Internal acoustic structuring in pied babbler recruitment cries specifies the form of recruitment

Sabrina Engesser [a,*], Amanda R. Ridley [b,c], Marta B. Manser [a], Andri Manser [d], Simon W. Townsend [a,e,f]

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

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

[c] Percy FitzPatrick Institute, University of Cape Town, University Avenue, Rondebosch 7701, South Africa

[d] Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, United Kingdom

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

[f] Comparative Communication and Cognition Group, Department of Comparative Linguistics, University of Zurich, Plattenstrasse 54, 8032 Zurich, Switzerland

[*] Corresponding author
Current address of corresponding author: Comparative Communication and Cognition Group, Department of Comparative Linguistics, University of Zurich, Plattenstrasse 54, 8032 Zurich, Switzerland; Email: sabrina.engesser@outlook.com; Phone: +41 (0)44 634 0223

Abstract

Language is inherently combinatorial, and parallels of this combinatorial capacity are found in non-human systems, with animals combining sounds and calls into larger meaningful structures. However, further analogue examples are central in unveiling the diversity, distribution and evolutionary drivers of combinatoriality. Here, we provide evidence for internal 'meaning-refining' acoustic variation within a larger stereotyped signal in pied babblers (*Turdoides bicolor*). Using acoustic analyses, we demonstrate that males produce two long, raucous, 'cry-like' structures, both starting with a wind-up segment grading into repetitions of A/single-note or AB/double-note motifs. Behavioral observations indicated that, consistent with similarities in their larger stereotyped structure, both variants function overall in recruiting group members during locomotion, but the internal A or AB sub-structure specifies the *precise* form of recruitment, from approaching the caller's announced location, to following it over longer distances. Playing back cries from a stationary loudspeaker further supported that the two variants elicit different responses, with more individuals approaching the loudspeaker in response to single-note compared to double-note cries. Additionally, despite similarities in overall

distance travelled, group movement was only directional for single-note, but undefined for double-note cries. We suggest that the overall structure of the two cry variants conveys the same general meaning, with embedded variation refining this meaning. We suggest these results further illustrate the variability of generative mechanisms outside of human language and lends support to the hypothesis that combinatorial structuring may have emerged in species with limited or fixed vocal repertoires in order to enhance communicative output.

Introduction

One of language's defining features is its expressive power. By flexibly combining a limited suite of acoustic elements, a theoretically infinite myriad of messages can be communicated (Hockett 1960; Chomsky 1981; Humboldt 1999). Identifying the potential selective drivers for combinatorial abilities is of fundamental significance in the field of human language evolution (Hauser et al. 2002; Christiansen and Kirby 2003). One dominant hypothesis derived from theoretical and computational modelling suggests that combining vocal elements is one evolutionary solution to vocal constraints that limit the number of distinct sounds that can be produced and perceived (Nowak and Krakauer 1999; Nowak et al. 1999; Nowak and Komarova 2001). Combining sounds can serve to create more easily distinguishable signals or words, and assembling such units into larger meaningful structures can further increase the amount of information that can be communicated (Nowak and Krakauer 1999; Nowak et al. 1999; Nowak and Komarova 2001).

Comparative research on animals provides a powerful tool to test this hypothesis (Hauser et al. 2002; Rendall 2013). Empirical studies have demonstrated that stringing both 'meaningless' sounds (i.e. sounds which are acoustically distinguishable and not produced in isolation) and 'meaningful' calls (i.e. functionally specific call units) together can function to increase and enhance the communicative output of a species (Marler 1977; Hurford 2012; Zuberbühler 2015). For example, chestnut-crowned babblers (*Pomatostomus ruficeps*) reuse and rearrange meaningless, acoustic elements to create functionally distinct and perceptually relevant flight and provisioning calls (Engesser et al. 2015). Banded mongooses (*Mungos mungo*) and Diana monkeys (*Cercopithecus diana*) produce two-compound utterances, with one unit encoding the caller's identity and the other relating to the caller's behavior or motivational state (Candiotti et al. 2012; Jansen et al. 2012; Coye et al. 2016). On a structurally higher, syntactic level, putty-nosed monkeys (*Cercopithecus nictitans*) combine predator-specific alarm calls into a sequence that functions to initiate group movement (Arnold and Zuberbühler 2006, 2008, 2012). Similarly, Campbell's monkeys (*Cercopithicus cambelli*) affix highly predator-specific eagle and leopard alarm calls with a meaning-modifying acoustic element which broadens both calls' meaning, by transforming eagle alarm calls into arboreal disturbance calls, and leopard alarm calls into general alert calls (Ouattara et al. 2009; Coye et al. 2015). Furthermore, two recent experimental studies in birds suggest these abilities are more widespread than previously thought; both Japanese great tits (*Parus major minor*) and pied babblers (*Turdoides bicolor*) have been shown to combine alert and recruitment calls

when mobbing predators, potentially linking information on the danger with an approach-request (Engesser et al. 2016; Suzuki et al. 2016). Such data are particularly important in helping to elucidate the various forms of combinatorial structures that exist in animal communication, and critically the mechanisms by which new meaning can be generated.

While previous studies have demonstrated that i) *acoustic variation within* the internal structure of animal calls can generate *qualitatively* different signals (Jansen et al. 2012; Engesser et al. 2015), and, ii) *external acoustic modifiers* can broaden the meaning of a discrete call in a *quantitative* way (Ouattara et al. 2009; Coye et al. 2015), we here extend this body of knowledge on the structural diversity of combinatorial mechanisms. Specifically, we investigated *meaning-refining modification* of *internal* acoustic motifs in a larger stereotyped pied babbler vocal sequence (Fig. 1a, b). Pied babblers are cooperatively breeding passerines that possess a restricted vocal repertoire of around 17 discrete call types, including alarm calls and an array of social calls (Radford and Ridley 2006, 2008; Bell et al. 2009; Hollén et al. 2011; Golabek and Radford 2013; Engesser et al. 2017). They live in stable groups of, on average, 5-6 individuals with reproduction primarily restricted to the dominant pair and subordinate helpers assisting in rearing the offspring (Ridley and Raihani 2007b; Nelson-Flower et al. 2011; Engesser et al. 2016). Members of the group cooperatively defend their territory (average territory size $0.75km^2$ (Humphries 2012)), and forage and travel in a highly cohesive manner (Ridley and Raihani 2007a; Golabek et al. 2012). Babblers spend most of the time on the ground (95% of their foraging time), with flight movements between foraging or resting sites typically being short and low to the ground (Ridley and Raihani 2007a).

When leading the group through their territory, for example to new foraging or resting sites, males (predominantly the dominant male of a group) produce two types of long and raucous, 'cry-like' structures: 'single-note recruitment cries' (SN cries) and 'double-note recruitment cries' (DN cries) (see also Golabek and Radford (2013) and Humphries et al. (2016)). Both cry types start with a wind-up segment which increases in amplitude and grades into repetitions of either A/single-note motifs, or AB/double-note motifs, with both motif types appearing to share the A note, and a B note being added to each A note in the case of double-note motifs (Fig. 1a, b). Previous work has demonstrated an approach-inducing function of A/single-note motifs, with babblers combining alert calls with A notes (or 'recruitment calls') during predator encounters, when calling from, and recruiting group members to the location of the threat (Engesser et al. 2016). In line with this work, further observations on naturally emitted cries suggest that SN cries function to recruit group members to a caller's broadcast location. DN cries, on the other hand, appear to initiate a cohesive group movement over longer distances, with the caller vocalizing and then subsequently departing, and receivers following the caller without approaching its original calling location. In accordance with these preliminary observations we systematically investigated the function of babbler 'recruitment cries' and hypothesized that both cry variants function to elicit group movement (i.e. "come"), with the cry's internal structure refining and modifying the cries' contextual specificity from soliciting approach (i.e. "come *to* me") to prompting following (i.e. "come *with* me").

Using acoustic analyses, behavioral observations, and playback experiments we investigated the structural similarity and contextual distinctiveness of the two recruitment cry variants. Firstly, acoustic analyses served to test whether DN cries are composed of two distinct note types (A and B notes), and whether the A notes of SN cries are
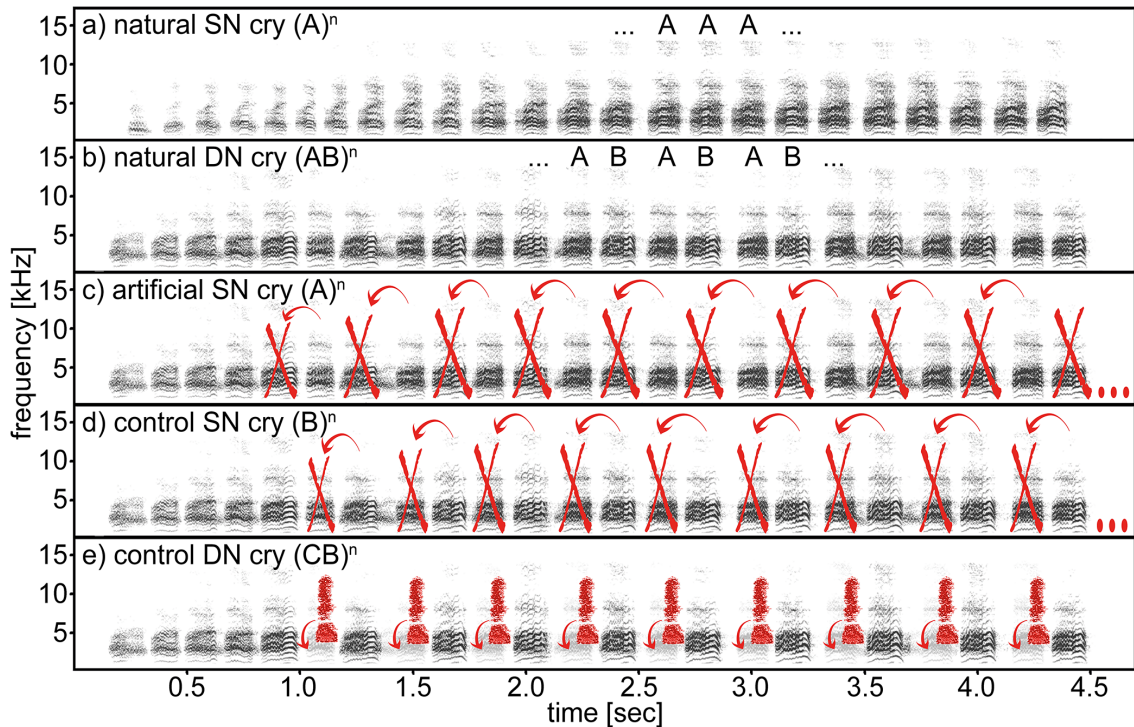
Figure 1. a, b) Spectrograms of a natural SN cry and a natural DN cry of the same dominant male. Capital letters denote the note type. c-e) Example spectrograms of manipulated cries used for playback experiments: c) artificial SN cry generated from a DN cry by deleting each B note; d) control SN cry generated from a DN cry by deleting each A note, e) control DN cry generated by substituting each A note with a chuck vocalization.

equivalent to the A notes of DN cries. Secondly, observations of the cry variants under natural conditions were conducted to determine (i) whether receivers differ in their probability to approach the location of the signaler when producing either of the two cry variants, and (ii) whether receivers (as well as the caller) differ in how far they move in response to SN and DN cries. We predicted receivers are more likely to approach the caller's announced location when producing SN cries, while for DN cries we expected no approach but, in contrast, a larger movement of the whole group following the caller. Thirdly, to experimentally confirm behavioral observations of natural cry events, we conducted playback experiments of natural, artificial, and control cries. We predicted receivers would approach the sound source when playing back natural, unmodified SN cries. In line with natural observations, one prediction would be that receivers travel a greater distance in response to DN cry playbacks, though this potentially requires providing receivers with additional visual information pertaining to the likely direction the signaler will travel. Whilst a playback set-up can simulate the vocalizations of a putative caller, simulating or initiating the caller's take-off and travel direction is logistically not possible. Nevertheless, we still expected DN cries to elicit a qualitatively different response compared to SN cries. Moreover, this playback setup actually allows the disentangling of whether receivers respond to cries based on behavioral (staying/moving) versus vocal factors (motif type), because in both conditions the simulated caller (i.e. speaker) remains stationary. If the two cry variants' information content would be deduced from the caller's behavior alone, we predicted that, similarly to SN cries, DN cries should also result in an approach to the sound source. If, on the other hand, the subjects decode the form of recruitment from the cry types, we expected DN cries to elicit a different response compared to SN cries, with receivers refraining

from approaching the loudspeaker. Furthermore, in line with the simulated caller (although allegedly announcing a movement) not taking off when playing back DN cries, we predicted the group to still show indications of movement, but without a clear travel direction.

Besides investigating the responses to natural recruitment cries, in order to demonstrate the same note (A note) is reused across SN and DN cries, artificial SN cries created out of DN cries (by deleting the cry's B notes; Fig. 1c) were played back. We predicted the same response to artificial SN cries as when exposed to natural SN cry playbacks. Lastly, in order to rule out that merely the cries' rhythmic patterns might account for their context specificity (i.e. any single or double-note motifs irrespective of the note types), therefore also confirming the reuse of A notes across the two cry types, control SN and DN cries were played back (Fig. 1d, e). Control SN cries were composed of B notes, and in control DN cries, each A note was substituted by a babbler contact call (see methods for rational). If rhythmic patterns or the cry's superstructure were solely responsible, irrespective of integrated note types, control sequences should elicit equivalent responses as to the natural SN and DN variants.

Methods

Study site and species

A free-living population of southern pied babblers was studied at the Kuruman River Reserve (26°58'S, 21°49'E), located in the semi-arid Kalahari Desert of South Africa (for details on the habitat see Raihani and Ridley (2007) and Hollén et al. (2011)). The birds of the study population have been monitored since 2003, and extensive life-history data are available for all individuals. All birds are habituated to human observers, enabling close observations (Radford and Ridley 2008). Unique combinations of three colored and one numbered metal ring allow the identification of all babblers at the study site (Radford and Ridley 2008). Pied babblers are sexually monomorphic, and individuals of the study population are sexed using DNA tests (Nelson-Flower et al. 2011).

Acoustic analyses

Acoustic analyses were conducted to verify that the same acoustic element (A note) is used in SN and DN cries. Vocalizations were recorded using a Rode NTG-2 directional microphone coupled with a Rode Blimp Suspension Windshield (Rode microphones) and a Roland R-26 portable recorder (Roland Corporation) (sampling frequency 48kHz, 24-bit accuracy). Due to the high vocal activity of the group's dominant male and his central role in leading the group to new areas (Engesser et al. 2017), in addition to controlling for potential rank or age effects, only dominant male vocalizations were considered. In order to prevent pseudoreplication (and thus the generation of erroneous P-value estimations), only one A note of a SN cry, as well as one A and one B note of a DN cry per individual were analyzed (Sharp et al. 2005). For each of 10 dominant males high signal-to-noise-ratio elements were chosen randomly, and the following acoustic parameters were extracted using Praat (v. 5.1.03): element duration, bandwidth of fundamental frequency, maximum fundamental frequency, time-point of maximum fundamental frequency, time-point of maximum amplitude, the frequency values at the upper limit of the first, second and third quartiles of energy, and the proportion of the

element that exhibited clear voiced/periodic structure (see supporting dataset for definitions of the acoustic parameters).

To assess the classification probabilities of element to note class (SN A note, DN A note, and DN B note), a cross-validated discriminant function analysis (DFA) was conducted using the MASS package in R (v. 3.2.3) (Venables and Ripley 2002; R-Core-Team 2014). With 3 element categories to be classified, the DFA created 2 discriminant functions by determining the linear combinations of the predictor variables that most adequately discriminate between the 3 categories (Mundry and Sommer 2007). In order to assess which acoustic parameters contributed most to the classification, their standardized coefficients were calculated (Coghlan 2014). Multivariate normality was assessed graphically, and only parameters with a variance inflation factor smaller than 10 were included in analyses, resulting in the exclusion of the frequency measurements at the first and third energy quartiles (Zuur et al. 2009; Fox and Weisberg 2011). Two-tailed binomial tests were conducted to assess whether the classification results of the DFA differed from that expected by chance (R stats package). According to the 3 element categories to be discriminated, a 33.3%-chance level of correct classifications was set.

Natural calling observations

Data collection

To investigate the context-specific production and subsequent receiver response to the two types of recruitment cries, observations on naturally produced recruitment cries were conducted from January to April/May in 2014 and 2015 on 20 babbler groups (average group size 4.9±1.3). A group was followed in the evening for approximately 2h until it had settled in a night roost. The subsequent morning, the group was followed from the roost for approximately 4h. Cry events of all male individuals, regardless of status, were recorded, documenting the identity of the caller, the cry type produced (SN or DN), and, if applicable, the distance moved by the caller. Since babblers are weak flyers, groups could be followed easily on foot. Additionally, we recorded the response of receivers, specifically whether the group approached the caller's broadcast location to within a minimum of 5 meters (yes/no), and the distance moved by the group in response to the cry. Movement was considered as receivers interrupting their foraging or resting activity and flying or moving fast along the ground to a new location (either to the caller's broadcast position or to the location the caller had moved to). The distances moved by both the caller and its group were measured using a handheld global positioning data-logger (GPS-logger, eTrex 10, Garmin). We recorded the caller's location and the location of the density-based center (estimated centroid) of the remaining group members during the cry production, as well as their locations after the group had switched to a new site in response to the cry. New sites were defined as locations where at least half (50%) of the group members moved to (for resting or to continue foraging). This 50% threshold was chosen given that, at times, not all individuals of a group arrived at a new site simultaneously, with some group members staying behind during the movement and joining shortly after. If no response occurred within 5 minutes after the cry production a distance of 0 meters was assigned.

Probability of approach in response to SN and DN cries

To investigate whether receivers were more likely to approach a caller producing SN or DN cries, a binomial generalized linear mixed model (GLMM) with a logit-link function was fitted (Bates et al. 2014), with group approach specified as the response variable (0=no approach, 1=approach to caller's broadcast position). Caller identity nested within group identity was fitted as a random term, and cry type (SN or DN) was the explanatory variable. To account for alternative parameters which might affect the response probability to each of the cry types, we additionally included a group's size (including all individuals of a group), the number of adult individuals (individuals >1 year), and the observation session (morning/evening), as well as their interaction term with the cry type as explanatory variables. Specifically, responses to the cries might be more flexible and hence less pronounced (i) in bigger groups and/or in groups with a larger proportion of independent adults (due to greater variation in individual preferences), as well as (ii) in the morning when preferences among individuals might vary more compared to the evening, when cohesion among group members is of particular importance for the group to settle in the same roost tree. To determine the percentage of occasions resulting in an approach by the group to the caller's broadcast location in response to SN and DN cries, respectively, the back-transformed model intercepts and the bootstrapped 95% confidence intervals of the difference of the mean (CI) were calculated (Burnham and Anderson 2002).

Distance moved in response to SN and DN cries

A linear mixed model was fitted (LMM) to test for differences in the distance a caller and its receivers' moved after producing and hearing SN and DN cries (Bates et al. 2014). The distance moved to a new location was fitted as the response variable, and square-root transformed to fulfil model assumptions. Explanatory variables were the cry type (SN/DN), identity of the subject (caller/receiver) and the interaction between them. The calling event, nested within caller identity and the caller's group was fitted as a random term. This allowed us to control for the dependency between the distance a caller and its receivers moved in response to one particular calling event (i.e. cry event 1 of individual X from group Y resulted in the caller moving *s* meters and its group moving *t* meters).

Playback experiments

Playback experiments on 10 babbler groups were conducted between March and May 2015. For consistency and due to their high vocal activity, only vocalizations of a group's dominant male were used to create the playback stimuli. High signal-to-noise-ratio vocalizations were selected and edited using Audition CS6 (Adobe, sampling frequency 48kHz, 24-bits accuracy). To verify the recruitment cries' context specificity, in each of 10 test groups, a natural SN cry and a natural DN cry of the dominant male were played back (Fig. 1a, b). To investigate whether the SN and DN cry are composed of the same acoustic element (the A note) and to complement acoustic analyses, we additionally played back an artificial SN cry, which was created by deleting each B note of the natural DN cry (Fig. 1c). The inter-element distance between the remaining A notes was adjusted to that found between A notes of the natural SN cry. While additionally testing the response to artificial DN cries (i.e. DN cries with A notes being substituted by A notes

originating from SN cries) might have yielded a more balanced design, we considered one test condition (i.e. the artificial SN cry) to be sufficient to probe the acoustic similarity and functional uniformity of A notes across SN and DN cries, whilst simultaneously avoiding over-exposing subjects to redundant playback experiments. To rule out that any single or double-note motifs, irrelevant of the note type, might elicit a recruitment response as observed to natural SN and DN cries, control stimuli were created. A control SN cry was created by deleting each A note within the natural DN cry, with the inter-element distance between the remaining B notes again matched to the distance found between the A notes of the natural SN cry (Fig. 1d). Selecting B notes for the creation of control SN cries further allowed testing the dissimilarity among A and B notes, which would be implied by different responses to played back natural and control SN cries. A control DN cry was created by substituting each natural DN cry's A note with a chuck vocalization (neutral babbler foraging contact call not encoding any urgency related features (Radford and Ridley 2008)) originating from the same individual, keeping the same distance between the replaced element and the B notes (Fig. 1e). Since chuck vocalizations can exhibit acoustic variation that correlates with the type and quality of the food source a babbler is processing (Golabek 2010), we repeated the same chuck exemplar (recorded while the bird was foraging on the ground not processing any food) so as to avoid presenting variable and contextually inconsistent chucks within one cry.

Thus, 5 playback stimuli were created from one SN cry, one DN cry, and one chuck vocalization, each originating from the same dominant male in each of the 10 test groups. In cases where elements were replaced, the maximum amplitude of the new element was matched to the one of the substituted element. All 5 stimuli for one group were adjusted to have approximately the same duration (4.7±0.5sec), with some cries being shortened and others being extended by duplicating elements within a cry. All stimuli were normalized and played back at a naturally occurring amplitude (~73dB at 4m distance, measured using a Cirrus CR261 sound level meter). Each cry was broadcast once using a portable speaker (Anchor AN-30) coupled with an iPod 3 (Apple Inc.). Each of the 10 groups received all 5 stimuli in a randomized order. A maximum of 2 stimuli were played back in one morning, with at least 2 days break in-between subsequent playback sessions to minimize the risk of habituation to playbacks. Since vocalizations of a group's dominant male were broadcast, and recruitment cries are individually distinct (Humphries et al. 2016), stimuli were played back from the location of the dominant male at the time of the playback. This served to prevent reactions to the stimuli based solely on the incongruence between the location of the playback and the location of the dominant male. Accordingly, the loudspeaker was placed below or next to a group's dominant male, with the rest of the group foraging at least 20 meters from the male (mean 25±4.5m).

To investigate whether receivers are more likely to approach the sound source in response to natural (and artificial) SN cries, we recorded the proportion of the group approaching to within 5 meters of the loudspeaker within 5 minutes post-playback. Responses were analyzed using a binomial GLMM with a logit-link function. To test for impacts of playback on movement behavior, we determined the distance the whole group had moved 5 minutes after the playback, as well as distance to the loudspeaker at 1, 2, 3, 4 and 5 minutes post-playback. If a group had passed the loudspeaker, negative values for the distance to the loudspeaker were assigned. Both response variables were analyzed using separate LMMs.

General statistical procedure

LMMs and GLMMs were fitted in R using the package lme4 (Bates et al. 2014). Normality of the data was assessed graphically. Collinearity among predictor variables could be excluded in the models with more than one explanatory variable, since variance inflation factors were always substantially smaller than 10 (Zuur et al. 2009). Variance inflation factors were calculated using the vif.mer function designed for mixed effects models in R (Frank 2011). Overdispersion of the binomial GLMM investigating the proportion of group members that approached the loudspeaker in response to played back cries was estimated by assessing the ratio of the sum of squared Pearson residuals to the degrees of freedom (each variance parameter representing one degree of freedom), and an observation-level random term (assigning each data point an individual random-effect level) was added to account for the over-dispersed data (Bolker et al. 2009; Harrison 2014). Top models were selected using the dredge function within the MuMIn package (Bartoń 2014). Based on the full model, the function ranks all models according to the Akaike's information criterion corrected for small sample sizes (AICc) (Burnham and Anderson 2002). A threshold difference (ΔAICc) of 2 to the top model was set. If more than one model fell into this range, the model with less degrees of freedom (DF) was chosen. To assess the significance of fixed effects, likelihood ratio tests were conducted comparing the top model with the model without the factor of interest (Zuur et al. 2009). To determine differences among factor levels, we assessed whether the bootstrapped CIs intersected with zero, with CIs not crossing zero representing significant differences between factor levels.

To investigate changes in a group's distance to the loudspeaker over time in response to the played back cries we fitted a polynomial LMM using the package nlme (Pinheiro et al. 2018). The interaction between treatment and time after playback onset was fitted as the sole explanatory variable, forcing the model predictions for each treatment to have the same starting point (intercept). The nlme package further allowed us to account for various sources of non-independence and heterogeneity of residuals (Zuur et al. 2009). Firstly, to control for repeated use of experimental groups across treatments, we allowed slopes to systematically vary across groups by fitting group identity as a random term (random slopes with identical intercepts, i.e. start points, for all groups). Secondly, investigating correlations among the residuals of a group's distance to the loudspeaker across different time points within each experimental trial revealed non-independence. Therefore, to account for temporal autocorrelation, we allowed residuals of time point t to be correlated with residuals of time point $t+1$ through a first-order autoregressive covariance structure (AR-1 function). Thirdly, strong differences in residual variance across time points (heterogeneity) was controlled for by allowing the variance to systematically differ between time points (varIdent function). Using a forward selection procedure and likelihood ratio tests, we then compared the polynomial models starting with the linear model (i.e. first-order model) until increasing the order did not significantly increase the fit of the model anymore. However, due to interpretability of the model and to avoid overfitting we set the third-order polynomial as limit.

Ethical statement

All work was conducted under the permission of the ethical committee for animal research of the University of Cape Town and the Northern Cape Conservation Authority,

South Africa.

Results

Acoustic analyses

The cross-validated DFA ($N_{elements}$=30, $N_{individuals}$=10) indicated that B notes were structurally distinct from A notes, never being misclassified as A notes and vice versa (correct classification: B notes=100%, P<0.001; Fig. 2, Table 1a). A notes originating from SN and DN cries were indistinguishable from each other (correct classifications: SN A note=40%, DN A note=30%, all P>0.05; with the remaining percentage of elements being classified as an A note of the opposing cry type). The first discriminant function accounted for 99.7% of the classification, with the bandwidth of the fundamental frequency and the element duration contributing most to the observed discriminability (Table 1b).

Table 1a. Classification probabilities of cross-validated discriminant function analyses on the notes comprising recruitment cries. N for each note type=10; random change level=10/30=33.3%. SN-A=SN cry A note, DN-A=DN cry A note, DN-B=DN cry B note.

	SN-A	DN-A	DN-B
SN-A	4	6	0
DN-A	7	3	0
DN-B	0	0	10

Table 1b. Standardized coefficients of linear discriminant functions (LD) of the discriminant function analysis. Greater magnitudes of the coefficients (absolute values) indicate a greater impact on the discriminant functions. [1] parameters excluded from the analysis due to high variance inflation factors.

| parameter	LD 1	LD 2
element duration	0.703	-0.454
bandwidth of fundamental frequency	0.833	0.018
maximum fundamental frequency	-0.013	0.007
time-point of maximum fundamental frequency	0.209	-0.259
time-point of maximum amplitude	0.133	-0.210
frequency value at first energy quartile	-[1]	-[1]
frequency value at second energy quartile	-0.413	-0.456
frequency value at third energy quartile	-[1]	-[1]
percentage of voiced structures	0.292	0.845
attribution [%] to DFA classification	99.7	0.3
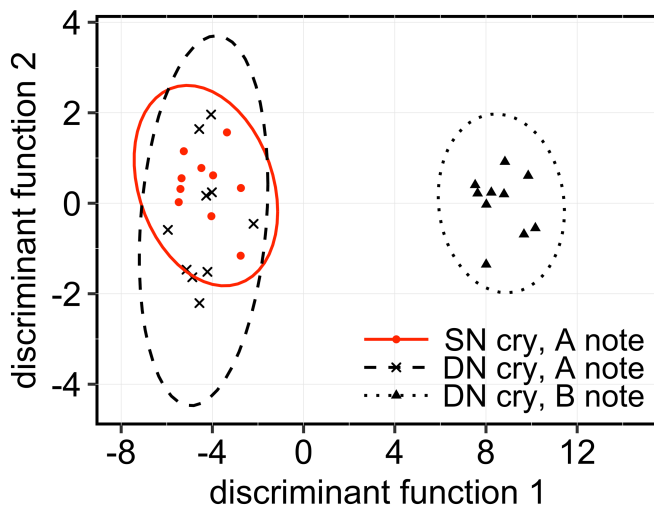
Figure 2. DFA output on SN cry A notes, and DN cry A and B notes of 10 dominant males each. Circles represent clusters assuming a multivariate normal distribution with a confidence level of 95%. Discriminant function 1 accounted for 99.7% of the classification probability, discriminant function 2 for 0.3%.

Natural calling observations

Probability of approach in response to SN and DN cries

Receivers were more likely to approach callers producing SN cries than DN cries (GLMM: χ^2=44.5, P<0.001, N=152 cries of 31 individuals in 20 groups; approach to 51 of 80 SN cries, and to 9 of 72 DN cries). For SN cries, receivers approached the caller in 63.7% (CIs: 53.7-74.2%) of occasions. In contrast, receivers approached the caller in response to DN cries in only 12.5% (CIs: 5.6-19.6%) of occasions. Neither the session (morning/evening), group size, or proportion of adult individuals in the group had an effect on the approach probability, since they were not retained during model selection (Table 2, S1).

Table 2. Effect of cry type, group size, proportion of adult individuals and session (morning/evening) on a group's probability to approach a caller naturally producing SN or DN cries. Presented are the full and intercept model, as well as the top model according to the AICc selection, and any models falling within a ΔAICc < 2 to the top model. All models included a caller's identity nested within its group affiliation as random term. Listed fixed effects represent the test statistics of the factors of models within a ΔAICc < 2 to the top model (assessed using likelihood ratio tests comparing models including and excluding the factor of interest). DF=degrees of freedom, AICc=Akaike's information criterion corrected for small sample sizes. (For full model list see Table S1).

Model		DF	AICc	Fixed effects	χ^2	P-value
approach ~	cry type	4	167.3			
				cry type	44.5	<0.001
approach ~	cry type + group size	5	169.2			
				cry type	44.1	<0.001
				group size	0.3	0.614
approach ~	cry type + group size + prop. adults + session + cry pattern x (group size + prop. adults + session)	10	179.1			
approach ~	1	3	209.7			

Parameter	Estimate	Standard error	Z-value	P-value
cry type				
SN cry	0.56	0.23	2.43	0.015
DN cry	-1.95	0.36	-5.46	<0.001

Distance moved in response to SN and DN cries

The distance callers and their receivers moved differed depending on the cry type produced (LMM: cry type [SN *vs*. DN] x subject [caller *vs*. receiver]: χ^2=16.7, P<0.001, N=295 observations originating from 153 calling events of 31 individuals in 20 groups). Specifically, when producing SN cries, callers moved less than receivers and generally remained stationary (64 out of 79 occasions), which is in accordance with our results demonstrating receivers approaching the caller's broadcast location in response to SN cries. When producing DN cries, callers and receivers moved the same distance, exceeding the distance receivers (and callers) moved in response to SN cries (Fig. 3).
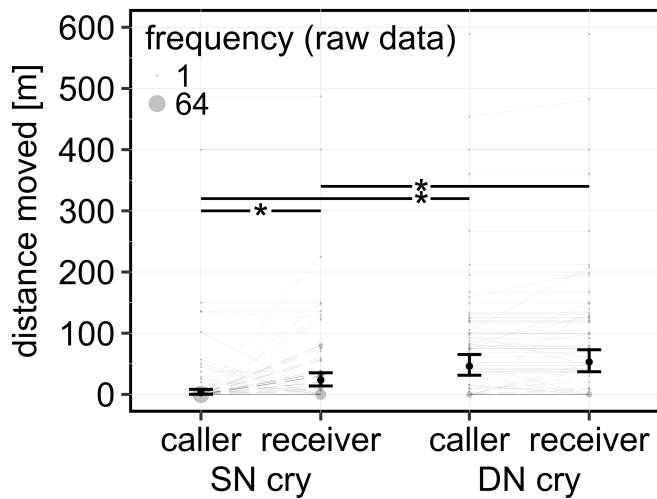
Figure 3. Distance that callers and receivers moved in response to naturally emitted SN and DN cries. N=295 observations originating from 153 calling events of 31 individuals in 20 groups. Bars illustrate the 95% confidence intervals of the difference, and points show the medians of the back-transformed data. Pale dots show the raw data, with the size corresponding to the frequency of occurrences. Lines connect calling events, where the distance a caller and its group moved in a specific calling event are paired-up. Asterisks indicate significant contrasts according to the CIs. Although not part of the statistical model investigating differences in the moved distance, but in order to provide a more comprehensive representation of the data, occasions where receivers approached the caller are additionally illustrated as dotted lines.

Playback experiments

The proportion of group members approaching the loudspeaker differed in response to playbacks of natural, artificial and control SN and DN cries (GLMM: χ^2=22.018, P<0.001, N=10 groups with 5 observations each; Fig. 4a). Specifically, artificial SN cries elicited the same response as natural SN cries. Fewer individuals approached the loudspeaker in response to natural DN cries compared to playbacks of natural or artificial SN cries. Ruling out rhythmicity effects, neither control SN cries elicited the same response as natural or artificial SN cries, nor did control DN cries elicit the same response as natural DN cries (Fig 4a).

No difference in response to the playback stimuli were found regarding the distance a group moved in the 5 minutes after the playback onset (LMM: χ^2=5.2, P=0.286; N=10 groups with 5 observations each). However, we found a significant effect of played back cries on the changes in distance to the loudspeaker over time (LMM treatment x time[3]: χ^2=69.8, P<0.001, N=10 groups with 5 observations each; Fig 4b, Table 3). Specifically, in response to natural, artificial, control SN and control DN cries, subjects overall moved towards the loudspeaker (significant first-order term). For control DN cries the movement speed towards the speaker was constant but slower than for any SN cries (second- and third-order term non-significant). For natural, artificial and control SN cries (i) the speed changed over time with subjects initially approaching the loudspeaker faster and subsequently decelerating (significant second-order term). DN cries elicited no directional changes (all polynomial terms non-significant).
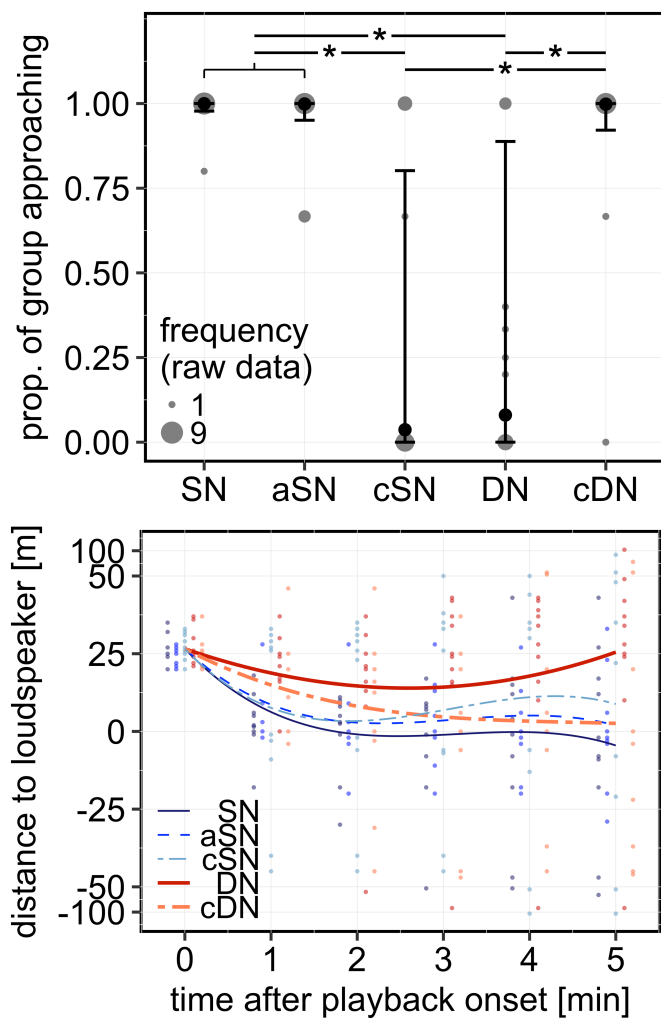
Figure 4. a) Proportion of the group approaching the loudspeaker in response to cry playbacks. Bars illustrate the 95% confidence intervals of the difference, and points show the medians of the back-transformed data. Pale dots show the raw data, with the size corresponding to the frequency of occurrences. Asterisks indicate significant contrasts according to the CIs. b) Group's distance to the loudspeaker at fixed time-intervals after the playback onset (third-order polynomial model). Negative values represent cases where the group was located opposite to the loudspeaker in relation to their start location. Dots represent the raw data. Note, above/below 50/-50m the Y-axis has been compressed for a better representation of the data. aSN=artificial SN cry, SN=natural SN cry, cSN=control SN cry, DN=natural DN cry, cDN=control DN cry. N=10 groups (within subject design).

Discussion

The results of this study provide empirical evidence for combinatorial structure in pied babbler recruitment cries. Acoustic analyses and playback experiments confirm babblers reuse the same acoustic element across the A/single-note motifs of SN cries and the AB/double-note motifs of DN cries. The A notes of both cries were statistically indistinguishable but distinct from B notes, and receivers engaged in the same heightened approach response to playbacks of natural and artificial SN cries. Neither playbacks of control SN or control DN cries elicited the same approach response as natural SN or natural DN cries, indicating that both motif pattern but also note type are important in differentiating the recruitment cries' meaning.

Regarding the function of SN cries, behavioral responses to both naturally produced cries and cry playbacks support the prediction that SN cries function to recruit group members to the caller's announced location. Natural calling observations further suggest DN cries function to induce the group to follow the caller. While playbacks of DN cries did not induce following of the simulated caller (likely a consequence of restricting caller/speaker movement in our experimental design), we still found differences in a group's movement behavior in response to the played back cry variants.

Table 3. Likelihood-ratio test output and summary output of the coefficients of the third-order polynomial model investigating the effect of played back cries on changes of a group's distance to the loudspeaker over time. All models included group as random term. Models are presented with increasing polynomial term. χ^2 and P-values represent the test statistic for comparisons of a given model with the polynomial of the next higher order. DF=degrees of freedom, AICc=Akaike's information criterion corrected for small sample sizes, SN=SN cry, aSN=artificial SN cry, cSN=control SN cry, DN=natural DN cry, cDN=control DN cry.

Model	DF	AICc	χ^2	P-value
distance ~ cry treatment x time	1	2430.5	28.6	<0.001
distance ~ cry treatment x time^2	2	2412.0	14.2	0.014
distance ~ cry treatment x time^3	3	2407.8		

Parameter	Estimate	Standard error	T-value	P-value
Intercept	26.57	0.96	27.59	<0.001
SN x time^1	-28.91	6.77	-4.27	<0.001
SN x time^2	9.61	3.43	2.80	0.006
SN x time^3	-1.01	0.46	-2.20	0.028
aSN x time^1	-25.88	6.77	-3.82	<0.001
aSN x time^2	8.88	3.43	2.59	0.010
aSN x time^3	-0.94	0.46	-2.04	0.042
cSN x time^1	-28.72	6.77	-4.24	<0.001
cSN x time^2	10.90	3.43	3.18	0.002
cSN x time^3	-1.17	0.46	-2.55	0.011
DN x time^1	-9.53	6.77	-1.41	0.160
DN x time^2	1.73	3.43	0.50	0.615
DN x time^3	0.03	0.46	0.06	0.952
cDN x time^1	-14.48	6.77	-2.14	0.033
cDN x time^2	3.08	3.43	0.90	0.370
cDN x time^3	-0.23	0.46	-0.50	0.619

Firstly, more individuals approached the speaker when playing back natural SN cries, and secondly, the movement pattern of the group (directionality and speed) differed between played back cries. More precisely, when broadcasting natural SN cries, receivers responded with a fast initial, but then decelerating, approach to the loudspeaker. In contrast, natural DN cries did not elicit any defined group movement. As predicted, the lack of a directed movement in response to natural DN cry playbacks (as well as the increased variation in the proportion of birds approaching the sound source within 5 meters) is likely due to receivers missing key visual information on the caller's take-off direction, which would naturally accompany DN cry production. Despite movement being undefined for natural DN cries, the fact that the total distance travelled in response to either of the cries did not differ, indicates that receivers still interpreted the cry as a signal associated with movement by showing increased locomotor activity. Our data therefore suggests that the two cry variants encode different information and rules out that cry meaning is simply deduced from the action of the caller. This conclusion is further supported by observations of rare instances where callers produced SN cries but then moved off, with receivers only following the moving caller in 33% of instances (5 out of 15 cases) and staying stationary in the remaining cases.

It is worth noting that while control DN cries elicited different movement responses than their natural counterparts, this was not the case for SN cries. However, this lack of difference between control and natural SN cries might be an artefact of the large variation in a group's distance to the sound source in response to control SN cries (spanning both a wide positive and negative range). In fact, the lack of a homogenous response is not surprising and may reflect receiver uncertainty regarding the unfamiliar/unnatural control sequences.

Our work suggests that the overall structure of both cry types (both being introduced by a wind-up followed by repetitions of reoccurring motifs, and both being of similar average duration and loudness) likely conveys the same intention of the signaler to recruit group members, with the internal motif pattern determining the precise form of recruitment. A/single-note motifs appear to address receivers to approach to the caller's broadcast location ("come *to* me"), whereas AB/double-note motifs intensify recruitment from approach to prompting following of the signaler over longer distances ("come *with* me"). Accordingly, the B notes might represent an acoustic modifier, which alone is devoid of function, but when combined with A notes modifies/intensifies the As' meaning.

Although further experimental work is necessary to clarify how babblers process the information in the combinatorial structure of these two recruitment cries, our work provides additional support for the use of combinatorial structures as a way to likely enhance signal transmission and perception in non-human communication systems. Parallels can be found in primate alarm call systems such as in Campbell's monkeys, where an acoustic modifier appears to broaden the meaning of predator-specific alarm calls (Ouattara et al. 2009; Coye et al. 2015). In contrast with Campbell's monkeys, however, the modifying element in pied babblers (i) refines meaning (rather than broadening), (ii) operates on a structurally different layer (i.e. internally within a stereotyped sequence, rather than externally with a modifier being added to a discrete call), and moreover is applied in only one instance (i.e. DN cries). Why, though, instead of A- and AB-motifs, do babblers not simply use stand-alone A- and B-motifs? We propose that even though A and B notes are acoustically distinguishable, these differences are subtle and potentially blur when broadcast over long distances, increasing the risk of signal misperception. Indeed, the large variation in response to control SN cries, rather than being caused by uncertainty in response to an unfamiliar stimulus, could alternatively have been the result of birds misperceiving control SN as natural SN cries. Stringing sounds together, can therefore act to create more easily distinguishable signals (Nowak et al. 1999), and counter the possibility of signal degradation hampering discriminability between otherwise similar sounds (i.e. A and B notes). Chestnut-crowned babblers have also been shown to combine acoustically distinct sounds to potentially increase signal discrimination (Engesser et al. 2015). However, in the chestnut-crowned babbler example the internal variation within the calls gives rise to signals with qualitatively different functions, whereas in pied babbler recruitment cries, the modification of the motif pattern represents a more quantitative change, serving to refine the function of the signal (akin to the meaning-modifying affix in Campbell's monkey alarm calls). Besides enhancing the discriminability among signals, from an evolutionary perspective, graded changes in a context-specific behavior such as recruitment, might be more easily achieved through an acoustic element modifying the meaning of a pre-

existing call, rather than through the generation of a completely new, perceptually distinct signal; particularly in species with limited vocal production capacities.

To conclude, pied babbler recruitment cries represent a combinatorial structure, with embedded acoustic variation refining the signal's function. We argue our work lends support to the theory that meaningful vocal combinations emerged to enhance signal transmission and to improve a species' communicative output (Nowak et al. 1999; Arnold and Zuberbühler 2006). Exactly what combinatorial mechanisms an animal implements (phonemic-like structuring, intensifications, affixation, compositionality etc. (Arnold and Zuberbühler 2006; Ouattara et al. 2009; Engesser et al. 2015; Engesser et al. 2016; Suzuki et al. 2016)), among other factors, may well depend on the context (social or predatory (Collier et al. 2017)) that vocalizations are given in and the potential costs incurred by misperception, as well as whether signals are bound to short-range or long-range communication (Engesser et al. 2015). Accordingly, further comparative work is essential in order to unveil the diversity and distribution of vocal combinatoriality, as well as to more accurately identify the ecological and social conditions that likely drive the emergence and, particularly, the variation of combinatorial mechanisms used by non-human animals.

References

Arnold K, Zuberbühler K. 2006. Language evolution: Semantic combinations in primate calls. Nature. 441:303.

Arnold K, Zuberbühler K. 2008. Meaningful call combinations in a non-human primate. Curr Biol. 18:R202-R203.

Arnold K, Zuberbühler K. 2012. Call combinations in monkeys: Compositional or idiomatic expressions? Brain Lang. 120:303-309.

Bartoń K. 2014. MuMIn: multi-model inference. R package version 1.10.5.

Bates D, Maechler M, Bolker B, Walker S. 2014. Fitting linear mixed-effects models using lme4. J Stat Softw. 67:1-48.

Bell MBV, Radford AN, Rose R, Wade HM, Ridley AR. 2009. The value of constant surveillance in a risky environment. Proc R Soc B. 276:2997-3005.

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MH, White JS. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol. 24:127-135.

Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. New York: Springer.

Candiotti A, Zuberbühler K, Lemasson A. 2012. Context-related call combinations in female Diana monkeys. Anim Cogn. 15:327-339.

Chomsky N. 1981. Knowledge of Language: Its Elements and Origins. Philos T R Soc B. 295:223-234.

Christiansen MH, Kirby S. 2003. Language evolution: consensus and controversies. Trends Cogn Sci. 7:300-307.

Coghlan A. 2014. A Little Book of R For Multivariate Analysis, Release 0.1. https://little-book-of-r-for-multivariate-analysis.readthedocs.io/en/latest/. Accessed 28 September 2016.

Collier K, Townsend SW, Manser MB. 2017. Call concatenation in wild meerkats. Anim Behav. 134:257-269.

Coye C, Ouattara K, Zuberbühler K, Lemasson A. 2015. Suffixation influences receivers' behaviour in non-human primates. Proc R Soc B. 282:20150265.

Coye C, Zuberbühler K, Lemasson A. 2016. Morphologically structured vocalizations in female Diana monkeys. Anim Behav. 115:97-105.

Engesser S, Crane JM, Savage JL, Russell AF, Townsend SW. 2015. Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System. PLoS Biol. 13:e1002171.

Engesser S, Ridley AR, Townsend SW. 2016. Meaningful call combinations and compositional processing in the southern pied babbler. Proc Natl Acad Sci USA. 113:5976-5981.

Engesser S, Ridley AR, Townsend SW. 2017. Element repetition rates encode functionally distinct information in pied babbler 'clucks' and 'purrs'. Anim Cog. 20:953-960.

Fox J, Weisberg W. 2011. An {R} Companion to Applied Regression. 2nd ed. Thousand Oaks CA: Sage.

Frank AF. 2011. R-hacks/mer-utils.R.

Golabek KA. 2010. Vocal communication and the facilitation of social behaviour in the southern pied babbler (*Turdoides bicolor*). Dissertation: University of Bristol.

Golabek KA, Radford AN. 2013. Chorus-call classification in the southern pied babbler: multiple call types given in overlapping contexts. Behaviour. 150:691-712.

Golabek KA, Ridley AR, Radford AN. 2012. Food availability affects strength of seasonal territorial behaviour in a cooperatively breeding bird. Anim Behav. 83:613-619.

Harrison XA. 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. PeerJ. 2:e616.

Hauser MD, Chomsky N, Fitch WT. 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? Science. 298:1569-1579.

Hockett CF. 1960. The Origin of Speech. Sci Am. 203:88-111.

Hollén LI, Bell MBV, Russell A, Niven F, Ridley AR, Radford AN. 2011. Calling by Concluding Sentinels: Coordinating Cooperation or Revealing Risk? PLoS ONE. 6:e25010.

Humboldt Wv. 1999. On Language. Translated by Peter Heath. Cambridge: Cambridge University Press.

Humphries D. 2012. The mechanisms and function of social recognition in the cooperatively breeding Southern pied babbler, *Turdoides bicolor*. Dissertation: Macquarie University.

Humphries D, Finch FM, Bell MB, Ridley AR. 2016. Vocal Cues to Identity: Pied Babblers Produce Individually Distinct But Not Stable Loud Calls. Ethology. 122:609-619.

Hurford J. 2012. Linguistics from an Evolutionary Point of View. In: Kempson R, Fernando M, Ashe N, editors. Handbook of the Philosophy of Science: Linguistics Amsterdam: Elsevier Science. p. 473-498.

Jansen DA, Cant MA, Manser MB. 2012. Segmental concatenation of individual signatures and context cues in banded mongoose (*Mungos mungo*) close calls. BMC Biol. 10:97.

Marler P. 1977. The structure of animal communication sounds. Recognition of complex acoustic signals: report of Dahlem workshop. Berlin: Abakon Verlagsgesellschaft.

Mundry R, Sommer C. 2007. Discriminant function analysis with nonindependent data: consequences and an alternative. Anim Behav. 74:965-976.

Nelson-Flower MJ, Hockey PAR, O'Ryan C, Raihani NJ, du Plessis MA, Ridley AR. 2011. Monogamous dominant pairs monopolize reproduction in the cooperatively breeding pied babbler. Behav Ecol. 22:559-565.

Nowak MA, Komarova NL. 2001. Towards an evolutionary theory of language. Trends Cogn Sci. 5:288-295.

Nowak MA, Krakauer DC. 1999. The evolution of language. Proc Natl Acad Sci USA. 96:8028-8033.

Nowak MA, Krakauer DC, Dress A. 1999. An error limit for the evolution of language. Proc R Soc B. 266:2131-2136.

Ouattara K, Lemasson A, Zuberbühler K. 2009. Campbell's Monkeys Use Affixation to Alter Call Meaning. PLoS ONE. 4:e7808.

Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC. 2018. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-131.1, https://cran.r-project.org/package=nlme. Accessed 22 Feb 2018.

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

Radford AN, Ridley AR. 2006. Recruitment Calling: A Novel Form of Extended Parental Care in an Altricial Species. Curr Biol. 16:1700-1704.

Radford AN, Ridley AR. 2008. Close calling regulates spacing between foraging competitors in the group-living pied babbler. Anim Behav. 75:519-527.

Raihani NJ, Ridley AR. 2007. Adult vocalizations during provisioning: offspring response and postfledging benefits in wild pied babblers. Anim Behav. 74:1303-1309.

Rendall D. 2013. Q&A: Cognitive ethology - inside the minds of other species. BMC Biol. 11:1-6.

Ridley AR, Raihani NJ. 2007a. Facultative response to a kleptoparasite by the cooperatively breeding pied babbler. Behav Ecol. 18:324-330.

Ridley AR, Raihani NJ. 2007b. Variable postfledging care in a cooperative bird: causes and consequences. Behav Ecol. 18:994-1000.

Sharp SP, McGowan A, Wood MJ, Hatchwell BJ. 2005. Learned kin recognition cues in a social bird. Nature. 434:1127-1130.

Suzuki TN, Wheatcroft D, Griesser M. 2016. Experimental evidence for compositional syntax in bird calls. Nat Commun. 7:10986.

Venables WN, Ripley BD. 2002. Modern Applied Statistics with S. 4th ed. New York: Springer.

Zuberbühler K. 2015. Linguistic capacity of non-human animals. WIREs Cogn Sci. 6:313-321.

Zuur A, Ieno EN, Walker N, Saveliev AA, Smith GM. 2009. Mixed Effects Models and Extensions in Ecology with R. New York: Springer.

Table S1. Full model set on the effect of cry type, group size, proportion of adult individuals and session (morning/evening) on a group's probability to approach a caller naturally producing SN or DN cries. All models included a caller's identity nested within its group affiliation as random term. DF=degrees of freedom, AICc=Akaike's information criterion corrected for small sample sizes.

Intercept	Group size	Cry type	Prop. adults	Session	Cry type x group size	Cry type x prop. adults	Cry type x session	DF	AICc	ΔAICc
0.565		+						4	167.3	0.0
0.182	0.080	+						5	169.2	1.9
0.615		+	-0.068					5	169.4	2.1
0.562		+		+				5	169.4	2.1
0.533	0.007	+			+			6	170.7	3.4
-0.251	0.114	+	0.365					6	171.3	4.0
0.647		+		+			+	6	171.3	4.0
0.161	0.081	+		+				6	171.4	4.0
0.739		+	-0.237			+		6	171.5	4.2
0.613		+	-0.076	+				6	171.6	4.3
0.143	0.038	+	0.324		+			7	172.9	5.6
0.570	0.003	+		+	+			7	172.9	5.6
0.301	0.067	+		+			+	7	173.3	6.0
-0.125	0.111	+	0.219			+		7	173.4	6.1
-0.251	0.115	+	0.368	+				7	173.5	6.2
0.661		+	-0.022	+			+	7	173.5	6.2
0.739		+	-0.256	+		+		7	173.7	6.4
0.961	-0.028	+	-0.356		+	+		8	174.6	7.3
0.675	-0.005	+		+	+		+	8	175.0	7.7
0.151	0.037	+	0.375	+	+			8	175.1	7.8
-0.107	0.101	+	0.363	+			+	8	175.5	8.2
-0.123	0.110	+	0.213	+		+		8	175.7	8.4
0.738		+	-0.139	+		+	+	8	175.7	8.4
0.966	-0.029	+	-0.309	+	+	+		9	176.8	9.5
0.261	0.028	+	0.369	+	+		+	9	177.2	9.9
-0.034	0.099	+	0.266	+		+	+	9	177.7	10.4
0.971	-0.029	+	-0.262	+	+	+	+	10	179.1	11.8
-3.328	0.307		2.056					5	209.5	42.2
-0.442								3	209.7	42.4
-1.109			0.974					4	210.6	43.3
-1.031	0.122							4	211.1	43.8
-3.336	0.312		2.160	+				6	211.5	44.2
-0.452				+				4	211.8	44.5
-1.094			1.023	+				5	212.7	45.4
-1.056	0.123			+				5	213.2	45.9