Meaningful call combinations and compositional processing in the southern pied babbler

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

[a]Pied Babbler Research Project, Kuruman River Reserve, Northern Cape 8467, South Africa; [b]Animal Behaviour, Department of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zurich, Switzerland; [c]Centre for Evolutionary Biology, School of Animal Biology, The University of Western Australia, Crawley, WA 6009, Australia; [d]Percy FitzPatrick Institute, University of Cape Town, Rondebosch 7701, South Africa; and [e]Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom

Language's expressive power is largely attributable to its compositionality: meaningful words are combined into larger/higher-order structures with derived meaning. Despite its importance, little is known regarding the evolutionary origins and emergence of this syntactic ability. Although previous research has shown a rudimentary capability to combine meaningful calls in primates, because of a scarcity of comparative data, it is unclear to what extent analog forms might also exist outside of primates. Here, we address this ambiguity and provide evidence for rudimentary compositionality in the discrete vocal system of a social passerine, the pied babbler (*Turdoides bicolor*). Natural observations and predator presentations revealed that babblers produce acoustically distinct alert calls in response to close, low-urgency threats and recruitment calls when recruiting group members during locomotion. On encountering terrestrial predators, both vocalizations are combined into a "mobbing sequence," potentially to recruit group members in a dangerous situation. To investigate whether babblers process the sequence in a compositional way, we conducted systematic experiments, playing back the individual calls in isolation as well as naturally occurring and artificial sequences. Babblers reacted most strongly to mobbing sequence playbacks, showing a greater attentiveness and a quicker approach to the loudspeaker, compared with individual calls or control sequences. We conclude that the sequence constitutes a compositional structure, communicating information on both the context and the requested action. Our work supports previous research suggesting combinatoriality as a viable mechanism to increase communicative output and indicates that the ability to combine and process meaningful vocal structures, a basic syntax, may be more widespread than previously thought.

call combination | compositionality | syntax | language evolution | southern pied babbler

Syntax is often considered one of the key defining features of human language (1). Through combining meaningful words together, larger sequences with related, compositional meaning can be constructed (2). One consequence of such productive compositional syntax in humans is that, with a finite inventory of words, an infinite range of ideas and concepts can be communicated (2, 3). Despite the central role that syntax plays in determining language's generativity, very little is known about its evolutionary origins or early, rudimentary forms (4, 5). Elucidating the proto forms of compositional syntax, although nontrivial (5, 6), represents a key step in understanding the evolution of language more holistically.

One means of investigating early forms and function of compositionality is to assess analog examples in animals (5, 7). Indeed, recent observational and experimental work on two related guenon monkeys has shown the propensity to combine context-specific, "meaningful" signals into sequences that resemble compositional structures in language (8–10). Male Campbell's monkeys (*Cercopithecus campbelli*), for example, produce predator-specific alarm calls that can be affixed with an acoustic modifier (8, 11). The affix acts to alter the "meaning" of the alarm calls in a predictable

way, transforming them into general disturbance calls (8, 11, 12). Similarly, male putty-nosed monkeys (*Cercopithecus nictitans*) combine two predator-specific alarm calls into a higher-order sequence (9, 13). Although the two calls are generally associated with the presence of aerial or terrestrial predators, the resultant combination initiates group movement in nonpredatory contexts (9, 13). Given the discrepancies between the responses elicited by the individual calls and the sequence, it remains unclear whether the putty-nosed monkey call sequence represents a basic form of compositional syntax or rather a combinatorial syntax, where the meaning of the whole is not directly related to the parts, akin to idiomatic expressions in language (i.e., "kick the bucket" for dying) (10, 13, 14). The existence of such "semantic combinations" (13) in primates has nevertheless been argued to support an evolutionarily ancient origin of human syntax rooted within the primate lineage (8, 15). However, it is unclear whether similar call concatenations and compositional processing of information might also exist in other lineages (see ref. 14 for review) and if so, whether they take analogous forms and serve analogous functions (1).

The last 50 y of comparative research have shown that a number of nonprimate animals, particularly songbirds, are capable of stringing sounds together into larger, often more structurally complex sequences (16–18). However, there is no indication that any of these song sequences are compositional in structure, because the individual sounds composing the songs of birds and other animals do not convey any independent meaning (16–18), ultimately precluding any attempt to test for proto-syntactic abilities in these species in the first place. Although the

Significance

Human language is syntactic in its nature: meaningful words are assembled into larger meaningful phrases or sentences. How unique this ability is to humans remains surprisingly unclear. A considerable body of work has indicated that birds are capable of combining sounds into large, elaborate songs, but there is currently no evidence suggesting that these structures are syntactic. Here, we provide important evidence for this ability in a highly social bird. Specifically, pied babblers combine two functionally distinct vocalizations into a larger sequence, the function of which is related to the function of its parts. Our work adds important evidence to the variation and distribution of combinatorial vocal mechanisms outside humans and provides insights into potentially early forms of human syntactic communication.

absence of compositional structure in songs might suggest that syntactic abilities are potentially confined to the primate lineage (8, 15), it may also be an artifact of limited focus on bird vocal systems other than song that are more likely to support the capacity for syntax (19).

Here, we address this ambiguity through investigating the prevalence of compositional vocal sequences in a highly social, nonsinging passerine bird that possesses a discrete vocal system: the cooperatively breeding southern pied babbler (*Turdoides bicolor*) (20, 21). Pied babblers are territorial and live in stable groups of 3–15 individuals (22). Reproduction is usually restricted to the dominant pair of the group (23), with subordinate individuals engaging in a number of helping behaviors, such as territorial and nest defense, daytime incubation, and feeding of the offspring during the nestling and postfledgling stages (22). Individuals of the cohesive foraging group spend most of the time on the ground searching for invertebrates hidden in the substrate, which they excavate using their bill (22, 24). Consequently, most of the time, pied babblers forage in a head-down position within and around forbs and shrubs and hence, rely heavily on vocalizations to keep track of changes in their surroundings (21, 25–29). As such, the pied babbler vocal system exhibits around 17 discrete vocalizations, including alarm calls and sentinel calls, as well as a diverse array of social calls produced during intra- and intergroup contexts (21, 25–29).

Observational work has indicated that pied babblers produce broadband, noisy alert calls in response to sudden but generally low-urgency threats (e.g., abruptly approaching animals) and more tonal, repetitive recruitment calls when recruiting group members to a new location or during locomotion, mainly in foraging or roosting contexts. Moreover, alert and recruitment calls can be combined into a sequence on encountering and mobbing, mainly terrestrial, predators (Fig. 1). Given the context in which the two independent calls are produced, we aimed to investigate whether the sequence might, therefore, function specifically to recruit group members in a dangerous situation (e.g., when mobbing a predator) by combining information on both the danger and the requested action. Accordingly, the combination of alert and recruitment calls (hereafter termed the "mobbing sequence") might constitute a rudimentary compositional structure, where the meaning of the whole is a product of the meaning of its parts (30).

To verify the context-specific information conveyed by the independent vocalizations and test whether pied babblers extract the meaning of the sequence in a compositional way, we conducted additional natural observations combined with acoustic analyses and experimental manipulations. First, acoustic analyses were applied to confirm that alert and recruitment calls constitute two distinct vocalizations. Second, to determine the contexts in which the individual calls and the call sequence are produced, we conducted natural observations and predator presentation experiments co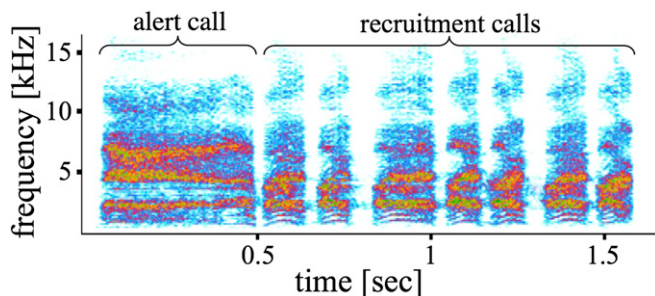mbined with audio recordings. Third, we carried out systematic natural, artificial, and control playback experiments to investigate whether birds perceive the sequence compositionally. Key support for compositionality requires that the context in which mobbing sequences are produced and the responses of receivers to playbacks of these sequences are related to the information encoded in alert and recruitment calls (30, 31).

Results

Acoustic Analysis. A discriminant function analysis (DFA) indicated that alert and recruitment calls could be statistically discriminated based on their structure alone ($N_{individuals} = 16$; $N_{calls} = 32$; correct classification: 97%; $P < 0.001$). When applying a leave one out cross-validated DFA, 94% were correctly classified, a classification higher than expected by chance (two-tailed binomial test; change level = 50%; $P < 0.001$).

Alert and Recruitment Calls: Natural Context. Natural observations combined with acoustic recordings were conducted to quantify the calls' context specificity. From a total of 36 alert calls recorded in 11 groups, 69% were elicited by suddenly appearing, nondangerous subjects (e.g., hares, antelopes, and researchers); 14% of alert calls were caused by inactive snakes or distant mongooses or foxes that did not present a direct threat to babblers. Another 6% were produced in response to alarm calls of con- or heterospecifics. For the remaining 11% of occasions, no obvious threat could be detected.

From a total of 196 recorded recruitment call events from 71 individuals in 20 groups, 60% resulted in other group members approaching the caller, and 6% resulted in overall group movement following the caller. In the remaining 34%, recipients either showed no response (44 of 67 occasions) or countercalled with recruitment or other loud calls (23 of 67 occasions) (29). All recorded recruitment calls were produced in nondangerous contexts in the absence of any predators. Thus, alert calls seem to encode information about low-urgency threats in a caller's imminent surrounding, and recruitment calls seem to function to recruit group members to a caller's current location.

Mobbing Sequences: Natural Context and Experimental Elicitation. We observed naturally elicited mobbing sequences on 39 occasions in 14 groups: 85% were produced in response to moving terrestrial predators (mongooses, snakes, or foxes), and 8% were produced in response to small perched raptors [pygmy falcon (*Polihierax semitorquatu*) and pearl spotted owl (*Glaucidium perlatum*)], which are assumed to only pose a threat to young, inexperienced babblers. In the remaining 8% of events, no clear context could be assigned. To experimentally confirm the context accompanying the production of mobbing sequences, babbler groups were presented with a model of a Cape cobra (*Naja nivea*), and their calling behavior was noted. From a total of 13 presentations in 10 groups, mobbing sequences were elicited 92% of the time.

Playback Experiment. To investigate the responses to mobbing sequences and their individual calls, we played back natural mobbing sequences as well as the constituent alert and recruitment calls to subjects. To rule out alternative explanations associated with the saliency of the stimulus (two vs. one call type) or priming effects (any call type preceding recruitment calls generates the same response), we implemented an additional important control condition, where we artificially replaced the alert call of a mobbing sequence with another acoustically distinct broadband babbler vocalization: the foraging "chuck" call (chuck recruitment sequence) (*Supporting Information*) (24, 32, 33). Finally, in line with previous studies (12, 13), to ensure that the key dimension for receivers was the combination of information and not any urgency-based acoustic variation encoded across the structure, as an additional control, artificial mobbing sequences

Fig. 1. Spectrogram of a mobbing sequence composed of one alert and seven recruitment calls.

EVOLUTION
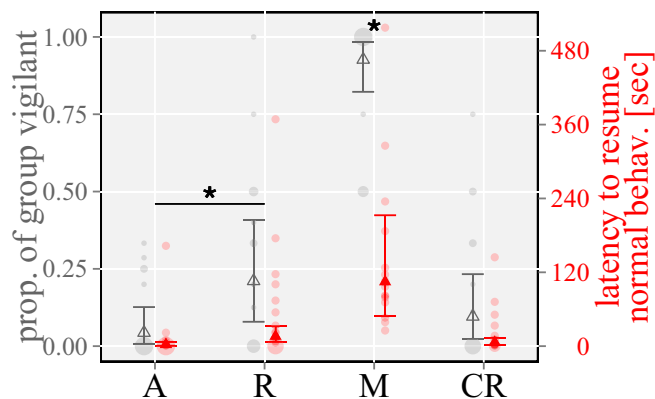
Fig. 2. Proportion of group vigilant (gray) and latency to resume normal behavior of the first reacting bird (red). Playback treatments: A, alert calls; CR, chuck recruitment sequences; M, mobbing sequences; R, recruitment calls. Bars illustrate the 95% confidence intervals, and points show the median over 16 groups of the back-transformed data. Pale circles show the raw data. *Significant differences according to the 95% confidence intervals of the difference.

were constructed from the independent calls and played back (*Supporting Information* and Table S1) (13, 32).

Our playbacks revealed differences in group attentiveness responses to the four playback conditions determined by the proportion of the group that became vigilant (treatment: $\chi^2 = 53.5$; $P < 0.01$; $n = 64$; 16 groups) (Fig. 2 and Table 1) and the latency to resume normal, nonvigilant behavior of the first reacting group member (treatment: $\chi^2 = 36.3$; $P < 0.001$; $n = 64$; 16 groups) (Fig. 2 and Table 1). Moreover, the movement patterns of a group relative to the sound source differed in response to the four stimuli (treatment: $\chi^2 = 97.2$; time: $\chi^2 = 34.9$; treatment × time: $\chi^2 = 23.6$; all $P < 0.001$; $n = 378$; 16 groups) (Fig. 3 and Table 1).

Specifically, alert calls played back in isolation did not result in noticeable changes in behavior (such as attentiveness) (Fig. 2, Table 1, and Table S2), and we found no effect of time on distance moved, with groups neither approaching nor retreating from the sound source (Fig. 3, Table 1, and Table S2). In accordance with the assumed function to recruit group members to a caller's location, in response to played back recruitment calls, babblers increased their attentiveness compared with playbacks of alert calls, likely as a way to locate the simulated recruiting caller, and slowly and steadily approached the sound source (Figs. 1 and 2, Table 1, and Table S2). Furthermore, in line with our central prediction of mobbing sequences functioning to

recruit group members in a dangerous situation, we found that subjects responded most strongly to playbacks of mobbing sequences, revealing the highest attentiveness and fastest approach toward the sound source (Figs. 2 and 3, Table 1, and Table S2). Ruling out priming or stimulus effects, playbacks of chuck recruitment control sequences did not elicit similar mobbing-like behaviors, with babblers neither approaching the sound source nor increasing their attentiveness compared with playbacks of mobbing sequences (Figs. 2 and 3, Table 1, and Table S2). These results support our hypothesis that the call sequence tested conforms to the definition of basic compositional syntax (1, 14), with the high vigilance response to mobbing sequences and the fast approach to the loudspeaker being directly related to the contextual information and function of both individual calls.

Discussion

Here, we provide key comparative data indicating that the cooperatively breeding pied babbler can extract rudimentary compositional information from combinations of acoustically distinct, context-specific vocalizations: alert and recruitment calls.

Systematic observational and experimental data implementing both natural and artificial playback experiments show that pied babbler alert calls encode information on existing or imminent low-urgency threats in the environment, whereas recruitment calls communicate the motivation to recruit group members to the caller's location. Combinations of these alert and recruitments calls, here called mobbing sequences, are produced when babblers encounter and mob predominantly terrestrial threats. In response to played back mobbing sequences, babblers reacted with an increased attentiveness (high proportion of the group being vigilant and long latency to resume nonvigilant behavior) and a rapid approach toward the sound source, potentially to support the simulated caller opposing the putative threat. The context accompanying the mobbing sequence, and particularly the responses to the playbacks, suggest that the information encoded in the combination is a direct product of the constituent calls (30). We are confident that we can rule out alternative explanations related to a sequential or additive processing of calls, because responses to played back mobbing sequences exceeded those elicited by the independent calls or their sum (33, 34). Furthermore, control experiments showed that potential superstimuli (two calls vs. one call) or simple priming effects that could otherwise explain the results can be excluded, because control sequences failed to elicit similar mobbing-like behavior (32, 33). In summary, our natural observations combined with the experimental manipulations indicate that babblers produce

Table 1. Effect of playback treatments on the three response variables

Model	AICc	df	Fixed effects	χ^2	P value	N
Proportion of group vigilant						
Basic = best model: treatment	156.9	6	Treatment	53.5	<0.01	64
Intercept model	203.3	3				
Latency to resume normal behavior						
Basic = best model: treatment	242.9	6	Treatment	36.3	<0.001	64
Intercept model	272.2	3				
Movement behavior						
Basic = best model: treatment + time + treatment × time	149.0	10	Treatment	97.2	<0.001	
			Time	34.9	<0.001	378
			Treatment × time	23.6	<0.001	
Intercept model	241.2	3				

Basic models included all fixed and random effects, intercept models included only the random effects, and best models included only the significant fixed effects based on AICc selection as well as the random effects. Listed fixed effects represent the test statistics of the significant factors according to the model selection. Playback experiments were conducted on 16 groups, with each group receiving all playback treatments.
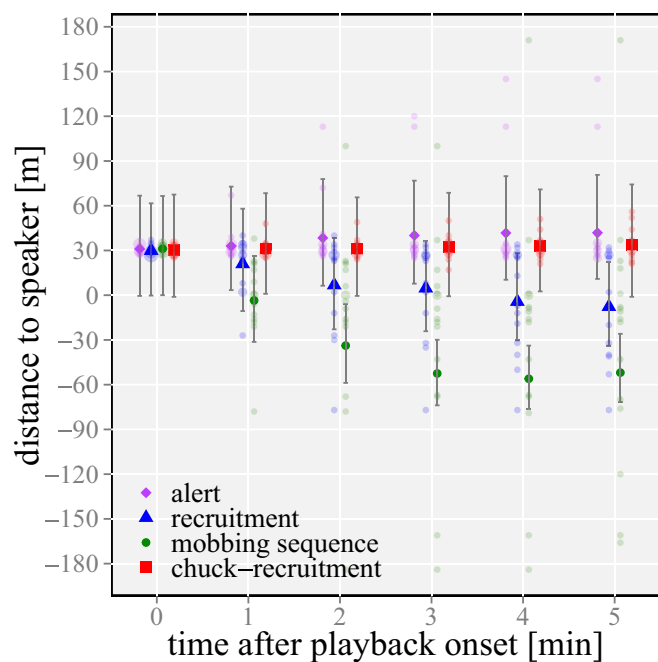
Fig. 3. Group's distance to the loudspeaker at the beginning and 1, 2, 3, 4, and 5 min after the playback start. Values over 30 m indicate a retreat from the loudspeaker. Negative values indicate that a group had passed the loudspeaker and continued moving in the same direction from where they originally heard the playback stimuli. Bars illustrate the 95% confidence intervals, and points show the median over 16 groups of the back-transformed data. Pale circles show the raw data.

and parse the sequence by linking information on the context (threat) and the requested action.

Our work, providing strong evidence for rudimentary compositionality in pied babblers, complements and extends previous research showing similar semantic combinations in primates and suggests that the basic capacity to combine meaningful calls into systematic higher-order structures may be more diverse and widespread than previously thought (8, 10, 14). Furthermore, these findings have important implications for understanding the evolutionary progression of human language. One dominant hypothesis posits that language's hierarchical syntactic system could have only evolved as part of a sudden evolutionary event, precluding the existence of intermediate protosyntactic forms (35). Alternatively, it has been suggested that syntax can be decomposed into more primitive layers, consisting of loose two- or few-word compounds that form the evolutionary and structural bases of syntactic systems (30, 36–38). Under this scenario, a sudden evolutionary leap is not necessary (30), because instead, language's syntactic complexity is hypothesized to have originally emerged out of simple but communicatively meaningful compositions. Support for this hypothesis can be found in language acquisition and newly emerging sign languages, where syntactic development initiates with simple two-word/sign compositions or "packages" (30, 38, 39) and gradually proceeds, in later stages, to more sophisticated multipackage compositions (30). Through providing comparative data for such two-signal constructs in the pied babbler vocal repertoire, our work contributes additional evidence that basic, intermediate compositional structures are viable, and hence, it supports the idea that syntax could have evolved by progressing gradually over time rather than spontaneously as an "all-or-nothing" package (37).

Exactly what evolutionary forces accompanied the progression of syntax remain elusive. Theoretical work conducted over the last two decades has aimed to disentangle the selective conditions

promoting the emergence of syntax (6, 40, 41). Specifically, mathematical modeling approaches have indicated that natural selection will favor a transition toward a syntactic communication system (from a nonsyntactic one) when the number of relevant events to be communicated exceeds the number of available calls (because of either production or perception constraints) (6, 41). Our work provides relevant empirical evidence that supports this claim. Given the pied babblers' constrained vocal repertoire paired with the extensive number of social and ecological contexts that require communication (22), compositional production and processing of vocalizations are likely adaptive for pied babblers, allowing them to coordinate key additional events than would be possible with a nonsyntactic system. Moreover, combining and processing signals in a compositional way may be cognitively less demanding than evolving and memorizing new signals (41) through, for example, reinforcement learning, on the condition that the informational aspects encoded in the signals are compatible with each other. Additional experimental work, particularly natural and artificial playbacks of combinatorial and compositional structures, including temporal manipulations of the two call types, will help to shed additional light on the cognitive mechanisms involved in the parsing of call sequences.

Ultimately, however, language's generativity is not solely concerned with syntactic constructions but also, the flexible and productive concatenation of meaningful signals (40). Distinct signals or words can, for example, reoccur freely in various syntactic constructs, in a myriad of ways, and when doing so, they retain their meaning, resulting in signal compounds with overlapping or similar meaning. Although here, we show evidence for one compound signal, preliminary data suggest that babblers also flexibly combine recruitment calls with at least two additional, functionally distinct call types. Other than alert calls, recruitment calls seem to be systematically combined with aerial alarm calls when mobbing large raptors or begging calls by dependent offspring when accompanying foraging helpers (Fig. S1). These preliminary data tentatively suggest that, rather than just memorizing a complex signal, pied babblers may apply a general combinatorial rule to encode multiple messages.

In conclusion, our work provides evidence for rudimentary compositional syntax in a social bird. We propose that, through studying highly social species with discrete, constrained vocal repertoires, additional light can be shed on the variation and distribution of combinatorial mechanisms outside of humans. We predict that comparative work will, in turn, help elucidate the evolutionary drivers promoting the emergence of syntactic communication in animals and ultimately, humans.

Materials and Methods

Study Site and Species. The study was conducted on a population of wild, free living southern pied babblers at the Pied Babbler Research Project, Kuruman River Reserve in the Kalahari Desert of South Africa (26°58′S, 21°49′E). The study site is characterized by sparse vegetation and a semiarid climate (42). The population is part of a long-term research project founded by A.R.R. in 2003. Individuals are habituated to human observers and can be followed at a distance of 1–2 m, enabling close observations (24). Colored rings allow individual identification of all members of the study population (24).

General Information. Natural observations were conducted from January to April/May 2014 and 2015. The rest of the study was performed between February and April, 2014. All audio recordings were conducted using a Rode NTG-2 Directional Microphone (sampling frequency of 48 kHz; 24-bits accuracy) coupled with a Rode Blimp Suspension Windshield (Rode Microphones) and a Roland R-26 Portable Recorder (Roland Corporation). The study was performed under the permission of the ethical committee for animal research of the University of Cape Town and the Northern Cape Conservation Authority, South Africa.

Acoustic Analysis. To verify that mobbing sequences are composed of two structurally distinct call types (i.e., alert and recruitment calls), we conducted basic acoustic analyses. To avoid erroneous *P* value estimation associated

with pseudoreplication, we only took 1 alert and 1 recruitment call per individual per group, totaling 32 calls from 16 different individuals belonging to 16 different groups (43). Calls were initially inspected and assessed for quality (signal to noise ratio), and both calls that were produced as part of a sequence as well as calls produced in isolation were included in the analyses. Because alert calls sometimes lacked a clear fundamental frequency, calls were compared based on parameters related more to time and energy distribution. Additionally, we assessed the percentage of the call that exhibited clear, tonal structures (i.e., did not exhibit noise or deterministic chaos) (44). The following acoustic measurements were recorded: call duration; 25%, 50%, and 75% energy quartiles; relative time of maximum intensity; amplitude variation; amplitude rate; shimmer; and percentage of voiced structures in the first and second halves of the call. Except the latter, all parameters were extracted using an automated, custom-built analysis script in Praat 5.1.03. To determine the classification probabilities of calls to call type (alert or recruitment), we first applied a DFA using SPSS (version 21.0; IBM). Only parameters with a variance inflation factor smaller than 10 were included in the analysis. Depending on the number of groups to be classified, the DFA creates one or more discriminant functions by identifying linear combinations of the predictor variables that best describe the discrimination between groups (45). A leave one out cross-validation procedure was applied for external validation. A two-tailed binomial test was used to estimate the overall significance of the classification of the DFA, with a corrected level of chance corresponding to the number of categories discriminated (two categories = 50%).

Natural Observations. To quantify the context in which alert calls, recruitment calls, and mobbing sequences are produced, natural observations combined with audio recordings were conducted. In 2014, we regularly visited 19 babbler groups with an average group size of 6.2 ± 2.3 individuals, and in 2015, we regularly visited 18 groups with an average group size of 5.1 ± 1.4 individuals. A specific group was followed in the evening for approximately 2 h until the group had settled down in a night roost. The next morning, the group was rejoined at the sleeping roost before dawn and then followed for around 4 h. Whole sessions were audio recorded, and they were annotated and analyzed using Cool Edit 2000 (Syntrillium Software Corporation) or Audition CS6 (Adobe), scanning for relevant events and vocalizations.

Predator Presentation Experiment. Presentation experiments were conducted to verify the context-specific production of mobbing sequences when mobbing predators. Ten babbler groups were exposed to a rubber snake simulating an active Cape cobra with an extended neck. The model was placed in a raised posture along the predicted path of a group either below vegetation or coiled around the trunk of a tree. The whole procedure was audio recorded and analyzed using Audition CS6 (Adobe) to determine whether mobbing sequences (i.e., combinations of alert and recruitment calls) were produced.

Playback Stimuli. For the creation of playback stimuli, high signal to noise ratio vocalizations of male or female subordinate group members from each of the test groups were selected. However, in one group only vocalizations from a dominant individual could be recorded. Playbacks were created and normalized with Audition CS6 (Adobe; sampling frequency of 48 kHz; 24-bits accuracy). To test whether the mobbing sequence derives its meaning from the meaning of its individual calls and verify behavioral observations suggesting context-specific production of the constituent calls, we played back natural mobbing sequences as well as alert calls and recruitment calls on their own to subjects. Because the response to the mobbing sequence could have been the result of simple priming or stimulus intensity effects, any acoustic element preceding recruitment calls or equally, any two call types in combination could have been sufficient to elicit the behavioral change (32, 33). To exclude these possibilities, we created a two-call control chuck recruitment sequence (*Supporting Information*). This control combination was created by replacing the alert call of the mobbing sequence with a chuck call [contact/close call produced during foraging (24)] of the same individual. The chuck call was, therefore, normalized to the amplitude of the substituted alert call, and the same interelement distance between the replaced element and the recruitment call was maintained.

To rule out that any urgency-based acoustic information encoded in the naturally occurring sequence might have elicited a mobbing-like response, we created two sets of stimuli versions for the playback experiments. The first set included natural mobbing sequences, the constituent alert and recruitment calls that were played back in isolation, as well as the chuck recruitment sequence created out of the natural mobbing sequence. The second set included artificially created mobbing sequences created by synthetically combining alert and recruitment calls (*Supporting Information*).

Playback Protocol and Response Variables. Stimuli were played back once at a naturally occurring, normalized amplitude (~73 dB at a 4 m distance) using an AN-30 Speaker Monitor (Anchor) coupled to an iPod 3 (Apple Inc.). Sound files were uploaded to the iPod which was remote controlled via Bluetooth using an iPhone 4 (Apple Inc.) and Tango Remote App (Blue Atlas Technology, LLC). Each of the 16 test groups was exposed to all four playback conditions in a randomized order, and only vocalizations of an existing group member were played back. All four treatment conditions were played back in one morning, except for in one occasion, when one condition had to be played back on a separate day because of experiment interruption by a predator. The loudspeaker was placed ~30 m from the target group and hidden by vegetation. Playbacks were conducted when no individual was on sentinel duty and when no major disturbances had occurred on the morning that the playbacks were undertaken. In line with our prediction of mobbing sequences functioning to recruit group members in a dangerous situation, we recorded subjects' vigilance responses as well as movement patterns. After the playback started, the proportion of individuals that became vigilant was recorded. Vigilance was classified as scanning the area or looking toward the location from where the stimulus was broadcast. To avoid including individuals that simply became attentive in response to an alert group member, only individuals that reacted immediately after the stimulus presentation were counted as vigilant. Additionally, the latency for the first responding bird to resume normal (nonvigilant) behavior was recorded. To evaluate differences in movement behavior (direction and speed), the distance from the spatial center of the group to the loudspeaker was recorded at the beginning of the playback and after 1, 2, 3, 4, and 5 min using a handheld global positioning system-logger (Garmin eTrex 10; Garmin Ltd.) and Garmin Basecamp software (Garmin Ltd.). All experiments were videotaped using a Sony Handycam (HDR-CX160). Videos were analyzed frame by frame using Audition CS6 (Adobe).

Statistical Analysis of Playback Experiment. Statistical analyses were conducted in R (version 3.1.1) (46). For the computation of linear mixed models (LMMs) and generalized LMMs, the packages lme4 (47) and MuMIn (48) were used. Model estimates were plotted using the packages ggplot2 (49) and gtable (50). Model selection was based on Akaike's information criterion corrected for small sample sizes (AICc), with a threshold difference (ΔAICc) of at least two to the next best model (51). If the difference between the model with the lowest AICc and subsequent models was less than two, the influence of each fixed factor on the response variable was assessed for each of the models within the specified range. The best model was then chosen by excluding the model(s) that included nonsignificant predictor variables. The significance of the fixed effects was assessed based on bootstrapping methods. Therefore, data were simulated based on the null model (best model according to model selection excluding the factor of interest). The full model (best model according to model selection) and the null model were then fitted to the simulated data, and their difference in deviance was calculated. Simulations and model fittings were iterated 10,000 times. The same procedure was repeated, but in this instance, the actual data was fit to the null and full models. The distributions of differences in deviances obtained with the simulated data and the actual data were then compared by applying a χ^2 test [see also R pbkrtest package (52)]. To investigate where the differences between the playback conditions are, the 95% confidence intervals of the difference were compared between each treatment conditions. If the confidence intervals intersected zero, differences were nonsignificant (51).

Model 1: Proportion of group vigilant. To test for an effect of the playback type, we fitted a generalized LMM with a binomial error distribution (0–1 = proportion of group vigilant), with the number of vigilant individuals representing the response term and group size representing the binomial denominator. Because of a possible zero inflation, overdispersion in the model was estimated by counting each variance parameter as one df. The data were considered overdispersed if the ratio of the sum of squared Pearson residuals to residual dfs was greater than one, which was true in our model (53). To correct for overdispersion, an observational-level random term was added to the model by serially numbering each observation (54). Accordingly, model 1 included the treatment type as a fixed effect and the group identity and the observation level as random effects.

Model 2: Latency to resume normal behavior. To examine whether the playback condition had an effect on the latency to resume normal, nonvigilant behavior of the first reacting bird, we fitted an LMM with treatment type as a fixed effect and group identity as a random effect. To achieve a normal distribution, the data were log-transformed.

Model 3: Movement behavior. To investigate differences in movement behavior over time between the playback conditions, a group's distance to the sound source was recorded at fixed time intervals. After a group had passed the loudspeaker and continued moving in the direction from where they

originally heard the stimuli, negative values for the distance to the speaker were assigned. An LMM was fitted with treatment type, time, and its interaction term as fixed effects and group identity as a random effect. To achieve a normal distribution, the data were log-transformed, with a constant value being added to the response variable to avoid transformation of negative values [i.e., $\log(x + 200)$] (55).

1. Hurford J (2011) Linguistics from an evolutionary point of view. Handbook of the Philosophy of Science: Linguistics, eds Kempson R, Fernando M, Ashe N (Elsevier Science, Amsterdam), Vol 14, pp 473–498.
2. Chomsky N (1981) Knowledge of language: Its elements and origins. *Philos Trans R Soc Lond B Biol Sci* 295(1077):223–234.
3. Humboldt Wv, *On Language*, trans Heath P (1999) (Cambridge Univ Press, Cambridge, United Kingdom).
4. Hauser MD, Chomsky N, Fitch WT (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science* 298(5598):1569–1579.
5. Christiansen MH, Kirby S (2003) Language evolution: Consensus and controversies. *Trends Cogn Sci* 7(7):300–307.
6. Nowak MA, Plotkin JB, Jansen VAA (2000) The evolution of syntactic communication. *Nature* 404(6777):495–498.
7. Harvey PH, Pagel MD (1991) *The Comparative Method in Evolutionary Biology* (Oxford Univ Press, Oxford).
8. Ouattara K, Lemasson A, Zuberbühler K (2009) Campbell's monkeys use affixation to alter call meaning. *PLoS One* 4(11):e7808.
9. Arnold K, Zuberbühler K (2008) Meaningful call combinations in a non-human primate. *Curr Biol* 18(5):R202–R203.
10. Arnold K, Zuberbühler K (2012) Call combinations in monkeys: Compositional or idiomatic expressions? *Brain Lang* 120(3):303–309.
11. Schlenker P, et al. (2014) Monkey semantics: Two 'dialects' of Campbell's monkey alarm calls. *Linguist and Philos* 37(6):439–501.
12. Coye C, Ouattara K, Zuberbühler K, Lemasson A (2015) Suffixation influences receivers' behaviour in non-human primates. *Proc Bio Sci* 282(1807):20150265.
13. Arnold K, Zuberbühler K (2006) Language evolution: Semantic combinations in primate calls. *Nature* 441(7091):303.
14. Collier K, Bickel B, van Schaik CP, Manser MB, Townsend SW (2014) Language evolution: Syntax before phonology? *Proc Biol Sci* 281(1788):20140263.
15. Zuberbühler K (2015) Linguistic capacity of non-human animals. *Wiley Interdiscip Rev Cogn Sci* 6(3):313–321.
16. Catchpole CK, Slater TLB (1995) *Bird Song: Themes and Variations* (Cambridge Univ Press, New York).
17. Berwick RC, Okanoya K, Beckers GJ, Bolhuis JJ (2011) Songs to syntax: The linguistics of birdsong. *Trends Cogn Sci* 15(3):113–121.
18. Rendall D (2013) Q&A: Cognitive ethology—inside the minds of other species. *BMC Biol* 11(108):108.
19. Suzuki TN, Wheatcroft D, Griesser M (2016) Experimental evidence for compositional syntax in bird calls. *Nat Commun* 7:10986.
20. Raihani NJ (2008) Cooperation and conflict in pied babblers. PhD thesis (University of Cambridge, Cambridge, United Kingdom).
21. Golabek KA (2010) Vocal communication and the facilitation of social behaviour in the southern pied babbler (Turdoides bicolor). PhD thesis (University of Bristol, Bristol, United Kingdom).
22. Ridley AR, Raihani NJ (2007) Variable postfledging care in a cooperative bird: Causes and consequences. *Behav Ecol* 18(6):994–1000.
23. Nelson-Flower MJ, et al. (2011) Monogamous dominant pairs monopolize reproduction in the cooperatively breeding pied babbler. *Behav Ecol* 22(3):559–565.
24. Radford AN, Ridley AR (2008) Close calling regulates spacing between foraging competitors in the group-living pied babbler. *Anim Behav* 75(2):519–527.
25. Bell MB, Radford AN, Rose R, Wade HM, Ridley AR (2009) The value of constant surveillance in a risky environment. *Proc Biol Sci* 276(1669):2997–3005.
26. Hollén LI, Bell MBV, Radford AN (2008) Cooperative sentinel calling? Foragers gain increased biomass intake. *Curr Biol* 18(8):576–579.
27. Radford AN, Ridley AR (2006) Recruitment calling: A novel form of extended parental care in an altricial species. *Curr Biol* 16(17):1700–1704.
28. Radford AN, Ridley AR (2007) Individuals in foraging groups may use vocal cues when assessing their need for anti-predator vigilance. *Biol Lett* 3(3):249–252.
29. Golabek KA, Radford AN (2013) Chorus-call classification in the southern pied babbler: Multiple call types given in overlapping contexts. *Behaviour* 150(7):691–712.
30. Hurford J (2011) *The Origins of Grammar* (Oxford Univ Press, Oxford).
31. Coye C, Zuberbühler K, Lemasson A (2016) Morphologically structured vocalizations in female Diana monkeys. *Anim Behav* 115:97–105.
32. Engesser S, Crane JM, Savage JL, Russell AF, Townsend SW (2015) Experimental evidence for phonemic contrasts in a nonhuman vocal system. *PLoS Biol* 13(6):e1002171.
33. Slocombe KE, Kaller T, Call J, Zuberbühler K (2010) Chimpanzees extract social information from agonistic screams. *PLoS One* 5(7):e11473.
34. Scott-Phillips TC, Gurney J, Ivens A, Diggle SP, Popat R (2014) Combinatorial communication in bacteria: Implications for the origins of linguistic generativity. *PLoS One* 9(4):e95929.
35. Berwick RC (2011) All you need is merge: Biology, computation, and language from the bottom-up. *The Biolinguistic Enterprise*, eds Sciullo AMD, Boeckx C (MIT Press, Cambridge, MA), pp 706–825.
36. Jackendoff R (1999) Possible stages in the evolution of the language capacity. *Trends Cogn Sci* 3(7):272–279.
37. Progovac L (2015) *Evolutionary Syntax* (Oxford Univ Press, Oxford).
38. Tomasello M (2003) *Constructing a Language: A Usage-Based Theory of Language Acquisition* (Harvard Univ Press, Cambridge, MA).
39. Senghas A, Kita S, Ozyürek A (2004) Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science* 305(5691):1779–1782.
40. Nowak MA, Komarova NL (2001) Towards an evolutionary theory of language. *Trends Cogn Sci* 5(7):288–295.
41. Nowak MA, Krakauer DC (1999) The evolution of language. *Proc Natl Acad Sci USA* 96(14):8028–8033.
42. Clutton-Brock TH, et al. (1998) Costs of cooperative behaviour in suricates (Suricata suricatta). *Proc Biol Sci* 265(1392):185–190.
43. Sharp SP, McGowan A, Wood MJ, Hatchwell BJ (2005) Learned kin recognition cues in a social bird. *Nature* 434(7037):1127–1130.
44. Fitch WT, Neubauer J, Herzel H (2002) Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Anim Behav* 63(3):407–418.
45. Mundry R, Sommer C (2007) Discriminant function analysis with nonindependent data: Consequences and an alternative. *Anim Behav* 74(4):965–976.
46. R Core Team (2014) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
47. Bates D, Maechler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67(1):1–48.
48. Bartón K (2014) MuMIn: Multi-model inference. R package. Available at CRAN.R-project.org/package=MuMIn. Version 1.10.5.
49. Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York).
50. Wickham H (2012) gtable: Arrange grobs in tables. R package. Available at CRAN.R-project.org/package=gtable. Version 0.1.2.
51. Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer, New York), 2nd Ed.
52. Halekoh U, Höjsgaard S (2014) A Kenward-Roger approximation and parametric bootstrap method for tests in linear mixed models - the R package pbkrtest. *J Stat Softw* 59(9):1–30.
53. Zuur A, Ieno EN, Walker N, Saveliev AA, Smith GM (2009) *Mixed Effects Models and Extensions in Ecology with R* (Springer, New York).
54. Harrison XA (2014) Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ* 2:e616.
55. LeBlanc DC (2004) *Statistics: Concepts and Applications for Science* (Jones and Bartlett Publishers, Mississauge, ON, Canada).

EVOLUTION