

1           **Title: African elephants address one another with individually specific calls**

2   **Authors:** Michael A. Pardo<sup>1\*</sup>, Kurt Fristrup<sup>2</sup>, David S. Lolchuragi<sup>3</sup>, Joyce Poole<sup>4</sup>, Petter Granli<sup>4</sup>,

3                           Cynthia Moss<sup>5</sup>, Iain Douglas-Hamilton<sup>3</sup>, George Wittemyer<sup>1,3</sup>

4   **Affiliations:**

5   <sup>1</sup>Department of Fish, Wildlife, and Conservation Biology, Colorado State University; Fort  
6   Collins, CO, USA

7   <sup>2</sup>Department of Electronic and Computer Engineering, Colorado State University; Fort Collins,  
8   CO, USA

9   <sup>3</sup>Save The Elephants; Nairobi, Kenya

10   <sup>4</sup>ElephantVoices; Sandefjord, Norway

11   <sup>5</sup>Amboseli Elephant Research Project; Nairobi, Kenya

12   \*Corresponding author. Email: map385@cornell.edu

13   **SUMMARY**

14   Personal names are a universal feature of human language, yet few analogs exist in other species.  
15   While dolphins and parrots address conspecifics by imitating the calls of the addressee<sup>1,2</sup>, human  
16   names are not imitations of the sounds typically made by the name's owner<sup>3</sup>. Labeling objects or  
17   individuals without relying on imitation of the sounds made by that object or individual is key to  
18   the expressive power of language. Thus, if non-imitative name analogs were found in other  
19   species, this could have important implications for our understanding of language evolution.  
20   Here, we show that wild African elephants address one another with individually specific calls  
21   without any evidence of imitating the receiver's vocalizations. A random forest model correctly  
22   predicted receiver identity from call structure better than expected by chance, regardless of  
23   whether the calls were more or less similar to the receiver's calls than typical for that caller.

24 Moreover, elephants differentially responded to playbacks of calls originally addressed to them  
25 relative to calls addressed to a different individual, indicating that they can determine from a  
26 call's structure if it was addressed to them. Our findings offer the first evidence for a non-human  
27 species individually addressing conspecifics without imitating the receiver.

## 28 **MAIN TEXT**

29 One of the hallmarks of spoken human language is the use of vocal labels, in which a  
30 learned sound refers to an object or individual (the “referent”) <sup>4</sup>. Many species produce  
31 functionally referential calls for food and predators <sup>5,6</sup>, but the production of these calls is  
32 typically innate <sup>7</sup>. Learned vocal labels allow for much more flexible communication than innate  
33 calls by making it possible to develop new labels for new referents. Thus, they are central to  
34 humans' ability to articulate symbolic thought and coordinate unusually sophisticated levels of  
35 cooperation <sup>8</sup>. However, few examples of learned vocal labeling are known in other species.  
36 Personal names are a type of vocal label that refer to another individual. Names must involve  
37 vocal learning, as an individual cannot be born knowing the names for all its future social  
38 affiliates. Thus, potential nonhuman analogs of personal names are highly relevant to  
39 understanding the evolution of language, and by extension, complex cognition and social  
40 behavior.

41 Most human words, including personal names, are arbitrary in structure; that is, they are  
42 not imitations of sounds typically made by the referent or tied to the physical properties of the  
43 referent <sup>3</sup>. Arbitrariness is crucial to language because it enables communication about objects  
44 and ideas that do not make any imitable sound. However, clear evidence for arbitrary analogs of  
45 names in other species is lacking. Bottlenose dolphins (*Tursiops truncatus*) and some parrots  
46 (Psittacidae) address individual conspecifics by imitating the receiver's “signature” call, a sound

47 that is most commonly produced by the receiver to signal individual identity <sup>1,2,9</sup>. When  
48 functioning as self-identification signals, these signature calls are indeed arbitrary <sup>10</sup>. However,  
49 when other individuals copy a conspecific's signature call to address them, it may be argued that  
50 the copied signature call is an iconic (non-arbitrary) label, since it is an imitation of a sound  
51 typically produced by the individual to whom the call refers. Non-imitative learned vocal  
52 labeling could allow communication about a wider range of referents than imitative labeling, but  
53 it may be more cognitively demanding, as it requires individuals to make an abstract connection  
54 between a sound and referent. Thus, if any non-human species were found to address individual  
55 conspecifics using labels that are not imitative of the receiver's own calls, this would indicate a  
56 novel and perhaps uniquely complex form of communication with important implications for our  
57 understanding of language evolution and cognition.

58       Elephants are among the few mammals capable of mimicking novel sounds, although the  
59 function of this vocal learning ability is unknown <sup>11,12</sup>. The most common call type produced by  
60 elephants is the rumble, a harmonically rich, low-frequency sound which is individually distinct  
61 <sup>13,14</sup>, distinguishable, <sup>15</sup> and produced across most behavioral contexts <sup>16</sup>. Contact rumbles  
62 (Supplementary Audio File S1) are long-distance calls produced when the caller is visually  
63 separated from one or more social affiliates and attempting to reinitiate contact, and greeting  
64 rumbles (Supplementary Audio File S2) are close-distance calls produced when one individual  
65 approaches another after a period of separation <sup>16</sup>.

66       We analyzed contact and greeting rumbles from female-offspring groups of wild African  
67 savannah elephants to assess whether they contain individual vocal labels. We only used calls for  
68 which we were able to identify the caller and apparent intended receiver (527 calls from the  
69 greater Samburu ecosystem, northern Kenya, 98 from Amboseli National Park, southern Kenya).

70 Receivers were identified as the individual who responded to the call by vocalizing or  
71 approaching the caller, the only adult member of the family group separated (>50m) from the  
72 caller when the caller produced a contact call, or the individual who approached/was approached  
73 by the caller when the caller produced a greeting call. We were able to determine which  
74 individuals were separated from the group at a given time by knowing the composition of each  
75 family group and by following the elephants for several hours each day and observing short-term  
76 fission and fusion events where some individuals split off from, lagged behind, and/or rejoined  
77 the rest of the group. Calls for which the receiver could not be identified or that appeared to be  
78 directed to multiple receivers (e.g., caller produced a contact call while separated from the whole  
79 family group) were excluded from analysis. We investigated (1) if elephants address conspecifics  
80 using receiver-specific vocal labels, (2) if the labels are imitative of the receiver's calls or  
81 arbitrary, and (3) if different callers share the same label for the same receiver (Extended Data  
82 Table 1).

83 Our dataset consisted of 114 unique callers and 119 unique receivers, with 1-46  
84 (median=2) calls per caller, 1-48 (median=2) calls per receiver, 1-9 (median=2) receivers per  
85 caller, and 1-10 (median=2) callers per receiver (Extended Data Fig. 1). For 597 of 625 calls, the  
86 caller and receiver belonged to the same family group. We measured two sets of acoustic  
87 features for each call (spectral and cepstral, see Supplementary Information; Extended Data Fig.  
88 2) and ran all statistical models separately for each set of features. Results reported in the text  
89 and figures are for the spectral features only (see tables for results with cepstral features, which  
90 were similar).

91 **Calls were specific to individual receivers**

92           We ran a random forest <sup>17</sup> with 6-fold cross-validation to predict the receiver of each of  
93 the 625 rumbles as a function of the acoustic features and compared the classification accuracy  
94 to a null distribution generated from 10,000 iterations of the same model with the acoustic  
95 features randomly permuted. We expected vocal labeling to only occur in contextually relevant  
96 calls, as humans and dolphins only use names or copied signature whistles in a minority of  
97 utterances <sup>18</sup>. However, we used all 625 rumbles for analysis as there was no way to determine *a*  
98 *priori* which calls (or what proportion of calls) might contain a vocal label. Call structure varied  
99 clearly with the identity of the targeted receiver (Extended Data Fig. 3) as would be expected if  
100 elephants use vocal labels for other individuals. Our model correctly identified the receiver for  
101 20.3% of calls analyzed, a significantly greater proportion than that of null models (permutation  
102 test, null models mean accuracy =  $7.6 \pm 0.75\%$  correct,  $P < 0.0001$ ) (Fig. 1, Table 1), indicating  
103 receivers of calls could be correctly identified from call structure statistically significantly better  
104 than chance.

105           To determine if this could be an artifact of the correlation between caller ID and receiver  
106 ID in our dataset, we controlled for caller ID by comparing the mean similarity of pairs of calls  
107 with the same caller and receiver to the mean similarity of pairs of calls with the same caller and  
108 different receivers, using proximity scores derived from the random forest as a metric of call  
109 similarity <sup>19</sup>. To control for the possibility that calls were specific to the type of relationship  
110 between the caller and receiver rather than to the individual receiver per se, we categorized social  
111 relationship based on relatedness and age (a proxy for dominance) (Extended Data Table 3), and  
112 only considered pairs of calls with the same type of relationship between caller and receiver.  
113 Calls with the same caller and same receiver were significantly more similar than calls with the  
114 same caller and different receivers, even after controlling for social relationship, behavioral

115 context, and recording date, further supporting the hypothesis that rumbles are specific to  
116 individual receivers (ANOVA,  $F_1=94.61$ ,  $P<0.0001$ , Cohen's  $D=0.412$ ) (Fig. 1, Extended Data  
117 Table 4). As calls in our dataset were predominantly between individuals in the same family  
118 group, our results only provide evidence for vocal labeling within family groups.

### 119 **Vocal labelling likely does not rely on imitation of receiver**

120 If calls are imitative of the receiver's calls, then callers should sound more like a given  
121 receiver when addressing her than when addressing other individuals. Pairs of calls in which the  
122 receiver of one call was the caller of the other call were slightly but significantly more similar on  
123 average than pairs in which this was not the case, suggesting possible imitation of the receiver's  
124 calls (ANOVA,  $F_1=11.70$ ,  $P=0.0006$ , Cohen's  $D=0.0037$ ) (Extended Data Table 5). However,  
125 given the exceedingly small effect size (0.78% of SD) and large sample size of call pairs  
126 ( $n=11,309$ ), this significant difference may not be biologically meaningful. Moreover, among the  
127 calls for which we had recordings of the receiver and recordings of the caller addressing other  
128 individuals ( $n=494$ ), 60.5% were divergent from the receiver's calls; that is, less similar to the  
129 receiver's calls than typical for that caller (see Supplementary Information). The classificatory  
130 model performed significantly better than the null model for both convergent and divergent calls  
131 (convergent calls: 17.2% correct, null models mean accuracy =  $4.9 \pm 1.1\%$ ,  $P<0.0001$ ; divergent  
132 calls: 32.4% correct, null models mean accuracy =  $13.1 \pm 1.4\%$ ,  $P<0.0001$ ) (Fig. 2, Table 1).  
133 Finally, among both convergent and divergent calls, calls with the same caller and same receiver  
134 were more similar than calls with the same caller and different receivers (ANOVA; convergent  
135 calls:  $F_1=15.30$ ,  $P=0.0001$ , Cohen's  $D=0.411$ ; divergent calls:  $F_1=8.67$ ,  $P=0.0033$ , Cohen's  
136  $D=0.262$ ) (Fig. 2, Extended Data Table 4). This suggests that vocal labeling in elephants likely  
137 does not rely on imitation of the receiver's calls. While we cannot rule out the possibility that

138 elephants imitated calls made by the receiver that were not included in our dataset, elephants are  
139 not known to produce discrete “signature” calls like dolphins and parrots; instead, the caller-  
140 specificity of elephant rumbles is likely a product of voice characteristics that are present across  
141 calls<sup>13,14</sup>.

### 142 **Mixed evidence for convergence among callers addressing same receiver**

143 In humans and bottlenose dolphins, different callers generally use the same label for a  
144 given receiver. To determine if different callers use similar labels to address the same receiver in  
145 elephants, we ran a random forest structured to predict receiver ID from different callers than the  
146 model was trained on. This model correctly classified 1.4% of calls, no better than the  
147 corresponding null models (permutation test, mean accuracy of null models= $1.4 \pm 0.40\%$  correct,  
148  $P=0.453$ ) (Fig. 3, Table 1). However, calls from different callers to the same receiver were  
149 significantly more similar on average than calls from different callers to different receivers  
150 (ANOVA,  $P<0.0001$ , Cohen’s  $D=0.134$ ) (Fig. 3, Extended Data Table 6). These mixed results  
151 may be due to the fact that rumbles simultaneously encode multiple messages<sup>13,16,20,21</sup>. If vocal  
152 labels account for only a small portion of the variation in rumbles, the random forest may have  
153 been influenced by context or caller-specific features, thus reducing its ability to predict receiver  
154 ID across callers, even if different callers address the same receiver with the same label. Further  
155 work to identify how vocal labels are encoded in elephant calls will be necessary to definitively  
156 determine if different callers use the same label for the same receiver.

### 157 **Elephants responded more strongly to playback of calls originally addressed to them**

158 To determine if elephants perceive and respond to the vocal labels in calls addressed to  
159 them, we compared reactions of 17 wild elephants to playback of a call that was originally  
160 addressed to them (test) relative to playback of a call from the same caller that was originally

161 addressed to a different individual (control). By using test and control stimuli from the same  
162 caller, we controlled for the possibility of the caller's relationship to the subject influencing the  
163 results. To control for the possibility that calls are specific to the type of relationship between the  
164 caller and receiver rather than to the individual receiver per se, we included the type of  
165 relationship between the caller and the original receiver of the call as a factor in the analysis.  
166 Further supporting the existence of vocal labels, subjects approached the speaker more quickly  
167 (Cox regression,  $\chi^2=6.8$ ,  $P=0.009$ ) and vocalized more quickly (Cox regression,  $\chi^2=7.9$ ,  
168  $P=0.005$ ) in response to test playbacks than control playbacks (Fig. 4, Table 2). They also  
169 produced more vocalizations in response to test playbacks, although this model failed to  
170 converge (Poisson regression,  $\chi^2=6.2$ ,  $P=0.013$ ) (Fig. 4, Table 2). There was no significant  
171 difference between test and control trials in latency to vigilance (Cox regression,  $\chi^2=3.1$ ,  
172  $P=0.08$ ) or in the change in vigilance duration before and after the playback (linear regression,  
173  $\chi^2=0.06$ ,  $P=0.81$ ), although there was a nonsignificant trend toward faster onset of vigilance in  
174 test trials (Table 2).

## 175 **Discussion**

176 To our knowledge, this study presents the first evidence for vocal addressing of  
177 conspecifics without imitation of the receiver's calls in nonhuman animals. Very few species are  
178 known to address conspecifics with vocal labels of any kind. Where evidence for vocal labels has  
179 been found, they are either clearly imitative<sup>1,2,9</sup> or of unknown structure<sup>22,23</sup>. Our data suggest  
180 that elephants label conspecifics without relying on imitation of the receiver's calls, a  
181 phenomenon previously known to occur only in human language.

182 The social behavior and ecology of elephants create an environment in which individual  
183 vocal labeling may be particularly advantageous. Due to their fission-fusion social dynamics,



184 elephants are often out of sight of their closely bonded social partners and produce contact  
185 rumbles to communicate over long distances <sup>16,24</sup>. Characteristic fission-fusion dynamics in  
186 elephants include coordinated movement to and from resources while proximately diffusing to  
187 avoid foraging competition <sup>25,26</sup>. Vocal labels could enhance coordinating ability when out of  
188 sight of one another. In contact calling scenarios, vocal labeling could allow callers to attract the  
189 attention of a specific intended receiver. While greeting rumbles are produced in close proximity  
190 when the caller and receiver typically have visual contact <sup>16</sup>, vocal labeling in greeting rumbles  
191 could possibly strengthen social bonds with specific individuals. Humans experience a positive  
192 affective response and increased willingness to comply with requests when someone remembers  
193 their name <sup>27</sup>.

194         Nonetheless, the fact that our random forest model correctly predicted receiver ID for  
195 only around a fifth of calls (albeit significantly better than random) suggests that vocal labels  
196 only occur in a minority of rumbles and thus are likely not necessary in all or even most  
197 contexts. For example, contact and greeting calls may occur in vocal sequences where labeling  
198 the receiver in each call would be redundant <sup>16</sup>, and in the dry season, when elephant families  
199 fission into smaller groups, there may be less ambiguity about the intended receiver in many  
200 scenarios <sup>26</sup>. Indeed, both humans and bottlenose dolphins only use individual vocal labels (i.e.,  
201 names or imitated signature whistles) in a small percentage of utterances <sup>18</sup>.

202         When vocal labels do occur, they are likely only one component among many in the call.  
203 Rumbles are recognized to simultaneously encode multiple messages, including but not limited  
204 to caller identity, age, sex, emotional state, and behavioral context <sup>13,16,20,21</sup>. Moreover, the top  
205 acoustic features for predicting receiver ID were not those that explained the most variation in  
206 the calls (see Supplementary Information), suggesting that vocal labels account for only a small

207 fraction of the total variation in rumbles. Rather than comprising a whole stand-alone call,  
208 elephant vocal labels may be embedded within a call that simultaneously conveys multiple  
209 additional messages. The richness in the information content of elephant vocalizations makes it  
210 difficult to identify the specific acoustic parameters that encode receiver ID. Unlike dolphin  
211 signature whistles<sup>18</sup>, elephant vocal labels cannot be discerned by visual inspection of the  
212 spectrogram and are likely encoded by a complex and subtle interaction among many acoustic  
213 parameters. As a result, we employed machine learning in this analysis, but innovative  
214 approaches in signal processing may be necessary to isolate the vocal labels within rumbles.

215 Both African and Asian elephants have a demonstrated capacity for vocal mimicry in  
216 captivity, but no prior study has documented a function of this ability in the wild<sup>11,12</sup>. We  
217 speculate that vocal labeling may be one, if not the primary, function of vocal production  
218 learning in wild elephants. Dolphins and parrots, which show evidence for individual vocal  
219 labeling via imitation of the receiver, are adept vocal learners. Another vocal learner, the  
220 Egyptian fruit bat (*Rousettus aegyptiacus*), produces calls that are specific to individual receivers  
221 and may be vocal labels as well, although it is currently unknown if the bats perceive this  
222 information<sup>23</sup>. Taken together, this raises the possibility that social selection pressures creating a  
223 need to address individual conspecifics may have led to multiple independent origins of vocal  
224 production learning.

225 The use of learned arbitrary labels is part of what gives human language its uniquely  
226 broad range of expression<sup>3</sup>. Our results suggesting that wild elephants also use arbitrary vocal  
227 labels for individual conspecifics provide an opportunity to investigate the selection pressures  
228 that may have led to the evolution of this rare ability in two divergent lineages. Moreover, these

229 findings raise intriguing questions about the complexity of elephant social cognition, considering  
230 the potential relevance of symbolic communication to their social decision making.

## 231 REFERENCES

- 232 1. King, S. L. & Janik, V. M. Bottlenose dolphins can use learned vocal labels to address  
233 each other. *Proc. Natl. Acad. Sci. USA* **110**, 13216–13221 (2013).
- 234 2. Balsby, T. J. S., Momberg, J. V. & Dabelsteen, T. Vocal imitation in parrots allows  
235 addressing of specific individuals in a dynamic communication network. *PLoS One* **7**,  
236 e49747 (2012).
- 237 3. Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H. & Monaghan, P.  
238 Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences* vol.  
239 19 603–615 (2015).
- 240 4. Fitch, W. T. The evolution of language: a comparative review. *Biol. Philos.* **20**, 193–230  
241 (2005).
- 242 5. Macedonia, J. M. & Evans, C. S. Variation among mammalian alarm call systems and the  
243 problem of meaning in animal signals. *Ethology* **93**, 177–197 (1993).
- 244 6. Clay, Z., Smith, C. L. & Blumstein, D. T. Food-associated vocalizations in mammals and  
245 birds: What do these calls really mean? *Anim. Behav.* **83**, 323–330 (2012).
- 246 7. Wheeler, B. C. & Fischer, J. Functionally referential signals: A promising paradigm  
247 whose time has passed. *Evol. Anthropol.* **21**, 195–205 (2012).
- 248 8. Smith, E. A. Communication and collective action: Language and the evolution of human  
249 cooperation. *Evol. Hum. Behav.* **31**, 231–245 (2010).
- 250 9. Dahlin, C. R., Young, A. M., Cordier, B., Mundry, R. & Wright, T. F. A test of multiple  
251 hypotheses for the function of call sharing in female budgerigars, *Melopsittacus*

- 252           *undulatus*. *Behav. Ecol. Sociobiol.* **68**, 145–161 (2014).
- 253    10.    Janik, V. M. & Sayigh, L. S. Communication in bottlenose dolphins: 50 years of signature  
254           whistle research. *J. Comp. Physiol. A Neuroethol. Sensory, Neural, Behav. Physiol.* **199**,  
255           479–489 (2013).
- 256    11.    Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S. & Watwood, S. Elephants are capable  
257           of vocal learning. *Nature* **434**, 455–456 (2005).
- 258    12.    Stoeger, A. S. *et al.* An Asian elephant imitates human speech. *Curr. Biol.* **22**, 2144–2148  
259           (2012).
- 260    13.    Soltis, J., Leong, K. & Savage, A. African elephant vocal communication II: Rumble  
261           variation reflects the individual identity and emotional state of callers. *Anim. Behav.* **70**,  
262           589–599 (2005).
- 263    14.    Clemins, P. J., Johnson, M. T., Leong, K. M. & Savage, A. Automatic classification and  
264           speaker identification of African elephant (*Loxodonta africana*) vocalizations. *J. Acoust.*  
265           *Soc. Am.* **117**, 956 (2005).
- 266    15.    McComb, K., Moss, C., Sayialel, S. & Baker, L. Unusually extensive networks of vocal  
267           recognition in African elephants. *Anim. Behav.* **59**, 1103–1109 (2000).
- 268    16.    Poole, J. H. Behavioral contexts of elephant acoustic communication. in *The Amboseli*  
269           *elephants: a long-term perspective on a long-lived mammal* (eds. Moss, C. J., Croze, H. &  
270           Lee, P. C.) 125–159 (University of Chicago Press, 2011).
- 271    17.    Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
- 272    18.    King, S. L., Sayigh, L. S., Wells, R. S., Fellner, W. & Janik, V. M. Vocal copying of  
273           individually distinctive signature whistles in bottlenose dolphins. *Proc. R. Soc. B* **280**,  
274           20130053 (2013).

- 275 19. Rhodes, J. S., Cutler, A. & Moon, K. R. Geometry- and accuracy-preserving random  
276 forest proximities. *arXiv* 1–18 (2022) doi:arXiv:2201.12682.
- 277 20. Baotic, A. & Stoeger, A. S. Sexual dimorphism in African elephant social rumbles. *PLoS*  
278 *One* **12**, e0177411 (2017).
- 279 21. Stoeger, A. S., Zeppelzauer, M. & Baotic, A. Age-group estimation in free-ranging  
280 African elephants based on acoustic cues of low-frequency rumbles. *Bioacoustics* **23**,  
281 231–246 (2014).
- 282 22. Wanker, R., Sugama, Y. & Prinage, S. Vocal labelling of family members in spectacled  
283 parrotlets, *Forpus conspicillatus*. *Anim. Behav.* **70**, 111–118 (2005).
- 284 23. Prat, Y., Taub, M. & Yovel, Y. Everyday bat vocalizations contain information about  
285 emitter, addressee, context, and behavior. *Sci. Rep.* **6**, 39419 (2016).
- 286 24. Garstang, M. Long-distance, low-frequency elephant communication. *J. Comp. Physiol. A*  
287 **190**, 791–805 (2004).
- 288 25. Charif, R. A. *et al.* Spatial relationships and matrilineal kinship in African savanna  
289 elephant (*Loxodonta africana*) clans. *Behav. Ecol. Sociobiol.* **57**, 327–338 (2005).
- 290 26. Wittemyer, G., Douglas-Hamilton, I. & Getz, W. M. The socioecology of elephants:  
291 analysis of the processes creating multitiered social structures. *Anim. Behav.* **69**, 1357–  
292 1371 (2005).
- 293 27. Howard, D. J., Gengler, C. & Jain, A. What’s in a name? A complimentary means of  
294 persuasion. *J. Consum. Res.* **22**, 200–211 (1995).

## 295 **METHODS**

### 296 **Field recording**

297 We collected audio recordings of wild female-calf groups in Amboseli National Park,  
298 Kenya in 1986-1990 and 1997-2006 and Samburu and Buffalo Springs National Reserves  
299 (hereafter, Samburu), Kenya in Nov 2019-Mar 2020 and Jun 2021-Apr 2022. Both populations  
300 have been continuously monitored for decades and all individuals can be individually identified  
301 by external ear morphology<sup>26,28</sup>. We recorded calls from a vehicle during daylight hours with  
302 all-occurrence sampling<sup>29</sup> using an Earthworks QTC1 microphone (4 Hz-40 kHz  $\pm$  1 dB) with a  
303 Nagra IV-SJ reel-to-reel tape recorder or an HHB PDR 1000 DAT recorder in Amboseli, and an  
304 Earthworks QTC40 microphone (3 Hz-40kHz  $\pm$  1 dB) with a Sound Devices MixPre3 or  
305 MixPre3-II digital recorder in Samburu. Recordings were recorded at a 48 kHz sampling rate  
306 with 16 bits of amplitude resolution and stored at 2 kHz in Amboseli and recorded and stored at  
307 44.1 kHz with 24 or 32 bits of amplitude resolution in Samburu.

308 When possible, we recorded for each call the identity of the caller, the behavioral context,  
309 and the identity of the receiver (criteria for identifying receiver defined in Main Text). The caller  
310 was identified using behavioral and contextual cues, such as an open mouth, flapping ears, or  
311 being the only individual of the right age class in the immediate vicinity<sup>16</sup>. We scored behavioral  
312 context according to a published methodology<sup>16</sup>. For each call, we recorded the certainty with  
313 which we knew the caller ID, behavioral context, and receiver ID as a number between 0 and 1  
314 (see Supplementary Information). In all statistical analyses, we weighted each call by the  
315 certainty of receiver ID, so calls with greater certainty about the identity of the receiver would  
316 have a proportionally greater impact on the model.

### 317 **Acoustic analysis**

318 We only included in analysis contact and greeting rumbles with certainty of caller ID,  
319 receiver ID, and behavioral context greater than 0, with no significant overlap with other calls or

320 other loud sounds in the same frequency range, and that were recorded close enough to the  
321 microphone for the first two formants to be clearly visible in the spectrogram (98 rumbles from  
322 Amboseli, 527 from Samburu). We performed all acoustic and statistical analyses in R version  
323 4.1.3<sup>30</sup>. We automatically detected the onset and offset of each call from the amplitude envelope  
324 using the function `segment()` in the package `soundgen`<sup>31</sup>, manually adjusting the detected times  
325 when necessary. We then measured two alternative sets of features: spectral and cepstral (see  
326 Supplementary Information). The spectral features consisted of the smoothed Hilbert amplitude  
327 envelope (350 ms moving average window, 90% overlap), the vectors of energy values in 26  
328 mel-frequency bands between 0-500 Hz (measured at 35 ms intervals), and the vectors of delta  
329 and delta-delta coefficients for the mel-frequency bands (79 vectors total) (Extended Data Fig.  
330 2). The cepstral features consisted of the amplitude envelope, the vectors of the first 12 mel-  
331 frequency cepstral coefficients measured at 35 ms intervals, and the vectors of delta and delta-  
332 delta coefficients for the cepstral coefficients.

333         As the raw acoustic vectors (mel spectral bands, MFCCs, and their delta and delta-delta  
334 values) represented a matrix of values for each call, it was necessary to calculate lower-  
335 dimensional derived features from these matrices as input variables for statistical models. We  
336 calculated derived features separately for the spectral and cepstral features. In brief, we scaled  
337 the acoustic vectors and decorrelated them with a robust principal components analysis using the  
338 `rpca` package in R, which decomposes the data into a robust matrix and a sparse matrix  
339 containing the outlier values ( $\lambda=0.00996$ )<sup>32</sup>. The final derived features we calculated were the  
340 median, robust skewness, minimum extent, and equivalent statistical extent of the sparse matrix,  
341 the means of the first  $n$  low-rank principal components required to explain 99.9% of the variation  
342 (74 for spectral features, 12 for cepstral features), and 8 measures of the spectral properties of the

343 low-rank principal components, calculated by treating each principal component as if it were a  
344 waveform (see Supplementary Information) (Extended Data Table 2).

### 345 **Statistical analysis of acoustic data**

346 *Are calls specific to individual receivers (hypothesis 1)?*

347 We ran a 6-fold cross-validated random forest model in the R package ranger<sup>33</sup> to predict  
348 the identity of the receiver of each call (receiver ID) as a function of the acoustic features. We  
349 stratified the cross-validation folds by caller ID and receiver ID to ensure as even a distribution  
350 as possible of all caller-receiver dyads across all folds. Thus, if calls contain acoustic cues to  
351 receiver ID, this model was expected to predict receiver ID better than chance regardless of  
352 whether the label for a given receiver is shared across callers (Extended Data Table 1, hypothesis  
353 1, prediction 1). The model used 625 observations, 500 trees, 6 variables per node, 60% of  
354 observations per tree, a minimum node size of 1, and no maximum tree depth, and observations  
355 were weighted by certainty of receiver ID. To increase the stability of the model's classification  
356 accuracy, we ran the model 2000 times and used the mean classification accuracy across the  
357 2000 runs. To determine if the model predicted receiver ID better than expected by chance, we  
358 ran the model 10,000 times with the acoustic features randomly permuted and compared the  
359 classification accuracy of the original model (averaged across 2000 runs) to the null distribution  
360 of classification accuracies generated by the 10,000 models with randomized acoustic features.

361 As caller ID and receiver ID were partially aliased in our dataset (Extended Data Fig. 1),  
362 the random forest could theoretically use acoustic cues to caller ID<sup>16</sup> to predict receiver ID, even  
363 if the calls did not contain any vocal label identifying the intended receiver. To disentangle the  
364 effects of caller ID and receiver ID on call structure, we compared the mean pairwise similarities  
365 between pairs of calls with the same caller and receiver and pairs with the same caller and



366 different receivers (Same Caller Pair Type). As a metric of call similarity, we extracted a  
367 proximity score for each pairwise combination of calls from a random forest trained to predict  
368 receiver ID as a function of the acoustic features on the full dataset (625 training observations,  
369 8000 trees, other hyperparameters and weighting same as above). The proximity score for a  
370 given pair of calls was the proportion of trees in which both calls were classified in the same  
371 terminal node, corrected for the size of each node, and represented the degree of similarity  
372 between the two calls in terms of the acoustic features most relevant to predicting receiver ID<sup>19</sup>.  
373 If calls are specific to individual receivers within a given caller, then pairs of calls with the same  
374 caller and same receiver should be more similar (have higher proximity scores) than pairs of  
375 calls with the same caller and different receivers (Extended Data Table 1, hypothesis 1,  
376 prediction 2).

377 Previous work has shown that elephants vary the structure of their rumbles when  
378 interacting with more dominant vs. more subordinate conspecifics<sup>13</sup>. To rule out the possibility  
379 that calls were specific to the type of relationship between caller and receiver rather than to  
380 individual receivers *per se*, we restricted the analysis of Same Caller Pair Type to pairs of calls  
381 that had the same type of relationship between caller and receiver. We defined caller-receiver  
382 relationship using 12 categories based on sex, family group membership, relative age, and  
383 mother-offspring relationship, reflecting the fact that dominance in elephants is primarily  
384 determined by age<sup>34,35</sup> and that mother-calf bonds are the strongest social bonds in elephants<sup>26,36</sup>  
385 (Extended Data Table 3). We also excluded pairs of calls that were recorded on the same date, as  
386 preliminary analyses indicated that calls recorded on the same day were more similar than calls  
387 recorded on different days, likely due to similarities in ambient conditions and/or autocorrelation  
388 within a calling bout (final sample size = 2391 call pairs). As calls from different behavioral

389 contexts differ in acoustic structure<sup>16</sup>, we categorized each pair of calls according to whether the  
390 two calls had the same or different behavioral contexts (“Same Context”) and included this  
391 variable as a factor in the analysis.

392 The proximity scores were highly skewed to the right, so we rank-transformed them and  
393 ran a Type III ANOVA with rank-transformed proximity score as the response variable and  
394 Same Caller Pair Type and Same Context as the factors. We weighted each observation (pair of  
395 calls) in the model by the minimum value of the certainty of caller ID and certainty of receiver  
396 ID for the two calls in the pair.

397 *Are vocal labels based on imitation of the receiver’s calls (hypothesis 2)?*

398 If elephants imitate the calls of the receiver that they are addressing, then callers should  
399 sound more like a given conspecific when they are addressing her than when they are addressing  
400 someone else (Extended Data Table 1, hypothesis 2, prediction 1). To assess whether this was  
401 the case, we classified each pair of calls into one of two types (hereafter, “Imitation Pair Type”):  
402 pairs in which the receiver of one call was the caller of the other call, and pairs in which this was  
403 not the case. We separately classified each call pair according to whether the two calls had the  
404 same relationship between caller and receiver (hereafter, “Same Relationship”). We also created  
405 a categorical variable Caller Dyad ID, which was an identifier for each unique combination of  
406 callers that comprised a call pair. We ran a Type III ANOVA with rank-transformed proximity  
407 score as the response variable and Imitation Pair Type, Same Relationship, Same Context, and  
408 Caller Dyad ID as factors. We weighted each observation (pair of calls) in the model by the  
409 minimum value of the certainty of caller ID and certainty of receiver ID for the two calls in the  
410 pair. By controlling for Caller Dyad ID in the model we assessed the effect of Imitation Pair  
411 Type within a given pair of callers; that is, whether calls from caller A to receiver B were more

412 similar to the receiver B's calls than calls from the caller A addressed to other receivers were to  
413 receiver B's calls. Pairs of calls that had the same caller or receiver, were recorded on the same  
414 day, were recorded from different family groups, or for which Caller Dyad ID did not occur with  
415 both levels of Imitation Pair Type were excluded from analysis (final sample size = 11,309 call  
416 pairs). Pairs of calls from different family groups were excluded because they comprised a small  
417 percentage of pairs where the receiver of one call was the caller of the other, and because it is  
418 possible that different families have different "dialects" which would influence call similarity.

419         If vocal imitation of the receiver occurs, it might or might not be the mechanism behind  
420 individual vocal labeling. To assess whether imitation of the receiver's calls was necessary for  
421 vocal labeling, we examined the calls in the dataset for which we had at least one recording of  
422 the receiver's calls and at least one recording of the caller addressing someone other than the  
423 receiver (n=494). For each of these calls, we calculated the mean proximity score between the  
424 focal call and all the calls made by the receiver (Mean Proximity to Focal Receiver When  
425 Targeting Focal Receiver) as well as the mean proximity score between each of the calls made  
426 by the focal caller to an individual other than the focal receiver and each of the calls made by the  
427 focal receiver (Mean Proximity to Focal Receiver When Targeting Others). Calls in which the  
428 Mean Proximity to Focal Receiver When Targeting Focal Receiver was greater than the Mean  
429 Proximity to Focal Receiver When Targeting Others were classified as "convergent" (n=195)  
430 and divergent otherwise (n=299). We then examined the proportion of convergent and divergent  
431 calls that were classified correctly by the random forest model with receiver ID and the acoustic  
432 features as input variables, and cross-validation folds stratified by caller ID and receiver ID. If  
433 vocal labeling relies on imitation of the receiver's calls, we expected only the convergent calls to  
434 be classified correctly more often than by the null model, but if imitation is not necessary for

435 vocal labeling, we expected both convergent and divergent calls to be classified correctly more  
436 often than by the null model (Extended Data Table 1, hypothesis 2, prediction 2). We also ran  
437 separate ANOVAs for the convergent calls and divergent calls, with rank-transformed proximity  
438 score as the response and Same Caller Pair Type and Same Context as the factors (excluding  
439 pairs of calls recorded on the same day). If vocal labeling relies on imitation of the receiver, we  
440 expected that there would only be an effect of Same Caller Pair Type among the convergent  
441 calls, but if imitation is not necessary for vocal labeling, we expected to observe an effect of  
442 Same Caller Pair Type among both sets of calls (Extended Data Table 1, hypothesis 2, prediction  
443 3).

444 *Do different callers use the same label for the same receiver (hypothesis 3)?*

445 To determine if different callers use the same label for the same receiver, we ran another  
446 6-fold cross-validated random forest model to predict receiver ID as a function of the acoustic  
447 features but partitioned the cross-validation folds such that all calls with the same caller and  
448 receiver were always allocated to the same fold (hyperparameters and weighting same as first  
449 model). This model tested whether receiver ID could be predicted independently of caller ID,  
450 which should only be possible if different callers use similar labels for a given receiver  
451 (Extended Data Table 1, hypothesis 3, prediction 1). We averaged the classification accuracy of  
452 the model across 2000 runs and compared this value to the distribution of classification  
453 accuracies generated by 10,000 iterations of the same model with the acoustic features randomly  
454 permuted.

455 If different callers use similar labels for the same receiver, then pairs of calls with  
456 different callers and the same receivers should be more similar than pairs of calls with different  
457 callers and different receivers (Extended Data Table 1, hypothesis 3, prediction 2). To test

458 whether this was the case, we ran another Type III ANOVA with rank-transformed proximity  
459 score as the response variable and Different Caller Pair Type (different callers/same receiver or  
460 different callers/same receiver), Same Relationship, and Same Context as the factors. As before,  
461 we excluded pairs of calls recorded on the same date or from different family groups (final  
462 sample size = 20,235 call pairs).

#### 463 *How are labels encoded in calls?*

464 To investigate which acoustic features encode receiver ID and caller ID we extracted  
465 variable importance scores (Supplementary Table S1) from a conditional inference random forest  
466 model in the R package “party”<sup>37</sup> trained on the full dataset to predict the response variable in  
467 question (receiver ID or caller ID) as a function of the acoustic features and weighted by the  
468 certainty of the response variable (625 training observations, 8000 trees, all other  
469 hyperparameters same as other random forests). We used a conditional inference forest because  
470 unlike traditional random forest, it is not biased towards correlated variables<sup>37</sup>. We only  
471 calculated variable importance scores for the spectral features, as cepstral coefficients are  
472 difficult to interpret intuitively. To assess the relative importance of the original acoustic  
473 contours, we weighted the loadings of the acoustic contours on each principal component by the  
474 variable importance score of the mean of the principal component in question, and then  
475 calculated the sum of the absolute values of these weighted loadings for each acoustic contour  
476 (Supplementary Table S2). Acoustic contours with a higher sum of the absolute values of the  
477 weighted loadings were deemed more important. This weighting process only considered the  
478 means of low-rank principal components, as it was not clear how to relate the other features back  
479 to the original acoustic contours. However, means of low-rank principal components accounted  
480 for the top 19 variables for the receiver ID model and top 33 variables for the caller ID model.

## 481 **Playback experimental design**

482 To determine if elephants respond more strongly to calls addressed to them (Extended  
483 Data Table 1, hypothesis 1, prediction 3), we played back rumbles with known adult female  
484 callers and known receivers to 17 elephants (15 adult females, one 9yo female, one 9-10yo male)  
485 in the Samburu study area. Fourteen subjects received one “test” playback of a call that was  
486 originally addressed to them and one “control” playback of a call from the same caller that was  
487 originally addressed to another individual. One subject received two sets of test and control  
488 playbacks from two different callers, one received only a test playback, and one received only a  
489 control playback (Extended Data Table 7). Most stimuli functioned as the test stimulus for one  
490 subject and the control stimulus for another, but no stimulus was used as the same experimental  
491 condition for more than one subject. Order of presentation was balanced across subjects, and we  
492 waited at least 7 days (mean =  $29.5 \pm 27.1$  days) between successive playbacks to the same  
493 subject.

## 494 **Playback stimuli**

495 Playback stimuli were recorded in Samburu and Buffalo Springs between January 2020  
496 and March 2022 from adult female callers. In all but two cases, the playback stimuli were contact  
497 calls. In one case we used a loud greeting call because we were unable to record a contact call  
498 from the caller in question, and in one case we used a call that was produced in a similar context  
499 to contact calls (caller and receiver >100 m apart and out of sight of each other), but was lower in  
500 amplitude than a typical contact call and was part of a lengthy antiphonal exchange between two  
501 individuals, and therefore was likely a “cadenced rumble”<sup>16</sup>. Three playback stimuli were  
502 elicited by another playback, and we assumed that the individual whose call was broadcast from  
503 the speaker was the intended receiver of the call that was produced in response to that playback.

504 We identified the receiver of natural calls as the only adult member of the family group who was  
505 separated from the caller during the call or the only individual who responded to the call. In one  
506 case, there were two adult females separated from the caller, and we assumed the receiver was  
507 the older of the two females who was in the lead and who rejoined the caller first (see Table  
508 S10). We prepared all playback stimuli in Audacity 3.0.2. Each stimulus consisted of a single  
509 rumble preceded by one second of background noise with a fade-in and followed by one second  
510 of background noise with a fade-out. In three cases, we applied a high-pass (5 Hz cutoff, 6 dB  
511 roll-off) or low-pass filter (1000 Hz cutoff, 6 dB roll-off) to remove excessive noise.

### 512 **Playback trial protocol**

513 The stimuli were played back as .wav files (uncompressed audio) from an iPhone SE  
514 (Apple Inc., Cupertino, CA) attached to QLXD1 wireless bodypack transmitter (Shure, Niles, IL)  
515 transmitting to a custom-built loudspeaker (Bag End Loudspeakers, Algonquin, IL) (see  
516 Supplementary Information). We placed the speaker 40.2-59.0 m from the subject (mean  $49.1 \pm$   
517 4.2 m), either on the ground in front of a tree or shrub and covered by camouflage netting or on  
518 the edge of the rear seat of a Toyota double cab Landcruiser facing the door with all four doors  
519 and windows and both roof hatches open. Re-recordings at 50 m revealed no obvious difference  
520 between sounds played with the speaker on the ground or inside the vehicle. We only conducted  
521 playbacks when the original caller and “alternate receiver” (the other subject receiving playbacks  
522 from the same caller) were >180 m from and out of sight of the subject (>270 m from the  
523 alternate receiver if she had not yet received all her playbacks). When the original caller’s  
524 location was known (19/34 trials) the speaker was placed in approximately the same direction  
525 relative to the subject as the original caller. In the remaining trials the caller could not be located  
526 after searching a ~300 m radius around the subject. Trials were redone after at least 7 days if the

527 speaker malfunctioned, the subject moved her head out of sight right before the playback started,  
528 or we discovered after the playback that the speaker was not in the correct location relative to the  
529 subject and the original caller. During each trial we filmed the subject from inside the vehicle for  
530 at least 1 min, then played the stimulus once, and continued filming for at least another 10 min.  
531 We also recorded audio with an Earthworks QTC40 microphone and Sound Devices MixPre3-II  
532 recorder. The observers were blind to the playback condition (test or control) until all trials were  
533 complete and all videos and audio recordings were scored.

### 534 **Statistical analysis of playback data**

535         From the video and audio recordings of each playback trial we measured the subject's  
536 Latency to Approach the speaker, Latency to Vocalize, Number of Vocalizations produced  
537 within 10 min following the playback, Latency to Vigilance, and Change in Vigilance Duration  
538 in the minute following the playback compared to the minute preceding the playback. Latencies  
539 were defined as the time from the start of the playback until the behavior of interest occurred and  
540 were censored when the subject moved out of sight or at 10 min, whichever came first. Vigilance  
541 was defined as lifting head above shoulder level, moving head from side to side, holding ears  
542 away from body without flapping, or lifting trunk while sniffing toward speaker. We ran a  
543 separate model for each response variable with Subject ID as a random effect and Treatment and  
544 the following covariates/factors as fixed effects: Caller-Original Receiver Relationship  
545 (relationship between the caller and the original receiver of the call; Extended Data Table 3),  
546 Distance (distance in meters between the speaker and the subject), dBC (amplitude of the  
547 playback stimulus in dBC at 1 m), Other Adults (whether other adults were within 50 m of  
548 subject during playback), Speaker Location (whether speaker was on ground or in vehicle), and  
549 Cumulative Playback Exposure (cumulative number of playbacks to which subject was exposed



550 at distance of 300 m or less, including trials that were redone and playbacks to other subjects).

551 We used Cox proportional hazards regression in the `coxme` package<sup>38</sup> for the latency variables, a  
552 generalized linear model with a Poisson error distribution in the `lme4` package<sup>39</sup> for Number of  
553 Vocalizations, and a linear model for Change in Vigilance Duration.

#### 554 **METHODS REFERENCES**

- 555 28. Moss, C. J. & Poole, J. H. Relationships and social structure in African elephants. in  
556 *Primate social relationships: an integrated approach* (ed. Hinde, R. A.) 315–325  
557 (Blackwell Science, Inc., 1983).
- 558 29. Altmann, J. Observational study of behavior: sampling methods. *Behaviour* **49**, 227–267  
559 (1974).
- 560 30. R Core Team. R: A language and environment for statistical computing. (2022).
- 561 31. Anikin, A. Soundgen: An open-source tool for synthesizing nonverbal vocalizations.  
562 *Behav. Res. Methods* **51**, 778–792 (2019).
- 563 32. Sykulski, M. `rpca`: RobustPCA: decompose a matrix into low-rank and sparse components.  
564 (2015).
- 565 33. Wright, M. N. & Ziegler, A. Ranger: A fast implementation of random forests for high  
566 dimensional data in C++ and R. *J. Stat. Softw.* **77**, 1–17 (2017).
- 567 34. Wittemyer, G. & Getz, W. M. Hierarchical dominance structure and social organization in  
568 African elephants, *Loxodonta africana*. *Anim. Behav.* **73**, 671–681 (2007).
- 569 35. Archie, E. A., Morrison, T. A., Foley, C. A. H., Moss, C. J. & Alberts, S. C. Dominance  
570 rank relationships among wild female African elephants, *Loxodonta africana*. *Anim.*  
571 *Behav.* **71**, 117–127 (2006).
- 572 36. Archie, E. A., Moss, C. J. & Alberts, S. C. Friends and relations: kinship and the nature of

- 573 female elephant social relationships. in *The Amboseli elephants: a long-term perspective*  
574 *on a long-lived mammal* (eds. Moss, C. J., Croze, H. & Lee, P. C.) 238–245 (University of  
575 Chicago Press, 2011). doi:10.7208/chicago/9780226542263.001.0001.
- 576 37. Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T. & Zeileis, A. Conditional variable  
577 importance for random forests. *BMC Bioinformatics* **9**, 1–11 (2008).
- 578 38. Therneau, T. M. *coxme: mixed effects cox models*. (2019).
- 579 39. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models  
580 using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).

## 581 **ACKNOWLEDGEMENTS**

582 We thank the Office of the President of Kenya, the Samburu, Isiolo, and Kajiado County  
583 governments, the Wildlife Research & Training Institute of Kenya, and Kenya Wildlife Service  
584 for permission to conduct fieldwork in Kenya. We thank Save The Elephants and the Amboseli  
585 Trust for Elephants for logistical support in the field, J. M. Leshudukule, D. M. Letitiya, and  
586 Norah Njiraini for assistance with the fieldwork, G. Pardo for blinding the playback stimuli, and  
587 S. Pardo for input on the statistical analyses. We thank J. Berger, W. Koenig, and A. Horn for  
588 comments on the manuscript. This project was funded by a Postdoctoral Research Fellowship in  
589 Biology to MP from the National Science Foundation (award no. 1907122) and grants to JP and  
590 PG from the National Geographic Society, Care for the Wild, and the Crystal Springs  
591 Foundation. Field work was supported by Save the Elephants.

## 592 **AUTHOR CONTRIBUTIONS**

593 MP conceived the study. MP and DL collected the data in Samburu and JP and PG collected the  
594 data in Amboseli. MP and KF performed the statistical analysis and MP created the figures. MP

595 drafted the manuscript and KF, JP, and GW edited it. CM, IDH, and GW provided resources and  
596 access to long-term datasets and GW supervised the study.

597 The authors declare no competing interests.

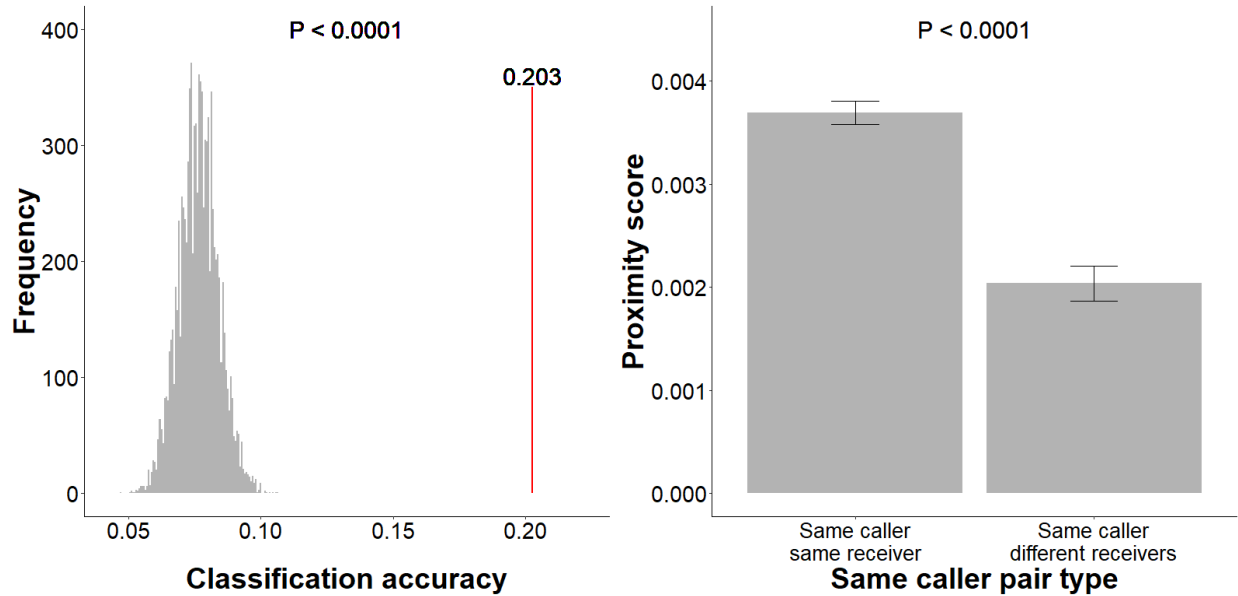
598 Supplementary Information is available for this paper.

599 Correspondence and requests for materials should be addressed to MP.

600 Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

601 Data are available at [doi:10.5061/dryad.hmgqnk9nj](https://doi.org/10.5061/dryad.hmgqnk9nj)

602 Code is available at [doi:10.5061/dryad.hmgqnk9nj](https://doi.org/10.5061/dryad.hmgqnk9nj)



603

604 **Figure 1. Evidence that calls are specific to individual receivers within a caller. Left:**

605 classification accuracy of random forest predicting receiver ID from acoustic features (red line)

606 was significantly higher than classification accuracies of 10,000 null models predicting receiver

607 ID from randomized acoustic features (gray histogram). Cross-validation folds were stratified so

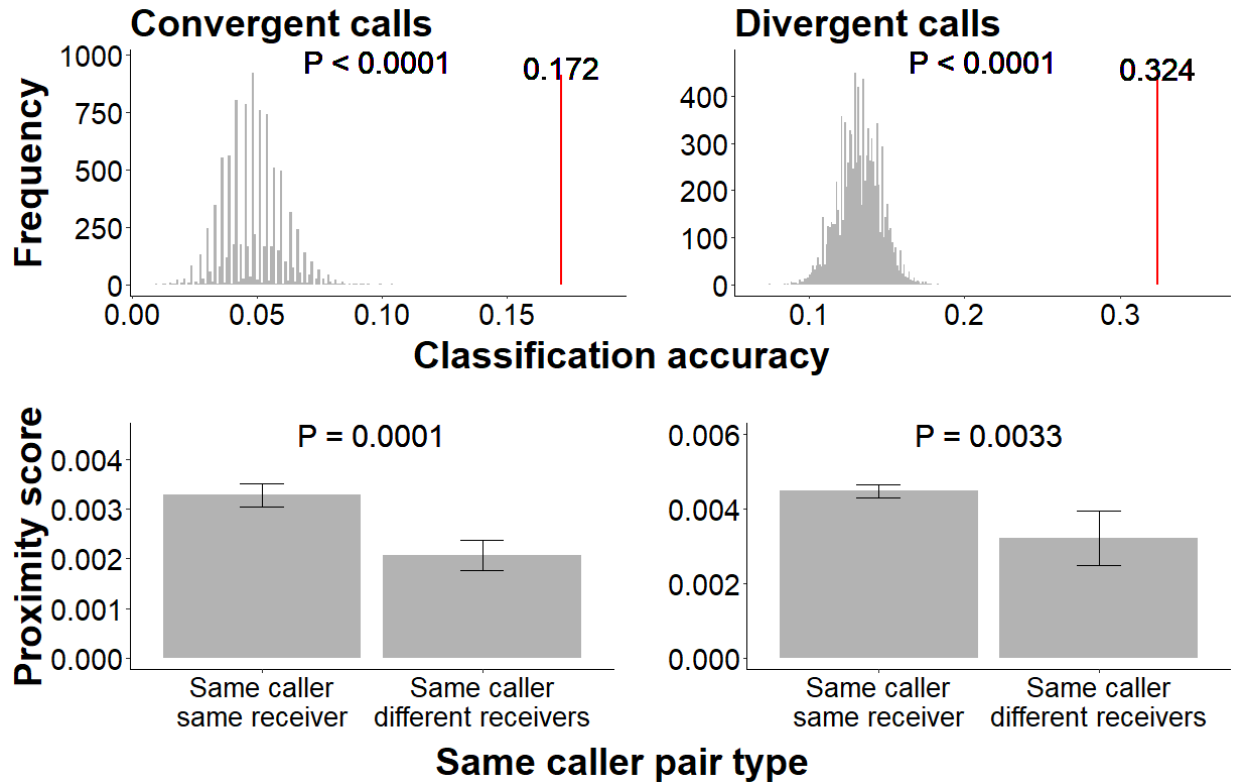
608 that model was trained and test on same combinations of caller and receiver; thus, classification

609 accuracy represents receiver specificity of calls within a caller. **Right:** calls with the same caller

610 and same receiver were significantly more similar (higher proximity score) than calls with the

611 same caller and different receivers (ANOVA on ranks). Error bars represent standard errors of

612 the mean.



613

614 **Figure 2. Evidence that vocal labeling likely did not rely on imitation of the receiver's calls.**

615 Random forest predicted receiver ID significantly better than null models both among calls that

616 were identified as convergent to receiver's calls (**top left**) and divergent from receiver's calls

617 (**top right**). Pairs of calls with the same caller and same receiver were more similar (higher

618 proximity score) than pairs of calls with the same caller and different receivers, both among calls

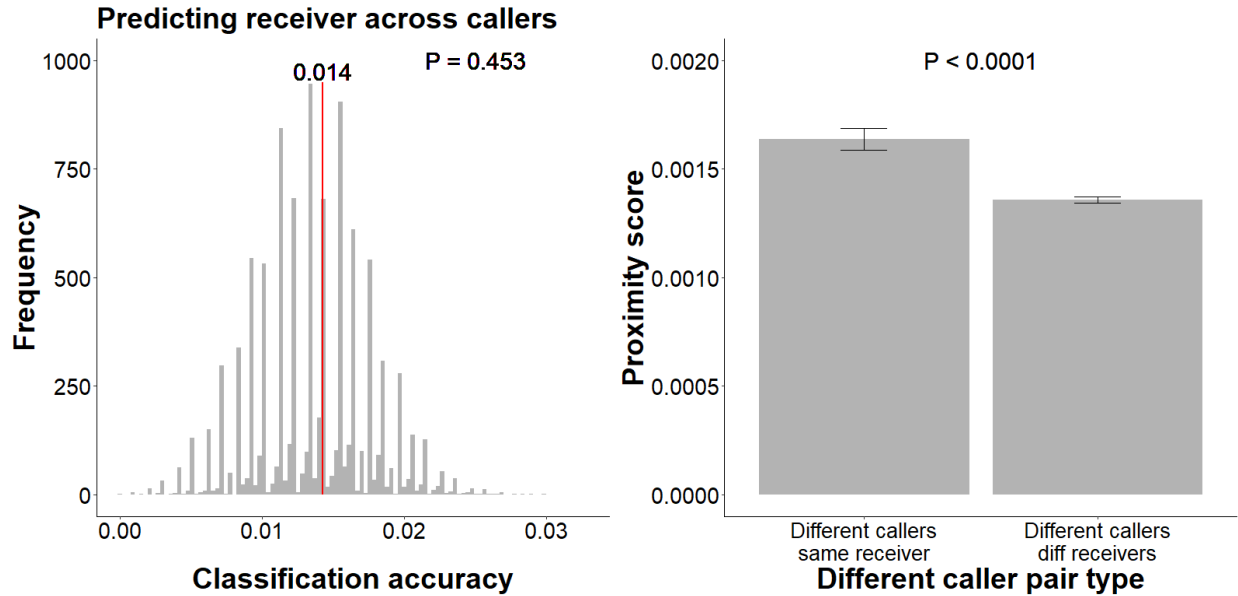
619 that were convergent to receiver's calls (**bottom left**) and calls that were divergent from

620 receiver's calls (**bottom right**) (ANOVA on ranks). In top row, red lines represent classification

621 accuracy of original random forest model and gray histograms represent distribution of

622 classification accuracies of null models with randomized acoustic features. In bottom row, error

623 bars represent standard errors of the mean.



624

625 **Figure 3. Mixed evidence that different callers use similar labels for the same receiver.**

626 **Left:** Classification accuracy (red line) of random forest designed to predict receiver ID from

627 acoustic features independently of caller ID (all calls with the same caller and receiver allocated

628 to the same cross-validation fold) was not significantly different from classification accuracies of

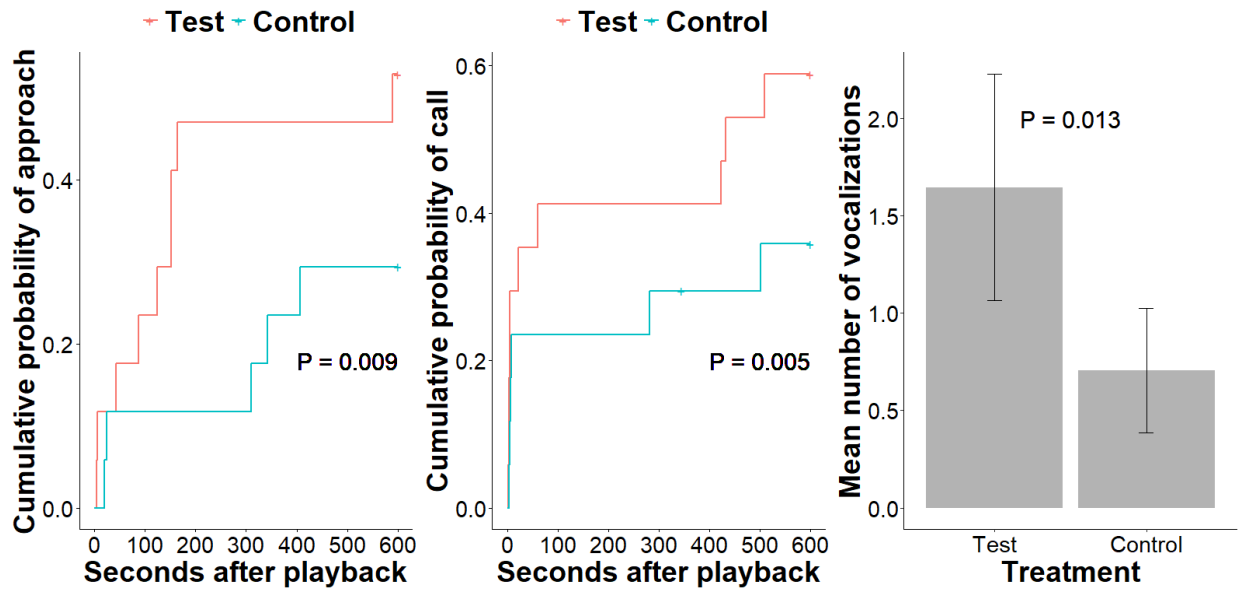
629 null models with randomized acoustic features (gray histogram). **Right:** Pairs of calls with

630 different callers and the same receiver were significantly more similar (higher proximity score)

631 than pairs of calls with different callers and different receivers (ANOVA on ranks). Error bars

632 represent standard errors of the mean.

633



634

635 **Figure 4. Response to playbacks of test stimuli (calls originally addressed to the subject) vs.**  
636 **control stimuli (calls from the same caller originally addressed to a different individual).**

637 Subjects approached the speaker more quickly (**left**; Cox regression), vocalized more quickly  
638 (**center**; Cox regression), and produced more vocalizations (**right**; Poisson GLM) in response to  
639 test playbacks than controls (note the model for number of vocalizations failed to converge).

640 Error bars in rightmost panel represent standard errors of the mean.

641

642 **Table 1. Results of random forest models predicting receiver ID as a function of the**  
 643 **acoustic features**

Hypothesis tested	Observations used	Data partitioning	Classification accuracy	Mean $\pm$ SD accuracy for null models	Permutation test <i>P</i> -value
<i>Spectral acoustic features</i>					
H1: calls are receiver specific	All (625)	Stratified by caller and receiver ID	20.3%	7.6 $\pm$ 0.75%	<0.0001
H2: labels are arbitrary	Convergent calls (195)	Stratified by caller and receiver ID	17.2%	4.9 $\pm$ 1.1%	<0.0001
H2: labels are arbitrary	Divergent calls (299)	Stratified by caller and receiver ID	32.4%	13.1 $\pm$ 1.4%	<0.0001
H3: labels shared across callers	All (625)	All calls with same caller and receiver in same fold	1.4%	1.4 $\pm$ 0.40%	0.453
<i>Cepstral acoustic features</i>					
H1: calls are receiver specific	All (625)	Stratified by caller and receiver ID	14.9%	6.3 $\pm$ 0.96%	<0.0001
H2: labels are arbitrary	Convergent calls (195)	Stratified by caller and receiver ID	13.4%	4.5 $\pm$ 1.4%	<0.0001
H2: labels are arbitrary	Divergent calls (299)	Stratified by caller and receiver ID	22.0%	10.0 $\pm$ 1.7%	<0.0001
H3: labels shared across callers	All (625)	All calls with same caller and receiver in same fold	1.4%	1.4 $\pm$ 0.48%	0.433

644 All random forests had 500 trees, 6 variables per node, 60% of observations per tree, minimum  
 645 node size = 1, and no maximum tree depth, and 6-fold for cross-validation. Observations were  
 646 weighted by the certainty of receiver ID. Classification accuracies were averaged across 2000



647 runs of the model to improve stability. To determine if the classification accuracy was higher  
648 than expected by chance, the model was run 10,000 times with randomly permuted acoustic  
649 variables, and the original classification accuracy was compared to the distribution of  
650 classification accuracies for these 10,000 null models.

651

652 **Table 2. Results for Type III Analyses of Deviance on playback experiment models**

<b>Response variable</b>	<b>Model type</b>	<b>Trtmnt</b>	<b>Reltnshp Caller to Org. Rcv.</b>	<b>Dist.</b>	<b>dBC</b>	<b>Other adults</b>	<b>Speaker location</b>	<b>Cumul. playback exposure</b>
Latency to approach	Cox	$\chi^2=6.8,$ $P=0.009$	$\chi^2=1.7,$ $P=0.80$	$\chi^2=2.4,$ $P=0.12$	$\chi^2=0.65,$ $P=0.42$	$\chi^2=0.41,$ $P=0.52$	$\chi^2=0.59,$ $P=0.44$	$\chi^2=0.11,$ $P=0.73$
Latency to vocalize	Cox	$\chi^2=7.9,$ $P=0.005$	$\chi^2=6.4,$ $P=0.17$	$\chi^2=0.97,$ $P=0.32$	$\chi^2=0.02,$ $P=0.90$	$\chi^2=0.64,$ $P=0.42$	$\chi^2=0.20,$ $P=0.66$	$\chi^2=0.10,$ $P=0.75$
Number of calls	Poisson	$\chi^2=6.2,$ $P=0.013$	$\chi^2=19.9,$ $P=0.0005$	$\chi^2=0.32,$ $P=0.57$	$\chi^2=0.48,$ $P=0.49$	$\chi^2=0.72,$ $P=0.40$	$\chi^2=0.13,$ $P=0.72$	$\chi^2=0.01,$ $P=0.91$
Latency to vigilance	Cox	$\chi^2=3.1,$ $P=0.08$	$\chi^2=10.1,$ $P=0.038$	$\chi^2=1.8,$ $P=0.18$	$\chi^2=1.9,$ $P=0.16$	$\chi^2=5.5,$ $P=0.019$	$\chi^2=0.55,$ $P=0.46$	$\chi^2=0.02,$ $P=0.88$
Vigilance duration after - before	Linear	$\chi^2=0.06,$ $P=0.81$	$\chi^2=2.1,$ $P=0.72$	$\chi^2=4.0,$ $P=0.045$	$\chi^2=0.02,$ $P=0.89$	$\chi^2=0.43,$ $P=0.51$	$\chi^2=0.33,$ $P=0.56$	$\chi^2=0.83,$ $P=0.36$

653 Subject ID (not shown) was also included as a random effect in each model. The Poisson

654 regression for Number of vocalizations failed to converge.

655

656 **TITLES AND LEGENDS FOR EXTENDED DATA**

657 **Extended Data Figure 1. Violin plots illustrating distribution of data with respect to callers**

658 **and receivers.** The dataset consisted of 625 total calls, 114 unique callers, and 119 unique  
659 receivers, but each caller only addressed a small number of the receivers in the dataset.

660 **Extended Data Figure 2. Schematic illustrating how spectral acoustic features were**

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

671 **Extended Figure 3. Scatterplots showing the separation in 3D space between calls from the**

672 **same caller to different receivers.** Axes are the three most important variables for predicting  
673 receiver ID (means of PCs 33, 23, and 48) as determined from the variable importance scores of  
674 a conditional inference random forest using the spectral acoustic features. Each plot represents a  
675 single caller, each point is a single call, and receiver IDs are coded by both color and shape. This  
676 figure only includes calls where certainty of caller ID and receiver ID were at least 0.5 (no more  
677 than 2 possible candidates) and the caller made at least 3 calls each to at least 2 different  
678 receivers.

679 **Extended Data Table 1. Hypotheses and predictions tested in this study**

680 **Extended Data Table 2. Acoustic features used in the random forest models**

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

690 **Extended Data Table 3. Definitions of social relationship categories between caller and**  
691 **receiver**

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

699 **Extended Data Table 4. Results for ANOVAs to test if calls with the same caller and**  
700 **receiver were more similar than calls with the same caller and different receivers**

701 Each observation was a pair of calls. ANOVA models were of the form Rank-transformed  
702 Proximity Score ~ Same Caller Pair Type (whether the two calls in a pair had the same caller and  
703 receiver or same caller and different receivers) + Same Context (whether the two calls in a pair  
704 had the same behavioral context). Pairs of calls recorded on the same date or where the two calls  
705 had a different type of caller-receiver relationship were excluded. Three models were run for  
706 each set of acoustic features (spectral and cepstral): all pairs of calls meeting above criteria  
707 (n=2391), pairs of calls in which both calls were convergent on the receiver's calls (n=252), and  
708 pairs of calls in which both calls were divergent from the receiver's calls (n=798). Convergent  
709 calls = calls from caller A to receiver B that were more similar to receiver B's calls than calls  
710 from caller A to other receivers were to receiver B's calls. Divergent calls = calls from caller A  
711 to receiver B that were less similar to receiver B's calls than calls from caller A to other receivers  
712 were to receiver B's calls.

713 **Extended Data Table 5. Results for ANOVAs to test if calls addressed to a given receiver**  
714 **were imitative of the receiver's calls**

715 Each observation was a pair of calls. ANOVA models were of the form Rank-transformed  
716 Proximity Score ~ Imitation Pair Type + Same Relationship + Same Context + Caller Dyad ID.  
717 Model was run once for each set of acoustic features: spectral and cepstral. Imitation Pair Type =  
718 whether the caller of one call in a pair was the receiver of the other call. Same Relationship =  
719 whether the callers of both calls in a pair had the same type of relationship to their respective  
720 receivers. Same Context = whether the two calls in a pair were recorded in the same behavioral  
721 context (contact/greeting). Caller Dyad ID = identifier for the two callers in a pair. Pairs of calls  
722 recorded on the same date, from callers in different social groups, or with the same caller or

723 receiver were excluded. We also excluded pairs of calls for which Caller Dyad ID only occurred  
724 with one level of Imitation Pair Type (final n=11,309).

725 **Extended Data Table 6. Results for ANOVAs to test if different callers used similar labels**  
726 **for the same receiver**

727 Each observation was a pair of calls. ANOVAs were of the form Rank-transformed Proximity  
728 Score ~ Different Caller Pair Type + Same Relationship + Same Context. Model was run  
729 separately for each set of acoustic features: spectral and cepstral. Different Caller Pair Type =  
730 whether the two calls in a pair had different callers and the same receiver or different callers and  
731 different receivers. Same Relationship = whether the two calls in a pair had the same type of  
732 relationship between caller and receiver. Same Context = whether the two calls in a pair were  
733 recorded in the same behavioral context (contact/greeting) or not. Pairs of calls recorded on the  
734 same date or from callers in different social groups were excluded (final n=20,235)

735 **Extended Data Table 7. Summary of playback trials for each subject**

736 All callers and subjects were adult females except M25.0012 (subadult male) and M9.9612  
737 (subadult female). The letter in parentheses after each caller ID represents a unique call (e.g.,  
738 R23 (a) and R23 (b) were different calls recorded from R23). Twelve trials were redone once or  
739 twice because the playback system malfunctioned, the subject went out of sight just as the  
740 playback began, or the speaker was accidentally placed >60 m away or in the wrong direction  
741 relative to the subject and the original caller. Trials that were later redone are not included in this  
742 table.