ORGANISMAL BIOLOGY

Titi monkeys combine alarm calls to create probabilistic meaning

Mélissa Berthet[1,2]*, Geoffrey Mesbahi[1], Aude Pajot[1], Cristiane Cäsar[3,4,5], Christof Neumann[1,6]†, Klaus Zuberbühler[1,3]†

Previous work suggested that titi monkeys *Callicebus nigrifrons* combine two alarm calls, the A- and B-calls, to communicate about predator type and location. To explore how listeners process these sequences, we recorded alarm call sequences of six free-ranging groups exposed to terrestrial and aerial predator models, placed on the ground or in the canopy, and used multimodel inference to assess the information encoded in the sequences. We then carried out playback experiments to identify the features used by listeners to react to the available information. Results indicated that information about predator type and location were encoded by the proportion of B-call pairs relative to all call pairs of the sequence (i.e., proportion of BB-grams). The results suggest that the meaning of the sequence is not conveyed in a categorical but probabilistic manner. We discuss the implications of these findings for current theories of animal communication and language evolution.

INTRODUCTION

One reason to study animal signals is to understand how linguistic reference has evolved. One relevant question is whether animals can use parts of their signal repertoire to refer to external events. Pioneering evidence has been provided by fieldwork with vervet monkeys (*1*), which triggered an important debate about whether animal signals really refer to external events or whether they are mere reflections of some unspecified internal states, elicited by the external events. This debate partly originates from the fact that little is known about whether or how animals represent the external world as mental concepts and whether this differs from the way humans do (*2*). More recently, an additional complexity has been added to the debate, due to the fact that some animal signals are organized sequentially (*3*), providing a further potential source of information based on the combinatorial properties of signal sequences.

Black-fronted titi monkeys *Callicebus nigrifrons* have contributed to this literature because adults produce two alarm calls, the A- and B-calls (fig. S1), which can be combined into complex sequences. A previous study (*4*) suggested that alarm call sequences varied not only with predator type (A-calls were mainly given to aerial predators, while B-calls were given to a large set of disturbances, including terrestrial predators) but also with predator location: When aerial predators were on the ground, B-calls were interspersed within the A-call sequences. When terrestrial predators were detected in the canopy, B-call sequences were always introduced by a single A-call. However, this study was based on a small sample size and investigated a few encoding mechanisms, and there was no experimental evidence that the encoded dual information (predator type and location) was perceived by receivers.

In the present study, we were interested in how titi monkeys produced and perceived information in their alarm sequences. To this

end, we carried out systematic predator model presentations following a 2 × 2 design (two predator types crossed with two locations) and playback experiments (four types of response sequences) with observer-habituated wild titi monkeys. We analyzed the alarm call sequences given in response to experimental stimuli by extracting 15 quantitative variables (referred to as "sequence metrics"; see table S1 and Materials and Methods) and assessed what information was conveyed by these metrics using multimodel inference. We compared behavioral responses to the broadcasting of different call sequences to determine which information and sequence metrics titi monkeys attended to.

RESULTS

What do alarm sequences encode?

In the first experiment, we presented models of two predator types (terrestrial and aerial predators) placed in two different locations (on the ground and in the canopy) to 34 individuals from six groups of monkeys. We obtained $n = 50$ alarm call responses and characterized each sequence by 15 different sequence metrics. We used multimodel inference (*5*) to investigate whether each metric conveyed information about predator type and/or location. We used model weights (*w*), derived from Akaike's information criterion (*6*), which represent the probability that each hypothesis (i.e., each predator type and location combination) is best supported by each metric, ranging from 0 (weak support) to 1 (strong support).

We found that several metrics encoded for predator type (Fig. 1, A and D, b), predator type combined with location (i.e., predator location acting in the same way for aerial and terrestrial predators; Fig. 1, A and D, c) and the interaction between predator type and location (i.e., predator location acting in different ways for aerial and terrestrial predators; Fig. 1, A and D, d). No metric encoded for location only, and for several metrics, the null models had the highest weights (Fig. 1A). Overall, these results suggest that titi monkeys mainly encode predator type with added or interactional information about predator location.

What information do monkeys attend to?

In a second experiment, we played back alarm call sequences of titi monkeys ($n = 28$ trials on 14 individuals), originally given in

[1]Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland. [2]Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, CNRS, PSL Research University, Paris, France. [3]School of Psychology and Neuroscience, University of St Andrews, St Andrews, UK. [4]Natural Sciences Museum PUC Minas, Belo Horizonte, Brazil. [5]Bicho do Mato Research Institute, Belo Horizonte, Brazil. [6]Laboratoire de sciences cognitives et de psycholinguistique, Département d'études cognitives, ENS, EHESS, CNRS, PSL University, Paris, France.
*Corresponding author. Email: melissa.berthet.ac@gmail.com
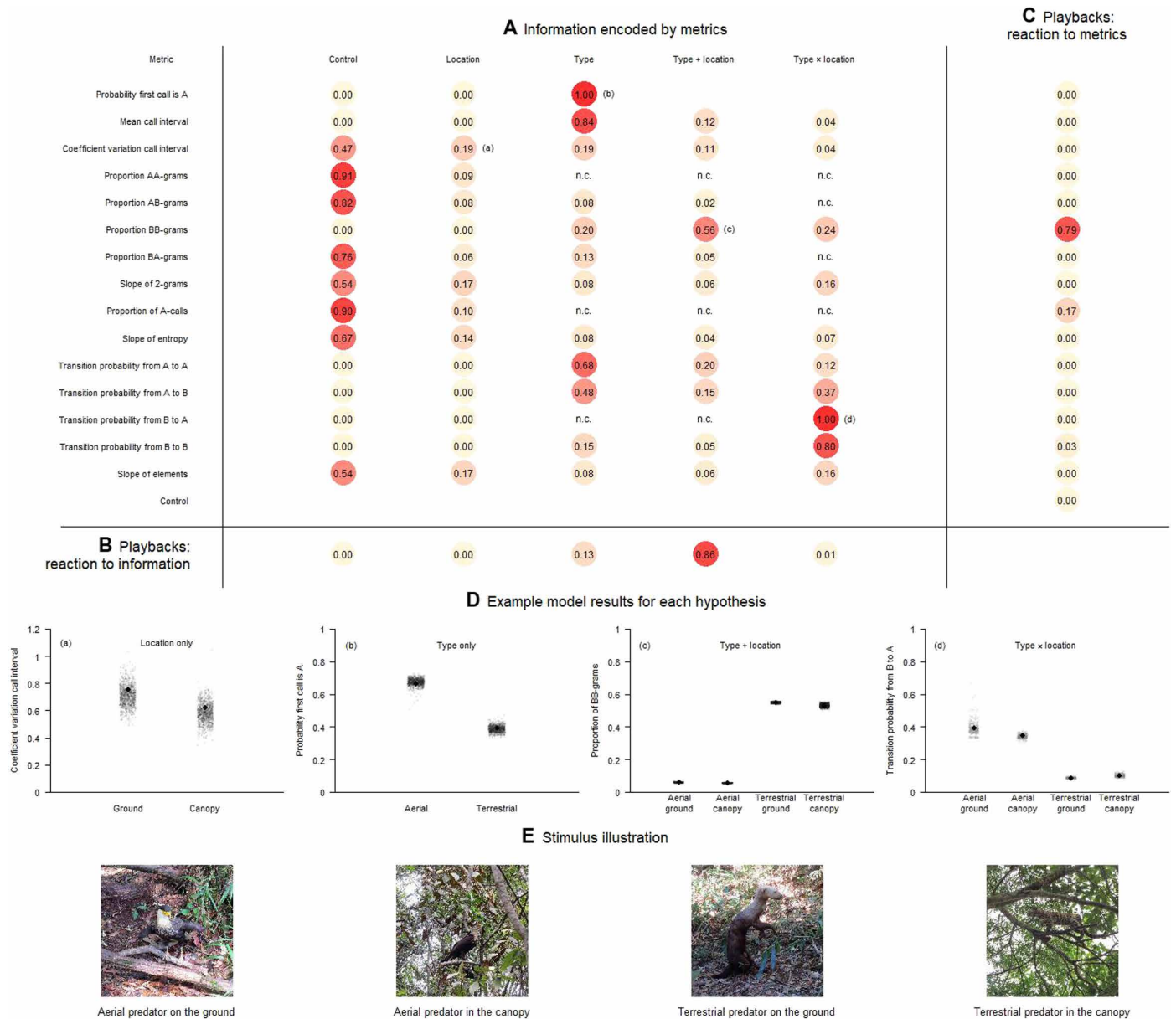†These authors contributed equally to this work.

Fig. 1. Results of the multimodel inference analyses. Circle colors in (**A**) to (**C**) refer to the Akaike's weight, i.e., the probability that a given model supports the hypothesis (white: $w = 0$, weak support; red: $w = 1$, strong support; n.c.: the model did not converge). (A) Information encoded in titi monkey alarm sequences: Metrics are presented row-wise, and information hypotheses are presented column-wise. For simplicity, the null and urgency models were combined as "control," and their weights were added. For the metric "probability that first call is A," models that addressed the possibility that predator type and location were encoded are not relevant because the first call can only be one of two possibilities and, thus, can only provide information about predatory type or location. (B) Gaze reaction of titi monkeys to the information contained within the playback stimuli sequences, i.e., the original condition during which broadcasted sequences were recorded. For a graphic representation of the best model (interaction between predator type and location), see Fig. 2. (**C**) Gaze reaction of the titi monkey to the metrics extracted from the playback stimuli sequences. For a graphic representation of the best model (proportion of BB-grams), see Fig. 4. (**D**) Illustration of sequence metrics that support each hypothesis. Letters refer to the corresponding model weights in (A). (**E**) Illustration of experimental design of the predator presentations.

response to natural or experimental predator encounters. Again, we used multimodel inference to investigate whether gaze direction of listeners was influenced by the origin of the sequence (i.e., sequence given to a terrestrial predator on the ground, a terrestrial predator in the canopy, an aerial predator on the ground, or an aerial predator in the canopy).

We found that monkeys attended most to information about predator type and location (model with main effects for type and location included, $w = 0.86$; Fig. 1B) and less to information about predator type only ($w = 0.13$). The remaining models representing information about the interaction between predator type and location, predator location only, or no information about predator type or
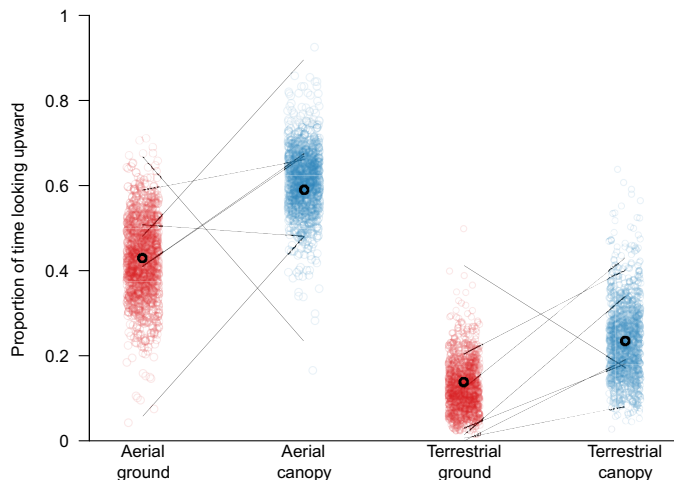
Fig. 2. Proportion of time the listener spent looking upward across original recording conditions of the playback stimuli. The figure shows raw data (one line per individual), as well as estimates per condition (black circles) and bootstrapped estimates (colored circles, 1000 bootstraps) of the model testing how gaze reaction depends on both predator type and location (main effects). Subjects looked more upward when they were presented with sequences elicited by an aerial predator (compared to a terrestrial predator) or elicited by a predator in the canopy (as opposed to a predator on the ground). For simplicity, we displayed the most salient reaction, i.e., looking upward. Results for other looking directions can be found in fig. S2.

Fig. 3. Proportion of BB-grams in the alarm call response depending on the eliciting stimulus. The figure shows estimates (black circles) and bootstrapped estimates per condition (colored circles, 1000 bootstraps) of the model testing how the proportion of BB-grams encodes both predator type and location (main effects). The proportion of BB-grams is higher in vocal responses to terrestrial predator than to aerial predators and higher when the predator is on the ground than when it is in the canopy.

location (null and urgency models) had a combined weight of 0.01 (Fig. 1B).

We then analyzed how the origin of the broadcasted sequence influenced the gaze reaction of the subjects. When hearing a sequence recorded from an encounter with an aerial predator, titi monkeys looked more upward and less toward the speaker than when the sequence was recorded from an encounter with a terrestrial predator (Fig. 2 and fig. S2). In addition, sequences recorded from encounters with predators in the canopy elicited more gazing upward and less toward the speaker than sequences recorded from predators on the ground (Fig. 2 and fig. S2). Looking upward is an appropriate response when expecting an aerial predator (that is usually in the air or in the canopy) or a predator located within the canopy. Looking toward the speaker is appropriate when expecting a terrestrial predator or a predator on the ground: Because of the density of the lower strata of the forest, spotting a predator on the ground can be difficult, and looking at the caller's behavior and gaze direction can provide cues about the exact location of the threat. Overall, these playback results suggest that titi monkeys can extract information about both predator type and location in an additive fashion from alarm sequences.

What sequential metrics do monkeys attend to?
In a final analysis, we assessed how the metrics characterizing the different call sequences used as playback stimuli affected the time listeners spent looking in predator-relevant directions, also using a multimodel inference approach. Here, we ignored the information content of the sequences (i.e., their origin) to focus on their sequence features only.

Model weights indicated that listeners reacted strongly to the proportion of BB-grams, i.e., the proportion of two contiguous B-calls among all the contiguous pairs of calls of the sequence ($w = 0.79$), and somewhat to the proportion of A-calls ($w = 0.17$) (Fig. 1C). All
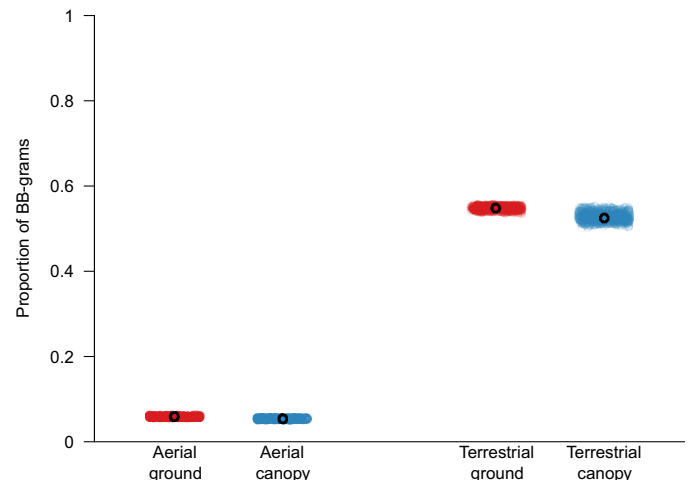
other models (including the null model) had a combined weight of 0.03 (Fig. 1C). Further inspection of the metric revealed that the proportion of BB-grams was substantially lower in sequences elicited by aerial predators than by terrestrial predators (Fig. 3). In addition, the proportion of BB-grams was slightly lower in sequences elicited by predators in the canopy than by predators on the ground (Fig. 3).

With regards to reactions toward playbacks, as the proportion of BB-grams in a sequence increased, listeners spent increasingly more time looking toward the speaker and increasingly less time looking upward, indicating that they expected more a terrestrial predator than an aerial predator and/or more a predator on the ground than in the canopy (Fig. 4). Thus, the playback results suggested that titi monkeys attended to the proportion of BB-grams to extract information about the predator type and location.

DISCUSSION
Our analysis shows that titi monkeys encode information about both predator type and location in their alarm sequences, albeit in ways that, to our knowledge, have not yet been described. Predator type and location were redundantly encoded by several sequence features, but none of the sequence metrics we investigated encoded for predator location only (Fig. 1A). To test whether recipients were able to attend to the information conveyed by these sequences, we carried out a playback experiment, with results showing that titi monkeys appeared to attend to the proportion of BB-grams (Figs. 1C and 4), i.e., the proportion of two contiguous B-calls, among all the contiguous pairs of calls of the sequence, that provided them with information of both predator type and location (Figs. 1A, 1B, 2, and 3). The proportion of BB-grams mainly encoded predator type and less predator location (Fig. 3), but our playbacks suggested that receivers were able to extract both information (Figs. 2 and 4).
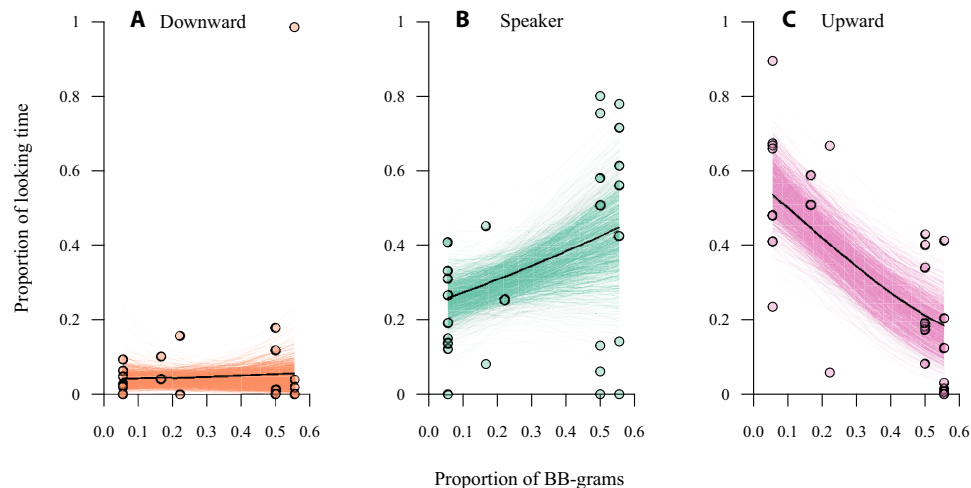
Fig. 4. Listener's gaze reaction, depending on the proportion of BB-grams of the alarm sequence. Proportion of time listeners spent looking downward (A), toward the speaker (B), and upward (C), depending on the proportion of BB-grams of the playback stimuli. The figure shows raw data (circles), as well as estimates (black lines) and bootstrapped estimates (colored lines, 1000 bootstraps) of the model testing how gaze reaction depends on the proportion of BB-grams. Listeners spent more time looking toward the speaker (B) and less time looking upward (C) when there were more BB-grams in the sequence. The time looking downward (A) was not affected.

These results corroborate earlier work that proposed that titi monkey alarm sequences encode predator location and type (4). Cäsar *et al.* (4) described three encoding mechanisms at the sequence level: the call rate and the proportion of A- and B-calls encoded for predator type only, and the insertion of either B-calls into an A-sequence or one single A-call at the beginning of a B-sequence (which is partly captured in the "transition probability from A to B" metric we used) encoded for both predator type and location. Our study corroborates these findings as we also found metrics that encode for predator type and for both predator type and location, but not for location alone (Fig. 1A), albeit by investigating a more comprehensive set of sequence features with an increased sample size. Building on these results, our study showed experimentally that titi monkeys extract this information but that the underlying mechanisms appear to be more complex than those proposed earlier (4).

The most relevant conclusion from our study, which contrasts earlier work on titi monkeys and other primates (7), is that information appeared to be conveyed probabilistically. The proportion of BB-grams, a continuous sequence feature, encoded categorical information about predator type and location. Receivers are likely to have extracted this information because they reacted in an appropriate but continuous fashion to playback experiments: the smaller the proportion of BB-grams, the more likely that subjects were looking upward, i.e., responding to an aerial predator or to a predator in the canopy, and the less likely that they were looking toward the speaker, i.e., responding to a terrestrial predator or a predator on the ground (Fig. 4) (8). Therefore, the proportion of BB-grams conveyed gradual information about a categorical event and elicited a graded reaction from the subjects.

Human and nonhuman animals (hereafter referred to as animals) live in environments where most stimuli appear in a continuous form, but perception is often categorical (9). For example, although rainbows consist of continuously changing wavelengths, they are perceived by humans as color bands. Similar effects are found in communication systems, including human speech. Acoustically, the human vocal tract can gradually alter the second formant of the syllable from the sound "b" (as in "beer") to "d" (as in "deer") and

then to "g" (as in "gear"), although they are perceived in sharply categorical ways by listeners (10). Another example comes from the American Sign Language, where the hand configuration gradually differs between the words "please" (the thumb and all the fingers are selected) and "sorry" (only the thumb is selected) but is perceived categorically by deaf signers (11).

Similarly, animal vocal repertoires often produce graded vocalizations [e.g., (12)], with evidence that these signal systems are perceived categorically by conspecific recipients (13). For example, female túngara frogs *Physalaemus pustulosus* categorize the mating calls of males as conspecific or not, although the calls exhibit graded variation in seven different acoustic parameters (14). By categorizing their environment, individuals can apply the same response to stimuli belonging to the same category, which results in an improvement of their fitness (e.g., by mating with potential sexual partners) and survival (e.g., by fleeing when exposed to a predator) (15). Thus, categorical perception is a crucial cognitive capacity with high fitness relevance in a physical world that is largely gradual.

Although the notion of categorical meaning is intuitively compelling, it is not necessarily the default mode of animal perception. Categorical perception has been a major theoretical pillar in animal communication research, particularly because of its intuitive link to linguistic theory. For example, Macedonia and Evans [(16), p. 179] presupposed that external events are processed in categorical terms ("...all eliciting stimuli must belong to a common category"). Although this approach has been fruitful and productive, it has also generated enigmas suggesting that the underlying theory may have to be revised. For example, in a seminal paper, Cheney and Seyfarth (17) were puzzled by the fact that animals appeared to have very few categorical semantic labels, mostly limited to predator classes and a few social events. One possibility is that graded meanings are the default way of animal communication [e.g., (18)], although this hypothesis has been much ignored and considered as less interesting than categorical perception (16). Our study suggests that explaining animal communication on categorical terms alone may be too restrictive and anthropocentric and may explain the struggle to extract meaning from some animal communication systems.

Our data show that the titi monkey alarm system most likely relies on call combinations at the sequence level, which potentially allows individuals to convey rich information with a limited set of calls (*3*). Since the listener needs to wait for the emission of enough calls to choose an appropriate reaction, this strategy may be seen as inefficient in predatory contexts where information should be quickly conveyed. When looking carefully at the alarm sequences of titi monkeys, it seems likely that predator type is the predominant information that can potentially be quickly extracted by the receivers: It is encoded by the first call in a sequence (A-calls for aerial predators and B-calls for terrestrial predators; Fig. 1, A and D, b) and is redundantly encoded later in the sequence through the proportion of BB-grams (Figs. 1A and 3). Predator location, on the other hand, seems to be secondary information: It is not encoded alone by any of the metrics we investigated (Fig. 1A) and only appears over the course of the sequence through the proportion of BB-grams (Fig. 3).

This imbalance of information can be explained by the fact that predator type and location typically are correlated (aerial predators attack from the canopy and terrestrial predators attack from the ground), suggesting that providing information about the predator type might be sufficient and would allow receivers to react quickly and efficiently to the threat in most predator detections. However, this system is not the most effective when a detected predator is not at its typical location (e.g., a bird of prey on the forest ground): In this case, titi monkeys add information about predator location at the sequence level using a call combinatory sequence feature (BB-grams), which elicits an appropriate reaction from the listeners (Figs. 2 and 4). Thus, alarm systems such as that of titi monkeys can provide some flexibility by conveying complex information with only few calls.

We have shown that information about predator type and location are encoded at the sequence level in a probabilistic manner. However, we only tested two locations (ground versus canopy), and further experiments might reveal whether titi monkeys also encode further predator locations (e.g., airborne). Moreover, at least two other encoding mechanisms can convey additional information about predation events. First, variation of spectral features of calls can convey rich information about external events (*12*, *19*) and were not addressed in the current study. Second, we did not investigate whether interactions among sequential and/or spectral metrics affected the information transfer and the probabilistic form of the alarm sequence. For example, spectral features could also convey information about predator type and location, in a fashion that allows the receiver to react more quickly and more efficiently to the threat than with the proportion of BB-grams. These possibilities remain to be tested in the future.

Our study on titi monkeys is, to our knowledge, unique in the way it provides empirical evidence of probabilistic meaning in an alarm call system. It is unclear whether this mechanism applies exclusively to titi monkeys and is absent in other taxa or whether other species have simply not been studied in the framework of probabilistic meaning attribution, something that will have to be resolved by future research. If common in other taxa, then a relevant next question to address is whether probabilistic meaning is the ancestral state and whether human categorical meaning evolved from it. An important general point emerging from this work is that the animal communication theory should be extended beyond the classic linguistic framework to encompass communicative capacities that are not commonly found in humans to better understand what makes language unique.

MATERIALS AND METHODS
Study subject and site
Our study was conducted from May 2015 to August 2016 at the "Reserva Particular do Patrimônio Natural Santuário do Caraça", an 11,000-ha private reserve in the Espinhaço Mountain range, State of Minas Gerais, Brazil (20°05′S, 43°29′W), where previous studies on titi monkeys already took place (*4*, *8*, *20*, *21*). The two Atlantic forests of interest, Tanque Grande and Cascatinha, are located 1 km apart from each other in the core of the reserve (transition zone between Cerrado, Atlantic forest, and Caatinga), with an elevation of around 1300 m.

Subjects were sampled from six groups of habituated black-fronted titi monkeys *C. nigrifrons*. Five of them (A, D, M, P, and R groups) were habituated to human presence between 2003 and 2008 (*20*); one additional group (S group) was habituated during the study period in 2015 (table S2). Titi monkeys typically live in family groups comprising an adult heterosexual pair and up to four offspring. Both sexes disperse after reaching sexual maturity, at around 3 to 4 years of age (*22*). Thus, the group compositions changed since 2003, with only some paired adults still present in our study (table S2). We considered an individual as an adult from the age of 30 months, as a sub-adult between 18 and 30 months, as a juvenile between 6 and 18 months, and as an infant if less than 6 months old [see (*20*)]. Recognition of individuals was based on morphological cues, such as size, fur pattern, and facial or corporal characteristics. The territories of the six habituated groups overlap with habituated groups and nonhabituated groups. This research was conducted in compliance with all relevant local and international laws and has the approval of the ethical committee CEUA/UNIFAL (Comissão de Ética no Uso de Animais da Universidade Federal de Alfenas), number 665/2015.

Predator presentations
The experiments followed a protocol developed by Cäsar *et al.* (*4*). Predator presentations were conducted between May 2015 and August 2016. We used the following four taxidermy predator models as stimuli: two models of caracaras *Caracara plancus* (aerial predator), one model of tayra *Eira Barbara*, and one of southern tiger cat *Leopardus guttulus* (terrestrial predators). The models were borrowed from the collection of the Natural Science Museum of the Pontifícia Universidade Católica de Minas Gerais. Each species was presented twice to each group, once in the canopy and once on the ground, i.e., 36 expected trials in total. The order of presentation was randomized across groups. Presentations were separated by at least 10 days for each group, and monkeys were monitored between trials. Before each trial (i.e., detection of the model by an individual), we monitored subjects for at least 30 min and, if possible, for another 30 min after the end of a trial (i.e., after the entire group had stopped calling or left the area). We made sure that no duet, group encounter, loud calls from a lost individual, or predator encounter occurred in the 30 min preceding the experiment; otherwise, the trial was aborted, and we waited for another 30 min to set up the equipment again.

For canopy presentations, we placed the model at 3 to 10 m off the ground (mean ± SD = 6.3 ± 1.6 m), depending on the structure of the arboreal strata. For ground presentations, we placed the model on the forest floor (i.e., at 0 m). We considered a trial as failed if more than one individual emitted the first 10 calls ($n = 1$) (this trial was removed from the dataset during the analyses and, thus, was not rerun), if the recording quality was insufficient (cicadas noise; $n = 1$), if model detection took place during setup ($n = 5$), if the model was detected by an individual of less than 2 years old ($n = 2$),

if another species gave alarm calls before visual detection by subjects ($n = 2$), if an individual bumped into the model before detection ($n = 1$), and if a real predator was encountered before detection of the model ($n = 1$). If a trial was scored as failed, we waited for at least 2 months before we retested the group, except for one case (35 days). Here, the monkeys responded to vegetation movement in the canopy (caused by the installation of the tayra model), although they probably did not see the model (M group). One experiment (Caracara in the canopy, D group) failed three times, and we decided to not rerun the experiment a fourth time. Therefore, the total number of successful trials was $n = 34$.

Vocal reactions were recorded in WAV (Wavesound Audio File) format with a Marantz solid-state recorder PMD661 (44.1-kHz sampling rate, 16-bit accuracy) and a directional microphone Sennheiser K6/ME66 or K6/ME67 (frequency response, 40 to 20,000 Hz ± 2.5 dB). Distance of detection (i.e., distance between the first individual to call and the model at the time of detection, in meters) and identity of the first caller were noted for each trial.

Vocal reaction dataset
Since we focused on sequences, we discarded responses composed of single calls ($n = 3$). We completed our own dataset with all alarm sequences recorded by Cäsar *et al.* (*4*) ($n = 20$) and another $n = 5$ sequences in response to the tayra model on the ground from Cäsar (*20*). For consistency, we discarded any sequence in which individuals were already calling at something else before detection of the model (flying bird, $n = 1$), if more than one individual emitted the first 10 calls ($n = 3$), if another species gave alarm call to the observers or to the model just before visual detection by the monkeys ($n = 1$), and vocal reaction consisted of only one call ($n = 1$). As a result, we included $n = 19$ sequences from Cäsar to our $n = 31$ sequences, i.e., the total dataset was composed of $n = 50$ sequences (table S3).

Some monkeys were probably present during both Cäsar's and our experiments (table S2) (potentially six individuals that emitted $n = 16$ sequences in total). However, groups were not systematically monitored between 2010 and 2015, so identification was not entirely reliable. Yet, since at least 5 years passed between the two sets of experiments, we found it unlikely that the responses to our stimuli were dependent on the monkeys' potential earlier experience with the paradigm. Thus, we considered these six callers as different between our study and Cäsar's study. In addition, in $n = 4$ sequences from Cäsar, the identity of the caller was unknown. For those, we considered the caller as a new individual that had not called in any other trials.

Stimuli preparation for playbacks
Broadcasted alarm sequences consisted of 10 calls recorded during predator presentations or during natural predator encounters. We did not broadcast sequences recorded by Cäsar because most of the group members were different or older from those recorded at that time, which could lead to bias in the experiment.

For the terrestrial predator in the canopy condition, we only managed to record two sequences corresponding to the pattern described by Cäsar *et al.* (*4*) out of 12 trials, and both were of poor quality. We thus created artificial sequences by adding an A-call from one given individual at the beginning of a B-call sequence from the same individual [as detailed in (*4*)]. The intercall intervals between the single A-call and the nine B-calls were measured on our recorded sequences and on two of Cäsar's sequences (*4*), and

the length of the silent gap for each of the artificial sequences was randomly chosen among these four measures. We sometimes had to replace bad quality calls with other calls from the same sequence (table S4). We filtered background noises and normalized all the sequences at −1 dB. We cut and edited the sequences using Praat 5.3.84 (*23*), Raven 1.5 (*24*), and Audacity 2.0.6. (*25*).

The total stimuli set was composed of 22 sequences: $n = 6$ aerial canopy, $n = 4$ aerial ground, $n = 6$ terrestrial canopy, and $n = 6$ terrestrial ground sequences. One terrestrial canopy sequence was of bad quality, so we removed the corresponding trials from the final dataset (tables S4 and S5).

Playback procedure
Seven females and seven males were tested from January to August 2016 (table S5). Each individual was exposed to one set of stimuli corresponding to a predator type in two different locations (aerial canopy, aerial ground, terrestrial canopy, and terrestrial ground), corresponding to a total of 28 trials. The presentation of the stimuli was randomized among individuals. No more than two trials were run on the same day within a given group and never for 2 days on a row to avoid habituation. No stimulus was broadcasted more than twice to limit pseudoreplication.

Stimuli sequences were recorded from a member of the family of the subject or from a member of one of the neighboring groups. There is no evidence that reactions of titi monkeys to others' alarm sequences is affected by the identity of the caller (*8*), possibly due to the fact that the pending danger requires a more urgent reaction than the caller identity. As it is still possible that monkeys recognize each other by spectral features, we made sure that if the playback sequence was from a member of the same group, then the caller was out of sight and the speaker was positioned so that the calls came from the direction of the caller. For neighboring alarm sequences, we played the stimuli in the overlap area between the subject's territory and the neighbor's territory to avoid bias due to intrusion, except in one case (sequence from the D group was played to the R group in the overlap between the S and the R groups' territories).

We monitored the group at least 30 min before and after the experiment. During the 30 min before a trial, we made sure that no duet, group encounter, loud calls from a lost individual, or predator encounter occurred; otherwise, we waited for another 30 min. We waited for the tested individual to be in low strata (1 to 8 m high) and in an open area to ensure a good visibility. The angle between the subject, the camera, and the speaker was about 90°, with the subject facing the camera. The speaker was covered with a camouflage net and held at the same height of the tested individual with a perch or, if not possible, at a maximum of 7 m high so that the angle between the horizontal line, the tested individual, and the speaker was less than 45° and as close as possible to 0° (mean = 8.1, SD = 7.1) (fig. S3). We made sure that no monkey was able to see the speaker. The reaction of the monkey was videotaped during twice the length of the broadcasted stimulus. Stimuli were played using an Anchor AN-Mini loudspeaker (audio output, 30 W; frequency response, 100 Hz to 15 KHz) connected to an iPhone 4.2.1, and videos were recorded using a camera Canon SX50 HS. We held the volume of the loudspeaker at a constant level matching the natural volume of a titi's vocalizations to a human hear. To test the setup, the territorial call of a white-shouldered fire-eye (*Pyriglena leucoptera*) was played. This bird call is common in the study area and elicits no reaction from the monkeys.

We considered a trial as failed if it was not possible to code most of the gazes of the monkey because it moved during the experiment ($n = 6$) or if the stimulus quality was too bad ($n = 2$; the stimulus was then removed from the analysis). If a trial failed, then we waited at least 8 days before rerunning it, except in one case (tested individual MR, aerial canopy trial: Only a few calls were played, so the subject did not hear the full stimulus and the trial was run again 4 days after) (table S5).

Vocal repertoire

We used the vocal repertoire established by Cäsar (*21*). The two main soft calls emitted during a predator encounter are the A-call, arch-shaped with a down-sweep modulation, and the B-call, S-shaped with an upsweep modulation (fig. S1). To estimate the accuracy of the call classification, we (M.B. and C.C.) tested between-rater reliability. We used a subset of 200 randomly selected calls that each of the two observers labeled. Between-rater agreement reached a sufficient level (Cohen's κ ≥0.8).

Metric extraction

We applied the same procedure to extract metrics from the sequences recorded during predator presentations and to the sequences broadcasted during playbacks. For the sequences recorded during predator presentations, we only focused on the first 10 calls of each sequence: The duration of emission of the first 10 calls ranges from 3.0 to 133.4 s (mean = 18.2, SD = 23.8), which we considered long enough to convey urgent information about a pending threat.

One observer (M.B.) labeled each of the calls and measured the duration of each call interval, i.e., the silence between each call, by using Praat 5.3.84 (*23*) (Spectrogram, Hanning window; time resolution, 5 ms; frequency resolution, 88 Hz).

On the basis of previous studies, we identified the 15 variables to characterize titi monkey alarm call sequences (table S1). Since proportions are often distorted by rare events and small sample sizes, we used a Bayesian approach to estimate the occurrence of rare and common events (*26*). The procedure is based on a two-step process, which starts with a theoretically motivated prior distribution of events (never or always observed), which is then updated to create an empirically motivated posterior distribution (values approaching 0 or 1). We used the Dirichlet distribution as the prior distribution with α = 1 [see (*26*) for more details on the technique]. The resulting Bayesian posterior mean for the occurrence of i is mean = count of event i + α/(total number of events + kα), where k is the number of possible events. In the Bayesian framework, the only probabilities being equal to 0 or 1 are those set by the design based on our prior assumptions and that correspond to impossible or mandatory events, respectively. Thus, the few metrics that have a counterpart in (*4*) and that were extracted using the Bayesian approach (*26*) are expected to display a lower value than in (*4*) if they are common events or a larger value if they are rare events.

We calculated 15 metrics for each sequence: (i) "Proportion of A-calls" using the Bayesian method. We chose this variable because it has been suggested to carry information about predator type (*4*). (ii) "Slope of elements" (the probability of observing an A-call at each place in the sequence, followed by a linear regression, with the coefficient representing the slope). Negative slopes indicate that A-calls are less likely to occur as the sequence progresses. (iii) "Mean call interval" of each sequence and (iv) "coefficient of variation of call interval" (SD/mean). Low coefficients indicate high regularity of

call emission. We chose this variable because temporal structures of sequences can convey context information (*19*). (v to viii) "Proportion of 2-grams". In two-signal systems, such as titi monkey alarm calling, the proportion of all four possible 2-grams (AA, AB, BA, and BB) can be determined as the number of each 2-gram/total number of 2-grams, followed by a Bayesian correction for small size sample. (ix) "Slope of 2-grams" [graphic representation of probability of each 2-gram (*27*, *28*) by decreasing probability and extraction of the coefficient of regression (later referred to as 2-gram slope)]. When the 2-gram slope is different from 0, then one 2-gram is more represented in the sequence. (x) "Slope of entropy". Shannon entropy uses principles of the information theory to measure complexity into a sequence and has been successfully used in animal communication (*29*, *30*). Entropy evaluates the unpredictability of a sequence, i.e., the degree of randomness in the sequence. Several values can be considered: The zero-order entropy evaluates the diversity of the vocal repertoire with H0 = $\log_2 N$, where N is the repertoire size; the first-order entropy assesses the proportion of different elements in the sequence, with H1 = $-\Sigma\, p(x) \log (x)$, where p(x) is the probability of a syllable x occurring in the sequence; the second-order entropy measures the proportion of different combinations of two elements in the sequence, with H2 = $-\Sigma\, p(xy) \log (xy)$, where $p(xy)$ is the probability of a syllable y following a syllable x in the sequence. If one plots the entropic values for the different orders (from 0 to 2), then the slope provides a measure of organizational complexity (*30*). A negative slope indicates an important sequential organization and, thus, high communication capacities, while a slope of zero indicates a random organization, with a low communicative capacity. (xi to xv) Transition probabilities. Markov chains are often used for sequence order analysis (*3*, *27*, *30*). The Markov paradigm assumes that probabilities of future events are dependent on a finite number of previous events. A transition matrix M can be derived from this assumption, in which $M_{i,j}$ represents the probability that an event j follows an element i. Chains of events are often represented with a state "Start" at the beginning and a state "End" in the end [e.g., (*26*)]. However, recent analysis suggests that Markov chains are not the most powerful tool to highlight structure in animal sequences (*27*). Moreover, Markov chains require exponential distribution of the durations, which is not our case. To address this issue, we conducted semi-Markov analysis (*31*). Semi-Markov analysis requires that the distribution of durations of the states is independent of the previous states or its place in the sequence. We verified with graphical assessments that the place of the call did not influence its duration. In our study, the titi sequences can be presented as a chain of events A- and B-calls with an artificial "Start" state at the beginning of the chain but no "End" state in the end, since we did not study the whole sequences. Then, we extracted the Bayesian transition probabilities from Start to A (also referred to as "probability that the first call is A"), A to A, A to B, B to A, and B to B for each sequence; Start to B was not considered here since it is negatively correlated with Start to A.

Two-grams and transition probabilities provide complementary information, the first one describing the probability of occurrence of a two-call syllable and the other one describing the probability that one call follows another one. For example, in a sequence AAAAABA, the BA-gram has a probability of occurrence of one of six, while the transition probability from B to A is of one. Metrics were extracted from each sequence by using the R software version 3.4.1 (*32*) and the cfp package (*33*).

Video analysis

The 28 videos recorded from the playback experiments were coded with the software Elan 4.9.4 (*34*). The reaction of the caller was analyzed during and after the playback experiment, for a total duration of twice the duration of the stimulus (i.e., the duration of the playback plus the same amount of time after the end of the stimulus). We extracted the duration (in seconds) and direction of each gaze, i.e., from the moment the subject looked to one direction until it looked to another direction. Directions of the gaze were categorized as (i) upward (the subject had the head orientated at least at 45° above the horizontal line and looked further than one body away from him), (ii) downward (the subject had the head orientated at least at 45° under the horizontal line and looked further than one body away from him), (iii) toward the speaker (the subject had the head orientated within 45° relative to the line between the subject and the speaker and looked further than one body away from him), and (iv) elsewhere (the subject looked in another direction or less than one body away from him (e.g., food, body part, etc.). When the eyes of the subject were not visible, the gaze direction was noted as "not visible" and excluded from calculations of proportions. The proportion of time looking in each direction was calculated as the duration the monkeys spent looking in each direction divided by the time the subject was visible.

Videos were analyzed by a coder blind to the experimental conditions (A.P.). To assess rater reliability, two raters (A.P. and M.B.) coded three videos (10% of the total dataset). We calculated Cohen's κ to assess the reliability in direction and duration coding of the gazes. An overlap matrix was created with the conditions (gaze directions) in rows and columns (*35*). Agreements were tailed on the table diagonal (same duration and same direction), and disagreements were tailed on off-diagonal cells: When one coder noted a duration as one gaze bout (e.g., "elsewhere" from 12 to 13 s, coder 1) and the other coded two (or more) gaze bouts for the same duration (e.g., "elsewhere" from 12 to 12.5 s and "down" from 12.5 to 13 s, coder 2), the gaze bout of the first coder was cut into two bouts to facilitate comparison with the other coder's results (e.g., "elsewhere" from 12 to 12.5 s and "elsewhere" from 12.5 to 13 s, coder 1; "elsewhere" from 12 to 12.5 s and "down" from 12.5 to 13 s, coder 2; agreement from 12 to 12.5 s and disagreement from 12.5 to 13 s). The level of between-rater agreement was considered as substantial (κ = 0.79) (*36*), but it should be stressed that this method has limits since a long agreement of several seconds counts as much as a short disagreement of half a second, so the statistical agreement is lower than reality. We thus considered that the inter-rater agreement was good.

Statistical analysis

We used multimodel inference within an information-theoretic framework (*5*). This approach can be used to compare relative support for each model in a set of models by using model weights *w*, derived from Akaike's information criterion (*6*). This weight gives the probability that a model is the best among the set of considered models, ranging from 0 (weak support for being the best model) to 1 (strong support).

To graphically represent statistical uncertainty around the model estimates, we used a nonparametric bootstrap procedure: We created 1000 datasets that were drawn from the original dataset by selecting observations with replacement so that each dataset comprised as many observations as the original dataset. For each dataset, we refitted the model and extracted and plotted model predictions.

All statistics were conducted using the R software version 3.4.1 (*32*). Linear mixed models (LMMs) were fit using the lme4 package (*37*)

and generalized LMMs (GLMMs) using the glmmADMB package (*38*), model selection was performed with the MuMIn package (*39*), and bootstraps were performed with a custom function (resamplefunction) from the cfp package (*33*). Collinearity of the variables was checked for each model using the package car (*40*).

What do alarm sequences encode?

To investigate whether each metric conveyed information about predator type and/or location, we created six models for each metric. Each of these six models corresponded to a combination of predator type and location. The first two models included only predator type or location as predictors, which addresses the possibility that sequences encoded for predator type or location only. The next two models addressed the possibility that sequences contained information about predator type and location: One model contained both main effects; the other model additionally contained the interaction term for location and type. In all these models, we controlled for distance of detection (in meters) to avoid a bias due to urgency. Last, in two control models, we considered the intercept only (null model) and the distance of detection only (urgency model). In all models, the sequence metric was the response variable. All models were mixed-effects models in which the identity of the caller was fitted as random intercept. Descriptions of the general set of models are given in table S6.

For five metrics and their corresponding model sets, we used LMMs. The remaining metrics were fitted as GLMMs with a beta, gamma, or binomial error structure (table S1). For each metric, we ranked the set of six candidate models using Akaike's weight *w*. If, for a metric, at least one model did not converge (*n* = 10 models, five metrics), then we performed the ranking with the Akaike's weight of the converging models only.

What information do monkeys attend to?

To assess how the combination of eliciting predator type and location of the played back sequences affected the time listeners spent looking in predator-relevant directions, we created six models. The first two models only included the predator type or location as predictors, respectively, which addressed the possibility that listeners only attended to either predator type or location. The second two models addressed the possibility that listeners attended to predator type and location: One model contained both main effects and the other additionally contained an interaction term for location and type. In all models, we controlled for the height of the listeners (i.e., the distance from the ground, in meters) to address perceived differences in urgency. Last, in two control models, we only considered direction of gaze (null model) and height of the individual and direction of gaze (urgency model). In all models, the response variable was the proportion of time the listeners looked to one direction. All models were mixed models (GLMMs) in which the identity of the listener and the broadcasted sequence were fitted as random intercepts with a binomial error structure (table S6). We ranked the set of six candidate models using Akaike's information criterion and interpreted model weights *w* (*5*).

What sequential metrics do monkeys attend to?

To assess how the metrics characterizing the call sequences used as playbacks affected the time listeners spent looking in predator-relevant directions, we created 15 models, each containing one metric as predictor variable. In all models, we controlled for the height of the listeners (in meters) to address perceived differences in urgency. We also designed two control models that only contained direction of gaze (null model) and height of the individual and direction of the gaze (urgency model) as predictor variables. In all models, the

response variable was the proportion of time the listeners looked to one direction. All models were mixed models (GLMMs) in which the identity of the listener and the broadcasted sequence were fitted as random intercepts with a binomial error structure (table S6). Again, we ranked the set of candidate models using Akaike's information criterion and interpreted model weights *w* (*5*).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/5/eaav3991/DC1

Fig. S1. Soft alarm calls of titi monkeys.
Fig. S2. Listener's gaze reaction depending on the eliciting stimulus of the sequence.
Fig. S3. Location of the speaker during the playback experiments.
Table S1. Design of the set of models for each metric.
Table S2. Composition of the six titi monkey groups during our study and that of Cäsar *et al.* (*4*).
Table S3. Description of the final dataset of predator presentations.
Table S4. Playback stimuli.
Table S5. Playback experiments schedule.
Table S6. Models formulas.

REFERENCES AND NOTES

1. R. M. Seyfarth, D. L. Cheney, P. Marler, Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Anim. Behav.* **28**, 1070–1094 (1980).
2. K. Zuberbühler, C. Neumann, in *APA handbook of comparative psychology: Basic concepts, methods, neural substrate, and behavior*, J. Call, G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, T. Zentall, Eds. (American Psychological Association, Washington, 2017); http://content.apa.org/books/15987-031.
3. A. Kershenbaum, D. T. Blumstein, M. A. Roch, Ç. Akçay, G. Backus, M. A. Bee, K. Bohn, Y. Cao, G. Carter, C. Cäsar, M. Coen, S. L. DeRuiter, L. Doyle, S. Edelman, R. Ferrer-i-Cancho, T. M. Freeberg, E. C. Garland, M. Gustison, H. E. Harley, C. Huetz, M. Hughes, J. Hyland Bruno, A. Ilany, D. Z. Jin, M. Johnson, C. Ju, J. Karnowski, B. Lohr, M. B. Manser, B. McCowan, E. Mercado III, P. M. Narins, A. Piel, M. Rice, R. Salmi, K. Sasahara, L. Sayigh, Y. Shiu, C. Taylor, E. E. Vallejo, S. Waller, V. Zamora-Gutierrez, Acoustic sequences in non-human animals: A tutorial review and prospectus. *Biol. Rev.* **91**, 13–52 (2016).
4. C. Cäsar, K. Zuberbühler, R. W. Young, R. W. Byrne, Titi monkey call sequences vary with predator location and type. *Biol. Lett.* **9**, 20130535 (2013).
5. K. P. Burnham, D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer, 2002).
6. D. R. Anderson, *Model Based Inference in the Life Sciences: A Primer on Evidence* (Springer Science, 2008).
7. K. Zuberbühler, Referential labelling in Diana monkeys. *Anim. Behav.* **59**, 917–927 (2000).
8. C. Cäsar, R. W. Byrne, W. Hoppitt, R. J. Young, K. Zuberbühler, Evidence for semantic communication in titi monkey alarm calls. *Anim. Behav.* **84**, 405–411 (2012).
9. R. L. Goldstone, A. T. Hendrickson, Categorical perception. *Wiley Interdiscip. Rev. Cogn. Sci.* **1**, 69–78 (2010).
10. A. M. Liberman, K. S. Harris, H. S. Hoffman, B. C. Griffith, The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* **54**, 358–368 (1957).
11. K. Emmorey, S. McCullough, D. Brentari, Categorical perception in American Sign Language. *Lang. Cognit. Process.* **18**, 21–45 (2003).
12. J. Fischer, K. Hammerschmidt, D. L. Cheney, R. M. Seyfarth, Acoustic features of female chacma baboon barks. *Ethology* **107**, 33–54 (2001).
13. J. D. Smith, A. C. Zakrzewski, J. M. Johnson, J. C. Valleau, B. A. Church, Categorization: The view from animal cognition. *Behav. Sci.* **6**, E12 (2016).
14. A. T. Baugh, K. L. Akre, M. J. Ryan, Categorical perception of a natural, multivariate signal: Mating call recognition in túngara frogs. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8985–8988 (2008).
15. M. D. Hauser, in *The Evolution of Communication* (The MIT Press, 1996), pp. 471–608.
16. J. M. Macedonia, C. S. Evans, Essay on contemporary issues in ethology: Variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethology* **93**, 177–197 (1993).
17. D. L. Cheney, R. M. Seyfarth, Why animals don't have language. *Tann. Lect. Hum. Values* **19**, 173–210 (1997).
18. C. N. Templeton, E. Greene, K. Davis, Allometry of alarm calls: Black-capped chickadees encode information about predator size. *Science* **308**, 1934–1937 (2005).
19. M. Berthet, C. Neumann, G. Mesbahi, C. Cäsar, K. Zuberbühler, Contextual encoding in titi monkey alarm call sequences. *Behav. Ecol. Sociobiol.* **72**, 8 (2018).
20. C. Cäsar, thesis, University of St Andrew (2011).
21. C. Cäsar, R. Byrne, R. J. Young, K. Zuberbühler, The alarm call system of wild black-fronted titi monkeys, *Callicebus nigrifrons. Behav. Ecol. Sociobiol.* **66**, 653–667 (2012).
22. J. C. Bicca-Marques, E. W. Heymann, in *Evolutionary Biology and Conservation of Titis, Sakis and Uacaris*, L. M. Veiga, A. A. Barnett, S. F. Ferrari, M. A. Norconk, Eds. (Cambridge Univ. Press, 2013), pp. 196–207.
23. P. Boersma, D. Weenink, Praat: Doing phonetics by computer (2009); www.praat.org/.
24. Bioacoustics Research Program, *Raven Pro: Interactive sound analysis software* (The Cornell Lab of Ornithology, Ithaca, 2014); http://www.birds.cornell.edu/raven.
25. Audacity Team, *Audacity* (2014); https://www.audacityteam.org/.
26. S. J. Alger, B. R. Larget, L. V. Riters, A novel statistical method for behaviour sequence analysis and its application to birdsong. *Anim. Behav.* **116**, 181–193 (2016).
27. D. Z. Jin, A. A. Kozhevnikov, A compact statistical model of the song syntax in Bengalese finch. *PLOS Comput. Biol.* **7**, e1001108 (2011).
28. A. Kershenbaum, E. C. Garland, Quantifying similarity in animal vocal sequences: Which metric performs best? *Methods Ecol. Evol.* **6**, 1452–1461 (2015).
29. A. Kershenbaum, Entropy rate as a measure of animal vocal complexity. *Bioacoustics* **23**, 195–208 (2014).
30. B. McCowan, S. F. Hanser, L. R. Doyle, Quantitative tools for comparing animal communication systems: Information theory applied to bottlenose dolphin whistle repertoires. *Anim. Behav.* **57**, 409–419 (1999).
31. V. R. Cane, Behaviour sequences as semi-Markov chains. *J. R. Stat. Soc. Ser. B Methodol.* **21**, 36–58 (1959).
32. R Development Core Team, *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2017); http://www.R-project.org/.
33. C. Neumann, *cfp: Christof's function package* (2018).
34. Max Planck Institute for Psycholinguistics, *Elan* (Nijmegen, Netherland, 2016); https://tla.mpi.nl/tools/tla-tools/elan/.
35. H. Holle, R. Rein, EasyDIAg: A tool for easy determination of interrater agreement. *Behav. Res. Methods* **47**, 837–847 (2015).
36. J. R. Landis, G. G. Koch, The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977).
37. D. Bates, M. Maechler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
38. D. A. Fournier, H. J. Skaug, J. Ancheta, J. Ianelli, A. Magnusson, M. N. Maunder, A. Nielsen, J. Sibert, AD Model Builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optim. Methods Softw.* **27**, 233–249 (2012).
39. K. Barton, *MuMIn: Multi-Model Inference* (2016).
40. J. Fox, S. Weisberg, *An R Companion to Applied Regression* (Sage, Thousand Oaks CA, ed. 2, 2011).

Science Advances

Titi monkeys combine alarm calls to create probabilistic meaning

Mélissa Berthet, Geoffrey Mesbahi, Aude Pajot, Cristiane Cäsar, Christof Neumann and Klaus Zuberbühler