4 package[58] for number of calls, and a linear model for change in vigilance duration. We applied analysis of deviance with type III sums of squares to each model to calculate a two-tailed P value for each fixed effect. For the Poisson regression modelling number of calls, the random effect of subject ID had a variance of 0, resulting in a near singular fit, so we removed the random effect from this model.

For the Cox regression models, we checked the assumption of proportional hazards with a Schoenfeld test, which tests the null hypothesis that there is no relationship between the scaled Schoenfeld residuals and time. This test was non-significant ($P > 0.05$) for all models, indicating no violation of the proportional hazards assumption. For the Poisson regression model, we checked for overdispersion using the AER package in R[59]. The dispersion parameter was estimated to be 1.1, which did not differ significantly from the ideal value of 1 ($P = 0.26$), indicating that a Poisson distribution was appropriate. For the linear regression model used to examine the change in vigilance duration before versus after playbacks, visual inspection of the histogram of the residuals indicated that the residuals were approximately normally distributed. For treatment, distance, dBC, speaker location and cumulative playback exposure, visual inspection of boxplots or residual plots indicated approximate homoscedasticity. Relationship of caller to original receiver and other adults were heteroscedastic. However, regardless of whether these covariates were included, treatment was not significant, so any potential issues with this model had no bearing on the conclusions of our study.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data are available at https://doi.org/10.5061/dryad.hmgqnk9nj.

Code availability

Code is available at https://doi.org/10.5281/zenodo.10576772.

References

1. Fitch, W. T. The evolution of language: a comparative review. *Biol. Philos.* **20**, 193–230 (2005).
2. Macedonia, J. M. & Evans, C. S. Variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethology* **93**, 177–197 (1993).
3. Clay, Z., Smith, C. L. & Blumstein, D. T. Food-associated vocalizations in mammals and birds: what do these calls really mean? *Anim. Behav.* **83**, 323–330 (2012).
4. Wheeler, B. C. & Fischer, J. Functionally referential signals: a promising paradigm whose time has passed. *Evol. Anthropol.* **21**, 195–205 (2012).
5. Smith, E. A. Communication and collective action: language and the evolution of human cooperation. *Evol. Hum. Behav.* **31**, 231–245 (2010).
6. Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H. & Monaghan, P. Arbitrariness, iconicity, and systematicity in language. *Trends Cogn. Sci.* **19**, 603–615 (2015).
7. King, S. L. & Janik, V. M. Bottlenose dolphins can use learned vocal labels to address each other. *Proc. Natl Acad. Sci. USA* **110**, 13216–13221 (2013).
8. Balsby, T. J. S., Momberg, J. V. & Dabelsteen, T. Vocal imitation in parrots allows addressing of specific individuals in a dynamic communication network. *PLoS ONE* **7**, e49747 (2012).
9. Janik, V. M. & Sayigh, L. S. Communication in bottlenose dolphins: 50 years of signature whistle research. *J. Comp. Physiol. A* **199**, 479–489 (2013).
10. Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S. & Watwood, S. Elephants are capable of vocal learning. *Nature* **434**, 455–456 (2005).
11. Stoeger, A. S. et al. An Asian elephant imitates human speech. *Curr. Biol.* **22**, 2144–2148 (2012).
12. Soltis, J., Leong, K. & Savage, A. African elephant vocal communication II: rumble variation reflects the individual identity and emotional state of callers. *Anim. Behav.* **70**, 589–599 (2005).
13. Clemins, P. J., Johnson, M. T., Leong, K. M. & Savage, A. Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations. *J. Acoust. Soc. Am.* **117**, 956–963 (2005).
14. McComb, K., Moss, C., Sayialel, S. & Baker, L. Unusually extensive networks of vocal recognition in African elephants. *Anim. Behav.* **59**, 1103–1109 (2000).
15. Poole, J. H. in *The Amboseli Elephants: A Long-Term Perspective on a Long-Lived Mammal* (eds Moss, C. J., Croze, H. & Lee, P. C.) 125–159 (Univ. Chicago Press, 2011).
16. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
17. Rhodes, J. S., Cutler, A. & Moon, K. R. Geometry- and accuracy-preserving random forest proximities. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 10947–10959 (2023).
18. Foley, N. M. et al. A genomic timescale for placental mammal evolution. *Science* **380**, eabl8189 (2023).
19. Dahlin, C. R., Young, A. M., Cordier, B., Mundry, R. & Wright, T. F. A test of multiple hypotheses for the function of call sharing in female budgerigars, *Melopsittacus undulatus*. *Behav. Ecol. Sociobiol.* **68**, 145–161 (2014).
20. Wanker, R., Sugama, Y. & Prinage, S. Vocal labelling of family members in spectacled parrotlets, *Forpus conspicillatus*. *Anim. Behav.* **70**, 111–118 (2005).
21. Prat, Y., Taub, M. & Yovel, Y. Everyday bat vocalizations contain information about emitter, addressee, context, and behavior. *Sci. Rep.* **6**, 39419 (2016).
22. Wittemyer, G., Douglas-Hamilton, I. & Getz, W. M. The socioecology of elephants: analysis of the processes creating multitiered social structures. *Anim. Behav.* **69**, 1357–1371 (2005).
23. Archie, E. A., Moss, C. J. & Alberts, S. C. The ties that bind: genetic relatedness predicts the fission and fusion of social groups in wild African elephants. *Proc. R. Soc. B* **273**, 513–522 (2006).
24. Howard, D. J., Gengler, C. & Jain, A. What's in a name? A complimentary means of persuasion. *J. Consum. Res.* **22**, 200–211 (1995).
25. King, S. L., Sayigh, L. S., Wells, R. S., Fellner, W. & Janik, V. M. Vocal copying of individually distinctive signature whistles in bottlenose dolphins. *Proc. R. Soc. B* **280**, 20130053 (2013).
26. Baotic, A. & Stoeger, A. S. Sexual dimorphism in African elephant social rumbles. *PLoS ONE* **12**, e0177411 (2017).
27. Stoeger, A. S., Zeppelzauer, M. & Baotic, A. Age-group estimation in free-ranging African elephants based on acoustic cues of low-frequency rumbles. *Bioacoustics* **23**, 231–246 (2014).
28. Zaman, S. R., Sadekeen, D., Alfaz, M. A. & Shahriyar, R. One source to detect them all: gender, age, and emotion detection from voice. In *Proc. IEEE 45th Annual Computers, Software, and Applications Conference* 338–343 (IEEE, 2021). Q13
29. Berg, K. S., Delgado, S., Cortopassi, K. A., Beissinger, S. R. & Bradbury, J. W. Vertical transmission of learned signatures in a wild parrot. *Proc. R. Soc. B* **279**, 585–591 (2012).
30. Stevens, S. S., Volkmann, J. & Newman, E. B. A scale for the measurement of the psychological magnitude pitch. *J. Acoust. Soc. Am.* **8**, 185–190 (1937).
31. Vernes, S. C. et al. The multi-dimensional nature of vocal learning. *Philos. Trans. R. Soc. B* **376**, 20200236 (2021).
32. Bradbury, J. W. & Balsby, T. J. S. The functions of vocal learning in parrots. *Behav. Ecol. Sociobiol.* **70**, 293–312 (2016).
33. Connor, R. C. Dolphin social intelligence: complex alliance relationships in bottlenose dolphins and a consideration of selective environments for extreme brain size evolution in mammals. *Philos. Trans. R. Soc. Lond. B* **362**, 587–602 (2007).
34. Bachorec, E. et al. Spatial networks differ when food supply changes: foraging strategy of Egyptian fruit bats. *PLoS ONE* **15**, 1–15 (2020).

35. Kerth, G., Perony, N. & Schweitzer, F. Bats are able to maintain long-term social relationships despite the high fission–fusion dynamics of their groups. *Proc. R. Soc. B* **278**, 2761–2767 (2011).
36. Moss, C. J. & Poole, J. H. In *Primate Social Relationships: An Integrated Approach* (ed. Hinde, R. A.) 315–325 (Blackwell Science, 1983).
37. Altmann, J. Observational study of behavior: sampling methods. *Behaviour* **49**, 227–267 (1974).
38. de Silva, S. Acoustic communication in the Asian elephant, *Elephas maximus maximus*. *Behaviour* **147**, 825–852 (2010).
39. R Core Team. R: a language and environment for statistical computing. *R Foundation for Statistical Computing* https://www.R-project.org (2022).
40. Sueur, J., Aubin, T. & Simonis, C. seewave, a free modular tool for sound analysis. *Bioacoustics* **18**, 213–226 (2008).
41. Ligges, U., Krey, S., Mersmann, O. & Schnackenberg, S. tuneR: analysis of music and speech. *R Project* https://CRAN.R-projet.org/package=tuneR (2018).
42. Anikin, A. Soundgen: an open-source tool for synthesizing nonverbal vocalizations. *Behav. Res. Methods* **51**, 778–792 (2019).
43. Heffner, R. S. & Heffner, H. E. Hearing in the elephant (*Elephas maximus*): absolute sensitivity, frequency discrimination, and sound localization. *J. Comp. Physiol. Psychol.* **96**, 926–944 (1982).
44. Ren, Y. et al. A framework for bioacoustic vocalization analysis using hidden Markov models. *Algorithms* **2**, 1410–1428 (2009).
45. Davis, S. B. & Mermelstein, P. Comparison of parametric representations for monosyllabic word recognition. *IEEE Trans. Acoust.* **28**, 357–366 (1980).
46. Sykulsi, M. rpca: RobustPCA: decompose a matrix into low-rank and sparse components. R package version 0.2.3. *R Project* https://CRAN.R-project.org/package=rpca. (2015).
47. Thomson, D. J. Spectrum estimation and harmonic analysis. *Proc. IEEE* **70**, 1055–1096 (1982).
48. Correll, J., Mellinger, C. & Pedersen, E. J. Flexible approaches for estimating partial eta squared in mixed-effects models with crossed random factors. *Behav. Res. Methods* **54**, 1626–1642 (2022).
49. Wright, M. N. & Ziegler, A. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **77**, 1–17 (2017).
50. Wittemyer, G. & Getz, W. M. Hierarchical dominance structure and social organization in African elephants, *Loxodonta africana*. *Anim. Behav.* **73**, 671–681 (2007).
51. Archie, E. A., Morrison, T. A., Foley, C. A. H., Moss, C. J. & Alberts, S. C. Dominance rank relationships among wild female African elephants, *Loxodonta africana*. *Anim. Behav.* **71**, 117–127 (2006).
52. Archie, E. A., Moss, C. J. & Alberts, S. C. in *The Amboseli Elephants: A Long-Term Perspective on a Long-Lived Mammal* (eds Moss, C. J., Croze, H. & Lee, P. C.) 238–245 (Univ. Chicago Press, 2011).
53. Blanca, M. J., Alarcón, R., Arnau, J., Bono, R. & Bendayan, R. Non-normal data: is ANOVA still a valid option? *Psicothema* **29**, 552–557 (2017).
54. Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T. & Zeileis, A. Conditional variable importance for random forests. *BMC Bioinform.* **9**, 1–11 (2008).
55. Poole, J. H., Payne, K., Langbauer, W. R. J. & Moss, C. J. The social contexts of some very low-frequency calls of African elephants. *Behav. Ecol. Sociobiol.* **22**, 385–392 (1988).
56. Poole, J. H. & Granli, P. in *The Amboseli Elephants: A Long-Term Perspective on a Long-Lived Mammal* (eds Moss, C. J., Croze, H. & Lee, P. C.) 109–124 (Univ. Chicago Press, 2011).
57. Therneau, T. M. coxme: mixed effects cox models. R package version 2.2-18.1. *R Project* https://CRAN.R-project.org/package=coxme (2019).
58. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
59. Kleiber, C. & Zeileis, A. *Applied Econometrics with R* (Springer, 2008).

Acknowledgements

We thank the Office of the President of Kenya, the Samburu, Isiolo and Kajiado County governments, the Wildlife Research & Training Institute of Kenya, and Kenya Wildlife Service for permission to conduct fieldwork in Kenya. We thank Save The Elephants and the Amboseli Trust for Elephants for logistical support in the field, J. M. Leshudukule, D. M. Letitiya and N. Njiraini for assistance with the fieldwork, G. Pardo for blinding the playback stimuli and S. Pardo for input on the statistical analyses. We thank J. Berger, W. Koenig, A. Horn and three anonymous reviewers for comments on the manuscript. This project was funded by a Postdoctoral Research Fellowship in Biology to M.A.P. from the National Science Foundation (award no. 1907122) and grants to J.P. and P.G. from the National Geographic Society, Care for the Wild, and the Crystal Springs Foundation. Fieldwork was supported by Save the Elephants.

Author contributions

M.A.P. conceived the study. M.A.P. and D.S.L. collected the data in Samburu, and J.P. and P.G. collected the data in Amboseli. M.A.P. and K.F. performed the statistical analysis, and M.A.P. created the figures. M.A.P. drafted the manuscript, and K.F., J.P. and G.W. edited it. C.M., I.D.-H. and G.W. provided resources and access to long-term datasets, and G.W. supervised the study.

Competing interests

The authors declare no competing interests.

Additional information

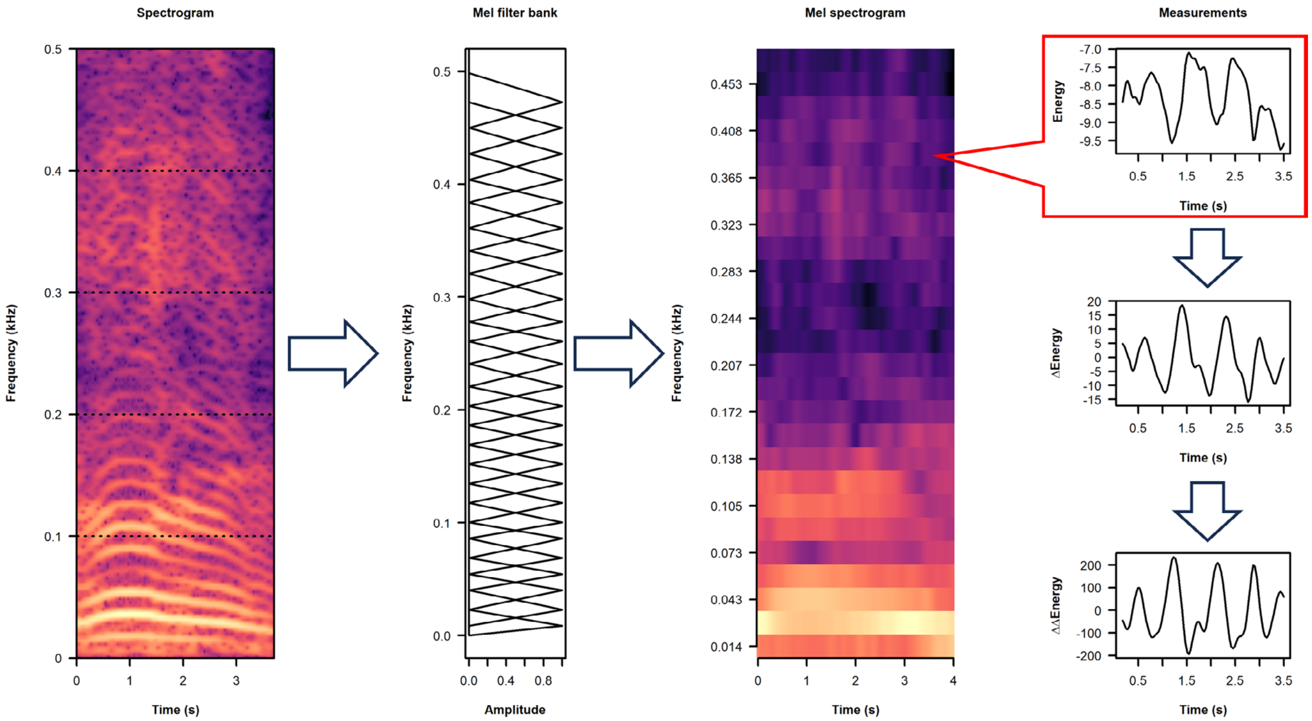
Extended data is available for this paper at https://doi.org/10.1038/s41559-024-02420-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-024-02420-w.

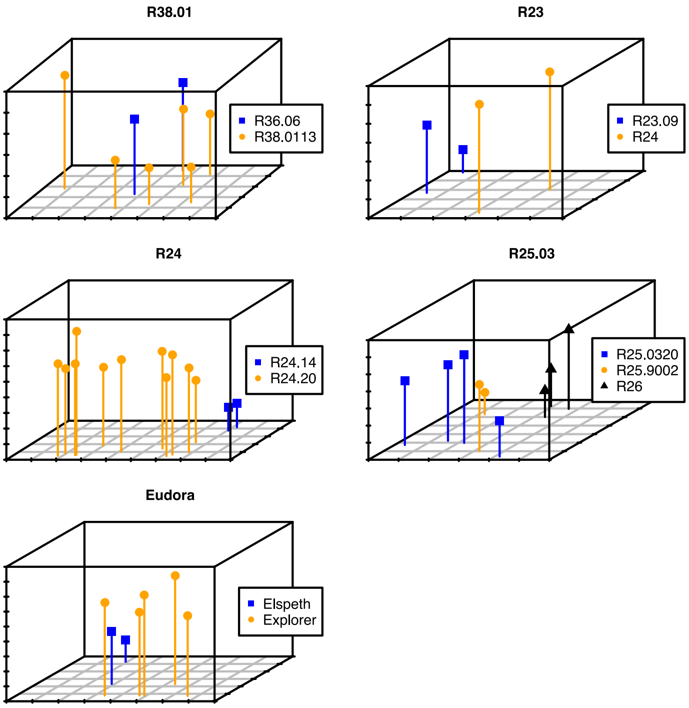
Correspondence and requests for materials should be addressed to Michael A. Pardo.

Peer review information *Nature Ecology & Evolution* thanks Kenna Lehmann and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

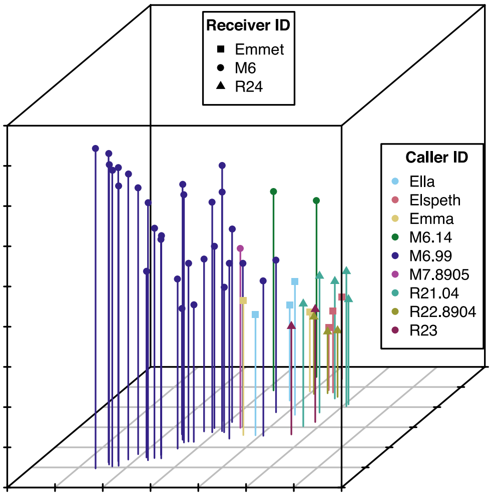
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Schematic illustrating how spectral acoustic features were measured. First, a spectrogram was calculated by applying a Fast Fourier Transform to the signal (Hamming window, 700 samples, 90% overlap). Then a mel filter bank with 26 overlapping triangular filters between 0-500 Hz was applied to each window of the spectrogram to produce a mel spectrogram. The mel spectrogram was then normalized by dividing the energy value in each cell by the total energy in that time window and these proportional energies were logit-transformed so they would not be limited to between 0 and 1. As features for the robust principal components analysis, we used the vector of energy in each of the 26 mel frequency bands as well as the vectors of delta and delta-delta values for each frequency band (representing the change and acceleration in energy over time, respectively). In the spectrogram and mel spectrogram in this figure, warmer colors indicate higher amplitudes (greater energy).



Extended Data Fig. 2 | Scatterplots illustrating the separation in 3D space between calls from the same caller to different receivers. Axes are the first three principal coordinates extracted from the proximity scores of a random forest trained to predict receiver ID. Each plot represents a single caller, each point is a single call, and receiver IDs are coded by both color and shape. This figure only includes calls where caller ID was known for certain, where the call was predicted correctly in at least 25% of random forest iterations, and where the caller made at least two such calls each to at least two different receivers.



Extended Data Fig. 3 | Scatterplot illustrating the clustering in 3D space of calls from different callers to the same receiver. Axes are the first three principal coordinates extracted from the proximity scores of a random forest trained to predict receiver ID. Each shape represents a different receiver and each color represents a different caller. This figure only includes calls where caller ID was known for certain, where the call was predicted correctly in at least 25% of random forest iterations, and where the receiver received at least one such call each from at least two different callers.

Extended Data Table 1 | Acoustic features used in the random forest models

Derived acoustic feature	Derived from	Description
Median	Sparse matrix	Median of all the values in sparse matrix
Robust skewness	Sparse matrix	Scaled median difference of left and right half of the distribution of values in sparse matrix. Estimate of skewness of sparse matrix
Minimum extent	Sparse matrix	Estimate of spread of sparse matrix, calculated as sum(x)/max(x)/length(x), where x is sparse matrix
Equivalent statistical extent	Sparse matrix	Estimate of spread of sparse matrix, calculated as sum(x)^2/sum(x^2)/length(x), where x is sparse matrix
Means of low-rank principal components explaining 99.9% of variation	Low-rank matrix	Arithmetic means of robust PCs required to explain 99.9% of variation in low-rank matrix (74 for spectral features, 12 for cepstral features)
Frequencies of unweighted PC spectral peaks 1-4	Low-rank matrix	Frequencies of 4 largest peaks aggregated across unweighted PC spectra, sorted from lowest to highest frequency
Magnitudes of unweighted PC spectral peaks 1-4	Low-rank matrix	Amplitudes of 4 largest peaks aggregated across unweighted PC spectra, sorted from lowest to highest frequency
Frequencies of weighted PC spectral peaks 1-4	Low-rank matrix	Frequencies of 4 largest peaks aggregated across weighted PC spectra, sorted from lowest to highest frequency. Spectra weighted by proportion of variation explained by each PC
Magnitudes of weighted PC spectral peaks 1-4	Low-rank matrix	Amplitudes of 4 largest peaks aggregated across weighted PC spectra, sorted from lowest to highest frequency. Spectra weighted by proportion of variation explained by each PC

All acoustic features were derived from either the sparse matrix or low-rank matrix of a robust principal components analysis performed on multiple acoustic contours of equal length that were measured directly from the signal. For the spectral acoustic features, the acoustic contours were the Hilbert amplitude envelope, the vector of energies in each of the 26 bands of a mel spectrogram, and the delta and delta-delta values of the mel spectral bands. For the cepstral acoustic features, the acoustic contours were the Hilbert amplitude envelope, first 12 mel-frequency cepstral coefficients, and the delta and delta-delta values of the first 12 cepstral coefficients. The principal components analysis was performed on a matrix of all the contours for each call stacked end-to-end.

Extended Data Table 2 | Results of random forest models predicting receiver ID as a function of the acoustic features

Hypothesis tested	Observations used	Data partitioning	Classification accuracy	Mean ± SD accuracy for permuted models	Permutation test *P*-value
Spectral acoustic features					
H1: calls are receiver specific	All with certain caller ID (437)	Stratified by caller and receiver ID	27.5%	8.0 ± 0.66%	0.000
H2: labels are arbitrary	Convergent calls (95)	Stratified by caller and receiver ID	20.1%	7.7 ± 1.3%	0.000
H2: labels are arbitrary	Divergent calls (141)	Stratified by caller and receiver ID	32.6%	17.9 ± 1.6%	0.000
H3: labels shared across callers	All with certain caller ID (437)	All calls with same caller and receiver in same fold	1.1%	1.4 ± 0.33%	0.896
Cepstral acoustic features					
H1: calls are receiver specific	All with certain caller ID (437)	Stratified by caller and receiver ID	19.8%	6.5 ± 0.95%	0.000
H2: labels are arbitrary	Convergent calls (95)	Stratified by caller and receiver ID	10.5%	4.0 ± 1.7%	0.001
H2: labels are arbitrary	Divergent calls (141)	Stratified by caller and receiver ID	28.3%	15.1 ± 2.3%	0.000
H3: labels shared across callers	All with certain caller ID (437)	All calls with same caller and receiver in same fold	0.83%	1.3 ± 0.44%	0.866

All random forests had 500 trees, 6 variables per node, 60% of observations per tree, minimum node size = 1, no maximum tree depth, and 7-fold cross-validation. Classification accuracies were averaged across 2000 runs of the model to improve stability. To determine if the classification accuracy was higher than expected by chance, the model was run 10,000 times with randomly permuted acoustic variables, and the original classification accuracy was compared to the distribution of classification accuracies for these 10,000 permuted models. *P*-values are one-tailed.

Extended Data Table 3 | Definitions of social relationship categories between caller and receiver

Relationship category	Code	Definition
Mother of adult female	Mom2AdDau	Caller is mother of receiver and receiver is adult female
Mother of juvenile	Mom2JuvOff	Caller is mother of receiver and receiver is juvenile
Adult daughter	Adult2Mom	Caller is adult daughter of receiver
Juvenile offspring	Juv2Mom	Caller is juvenile offspring of receiver
Older adult female to younger adult female	Ad2YngrAd	Caller and receiver both adult females from same family group. Caller ≥6 years older than receiver and not mother of receiver
Adult (female) age mates	AdAgeMates	Caller and receiver <6 years apart, from same family group, at least one is adult female, neither is adult male
Younger adult female to older adult female	Ad2OldrAd	Caller is adult female, ≥6 years younger than receiver, not daughter of receiver, both from same family group
Adult female to juvenile	Adult2Juv	Caller is adult female and not mother of receiver, receiver is juvenile and ≥6 years younger than caller, both from same family group
Juvenile to adult female	Juv2Adult	Caller is juvenile and ≥6 years younger than caller, receiver is adult female and not mother of caller, both from same family group
Juvenile to juvenile	Juv2Juv	Caller and receiver both juveniles from same family group
Different groups	FemDiffGrps	Caller and receiver from different family groups; neither is adult male
Adult male	AdultMale	Caller and/or receiver is adult male

Categories were defined based on sex, age, and mother-offspring status, the most important factors influencing dominance and bond strength within an elephant family group. Females were defined as adults if ≥10 years old, and males were defined as adults if independent from their natal group. All non-adults under this definition were classified as juveniles. Six years was chosen as the cutoff for different age classes because it is between 1-2x the average inter-birth interval, so a female ≥6 years older than another individual could have been that individual's allomother.

Extended Data Table 4 | Results for linear mixed model assessing whether calls are specific to individual receivers or the type of relationship between caller and receiver

Model term	Param. estimate	Std. error	χ^2 statistic	DF	*P*-value	Partial η^2	Std. dev. (rand. eff.)
Spectral acoustic features							
Same Caller Pair Type	-214.5	59.6	13.0	1	0.0003	0.063	--
Same Context	83.8	31.6	7.0	1	0.008	0.014	--
Same Date	262.3	22.1	140.9	1	0.0000	0.163	--
Pair ID	--	--	--	--	--	--	174.6
Cepstral acoustic features							
Same Caller Pair Type	-155.3	59.8	6.7	1	0.009	0.026	--
Same Context	69.8	32.7	4.6	1	0.033	0.019	--
Same Date	230.9	23.0	101.3	1	0.0000	0.139	--
Pair ID	--	--	--	--	--	--	172.8

Each observation was a pair of calls and the response variable was rank-transformed proximity score. Same Caller Pair Type = whether the two calls in a pair had the same caller and receiver (reference level) or same caller and different receivers with the same type of relationship to the caller; Same Context = whether the two calls in a pair had the same behavioral context (reference level = no); Same Date = whether the two calls in a pair were recorded on the same day; Pair ID = unique combination of callers and receivers (random effect). Pairs of calls recorded from different groups and levels of Pair ID that only occurred once were excluded (*n*=1105 call pairs with same receiver, 179 with different receivers who had the same type of relationship to the caller). *P*-values are two-tailed.

Extended Data Table 5 | Results for mixed effects logistic regression modeling the probability of a call being correctly classified

Model term	Odds ratio	χ^2 statistic	DF	*P*-value	Std. dev. (random effect)
Spectral acoustic features					
Context	Care/contact = 6.4 Care/greet = 48.9 Contact/greet = 7.6	9.0	2	0.011	--
Caller age class	0.067	6.5	1	0.011	
Position in bout	2.9	3.8	1	0.0498	--
Num. calls for receiver	1.4	7.6	1	0.006	--
Receiver ID	--	--	--	--	3.39
Cepstral acoustic features					
Context	Care/contact = 4.0 Care/greet = 7.7 Contact/greet = 1.9	2.6	2	0.278	--
Caller age class	1.3e-13	0.0	1	1.00	--
Position in bout	2.9	7.0	1	0.008	--
Num. calls for receiver	1.4	16.0	1	0.0000	--
Receiver ID	--	--	--	--	2.06

Odds ratios, χ^2 statistics, degrees of freedom, two-tailed *P*-values, reported for fixed effects. Standard deviations (square root of the variance explained) reported for the random effect. Odds ratios for Context were calculated from the estimated marginal means. χ^2 statistics, degrees of freedom, two-tailed *P*-values were calculated from Type III Analysis of Deviance on the full model. Receivers that only occurred once were excluded. Cepstral features model had warning message indicating convergence issues when Caller age class was included. Context: *n*=138 contact rumbles, 127 greeting rumbles, 62 caregiving rumbles. Caller age class: *n*=274 calls from adults, 53 juvenile calls from juveniles.

Extended Data Table 6 | Results for linear mixed model assessing whether calls addressed to a receiver imitate the receiver's calls

Model term	Param. estimate	Std. error	χ^2 statistic	DF	*P*-value	Partial η^2	Std. dev. (rand. eff.)
Spectral acoustic features							
Imitation Pair Type	-103.7	54.3	3.7	1	0.056	7.7e-4	--
Same Relationship	-43.5	73.2	0.35	1	0.552	-9.5e-4	--
Same Context	-34.9	37.2	0.88	1	0.347	1.4e-4	--
Same Date	17.53	74.4	0.06	1	0.814	-8.3e-4	--
Caller Dyad ID	--	--	--	--	--	--	220.5
Pair ID	--	--	--	--	--	--	367.0
Cepstral acoustic features							
Imitation Pair Type	-71.4	50.7	2.0	1	0.159	4.6e-4	--
Same Relationship	-19.9	67.7	0.09	1	0.768	-1.7e-3	--
Same Context	69.5	38.2	3.32	1	0.069	-7.4e-4	--
Same Date	204.1	77.2	7.0	1	0.008	-1.3e-3	--
Caller Dyad ID	--	--	--	--	--	--	189.6
Pair ID	--	--	--	--	--	--	316.9

Each observation was a pair of calls and the response variable was rank-transformed proximity score. Imitation Pair Type = whether the receiver of one call in a pair was the caller of the other call (reference level = yes); Same Relationship = whether the callers of both calls in a pair had the same type of relationship to their respective receivers (reference level = no); Caller Dyad ID = unique combination of callers (random effect). Same Context, Same Date, and Pair ID same as in Extended Data Table 4. Pairs of calls recorded from different groups, pairs with the same caller or receiver, levels of Caller Dyad ID that only occurred with one level of Imitation Pair Type, and levels of Pair ID that only occurred once were excluded (n=943 call pairs where receiver of one call was the caller of the other, 1553 where this was not the case). *P*-values are two-tailed.

Extended Data Table 7 | Results for linear mixed model assessing whether different callers use similar labels for same receiver

Model term	Param. estimate	Std. error	χ^2 statistic	DF	*P*-value	Partial η^2	Std. dev. (rand. eff.)
Spectral acoustic features							
Different Caller Pair Type	-577.0	176.7	10.7	1	0.001	3.9e-3	--
Same Relationship	245.6	134.3	3.3	1	0.068	9.6e-4	--
Same Context	101.0	68.1	2.2	1	0.138	4.2e-4	--
Same Date	547.5	157.1	12.1	1	0.0005	-3.9e-4	--
Pair ID	--	--	--	--	--	--	1385.0
Cepstral acoustic features							
Different Caller Pair Type	-46.4	165.9	0.08	1	0.780	-3.4e-4	--
Same Relationship	243.2	125.8	3.7	1	0.053	1.3e-3	--
Same Context	233.5	69.1	11.4	1	0.001	1.9e-3	--
Same Date	435.1	160.8	7.3	1	0.007	6.3e-4	--
Pair ID	--	--	--	--	--	--	1219.0

Each observation was a pair of calls and the response variable was rank-transformed proximity score. Different Caller Pair Type = whether the two calls in a pair had different callers and the same receiver (reference level) or different callers and different receivers; Same Relationship, Same Context, Same Date, and Pair ID same as in Extended Data Tables 4 and 6. Pairs of calls recorded from different groups and levels of Pair ID that only occurred once were excluded (n=693 call pairs with same receiver, 7522 with different receivers). *P*-values are two-tailed.

QUERY FORM

Manuscript ID	**[Art. Id: 2420]**
Author	**Michael A. Pardo**

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	*Nature of Query*
Q1:	Please check your article carefully, coordinate with any co-authors and enter all final edits clearly in the eproof, remembering to save frequently. Once corrections are submitted, we cannot routinely make further changes to the article.
Q2:	Note that the eproof should be amended in only one browser window at any one time; otherwise changes will be overwritten.
Q3:	Author surnames have been highlighted. Please check these carefully and adjust if the first name or surname is marked up incorrectly, as this will affect indexing of your article in public repositories such as PubMed. Also, carefully check the spelling and numbering of all author names and affiliations, and the corresponding author(s) email address(es). Please note that email addresses should only be included for designated corresponding authors, and you cannot change corresponding authors at this stage except to correct errors made during typesetting.
Q4:	You cannot alter accepted Supplementary Information files except for critical changes to scientific content. If you do resupply any files, please also provide a brief (but complete) list of changes. If these are not considered scientific changes, any altered Supplementary files will not be used, only the originally accepted version will be published.
Q5:	Please check Figures for accuracy as they have been relabelled. Please markup minor changes in the eProof. For major changes, please provide revised figures. (Please note that in the eProof the figure resolution will appear at lower resolution than in the pdf and html versions of your paper.)
Q6:	If applicable, please ensure that any accession codes and datasets whose DOIs or other identifiers are mentioned in the paper are scheduled for public release as soon as possible, we recommend within a few days of submitting your proof, and update the database record with publication details from this article once available.
Q7:	Please note, we reserve 'significant' and its derivatives for statistical significance. Please reword where this is not the intended meaning (for example to important, notable, substantial); there are 23 instances throughout your text.
Q8:	In Fig. 4 caption, please expand GLM.

QUERY FORM

Manuscript ID	**[Art. Id: 2420]**
Author	**Michael A. Pardo**

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	*Nature of Query*
Q9:	In the sentence beginning 'Our discovery of individual vocal ...', please confirm the expansion of MYA.
Q10:	In the sentence beginning 'For analyses that examined behavioural...', please expand LMMs.
Q11:	In the sentence beginning 'We manually drew a selection box ...', please expand DFT (two instances).
Q12:	In the sentence beginning 'We did not have access to...', please define dBC.
Q13:	For ref. 28, please add editor(s) if appropriate.

nature portfolio

Corresponding author(s): Michael A. Pardo

Last updated by author(s): Apr 2, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used to collect data in this study.
Data analysis	Rough segmentation of calls was performed in Raven Pro 1.5 (Cornell Lab of Ornithology, Ithaca, NY, USA). All other acoustic and statistical analyses were performed in R version 4.1.3. The following R packages were used: AER: testing overdispersion of Poisson GLM car: type III ANOVA caret: data partitioning for machine learning coxme: mixed-effects Cox regression data.table: data wrangling dplyr: data wrangling emmeans: post-hoc comparisons ggplot2: plotting gridExtra: combining plots lme4: mixed effects models lubridate: handling dates in R moments: skewness and kurtosis multitaper: multi-taper spectral estimation (for deriving some acoustic features) patchwork: combining plots party: conditional inference random forest (for variable importance scores) ranger: fast random forest robustbase: calculating robust skewness

rsvd: robust principal components
Rraven: importing Raven Pro selection tables into R
rsvd: robust principal components analysis (for derived acoustic features)
runner: control running operations
scatterplot3d: 3D plotting
seewave: acoustic analysis
soundgen: acoustic analysis
stringr: string manipulation
survival: cox regression
survminer: plotting survival curves
tuneR: acoustic analysis
viridis: more color palettes (for spectrogram)

We did not create any new software or R packages for this study. All of our code is available on Zenodo at this link: doi:10.5281/zenodo.10576772

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data are available on Dryad at the following link: doi:10.5061/dryad.hmgqnk9nj

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.* *Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.* *Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We investigated the hypothesis that elephants address individual members of their family group with name-like calls. We recorded contact and greeting calls from wild African elephants in Samburu & Buffalo Springs National Reserves, northern Kenya and Amboseli National Park, southern Kenya, noting when possible the identity of the caller and the identity of the receiver. We measured a suite of acoustic features on each call (n=469 calls) and used a random forest model to show that calls could be assigned to individual receivers based on acoustic structure with greater than chance accuracy. To determine if elephants rely on imitation of the receiver's calls to address receiver, we examined random forest classification accuracies separately for calls that were more similar to the receiver's calls than typical for that caller (convergent calls, n=95) and calls that were less similar to the receiver's calls than typical for that caller (divergent calls, n=141). We found that calls could be assigned to receiver ID with greater than chance accuracy regardless of whether they were convergent with or divergent from the receiver's calls. We calculated pairwise proximity scores between each call in the dataset and ran an ANOVA which showed that call pairs with the same caller and same receiver were more similar on average than call pairs with the same caller and different receivers who had the same type of relationship with the caller. We ran a logistic regression to assess the factors influencing the probability that the random forest would correctly predict the receiver for a call. We found that the receiver was more likely to be correctly predicted for contact rumbles and caregiving rumbles than for greeting rumbles and more likely to be correctly predicted for adult callers than for juvenile callers. This suggests that contact and caregiving rumbles may be more likely to contain a vocal label than greeting rumbles and adults may be more likely than juveniles to use vocal labels. To determine if elephants imitated the calls of the receiver they were addressing, we ran another ANOVA to test if call pairs in which the receiver of one call produced the other call had higher proximity scores than call pairs in which this was not the case. There was no significant difference, indicating no evidence for imiation. To determine if different callers use the same label to address a given receiver (i.e., if calls could be assigned to receiver ID independent of caller ID), we ran a second random forest with the training and test sets partitioned so the model was trained and tested on different callers. This random forest failed to assign calls to receiver ID any better than chance, suggesting that different callers do not use the same label for the same receiver. However, an ANOVA showed that call pairs with different callers and the same receiver were more similar (had higher proximity scores) on average than call pairs with different callers and different receivers, suggesting that different callers do use similar labels for the same receiver. Finally, we conducted a playback experiment to determine if elephants perceive and respond to putative labels in their calls. We played 17 elephants a recording of a call that was originally addressed to them (test) and a recording of a call from the same caller that was originally addressed to someone else (control). One subject received two different sets of test and control playbacks, one subject received just 1 test playback (no control) and one subject received just one control playback (no test). All other subjects received exactly one test playback and one control playback each. Subjects approached the speaker more quickly, vocalized more quickly, and produced more vocalizations in response to test playbacks than controls, further supporting the hypothesis that calls are specific to individual receivers.
Research sample	Subjects were wild African savannah elephants (Loxodonta africana) from two Kenyan populations: Samburu & Buffalo Springs (northern Kenya) and Amboseli (southern Kenya). Acoustic analyses were conducted on 371 rumbles from 52 adult females, 16 juvenile females, 2 females recorded as both juveniles and adults (cutoff for adulthood was 10 years of age), and 14 juvenile males in Samburu, as well as 98 rumbles from 13 adult females, 3 juvenile females, and 1 juvenile male in Amboseli. Playbacks were conducted to 17 individuals in Samburu (15 adult females, 1 adolescent female, and 1 adolescent male).
Sampling strategy	Calls were recorded using all-occurrence sampling. There was no predetermined sample size as we attempted to record as many calls as possible. Subjects for playbacks were chosen based on which individuals we were able to record a test stimulus and control stimulus for. We did not predetermine the sample size for playbacks and instead did as many playbacks as we were able to given what recordings were available.
Data collection	Calls were recorded during daylight hours from a vehicle using a handheld Earthworks microphone. Callers and receivers were identified using behavioral cues, and elephants were identified individually using naturally-occurring marks on the ears and other distinct physical features. Playbacks were conducted from 50 meters away from a loudspeaker placed on the ground or in a Landcruiser with all the doors and windows open. Data in Samburu (recordings and playbacks) were collected by MP and DL. Data in Amboseli (recordings) were collected by JP and PG.
Timing and spatial scale	Calls were recorded in Samburu in Nov 2019-Mar 2020 and Jun 2021-Apr 2022. Calls were recorded in Amboseli in 1986-1990 and 1997-2006. Playbacks were conducted from Oct 2021 to Apr 2022. Playbacks to the same subject were spaced apart by at least 7 days which previous studies on elephants have suggested as a rule of thumb to minimize the risk of habituation. Samburu and Buffalo Springs National Reserves cover an area of about 296 km2 and Amboseli covers an area of about 392 km2.
Data exclusions	We only analyzed rumbles that were produced in the contexts of contact calling , greeting, and caregiving. We also only included calls with minimal overlapping sounds, a high enough signal-to-noise ratio for the first two formants to be clearly visible in the spectrogram. Finally, we only included calls where the identity of the receiver was known for certain and for which there was only one receiver. For analyses involving caller ID or behavioral context, we also made sure that the identity of the caller/behavioral context was known for certain.
Reproducibility	Due to the logistical constraints of conducting this type of experiment in the field and the time constraints of available funding, we did not attempt to replicate the experiment.
Randomization	For the playback experiment, we attempted to conduct both a test playback and control playback to each individual (within-subjects design), only failing to do so for 2/17 subjects. The order of presentation of test and control playbacks was balanced across subjects.

	Subjects were randomly assigned to receive the test or control playback first, with the constraint that 50% of subjects should receive the test first and 50% should receiver the control first.
Blinding	The experimenters were blind to the condition of each playback trial until after all playback trials had been conducted and all videos of those trials were scored. The same observer (MP) conducted the playback trials and scored the videos.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	The habitat of both field sites is a mixture of open grassland, bushy shrubs, and patches of woodland and permanent swamp. Both sites are semi-arid, receiving an average of about 350 mm of rain per year with peaks in November and April. Fieldwork was conducted in both wet and dry seasons. Average annual temperature is about 21.6 degrees Celsius in Amboseli and 24.5 degrees Celsius in Samburu.
Location	Samburu & Buffalo Springs: (0.61 N, 37.5 E), 800-1230 m above sea level Amboseli National Park: (2.7 S, 37.3 E), 1100-1200 m above sea level.
Access & import/export	Permits were obtained from the Wildlife Research & Training Institute (WRTI) of Kenya and the National Commission for Science, Technology, and Innovation (NACOSTI) of Kenya, in consultation with local county governments (Samburu, Isiolo, and Kajiado counties). Permit numbers: NACOSTI/P/19/2735, WRTI-0061-06-21, NACOSTI/P/21/14091.
Disturbance	Elephants were not physically handled as part of this study. They may have been temporarily and slightly disturbed by playback stimuli. To minimize potential disturbance, we only played back a single call in any given trial and waited a minimum of 7 days between playbacks to the same subjects. Subjects did not always exhibit any response to playbacks, and when they did, they returned to baseline behavior in <10 min. The elephants in Samburu and Amboseli are habituated to research vehicles so it is unlikely that they were disturbed in any substantial way by our presence. To avoid damage to vegetation, we only drove off road when absolutely necessary to access the elephants and returned to an existing road as soon as possible.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	This study did not involve laboratory animals.
Wild animals	This study involved wild African savannah elephants (Loxodonta africana). No elephants were captured or handled as part of this study. We used audio recordings from 65 adult females, 19 juvenile females, and 15 juvenile males, as well as 2 females who were considered juveniles (<10 yo) in earlier recordings and adults (>10 yo) in later recordings. Playbacks were conducted to 17 individuals in Samburu (15 adult females, 1 adolescent female, and 1 adolescent male).
Reporting on sex	We focused on female-calf groups for this study because females and calves are much more vocal than adult males in elephants. As most of the elephants (and all the adults) in our study were female, these results may only be applicable to females. We did not conduct a sex-based analysis because we did not have sufficient data from males to consider them separately from females.
Field-collected samples	This study did not involve samples collected from the field (only audio and video recordings)
Ethics oversight	This study was approved by the Institutional Animal Care and Use Committee of Colorado State University (protocol #19-9229A)

Note that full information on the approval of the study protocol must also be provided in the manuscript.