1   Classification: Biological Sciences, Evolution

2   Short Title: Compositional Calling in Pied Babbler

3

4   **Meaningful Call Combinations and Compositional Processing in the Southern**
5   **Pied Babbler**

6

7   Sabrina Engesser [a,b,1], Amanda R. Ridley [a,c,d], Simon W. Townsend [a,b,e,1]

8

9   [a] Pied Babbler Research Project, Kuruman River Reserve, Van Zylsrus, South Africa

10 [b] Animal Behaviour, Department of Evolutionary Biology and Environmental Studies,
11   University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

12 [c] Centre for Evolutionary Biology, School of Animal Biology, The University of
13   Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia

14 [d] Percy Fitz Patrick Institute, University of Cape Town, Rondebosch 7701, Cape
15   Town, South Africa

16 [e] Department of Psychology, University of Warwick, University Road, Coventry
17   CV4 7AL, United Kingdom

18

19 [1] Corresponding authors:

20 Sabrina Engesser, Animal Behaviour, Department of Evolutionary Biology and
21   Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057
22   Zurich, Switzerland
23   Email: sabrina.engesser@ieu.uzh.ch
24   Phone: +41 (0)44 635 52 72

25 Simon W. Townsend, Department of Psychology, University of Warwick, University
26   Road, Coventry CV4 7AL, United Kingdom
27   Email: simon.w.townsend@warwick.ac.uk
28   Phone: +44 (0)24 765 746 43

29 **Abstract**

30 Language's expressive power is largely attributable to its compositionality:

31 meaningful words are combined into larger/higher-order structures with derived

32 meaning. Despite its importance, little is known regarding the evolutionary origins

33 and emergence of this syntactic ability. Whilst previous research has demonstrated a

34 rudimentary capability to combine meaningful calls in primates, due to a scarcity of

35 comparative data, it is unclear whether analogue forms might also exist outside of

36 primates. Here we address this ambiguity and provide evidence for rudimentary

37 compositionality in the discrete vocal system of a social passerine, the pied babbler

38 (*Turdoides bicolor*). Natural observations and predator presentations revealed

39 babblers produce acoustically distinct alert calls in response to close, low-urgency

40 threats, and recruitment calls when recruiting group members during locomotion.

41 Upon encountering terrestrial predators both vocalisations are combined into a

42 'mobbing-sequence', potentially to recruit group members in a dangerous situation.

43 To investigate whether babblers process the sequence in a compositional way, we

44 conducted systematic experiments, playing back the individual calls in isolation, as

45 well as naturally occurring and artificial sequences. Babblers reacted most strongly to

46 mobbing-sequence playbacks, showing a greater attentiveness and a quicker approach

47 to the loudspeaker, compared to individual calls or control sequences. We conclude

48 the sequence constitutes a compositional structure, communicating information on

49 both the context and the requested action. Our work supports previous research

50 suggesting combinatoriality as a viable mechanism to increase communicative output,

51 and indicates that the ability to combine and process meaningful vocal structures, a

52 basic syntax, may be more widespread than previously thought.

53

54   Key words: call combination, compositionality, syntax, language evolution, southern

55   pied babbler

56

57   **Significance Statement**

58   Human language is syntactic in its nature: meaningful words are assembled into larger

59   meaningful phrases or sentences. How unique this ability is to humans remains

60   surprisingly unclear. A considerable body of work has indicated birds are capable of

61   combining sounds into large elaborate songs, but there is currently no evidence

62   suggesting these structures are syntactic. Here, we provide important evidence for this

63   ability, in a highly social bird. Specifically, pied babblers combine two functionally

64   distinct vocalisations into a larger sequence, the function of which is related to the

65   function of its parts. Our work adds important evidence to the variation and

66   distribution of combinatorial vocal mechanisms outside humans, and provides insights

67   into potentially early forms of human syntactic communication.

68   \body

69   **Introduction**

70   Syntax is often considered one of the key defining features of human language (1).

71   Through combining meaningful words together, larger sequences with related,

72   compositional meaning can be constructed (2). One consequence of such

73   compositional syntax in humans is that with a finite inventory of words, an infinite

74   range of ideas and concepts can be communicated (2, 3). Despite the central role

75   syntax plays in determining language's generativity, very little is known about its

76   evolutionary origins or early forms (4, 5). Elucidating the proto forms of

77   compositional syntax, whilst non-trivial (5, 6), represents a key step in understanding

78   the evolution of language more holistically.

79   One means of investigating early forms and function of compositionality is to

80   assess analogue examples in animals (5, 7). Indeed, recent observational and

81   experimental work on two related guenon monkeys has demonstrated the propensity

82   to combine context-specific, 'meaningful' signals into sequences that resemble

83   compositional structures in language. Male Campbell's monkeys (*Cercopithecus*

84   *campbelli*), for example, produce predator-specific alarm calls that can be affixed

85   with an acoustic modifier (8). The affix acts to alter the 'meaning' of the alarm calls

86   in a predictable way, transforming them into general disturbance calls (8). Similarly,

87   male putty-nosed monkeys (*C. nictitans*) combine two predator-specific alarm calls

88   into a higher-order sequence (9, 10). While the two calls are associated with the

89   presence of aerial, or terrestrial predators, respectively, the resultant combination

90   initiates group movement in non-predatory contexts (9, 10). Given the discrepancies

91   between the responses elicited by the individual calls and the sequence, it remains

92   unclear whether the putty-nosed monkey call sequence represents a form of

93   compositional syntax, or rather a combinatorial syntax, where the meaning of the

94   whole is not directly related to the parts, akin to idiomatic expressions in language

95   (i.e. "kick the bucket" for dying) (9, 11, 12). The existence of such 'semantic

96   combinations' (9) in primates has nevertheless been argued to support an

97   evolutionarily ancient origin of human syntax, rooted within the primate lineage (8,

98   13). However, it is unclear whether similar call concatenations and compositional

99   processing of information might also exist in other lineages and, if so, whether they

100   take analogous forms and serve analogous functions (1).

101   The last 50 years of comparative research has demonstrated a number of non-

102   primate animals, particularly songbirds, are capable of stringing sounds together into

103   larger, often more structurally complex, sequences (14-16). Yet, there is no indication

104   that any of these song sequences are compositional in structure, since the individual

105   sounds composing the songs of birds and other animals do not convey any

106   independent meaning (14-16); ultimately precluding any attempt to test for

107   proto-syntactic abilities in these species in the first place. While this might suggest

108   that syntactic abilities are potentially confined to the primate lineage (8, 13), it may

109   also be an artefact of limited focus on bird vocal systems, other than song, that are

110   more likely to support the capacity for syntax.

111   Here we address this ambiguity through investigating the prevalence of

112   compositional vocal sequences in a highly social, non-singing passerine bird that

113   possesses a discrete vocal system: the cooperatively breeding southern pied babbler

114   (*Turdoides bicolor*) (17, 18). Pied babblers are territorial and live in stable groups of

115   three to 15 individuals (19). Reproduction is usually restricted to the dominant pair of

116   the group (20), with subordinate individuals engaging in a number of helping

117   behaviours, such as territorial and nest defence, daytime incubation, and feeding of

118   the offspring during the nestling and post-fledgling stage (19). Individuals of the

119   cohesive foraging group spend most of the time on the ground searching for

120   invertebrates hidden in the substrate, which they excavate using their bill (19, 21).

121   Consequently, most of the time pied babblers forage in a head-down position within

122   and around forbes and shrubs and hence rely heavily on vocalisations to keep track of

123   changes in their surroundings (18, 22-26). As such, the pied babbler vocal system

124   exhibits around 17 discrete vocalisations including alarm calls, sentinel calls, as well

125   as a diverse array of social calls produced during intra- and inter-group contexts (18,

126   22-26).

127   Observational work has indicated that pied babblers produce broad-band,

128   noisy alert calls in response to sudden, but generally low urgency, threats (e.g.

129   abruptly approaching animals), and more tonal, repetitive, recruitment calls when

130   recruiting group members to a new location or during locomotion, mainly in foraging

131   or roosting contexts. Moreover, alert and recruitment calls can be combined into a

132   sequence upon encountering and mobbing, mainly terrestrial, predators (Fig. 1).

133   Given the context in which the two independent calls are produced, we aimed to

134   investigate whether the sequence might therefore function specifically to recruit group

135   members in a dangerous situation (e.g. when mobbing a predator) by combining

136   information on both the danger and the requested action. Accordingly, the

137   combination of alert and recruitment calls (hereafter termed 'mobbing sequence') may

138   constitute a rudimentary compositional structure, where the meaning of the whole is a

139   product of the meaning of its parts (27).

140   To verify the context-specific information conveyed by the independent

141   vocalisations, and to test whether pied babblers extract the meaning of the sequence in

142   a compositional way, we conducted further natural observations in combination with

143   acoustic analyses and experimental manipulations. First, acoustic analyses were

144   applied to confirm that alert and recruitment calls constitute two distinct vocalisations.

145   Second, to determine the contexts in which the individual calls and the call sequence

146   are produced, we conducted natural observations and predator presentation

147   experiments in combination with audio recordings. Third, we carried out systematic

148   natural, artificial and control playback experiments to investigate whether birds

149   perceive the sequence compositionally. Key support for compositionality requires that

150   the contexts in which mobbing sequences are produced and the responses of receivers

151   to playbacks of these sequences are related to the information encoded in alert and

152   recruitment calls (27).

153

154   **Results**

155   **Acoustic Analysis**. A Discriminant Function Analysis (DFA) indicated that alert and

156   recruitment calls could be statistically discriminated based on their structure alone

157   ($N_{individuals}$=16, N_{calls}=32, correct classification: 97%, p<0.001). When applying a

158   leave-one-out cross-validated DFA, 91% were correctly classified, a classification

159   higher than expected by chance (two-tailed binomial test, change level=50%,

160   p<0.001).

161

162   **Alert and Recruitment Calls: Natural Context.** Natural observations in

163   combination with acoustic recordings were conducted to quantify the calls' context

164   specificity. From a total of 36 alert calls recorded in 11 groups, 69% were elicited by

165   suddenly appearing, non-dangerous subjects (e.g. hares, antelopes, researchers). 14%

166   of alert calls were caused by inactive snakes, or by distant mongooses or foxes that

167   did not present a direct threat to babblers. A further 6% were produced in response to

168   alarm calls of con- or hetero-specifics. For the remaining 11% of occasions no

169   obvious threat could be detected.

170   From a total of 196 recorded recruitment call events from 71 individuals in 20

171   groups, 60% resulted in other group members approaching the caller, and 6% in

172   overall group movement, following the caller. In the remaining 34%, recipients either

173   showed no response (44 out of 67 occasions) or counter-called with recruitment or

174   other loud calls (23 out of 67 occasions) (26). All recorded recruitment calls were

175   produced in non-dangerous contexts, in the absence of any predators. Thus, while

176   alert calls seem to encode information about low-urgency threats in a caller's

177   imminent surrounding, recruitment calls appear to function to recruit group members

178   to a caller's current location.

179

180   **Mobbing Sequences: Natural Context and Experimental Elicitation.** We observed

181   naturally elicited mobbing sequences on 39 occasions in 14 groups: 85% were

182   produced in response to moving terrestrial predators (mongooses, snakes, foxes), and

183   8% in response to small perched raptors (pygmy falcon (*Polihierax semitorquatu*),

184   pearl spotted owl (*Glaucidium perlatum*)) which are assumed to only pose a threat to

185   young, inexperienced babblers. In the remaining 8% of events, no clear context could

186   be assigned. To experimentally confirm the context accompanying the production of

187   mobbing sequences, babbler groups were presented with a model of a Cape cobra

188   (*Naja nivea*) and their calling behaviour was noted. From a total of 13 presentations in

189   10 groups, mobbing sequences were elicited 92% of the time.

190

191   **Playback Experiment.** To investigate the responses to mobbing sequences and their

192   individual calls, we played back natural mobbing sequences, as well as the constituent

193   alert and recruitment calls to subjects. To rule out alternative explanations associated

194   with the saliency of the stimulus (two vs. one call type) or priming effects (any call

195   type preceding recruitment calls generates the same response), we implemented an

196   additional important control condition, where we artificially replaced the alert call of

197   a mobbing sequence with another acoustically distinct broad-band babbler

198   vocalisation, the foraging 'chuck' call (chuck-recruitment sequence, see Supporting

199   Online Material: Acoustic Analysis of Chuck, Alert, and Recruitment Calls) (21, 28,

200   29). Finally, in line with previous studies (9, 30), to really ensure the key dimension

201   for receivers was the combination of information, and not any urgency-based acoustic

202   variation encoded within the structure, as a further control, artificial mobbing

203   sequences were constructed from the independent calls and played back (see

204   Supporting Online Material: Stimuli Sets) (9, 28).

205   Our playbacks revealed differences in group attentiveness responses to the

206   four playback conditions, determined by the proportion of the group that became

207   vigilant (treatment: χ^2=53.5, P<0.01, N=64, 16 groups, Fig. 2, Table 1), and the

208   latency to resume normal, non-vigilant behaviour of the first reacting group member

209   (treatment: χ^2=36.3, P<0.001, N=64, 16 groups, Fig. 2, Table 1). Moreover, the

210   movement patterns of a group relative to the sound source differed in response to the

211   four stimuli (treatment: χ^2=97.2, time: χ^2=34.9, treatment*time: χ^2=23.6; all

212   P<0.001, N=378, 16 groups, Fig. 3, Table 1).

213   Specifically, alert calls played back in isolation did not result in noticeable

214   changes in behaviour (such as attentiveness, Fig. 2, Table 1 & S1), and we found no

215   effect of time on distance moved, with groups neither approaching or retreating from

216   the sound source (Fig. 3, Table 1 & S1). In accordance with the assumed function to

217   recruit group members to a caller's location, in response to played back recruitment

218   calls, babblers increased their attentiveness compared to playbacks of alert calls,

219   likely as a way to locate the simulated recruiting caller, and slowly, steadily

220   approached the sound source (Fig. 1 & 2, Table 1 & S1). Furthermore, in line with our

221   central prediction of mobbing sequences functioning to recruit group members in a

222   dangerous situation, we found that subjects responded most strongly to playbacks of

223   mobbing sequences revealing the highest attentiveness and fastest approach towards

224   the sound source (Fig. 2 & 3, Table 1 & S1). Ruling out priming or stimulus effects,

225   playbacks of chuck-recruitment control sequences did not elicit similar mobbing-like

226   behaviours, with babblers neither approaching the sound source, nor increasing their

227   attentiveness, compared to playbacks of mobbing sequences (Fig. 2 & 3, Table 1 &

228   S1). These results support our hypothesis that the call sequence tested conforms to the

229   definition of basic compositional syntax, with the high vigilance response to mobbing

230   sequences and the fast approach to the loudspeaker being directly related to the

231   contextual information and function of both individual calls.

232

233   **Discussion**

234   Here we provide key comparative data indicating the cooperatively breeding pied

235   babbler can extract meaningful rudimentary compositional information from

236   combinations of acoustically distinct, context-specific vocalisations: alert and

237   recruitment calls.

238   Systematic observational and experimental data implementing both natural

239   and artificial playback experiments demonstrate that pied babbler alert calls encode

240   information on existing or imminent, low-urgency threats in the environment, whilst

241   recruitment calls communicate the motivation to recruit group members to the caller's

242   location. Combinations of these alert and recruitments calls, here-called 'mobbing

243   sequences', are produced when babblers encounter and mob predominantly terrestrial

244   threats. In response to played back mobbing sequences, babblers reacted with an

245   increased attentiveness (high proportion of the group being vigilant and long latency

246   to resume non-vigilant behaviour), and a rapid approach toward the sound source,

247   potentially to support the simulated caller opposing the putative threat. The context

248   accompanying the mobbing sequence and particularly the responses to the playbacks,

249   suggests the information encoded in the combination is a direct product of the

250   constituent calls (27). We are confident we can rule out alternative explanations

251   related to a sequential or additive processing of calls, as responses to played-back

252   mobbing sequences exceeded those elicited by the independent calls (29, 31).

253   Furthermore, control experiments demonstrated that potential super-stimuli (two calls

254   vs. one call) or simple priming effects that could otherwise explain the results can be

255   excluded, since control sequences failed to elicit similar mobbing-like behaviour (28,

256   29). In summary, our natural observations combined with the experimental

257   manipulations indicate that babblers produce and parse the sequence by linking

258   information on the context (threat) and the requested action.

259   Our work, providing strong evidence for a rudimentary compositional syntax

260   in birds, complements and extends previous research demonstrating similar semantic

261   combinations in primates and suggests that the basic capacity to combine

262   'meaningful' calls into systematic higher-order structures may be more diverse and

263   widespread than previously thought (8, 11). Furthermore, these findings have

264   important implications for understanding the evolutionary progression of human

265   language. One dominant hypothesis posits that language's hierarchical syntactic

266   system could have only evolved as part of a sudden evolutionary event, precluding the

267   existence of intermediate proto-syntactic forms (32). Alternatively, it has been

268   suggested that syntax can be decomposed into more primitive layers, consisting of

269   loose two- or few-word compounds which form the evolutionary and structural basis

270   of syntactic systems (27, 33-35). Under this scenario, a sudden evolutionary leap is

271   not necessary (27), as instead, language's syntactic complexity is hypothesized to

272   have originally emerged out of simple, but communicatively meaningful

273   compositions. Support for this hypothesis can be found in language acquisition and

274   newly emerging sign languages, where syntactic development initiates with simple

275   two word/sign compositions or "packages" (27, 35, 36), gradually proceeding, in later

276   stages, to more sophisticated multi-package compositions (27). Through providing

277   comparative data for such two-signal constructs in the pied babbler vocal repertoire,

278   our work contributes further evidence that basic, intermediate compositional

279   structures are viable, and hence supports the idea that syntax could have evolved by

280   progressing gradually over time, rather than spontaneously as an 'all-or-nothing'

281   package (34).

282   Exactly what evolutionary forces accompanied the progression of syntax

283   remain elusive. Theoretical work conducted over the last two decades has aimed to

284   disentangle the selective conditions promoting the emergence of syntax (6, 37, 38).

285   Specifically, mathematical modelling approaches have indicated that natural selection

286   will favour a transition toward a syntactic communication system (from a non-

287   syntactic one) when the number of relevant events to be communicated exceeds the

288   number of available calls (either due to production or perception constraints) (6, 38).

289   Our work provides important empirical evidence that support this claim. Given the

290   pied babblers' constrained vocal repertoire, paired with the extensive number of social

291   and ecological contexts that require communication (19), compositional production

292   and processing of vocalisations is likely adaptive for pied babblers, allowing them to

293   coordinate key additional events than would be possible with a non-syntactic system.

294   Moreover, combining and processing signals in a compositional way may be

295   cognitively less demanding than evolving and memorizing new signals (38), through

296   for example reinforcement learning, on the condition that the informational aspects

297   encoded in the signals are compatible with each other. Further experimental work,

298   particularly natural and artificial playbacks of combinatorial and compositional

299   structures will help shed additional light on the cognitive mechanisms involved in the

300   parsing of call sequences.

301   Ultimately, however, language's generativity is not solely concerned with

302   syntactic constructions but also the flexible concatenation of meaningful signals (37).

303   Distinct signals, or words, can, for example, re-occur freely in various syntactic

304   constructs and when doing so retain their meaning, resulting in signal compounds

305   with overlapping or similar meaning. Whilst here we demonstrate evidence for one

306   compound signal, preliminary data suggests that babblers also flexibly combine

307   recruitment calls with at least two additional, functionally distinct, call types. Besides

308   alert calls, recruitment calls seem to be systematically combined with aerial alarm

309   calls when mobbing large raptors, or with begging calls by dependent offspring when

310   accompanying foraging helpers (see Fig. S1). These preliminary data tentatively

311   suggest that, rather than just memorizing a complex signal, pied babblers apply a

312   general combinatorial rule to encode multiple messages.

313   In conclusion, our work provides evidence for semantically compositional

314   syntax in a social bird. We propose that through studying highly social species with

315   discrete, constrained vocal repertoires, further light can be shed on the variation and

316   distribution of combinatorial mechanisms outside of humans. We predict this will, in

317   turn, help elucidate the evolutionary drivers promoting the emergence of syntactic

318   communication in animals and ultimately humans.

319

320   **Material and Methods**

321   **Study Site and Species.** The study was conducted on a population of wild, free living

322   southern pied babblers at the Pied Babbler Research Project, Kuruman River Reserve

323   in the Kalahari Desert of South Africa (26°58S, 21°49E). The study site is

324   characterised by sparse vegetation and a semi-arid climate (39). The population is part

325   of a long-term research project founded by ARR in 2003. Individuals are habituated to

326   human observers and can be followed at a distance of 1-2m, enabling close

327   observations (21). Coloured rings allow individual identification of all members of the

328   study population (21).

329

330   **General Information.** Natural observations were conducted between January to

331   April/May 2014 and 2015. The rest of the study was performed between February and

332   April 2014. All audio recordings were conducted using a Rode NTG-2 directional

333   microphone (sampling frequency 48 kHz, 24-bits accuracy) coupled with a Rode

334   blimp suspension windshield and a Roland R-26 portable recorder (Roland

335   Corporation, Japan).

336

337   **Acoustic Analysis.** In order to verify that mobbing sequences are composed of two

338   distinct call types, i.e. alert and recruitment calls, we conducted acoustic analyses. To

339   avoid erroneous p-value estimation associated with pseudo-replication, we only took

340   one alert, and one recruitment call per individual totalling 32 calls from 16 different

341   individuals belonging to 16 different groups (40). Calls were initially inspected and

342   assessed for quality (signal-to-noise ratio), and both calls that were produced as part

343   of a sequence, as well as calls produced in isolation, were included in the analyses.

344   Since most alert calls lack a clear fundamental frequency, calls were compared based

345   on parameters related to the energy distribution. Additionally, we assessed the

346   percentage of the call that exhibited clear, tonal structures (i.e. did not exhibit noise or

347   deterministic chaos) (41). The following acoustic measurements were recorded: call

348   duration, 25%-, 50%- & 75%-energy quartiles, peak frequency, effective peak

349   frequency, percentage of effective peak frequency, relative time of maximum

350   intensity, amplitude variation, amplitude rate, shimmer, and percentage of voiced

351   structures in the first and second half of the call. Except the latter, all parameters were

352   extracted using an automated analysis script in Praat 5.1.03. In order to determine the

353   classification probabilities of calls to call type (alert or recruitment), we first applied a

354   Discriminant Function Analysis (DFA) using SPSS (IBM, version 21.0.0.0).

355   Depending on the number of groups to be classified, the model creates one or more

356   discriminant functions by identifying linear combinations of the predictor variables

357   that best describe the discrimination between groups (42). A leave-one-out cross-

358   validation procedure was applied for external validation. A two-tailed binomial test

359   was used to estimate the overall significance of the classification of the DFA, with a

360   corrected level of chance corresponding to the number of categories discriminated

361   (two categories = 50%).

362

363   **Natural Observations**. To quantify the context in which alert calls, recruitment calls

364   and mobbing sequences are produced, natural observations, in combination with audio

365   recordings, were conducted. In 2014 we regularly visited 19 babbler groups with an

366   average group size of 6.2 ±2.3 individuals, and in 2015, 18 groups with an average

367   group size of 5.1 ±1.4 individuals. A specific group was followed in the evening for

368   approximately two hours until the group had settled down in a night roost. The next

369   morning, the group was re-joined at the sleeping roost before dawn and was then

370   followed for around four hours. Whole sessions were audio recorded, and annotated

371   and analysed using Cool Edit 2000 (Syntrillium Software Corporation) or Audition

372   CS6 (Adobe), scanning for relevant events and vocalisations.

373

374   **Predator Presentation Experiment.** Presentation experiments were conducted to

375   verify the context specific production of mobbing sequences when mobbing

376   predators. Ten babbler groups were exposed to a rubber snake simulating an active

377   Cape cobra, with an extended neck. The model was placed in a raised posture along

378   the predicted path of a group either below vegetation, or coiled around the trunk of a

379   tree. The whole procedure was audio recorded and analysed using Audition CS6

380   (Adobe), to determine whether mobbing sequences (i.e. combinations of alert and

381   recruitment calls) were produced.

382

383   **Playback Stimuli.** For the creation of playback stimuli, high signal-to-noise ratio

384   vocalisations of male or female subordinate group members, from each of the test

385   groups, were selected. In one group this was not possible as only vocalisations from a

386   dominant individual could be recorded. Playbacks were created and normalised with

387   Audition CS6 (Adobe, sampling frequency 48 kHz, 24 bits accuracy). In order to test

388   whether the mobbing sequence derives its meaning from the meaning of its individual

389   calls, and to verify behavioural observations suggesting context-specific production of

390   the constituent calls, we played back natural mobbing sequences, as well as alert calls

391   and recruitment calls, on their own, to subjects. To match the natural variation,

392   mobbing sequences were composed of 1-2 alert calls (2 calls in cases where alert calls

393   were particularly short in duration (2 instances)) and 4-7 recruitment calls. Since the

394   response to the mobbing sequence could have been the result of simple priming or

395   stimulus intensity effects, any acoustic element preceding recruitment calls, or

396   equally, any two call types in combination, could have been sufficient to elicit the

397   behavioural change (28, 29). To exclude these possibilities, we created a two-call

398   control chuck-recruitment sequence (see Supporting Online Material: Acoustic

399   Analysis of Chuck, Alert, and Recruitment Calls). This control combination was

400   created by replacing the alert call of the mobbing sequence with a chuck call

401   (contact/close call produced during foraging (21)) of the same individual. The chuck

402   call was therefore normalised to the amplitude of the substituted alert call, and the

403   same inter-element distance between the replaced element and the recruitment call

404   was maintained.

405   To rule out that any urgency-based acoustic information encoded in the

406   naturally occurring sequence might have elicited a mobbing-like response, we created

407   two sets of stimuli versions for the playback experiments. The first set included

408   natural mobbing sequences, the constituent alert and recruitment calls which were

409   played back in isolation, as well as the chuck-recruitment sequence created out of the

410   natural mobbing sequence. The second set included artificially created mobbing

411   sequences, created by combining single alert and recruitment calls (see Supporting

412   Online Material: Stimuli Sets).

413

414   **Playback Protocol and Response Variables.** Stimuli were played back once at a

415   naturally occurring, normalised amplitude (~73dB at 4m distance), using an AN-30

416   Speaker Monitor (Anchor, USA) coupled to an iPod 3 (Apple Inc.). Each of the 16

417   test groups was exposed to all four playback conditions in a randomised order, and

418   only vocalisations of an existing group member were played back. All four treatment

419   conditions were played back in one morning, except for one occasion where one

420   condition had to be played back on a separate day, due to experiment interruption by a

421   predator. The loudspeaker was placed at approximately 30 meters from the target

422   group and was hidden by vegetation. The sound files were uploaded on an iPod,

423   which was controlled via Bluetooth using an iPhone 4 (Apple Inc.) and Tango Remote

424   App (Blue Atlas Technology, LLC). Playbacks were conducted when no individual

425   was on sentinel duty and when no major disturbances had occurred on the morning

426   the playbacks were undertaken. In line with our prediction of mobbing sequences

427   functioning to recruit group members in a dangerous situation, we recorded subjects'

428   vigilance responses, as well as movement patterns. Once the playback started, the

429   proportion of individuals that became vigilant was recorded. Vigilance was classified

430   as scanning the area or looking towards the location from where the stimulus was

431   broadcast. In order to avoid including individuals that simply became attentive in

432   response to an alert group member, only individuals that reacted immediately after the

433   stimulus presentation were counted as vigilant. Additionally, the latency for the first

434   responding bird to resume normal (non-vigilant) behaviour was recorded. In order to

435   evaluate differences in movement behaviour (direction and speed), the distance from

436   the spatial centre of the group to the loudspeaker was recorded at the beginning of the

437   playback and after 1, 2, 3, 4 and 5 minutes, using a handheld GPS logger (Garmin

438   eTrex® 10, Garmin Ltd.) and Garmin® Basecamp® software (Garmin Ltd.), All

439   experiments were video-taped using a Sony Handycam (HDR-CX160). Videos were

440   analysed frame-by-frame using Audition CS6 (Adobe).

441

442   **Statistical Analysis of Playback Experiment.** Statistical analyses were conducted in

443   R (version 3.1.1) (43). For the computation of linear and generalised linear mixed

444   models the packages lme4 (44) and MuMIn (45) were used. Model estimates were

445   plotted using the packages ggplot2 (46) and gtable (47). Model selection was based on

446   Akaike's information criterion corrected for small sample sizes (AICc), with a

447   threshold difference (Δ AICc) of at least two to the next best model (48). If the

448   difference between the model with the lowest AICc and subsequent models was less

449   than two, the influence of each fixed factor on the response variable was assessed for

450   each of the models within the specified range. The best model was then chosen by

451   excluding the model(s) that included non-significant predictor variables. The

452   significance of the fixed effects was assessed based on bootstrapping methods.

453   Therefore, data was simulated on the basis of the null model (best model according to

454   model selection excluding the factor of interest). The full model (best model

455   according to model selection) and the null model were then fitted to the simulated

456   data, and their difference in deviance was calculated. Simulations and model fittings

457   were iterated 10000 times. The same procedure was repeated but in this instance,

458   fitting the actual data to the null and full model. The distribution of differences in

459   deviances obtained with the simulated data and the actual data were then compared by

460   applying a χ^2 test (see also R pbkrtest package (49)). To investigate where the

461   differences between the playback conditions lay, the 95% confidence-intervals of the

462   difference (CI) were compared between each treatment conditions. If the confidence

463   intervals intersected zero, differences were non-significant (48).

464   *Model 1) Proportion of group vigilant.* To test for an effect of the playback

465   type, we fitted a generalised linear mixed model (GLMM) with a binomial error

466   distribution (0-1 = proportion of group vigilant), with number of vigilant individuals

467   representing the response term and group size the binomial denominator. Due to a

468   possible zero-inflation, overdispersion in the model was estimated by counting each

469   variance parameter as one degree of freedom. The data was considered overdispersed

470   if the ratio of the sum of squared Pearson residuals to residual degrees of freedom was

471   greater than one, which was true in our model (50). To correct for this, an

472   observational-level random term was added to the model, by serially numbering each

473   observation (51). Accordingly, model 1 included the treatment-type as a fixed effect,

474   and group-identity and the observation-level as random effects.

475   *Model 2) Latency to resume normal behaviour*. To examine whether the

476   playback condition had an effect on the latency to resume normal, non-vigilant

477   behaviour of the first reacting bird, we fitted a LMM (linear mixed model) with

478   treatment-type as a fixed effect and group-identity as a random effect. In order to

479   achieve a normal distribution the data were log-transformed.

480   *Model 3) Movement behaviour*. In order to investigate differences in

481   movement behaviour over time between the playback conditions, a group's distance

482   to the sound source was recorded at fixed time intervals. Once a group had passed the

483   loudspeaker and continued moving in the direction from where they originally heard

484   the stimuli, negative values for the distance to the speaker were assigned. A LMM

485   was fitted with treatment-type, time, and its interaction term as fixed effects, and

486   group-identity as a random effect. To achieve a normal distribution, the data were log-

487   transformed, with a constant value being added to the response variable to avoid

488   transformation of negative values (i.e. $\log(x+200)$) (52).

489

505   **Ethical Statement.** The study was performed under the permission of the ethical

506   committee for animal research, University of Cape Town and the Northern Cape

507   Conservation Authority, South Africa.

508   **Data accessibility.** Supporting data can be found in the online material.

509

510   **References**

511   1. Hurford J (2011) Linguistics from an Evolutionary Point of View. *Handbook of*
512   *the Philosophy of Science: Linguistics*, eds Kempson R, Fernando M, & Ashe N
513   (Elsevier Science, Amsterdam), Vol 14, pp 473-498.

514   2. Chomsky N (1981) Knowledge of Language: Its Elements and Origins. *Philos T R*
515   *Soc B* 295(1077):223-234.

516   3. Humboldt Wv (1999) On Language. Translated by Peter Heath. (Cambridge
517   University Press, Cambridge).

518   4. Hauser MD, Chomsky N, & Fitch WT (2002) The Faculty of Language: What Is
519   It, Who Has It, and How Did It Evolve? *Science* 298(5598):1569-1579.

520   5. Christiansen MH & Kirby S (2003) Language evolution: consensus and
521   controversies. *Trends Cogn Sci* 7(7):300-307.

522   6. Nowak MA, Plotkin JB, & Jansen VAA (2000) The evolution of syntactic
523   communication. *Nature* 404(6777):495-498.

524   7. Harvey PH & Pagel MD (1991) The Comparative Method in Evolutionary
525   Biology. (Oxford University Press, Oxford).

526   8. Ouattara K, Lemasson A, & Zuberbühler K (2009) Campbell's Monkeys Use
527   Affixation to Alter Call Meaning. *PLoS ONE* 4(11):e7808.

528   9. Arnold K & Zuberbühler K (2006) Language evolution: Semantic combinations in
529   primate calls. *Nature* 441(7091):303.

530   10. Arnold K & Zuberbühler K (2008) Meaningful call combinations in a non-human
531   primate. *Curr Biol* 18(5):R202-R203.

532   11. Arnold K & Zuberbühler K (2012) Call combinations in monkeys: Compositional
533   or idiomatic expressions? *Brain Lang* 120(3):303-309.

534   12. Collier K, Bickel B, van Schaik CP, Manser MB, & Townsend SW (2014)
535   Language evolution: syntax before phonology? *Proc R Soc B*
536   281(1788):20140263.

537   13. Zuberbühler K (2015) Linguistic capacity of non-human animals. *WIREs Cogn*
538   *Sci* 6(3):313-321.

539   14. Catchpole CK & Slater TLB (1995) Bird Song: Themes and variations.
540   (Cambridge University Press, New York).

541   15. Berwick RC, Okanoya K, Beckers GlJL, & Bolhuis JJ (2011) Songs to syntax: the
542   linguistics of birdsong. *Trends Cogn Sci* 15(3):113-121.

543   16. Rendall D (2013) Q&A: Cognitive ethology - inside the minds of other species.
544   *BMC Biol* 11(108):1-6.

545   17. Raihani NJ (2008) Cooperation and conflict in pied babblers. (University of
546   Cambridge, PhD).

547   18. Golabek KA (2010) Vocal communication and the facilitation of social behaviour
548   in the southern pied babbler (Turdoides bicolor). (University of Bristol, PhD).

549   19. Ridley AR & Raihani NJ (2007) Variable postfledging care in a cooperative bird:
550   causes and consequences. *Behav Ecol* 18(6):994-1000.

551   20. Nelson-Flower MJ, Hockey PAR, O'Ryan C, Raihani NJ, & du Plessis MA (2011)
552   Monogamous dominant pairs monopolize reproduction in the cooperatively
553   breeding pied babbler. *Behav Ecol* 22(3):559-565.

554   21. Radford AN & Ridley AR (2008) Close calling regulates spacing between
555   foraging competitors in the group-living pied babbler. *Anim Behav* 75(2):519-527.

556 22. Bell MBV, Radford AN, Rose R, Wade HM, & Ridley AR (2009) The value of
557   constant surveillance in a risky environment. *Philos T R Soc B* 276(1669):2997-
558   3005.

559 23. Hollén LI, Bell MBV, & Radford AN (2008) Cooperative Sentinel Calling?
560   Foragers Gain Increased Biomass Intake. *Curr Biol* 18(8):576-579.

561 24. Radford AN & Ridley AR (2006) Recruitment Calling: A Novel Form of
562   Extended Parental Care in an Altricial Species. *Curr Biol* 16(17):1700-1704.

563 25. Radford AN & Ridley AR (2007) Individuals in foraging groups may use vocal
564   cues when assessing their need for anti-predator vigilance. *Biol Lett* 3(3):249-252.

565 26. Golabek KA & Radford AN (2013) Chorus-call classification in the southern pied
566   babbler: multiple call types given in overlapping contexts. *Behaviour* 150(7):691-
567   712.

568 27. Hurford J (2011) The origins of grammar. (Oxford University Press, Oxford).

569 28. Engesser S, Crane JM, Savage JL, Russell AF, & Townsend SW (2015)
570   Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System.
571   *PLoS Biol* 13(6):e1002171.

572 29. Slocombe KE, Kaller T, Call J, & Zuberbuhler K (2010) Chimpanzees Extract
573   Social Information from Agonistic Screams. *PLoS ONE* 5(7):e11473.

574 30. Coye C, Ouattara K, Zuberbühler K, & Lemasson A (2015) Suffixation influences
575   receivers' behaviour in non-human primates. *Proc R Soc B* 282(1807):20150265.

576 31. Scott-Phillips TC, Gurney J, Ivens A, Diggle SP, & Popat R (2014) Combinatorial
577   communication in bacteria: implications for the origins of linguistic generativity.
578   *PLoS ONE* 9(4):e95929.

579 32. Berwick RC (2011) All you need is merge: biology, computation, and language
580   from the bottom-up. *The Biolinguistic Enterprise*, eds Sciullo AMD & Boeckx C
581   (The MIT Press, Cambridge, MA), pp 706–825.

582 33. Jackendoff R (1999) Possible stages in the evolution of the language capacity.
583   *Trends Cogn Sci* 3(7):272-279.

584 34. Progovac L (2015) Evolutionary Syntax. (Oxford University Press, Oxford).

585 35. Tomasello M (2003) Constructing a language: a usage-based theory of language
586   acquisition. (Harvard Univerity Press, Cambridge, MA).

587 36. Senghas A, Kita S, & Ozyurek A (2004) Children creating core properties of
588   language: evidence from an emerging sign language in Nicaragua. *Science*
589   305(5691):1779-1782.

590 37. Nowak MA & Komarova NL (2001) Towards an evolutionary theory of language.
591   *Trends Cogn Sci* 5(7):288-295.

592   38. Nowak MA & Krakauer DC (1999) The evolution of language. *PNAS*
593   96(14):8028-8033.

594   39. Clutton-Brock TH*, et al.* (1998) Costs of cooperative behaviour in suricates
595   (Suricata suricatta). *Proc R Soc B* 265(1392):185-190.

596   40. Sharp SP, McGowan A, Wood MJ, & Hatchwell BJ (2005) Learned kin
597   recognition cues in a social bird. *Nature* 434(7037):1127-1130.

598   41. Fitch WT, Neubauer J, & Herzel H (2002) Calls out of chaos: the adaptive
599   significance of nonlinear phenomena in mammalian vocal production. *Anim*
600   *Behav* 63(3):407-418.

601   42. Mundry R & Sommer C (2007) Discriminant function analysis with
602   nonindependent data: consequences and an alternative. *Anim Behav* 74(4):965-
603   976.

604   43. R-Core-Team (2014) R: A language and environment for statistical computing. R
605   Foundation for Statistical Computing, Vienna, Austria.

606   44. Bates D, Maechler M, Bolker B, & Walker S (2014) lme4: _lme4: Linear mixed-
607   effects models using Eigen and S4_. R package version 1.1-8.

608   45. Barto'n K (2014) MuMIn: Multi-model inference. R package version 1.10.5.

609   46. Wickham H (2009) ggplot2: elegant graphics for data analysis. (Springer, New
610   York).

611   47. Wickham H (2012) gtable: Arrange grobs in tables. R package version 0.1.2.

612   48. Burnham KP & Anderson DR (2002) Model selection and multimodel inference:
613   a practical information-theoretic approach. (Springer, New York), 2nd Ed.

614   49. Halekoh U & Höjsgaard S (2014) pbkrtest : A Kenward-Roger Approximation
615   and Parametric Bootstrap Methods for Tests in Linear Mixed Models. R package
616   version 0.4-5.

617   50. Zuur A, Ieno EN, Walker N, Saveliev AA, & Smith GM (2009) Mixed Effects
618   Models and Extensions in Ecology with R. (Springer, New York).

619   51. Harrison XA (2014) Using observation-level random effects to model
620   overdispersion in count data in ecology and evolution. *PeerJ* 2:e616.

621   52. LeBlanc DC (2004) Statistics: Concepts and Applications for Science. (Jones and
622   Bartlett Publishers, Mississauge).
623

624   **Figure Legends**

625   Figure 1: Spectrogram of a mobbing sequence composed of one alert and seven

626   recruitment calls.

627

628   Figure 2: Proportion of group vigilant (grey illustration) and latency to resume normal

629   behaviour of the first reacting bird (red illustration). Playback treatments:

630   M=mobbing sequences, A=alert calls, R=recruitment calls, CR=chuck-recruitment

631   sequences. Asterisks indicate significant differences according to the 95% confidence

632   intervals of the difference. Bars illustrate the 95% confidence-intervals and points the

633   median over 16 groups of the back-transformed data. Pale dots show the raw data.

634

635   Figure 3: Group's distance to the loudspeaker at the beginning and 1, 2, 3, 4 and 5

636   minutes after the playback start. Values over 30 metres indicate a retreat from the

637   loudspeaker. Negative values indicate that a group had passed the loudspeaker and

638   continued moving in the same direction from where they originally heard the

639   playback stimuli. Bars illustrate the 95% confidence-intervals and points the median

640   over 16 groups of the back-transformed data. Pale dots show the raw data.