Morphologically structured vocalisations in female Diana monkeys

Camille Coye\textsuperscript{1,2,3}, Klaus Zuberbühler\textsuperscript{2,3,4}, Alban Lemasson\textsuperscript{1}

\textsuperscript{1}\textit{Université de Rennes 1, Ethologie animale et humaine - EthoS, U.M.R 6552-C.N.R.S, Station Biologique, 35380 Paimpont, France};\textsuperscript{2}\textit{University of St Andrews, School of Psychology and Neuroscience, St Andrews, Fife KY16 9AJ, UK};\textsuperscript{3}\textit{Centre Suisse de Recherches Scientifiques, Taï Monkey Project, Adiopodoumé, Abidjan Ivory Coast};\textsuperscript{4}\textit{Université de Neuchâtel, Institut de Biologie, rue Emile Argand 11, 2000 Neuchâtel, Switzerland}

Corresponding author: E-mail: camille.coye@univ-rennes1.fr; Phone: 00332.99.61.81.55
Postal address: Station Biologique de Paimpont, 35380 Paimpont

Social complexity is often thought of as a driving force in the evolution of communication and cognition, but this is at odds with the fact that non-human primates generally display only very limited flexibility in vocal production. Some primates partially overcome their limited vocal flexibility by combining two or more acoustically inflexible calls into complex sequences. Equally relevant is that some primate calls consist of separable morphological elements whose combinations create different meanings. Here, we focus on the vocal system of wild female Diana monkeys (\textit{Cercopithecus diana diana}), who produce three call units (R, L, A) either singly or merged as RA or LA call combinations. Previous work has shown that R and L convey information about external events, while A conveys information about caller identity. We tested this hypothesis experimentally, by broadcasting artificially combined utterances to eight adult females. To test the significance of the R and L ‘event’ units, we merged them with the A ‘identity’ unit of a group member. To test the significance of the ‘identity’ unit, we merged an R ‘event’ unit with an ‘identity’ unit from a group member or a neighbouring individual. Subjects responded in ways that suggested that both event and identity units were relevant, suggesting that Diana monkeys’ social calls possess morpho-semantic features. We discuss this finding in relation to the co-evolution of communication and social complexity in primates.

Keywords: acoustic playback, call combination, field experiment, guenons, morphology, social communication
INTRODUCTION

The evolution of vocal complexity in animals appears to be largely driven by social complexity as well as by visually difficult and ecologically challenging habitats (Bouchet, Blois-Heulin, & Lemasson, 2013; Dunbar, 1993; Dunbar, 1998; Marler, 1967; McComb & Semple, 2005). Many vertebrates and most primates live in social groups with complex and dynamic social networks and long-term bonds (de Waal, 1987; Lehmann, Korstjens, & Dunbar, 2007; Wrangham, 1987). As a result, primates are constantly challenged to maintain cohesion during travel and other activities to optimise foraging, to compete with neighbouring groups and to protect themselves against predators (Dunbar & Shultz, 2007; Lehmann et al., 2007; van Schaik, 1983; van Schaik & van Hooff, 1983). To this end, many species evolved specific vocalisations to maintain cohesion and synchronise within-group activities (Gautier & Gautier, 1977; Oda, 1996; Uster & Zuberbühler, 2001). Calls are often individually distinct and function to advertise individual identity or membership to specific social units (Bouchet, Pellier, Blois-Heulin, & Lemasson, 2010; Crockford, Herbinger, Vigilant, & Boesch, 2004; Dunbar, 2003; Neumann, Assahad, Hammerschmidt, Perwitasari-Farajallah, & Engelhardt, 2010; Rendall, Rodman, & Emond, 1996).

In light of this, it is surprising that non-human primates are thought to have relatively limited, species-specific vocal repertoires with a fixed set of call types that remain largely unchanged throughout adult life (review by Bouchet et al., 2013) and little signs of flexibility or voluntary control in call production (Hammerschmidt & Fischer, 2008). However, a more recent line of research has continued to demonstrate a previously under-described source of communicative complexity, namely the ability of individuals to assemble fixed acoustic units of their repertoire into more complex utterances. There is now good evidence that several non-human primate species produce calls in non-random sequences, with the information changing depending on
the order or temporal structure of call sequences (vervet monkeys *Chlorocebus aethiops:*
Seyfarth, Cheney, & Marler, 1980; Campbell’s monkeys *Cercopithecus Campbelli:* Lemasson,
Ouattara, Bouchet, & Zuberbühler, 2010; Ouattara, Lemasson, & Zuberbühler, 2009a;
Zuberbühler, 2001; white-handed gibbons *Hylobates lar:* Clarke, Reichard, & Zuberbühler,
2006; bonobos *Pan paniscus* and chimpanzees *Pan troglodytes:* Clay & Zuberbühler, 2011;
Slocombe & Zuberbühler, 2005; Diana monkeys *Cercopithecus diana diana:* Zuberbühler,
2000). One argument has been that these combinatorial capacities evolved in primates to enable
more complex communication (Zuberbühler & Lemasson, 2014).

So far, however, most examples of call combinations are from studies on male primate alarm
calls, which is surprising because social events may be at least as complex as dealing with
predator encounters. Hence, if complex vocal abilities have evolved to deal with social
complexity, we should find combinatorial phenomena in vocal behaviour during social
interactions and also in females, the social core of primates species (Buzzard & Eckardt, 2007;
Smuts, Cheney, Seyfarth, Wrangham, & Struhsaker, 1987).

One particularly promising candidate is the contact call of some forest living female guenons.
In Campbell’s monkeys, for instance, adult females produce a short, low-pitched contact call,
either as a single unit or merged with a second long, arched, and frequency-modulated unit to
form multi-unit utterances (Lemasson & Hausberger, 2011). The acoustic structure of the first
unit varies depending on the degree of arousal experienced by the caller (Lemasson, Remeuf,
Rossard, & Zimmermann, 2012) and contains fewer identity cues than the second, arched unit,
which strongly relates to the caller’s identity (Lemasson & Hausberger, 2011; Lemasson,
Hausberger, & Zuberbühler, 2005; Lemasson, Ouattara, Petit, & Zuberbühler, 2011). The
second unit is never uttered alone but functions as an affixation to the first unit, which can also be uttered alone.

A second relevant example is the contact calls of female Diana monkeys, an arboreal forest-dwelling primate living in groups of one adult male and seven to thirteen adult females with their offspring (McGraw, Zuberbühler, & Noë, 2007). As in most primates, the females are the philopatric sex and constitute the social core of the group (Candiotti et al., 2015). They produce, amongst others, three acoustically distinct social calls (L, R and A) depending on context (Candiotti, Zuberbühler, & Lemasson, 2012a, 2012b; Uster & Zuberbühler, 2001): L calls are mostly given in socio-positive and neutral events (e.g., foraging, affiliative interactions). R calls are mostly given in socio-negative events and mild danger (e.g., conflict within or between groups, walking on the ground), suggesting that these calls relate to the external events or emotional valences experienced by the caller. A calls, finally, are produced in unspecific ways to a large variety of events, but here the acoustic structure varies substantially between individuals, suggesting they function to signal the caller’s identity, similar to what has been found in Campbell’s monkeys. The three call types can be emitted alone (A, L, R) or merged as two combined utterances, either LA, or RA. Combined structures thus contain information about the external event (L or R) and the caller’s identity (Candiotti et al., 2012a, 2012b), with some interesting parallels to the function of morphemes in human speech (Collier, Bickel, Schaik, Manser, & Townsend, 2014; Hurford, 2008; Tellier, 2008; Veselinovic, Candiotti, & Lemasson, 2014).

Here, we tested experimentally whether the information conveyed by complex calls of Diana monkeys is compositional, i.e., whether the combined calls relate linearly to the information conveyed by the units given singly, as suggested by Candiotti et al.’s (2012a) observational data. To this end, we broadcast artificially combined calls to different subjects, eight female
Diana monkeys belonging to a study group habituated to human presence. We created experimental stimuli by manipulating either the initial or final call unit using recordings from group members, neighbours and completely unfamiliar individuals. Our goal was to test the significance of the L and R ‘event’ and A ‘identity’ units. To this end, we merged L and R units with A ‘identity’ units from familiar group members or neighbouring individuals. We predicted that, if combinations of call units were meaningful to receivers, L and R units should cause significant behavioural differences, particularly in terms of vocal responses, vigilance and exploratory behaviours. Because Diana monkeys are highly territorial, we also predicted different behavioural responses to identity-encoding A units, depending on whether they originated from a group member or neighbour.

**MATERIAL AND METHODS**

**Study site and subjects**

Field experiments were conducted between June and September 2014 in Taï National Park, Ivory Coast (5°50’N, 7°21’W). The experimenter (CC) and two field assistants (FB and FG) conducted playback experiments in a free-ranging group of Diana monkeys (*Cercopithecus diana diana*) with individually known subjects habituated to human presence for more than twenty years. At the time of the experiments, the group consisted of one adult male and eight adult females with their offspring. The experiment involved all the adult females of the habituated group.

**Playback stimuli**

All calls (stimuli and subject’s reaction to the playbacks) were recorded using a Sennheiser K6/ME66 directional microphone connected to a Marantz PMD660 recorder (sampling rate 44.1 kHz, resolution 16 bits, WAV sound format) in Taï National Park. Calls from group
members were recorded in May 2014; calls from unfamiliar and neighbouring Diana monkeys were recorded in June-July 2013 and February-June 2010. All calls were recorded under similar environmental conditions and distances to ensure high quality, low background noise, and no overlap with any other sound.

When creating the playback stimuli, we followed Candiotti et al., (2012a) classification (Fig. 1), who define L call units as continuous low-pitched trills with a general ascending frequency modulation (duration±SD: 409±106 ms, Minimum fundamental frequency±SD: 247±84 Hz, Maximum fundamental frequency±SD: 654±354 Hz), R call units as rapid repetitions of one to four short atonal units separated by brief periods of silence (duration±SD: 82±29 ms, Minimum fundamental frequency±SD: 331±170 Hz, Maximum fundamental frequency±SD: 429±199 Hz, First unit duration±SD: 28±11 ms, First inter-unit silence±SD: 46±18 ms) and A call units as tonal, arch-shaped frequency modulations (duration±SD: 298±105 ms, Minimum fundamental frequency±SD: 324±233 Hz, Maximum fundamental frequency±SD: 3090±696 Hz).
We created twenty-four different playback stimuli to generate the following three categories (Fig. 2): LA\(_G\): combination of an unfamiliar individual’s L merged with an A from an adult female group member (N=8); RA\(_G\): combination of an unfamiliar individual’s R merged with an A from an adult female group member (N=8); RA\(_N\): combination of an unfamiliar individual’s R merged with an A from an adult female from a neighbouring group (N=8).

<table>
<thead>
<tr>
<th>Introductory Unit</th>
<th>+</th>
<th>Affix</th>
<th>Stimulus</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="#" alt="Circle L" /></td>
<td></td>
<td><img src="##" alt="A(_G)" /></td>
<td><img src="##" alt="Circle L A(_G)" /></td>
</tr>
<tr>
<td><img src="#" alt="Triangle R" /></td>
<td></td>
<td><img src="##" alt="A(_G)" /></td>
<td><img src="##" alt="Triangle R A(_G)" /></td>
</tr>
<tr>
<td></td>
<td></td>
<td><img src="##" alt="A(_N)" /></td>
<td><img src="##" alt="Triangle R A(_N)" /></td>
</tr>
</tbody>
</table>

Figure 2: Stimulus creation. The geometric shapes indicate the valence of the call. Circle and triangle represent context-dependent units, respectively L (associated with positive and neutral events) and R (associated with negative events). Rectangles represent identity-dependent units A. The shading indicates the origin of the call recorded: white: unfamiliar individuals; black: group-members; grey: neighbouring individuals.

Each of the eight subjects received its own set of LA\(_G\), RA\(_G\) and RA\(_N\) call combinations. Within a given set, we systematically used the same R call unit and the same A call unit to create paired stimuli (i.e., L\(_A\)_G-R\(_A\)_G and R\(_A\)_G-R\(_A\)_N) to allow the comparison of the changes in subject’s reaction due to changes in only one part of the call. L and R call units were systematically
extracted from naturally produced LA and RA call unit combinations. We further made sure that all unfamiliar L and R call units came from different individuals by using recordings from a different group. Only R calls composed of double units were used. ‘A’ call units were from identified and habituated adult females of the focal group (Ag) or the neighbouring group (An). For each focal female, group or neighbour identities were pseudo-randomly attributed. This was done to avoid complete pair-matching between group members: if the call from female A was used as a stimulus for female B, then the call from female B was not selected to serve as stimulus for female A, to prevent any particular social relationship between two individuals to be over-represented. ‘A’ call units can be subdivided into full arches (Af subtype) or broken arches (Ab subtype) (Fig. 1). Females differ in how they make use of this feature, with some females mainly using ‘Af’ or ‘Ab’ subtypes (Candiotti et al., 2012a). When editing playback stimuli, we used a group-member’s most typical A subtype which was then matched with the corresponding subtype for the neighbouring female stimulus.

Playback stimuli were created using Raven Pro 1.4 Software. Call exemplars were selected based on recording quality and call duration (in seconds: mean ± SE: L = 0.170 ± 0.012 s, R = 0.108 ± 0.002 s, A = 0.293 ± 0.014 s). We amplified stimuli when necessary to obtain a naturalistic intensity. Final tests of broadcast intensity were made in the Taï National Park to make sure that stimuli’s intensity was homogeneous and appropriate according to the natural background noise.

Experimental protocol

Three observers followed the study group from 9:00 to 17:00 local time. Before the first experiment, we ran a habituation phase on several consecutive days during which the equipment was installed underneath the middle of the group in an open area so that it was fully visible to
the subjects. We repeated this until the subjects stopped giving alert calls and lost interest in the equipment. Before starting an experimental trial, we ensured that the group was not travelling or foraging 30 m or higher, that no neighbours were in the vicinity and that no male loud calls had been produced for at least 15 min. The experimenters then selected the subject and positioned the playback equipment at an elevation of 4 to 6 m above ground using a telescopic perch, either in periphery of the group (stimuli made from neighbouring female calls) or within the group. For within group trials, we kept the speaker about 5 m to 10 m away from the A call unit providing female. We did not wait for the subject to move to a specific position within the group and tested her where she was. Hence, playbacks of intra-group calls have been given from varied positions (more or less peripheral) in the group, without any obvious consequence.

For each playback, CC continuously observed the subject, while FB and FG followed the call provider (in-group trials) and handled the equipment, respectively. Before initiating a trial, we ensured that (1) the subject was fully visible, (2) the call provider was 5 to 10 m away from the speaker, and (3) no call was given by any group member for at least 8 s.

Playback stimuli were broadcast from a Marantz PMD660 solid-state recorder connected to a Nagra DSM speaker/amplifier and a Bose 151 Environmental speaker that had been mounted to a telescopic perch. We recorded the behaviour of the subject for 30 s and the group’s vocal behaviour for 60 s following each trial.

We never performed more than two trials per day and never for more than two days in a row. We ensured that we never broadcast two combined calls of the same category within the same day and we did not test the same individual in two consecutive experiments. Each day, we performed one to three “mock” experiments (even on days when no experiment was scheduled)
by executing the full experimental protocol, but no sound diffusion, to prevent subjects from anticipate a trial.

Dependent variables

During the 30 s post-playback period, we described the behaviour of the subject as the total duration (s) and frequency (i.e., the number of times a behaviour was observed) of its posture (i.e., sitting, standing or in vigilance posture), locomotion (i.e., immobile, walking, running or jumping) and direction of gaze (i.e., looks at the speaker, above, under, at a conspecific, at the observer and scans the environment). We also scored the latency in seconds for four behaviours: adopting a vigilance posture, sitting, walking and looking at the speaker. Overall, this resulted in 32 variables to describe each subject’s behavioural response.

During the 1 min post-playback period, we also quantified the group’s vocal activity using the following variables: the latency to give first call (any call type), the number of social calls (Coye, Ouattara, Zuberbühler, & Lemasson, 2015), the number of alert calls (Coye et al., 2015), the number of calls given during vocal exchanges (a sequence of any calls separated by less than 3 s of silence), the number of isolated calls (any call given more than 3 s before or after another call), as well as the number of vocal exchanges (involving any call type) and the average number of calls involved in a vocal exchange.

Statistical analysis

We expected the subjects’ responses to differ in several subtle behavioural indicators simultaneously (notably locomotion, vocalizations and direction of gaze), suggesting that multivariate testing was most appropriate to deal with the potential co-variation of the variables. Any pair of variables with an R Pearson’s correlation coefficient above 0.7 was considered
colinear in case of which we systematically deleted one member of the colinear pair (Dormann et al., 2013; Katz, 2011). Then, we used a Linear Discriminant Analysis as a preliminary guide for variables selection but did not use this method for further statistical analysis due to repeated measures present in our data (see Mundry & Sommer (2007) for details on LDA and discussion of the case of repeated measures). This first, exploratory, step led to the selection of a subset of seven biologically relevant variables susceptible to represent subjects’ reaction (group’s vocal behaviour, subject locomotion and gaze direction) across the experimental conditions. We used five quantitative variables: Number of isolated calls (given more than 3 seconds before or after another call), Latency to give first call (s), Time spent walking (s), Latency before locomotion (s), Duration of first look to the speaker (s), and two binary variables: Presence/absence of Look towards the observer, and Presence/absence of Visual scanning of the environment (Fig. 3).

We calculated Gower’s dissimilarity index between samples in the dataset (daisy{cluster}, R statistical software, Maechler et al., 2015). This index “summarises” the difference between two samples into a measure of distance, based on the samples’ values for each variable included in the analysis. Gower’s distance is a common method which allows the use of various types of variables (binary, ordinal, nominal and quantitative variables) (Gower, 1971; Oksanen et al., 2007; Podani, 1999). The two binary variables were treated as symmetric variables and no standardisation of variables was applied.

To study the impact of the introductory unit (L or R units) and the impact of the affix (A units from a group-member or a neighbour), we performed two separate non-parametric MANOVAs (Adonis{vegan}, R statistical software, Oksanen et al., 2007) on the matrixes of Gower’s dissimilarity index, giving the distance between trials in LA\textsubscript{G}-RA\textsubscript{G} and RA\textsubscript{G}-RA\textsubscript{N} conditions respectively. Both NPMANOVA\textsubscript{s} were two-tailed, included the type of stimulus and the...
identity of the subject as factors and were conducted using free permutation of the distance
matrixes, as suggested by Anderson (2001) and Gonzalez and Manly (1998) for small datasets.

NPMANOVA is a non-parametric multivariate method involving the calculation of an F-ratio
on an index of distances between samples. The computation of a p-value, like any other
permutational test, involves a comparison of the test value (i.e., the F-ratio) obtained on the
original dataset (i.e., the distance matrix) with test values computed on random permutations of
the same dataset. See Anderson (2001) for detailed explanation of the method and equations,
as well as Adams and Anthony (1996) for a discussion of the use of permutational tests on
behavioural data.

We completed the analysis with graphic representation of the results to describe the nature of
the behavioural changes (Fig. 3). We computed effect size for each variable included in the
multivariate testing. We used Cliff’s delta for the quantitative (i.e., count and continuous)
variables (cliff.delta{effsize}, R statistical software, (Torchiano, 2015) using the original
formula proposed by (Cliff, 2014). And we used risk difference (riskdifference{fmsb}, R
statistical software (Nakazawa, 2015) with a 95% confidence interval for the binary variables.

Ethical note

Ethics approval was given by the St Andrews’ University Ethics Boards; the research protocol
was authorised in Côte d’Ivoire, by the Minister of Scientific Research and the ‘Office Ivoirien
des Parcs et Réserves’ (OIPR). This study does not raise major issues regarding animal welfare.
Study groups have been habituated to human presence and followed on a regular basis since
1990 while the continued presence of researchers and field assistants has had a significant
impact on decreasing firearms-based poaching activities in the area. The habituation to the
Playback equipment was conducted smoothly. Moreover, the call types broadcast during the playbacks are naturally given at relatively high frequency: LA calls: 19.8 calls per hour, RA calls: 2.7 calls per hour (Candiotti et al. 2012a). Intergroup encounters, as simulated by playbacks of RA_N combinations (involving A calls from a neighbour), occur on average once every three days (McGraw et al., 2007 p59). No playback enhanced male alarm calling behaviour or triggered any sign of group panic or other abnormal behaviour.

RESULTS

Impact of the Introductory Unit on subjects’ reaction:

When analysing the impact of the introductory unit, the NPMANOVA showed a significant impact of the type of stimulus ($F_{1,7}= 3.37, P=0.043$) and no significant effect of the subject’s identity ($F_{7,7}= 1.71, P=0.142$). Graphic representation of the variables measured, combined with measures of effect size, show that test subjects expressed distinct behavioural patterns in the different experimental conditions (Fig. 3). Playbacks of LA_G (‘positive’ introduction, A from a group member) and RA_G (‘negative’ introduction, A from a group member) stimuli caused differences in locomotion, vocal behaviour and gaze direction (Fig. 3). Latency before locomotion appears shorter after playbacks of RA_G than LA_G (medium effect size: $N=16, Cliff’s delta=-0.47$), although we found no clear difference in the time spent walking (negligible effect size: $N=16, Cliff’s delta=-0.125$). The group gave more isolated calls (i.e., calls not part of a vocal exchange) in the RA_G than in the LA_G condition, with a medium effect size ($N=16, Cliff’s delta=-0.47$). The latency to give a first call was shorter and much less variable in the RA_G than in the LA_G condition although only a negligible effect was detected ($N=16, Cliff’s delta=0.125$). Finally, the duration of the first look towards the speaker ($N=16, Cliff’s delta=-0.031$, negligible effect) and the presence of looks towards the observer ($N=16, RD=-0.125, P=0.285$)
did not change between LA<sub>G</sub> and RA<sub>G</sub> conditions (0% of the LA<sub>G</sub> trials, 12.5% of the RA<sub>G</sub> trials), but subjects scanned the environment more after the playback of the negative (i.e., RA<sub>G</sub>) stimulus (37.5% of the trials) than after playbacks of LA<sub>G</sub> (0% of the trials; N=16, RD=0.375, P=0.028).

Figure 3: Graphic representation comparing subjects’ reaction to LA<sub>G</sub> and RA<sub>G</sub> stimuli (N=8 for each condition). Box and Whisker plots show first quartile, median and third quartile, whiskers show the minimum and maximum values except for outliers (i.e. values higher than 1.5 times the interquartile range). Plots (a-e) show respectively time spent walking, latency before locomotion, number of isolated calls, latency before the first call and duration of the first look towards the speaker. See Methods section 2. and 4. for details on experimental conditions and variables.
Impact of the affix on subjects’ reaction:

When analysing the impact of the affix, the NPMANOVA showed a significant impact of the type of stimulus involved ($F_{1,7}=4.29, P=0.02$) and no significant role of subject’s identity ($F_{7,7}=1.66, P=0.155$). Graphic representation of the variables, combined with measures of effect size, show that test subjects expressed distinct behavioural patterns in the experimental conditions (Fig. 4): Comparison between responses to RA$_G$ (‘negative’ introduction, A from group member) and RA$_N$ (‘negative’ introduction, A from a neighbour) stimuli highlighted differences in locomotion, vocal behaviour and gaze direction but with a different pattern from the one found for the impact of the introductory unit (Fig. 4). Subjects spent less time walking ($N=16$, Cliff’s $delta=0.31$, small effect size) and had a greater latency before locomotion ($N=16$, Cliff’s $delta=-0.38$, medium effect size) after playback of RA$_N$ than RA$_G$ stimuli. The results also suggest slight differences in group’s vocal behaviour: groups gave fewer isolated calls ($N=16$, Cliff’s $delta=0.33$) and displayed slightly greater and more variable latencies when giving the first calls ($N=16$, Cliff’s $delta=-0.23$) after playbacks of RA$_N$ than RA$_G$ stimuli. Finally, the pattern of gaze direction differed strongly between RA$_G$ and RA$_N$ stimuli: after playbacks of RA$_N$ stimuli, subjects’ first looks to the speaker were longer ($N=16$, Cliff’s $delta=-0.73$) with a large effect size. Subjects looked more at the observer ($N=16$, $RD=-0.375$, $P=0.077$) in the RA$_N$ condition (50% of the RA$_N$ trials vs 12.5% of the RA$_G$ trials) but visual scanning of the environment did not seem to differ strongly between the playback of RA$_G$ (37.5% of the trials) and RA$_N$ (12.5% of the trials) stimuli ($N=16$, $RD=0.25$, $P=0.23$).
Figure 4: Graphic representation comparing subjects’ reaction to RA\textsubscript{G} and RA\textsubscript{N} stimuli (N=8 for each condition). Box and Whisker plots show first quartile, median and third quartile, whiskers show the minimum and maximum values except for outliers (i.e. values higher than 1.5 times the interquartile range). Plots (a-e) show respectively time spent walking, latency before locomotion, number of isolated calls, latency before the first call and duration of the first look towards the speaker. See Methods section 2. and 4. for details on experimental conditions and variables.
In this study, we demonstrated experimentally that Diana monkeys responded differently to social calls composed of different morphological units in ways that suggested that at least two levels of information were conveyed. Morphological compounds consisted of L or R units, which related to different external events experienced by the caller (Candiotti et al., 2012a), and of A units, which related to caller identity. Response pattern suggested that recipients attended to these different levels of information conveyed by the call compounds. Specifically, our findings supported the idea that the initial morphemic unit of a combined call (L or R) encodes information about the social context experienced by the caller. R call units are typically associated with negative events, such as the detection of mild danger. Here, subjects responded with isolated social calls, prolonged latency before locomotion and scanning of the environment compared to L call units (typically associated with neutral and positive events). Overall, these results suggest that recipients associated the R and L call units with distinct socio-environmental contexts and adapted their behaviour accordingly.

The ability to reveal one’s motivational states is a well-known function of animal communication (Briefer, 2012; Lemasson et al., 2012; Schehka & Zimmermann, 2009; Taylor & Reby, 2010), although it is often difficult to make a compelling argument about the exact nature of the underlying inner processes. Interestingly, the acoustically homologous call of Diana’s monkey L unit in Campbell’s monkeys increases in duration and frequency according to presumed differences in arousal (Lemasson et al., 2012). Here, we confirm the importance of this acoustic component to convey information about the emotional context, but we also show a different use in Diana monkey, whose social calls (L vs R) relate to the general valence of the external world as perceived by the caller. Importantly, L and R call units can be emitted singly or, more often, combined with A units into a compound call. More detailed contextual
analyses are required to determine which social situations are associated with single or combined calls. For example, it is possible that the distance between the caller and the receiver determines whether an A unit is affixed. Another possible explanation lies in variations in the degree of visibility in the habitat (Candiotti et al., 2012a, 2012b).

Call compounds that contained A units from a neighbouring individual (‘RA_n’) triggered a high decrease of locomotion, a slight decrease of vocal activity and an increased visual scanning towards the presumed caller compared to call compounds that contained A units from a group-member (‘RA_g’). This pattern is similar to what has been observed in other primates reacting to unexpected stimuli (Bergman, Beehner, Cheney, & Seyfarth, 2003; Briseño-Jaramillo, Estrada, & Lemasson, 2014; Zuberbühler & Wittig, 2011). These behavioural patterns suggest that both RA stimuli were perceived as urgent but that subjects based decisions on differences in the Affix.

Individual acoustic variations and auditory discrimination by receivers have been reported in many primate species (putty-nosed monkeys Cercopithecus nictitans: Price, Arnold, Zuberbühler, & Semple, 2009; marmosets Callithrix jacchus: Miller & Thomas, 2012; Olive baboons Papio hamadryas anubis: Lemasson, Palombit, & Jubin, 2008; Japanese macaques Macaca fuscata: Ceugniet & Izumi, 2003; squirrel monkey Saimiri sciureus: Kaplan, Winship-Ball, & Sim, 1978), suggesting that providing identity cues in primate calls is of considerable biological importance (Blumstein, Verneyre, & Daniel, 2004; Lemasson et al., 2007; Seyfarth et al., 2010; Seyfarth & Cheney, 2010).

This is likely so in this species in which vocal exchanges play an important role to ensure social cohesion and in which females differ in their vocal activity depending on their social integration within the group (Candiotti et al., 2015). In line with this, a previous study on Campbell’s
monkeys found that familiar calls (using homologous calls of Diana monkeys’ LA) elicited more affiliative calling and vocal responses than unfamiliar calls (Lemasson et al., 2005). Taken together, the responses given by the subjects to the different types of stimuli suggest that the first unit (i.e., L or R) allows the receiver to get information about the direct social and physical environment—probably by associative learning—while the identity conveyed by the second unit (A) may influence receiver’s decision regarding the behaviour to adopt in line with their respective positions in the social network. For example, receivers may have different reactions depending on the identity of the caller when hearing an RA call which signals that a given individual has spotted something disturbing. But it seems premature to draw stronger conclusions about the relative importance of both types of information conveyed as here no playback of LA\textsubscript{N} call was done. In future experiments, it would therefore be necessary to test subjects’ reaction to L call units combined with A call units from neighbours as well as L call units combined with A calls from immature or more or less affiliated group members.

Our study has high external validity because the data are from spontaneous reactions of untrained and free-ranging animals living in their natural habitat. The results obtained here suggest that the main social calls given by the adult females are linear combinations of different morphological units that convey information about the social context and the identity of the caller. In a related study based on analysis tools from formal linguistics (Veselinovic et al., 2014), call sequences of wild adult females Diana monkeys were analysed which revealed non-random patterns in terms of the order and type of calls units that were merged. The authors concluded that calls consisting of combinations of call units functioned as single calls rather than rapid sequences of independent units (Veselinovic et al., 2014).
Nevertheless, this study remains only a first step towards understanding the use of complex calls and combinatorial abilities. Several technical and conceptual limitations must be acknowledged. Firstly each combined stimulus was created from call units taken from two distinct individuals and we do not know if the same reactions would be observed if we combined calls from the same caller. This choice was based on evidence in Campbell’s monkey, a closely related species, that calls homologous to L and R calls in Diana monkeys (i.e., SH and RRC calls) relate to a much lesser extent to caller’s identity than the arched structure homologous to Diana monkeys’ A calls (i.e., CH) (Lemasson and Hausberger, 2011). Hence, if those data suggest that receivers discriminate caller’s identity mostly from the arched part of the call, its actual importance in L and R calls remains to be tested in Diana monkeys. However, the fact that LA_G did not trigger any reaction showing that subjects were disturbed supports our hypothesis.

Moreover, in our experimental design, the L and R units of stimuli were taken from combined calls but A units were taken from calls emitted alone (i.e., not combined to another unit). The question remains whether the acoustic structure of A calls (when combined or not) differs slightly. But again, the fact that LA_G (socio-positive calls) did not trigger disturbed reactions in subjects suggests that this question may remain peripheral for the results obtained here. Testing this question could generate an interesting comparison with males Campbell’s monkeys in which previous work found similar responses to natural Krak calls and artificial Krak calls created from Krak-oo calls (Coye et al., 2015).

An alternative solution to determine both the potential of the first unit (i.e., L or R) to signal caller’s identity, and the influence of a possible variation in the structure of A calls’ between combined and single calls would be to develop a playback experiment comparing sets of artificial stimuli. Notably a set of artificially combined calls created from single calls (i.e., L, R
and A calls combined in LA and RA complex calls) and a set of artificial ‘single calls’ taken from complex calls (i.e., breaking down LA and RA calls into L, R and A units), using either calls from a group-member or calls from a neighbour.

Interestingly, combinatorial abilities have been found in several animal species both at the call level and at the sequence level, in both males and females, and notably in species where males are more integrated socially (Bouchet et al., 2013; Bouchet, Laporte, Candiotti, & Lemasson, 2014; Bouchet et al., 2010; Coye et al., 2015; Lemasson, 2011; Ouattara, Lemasson, & Zuberbühler, 2009b). Although most studies based their conclusions on contextual and acoustic (non-experimental) analyses, they all suggest that combinatorial phenomena are an evolutionary adaptive response to an increased need for complex communication, which may be more widespread than initially thought. Whether the combinatorial abilities of non-human and human primates originated in an ancestral capacity or result from convergent evolution remains unclear and will require further comparative studies, notably to investigate phylogenetic and cognitive aspects of the evolution of combinatorial phenomena.

**Acknowledgements**

Research has been funded by the French Ministry of Research, Institut Universitaire de France, ANR ‘Orilang’ and the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 283871. The Ethics approval was given by the St Andrews’ University Ethics Boards (UK). In Côte d'Ivoire, we thank the Minister of Scientific Research and the ‘Office Ivoirien des Parcs et Réerves’ (OIPR) for permission to conduct research in Taï National Park. We thank A. Bitty, the Centre Suisse de Recherches Scientifiques and B. Diero for logistic support, our field assistants F. Bélé and F. Gnepa for their invaluable help with data collection, the Taï Chimpanzee Project (TCP) and the ‘Centre de Recherche en Ecologie’ (CRE)
for their support in the field. We thank Brandon Wheeler and an anonymous referee for very relevant comments which greatly improved this article. We are very grateful for comments and fruitful discussions from C. Neumann and C. Rochais.

REFERENCES


Veselinovic, D., Candiotti, A., & Lemasson, A. (Under revision). Female Diana monkeus (Cercopithecus diana) have complex calls. *LI Squibs.*


