Distribution Agreement

Atlas S. Moss April 10, 2023

Vocal Recognition and Social Knowledge in captive Tufted Capuchin Monkeys (*Sapajus apella*)

by

Atlas S. Moss

Marcela E. Benítez
Adviser

Anthropology

Marcela E. Benítez

Adviser

Elizabeth Lonsdorf

Committee Member

Benjamin Wilson

Committee Member

Aubrey Kelly

Committee Member

2023

Vocal Recognition and Social Knowledge in captive Tufted Capuchin Monkeys (*Sapajus apella*)

By

Atlas S. Moss

Marcela E. Benítez

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Anthropology

2023

Abstract

Vocal Recognition and Social Knowledge in captive Tufted Capuchin Monkeys (*Sapajus apella*)

By Atlas Moss

Individual recognition is a key ability for social animals, particularly primates. Recognizing individuals, tracking changing social relationships, deceiving others, and theory of mind may confer a fitness advantage to individuals living in complex groups and is thus considered an important factor in the development of the large brains in primates (Emery et al., 2007). In this study, I examine whether captive tufted capuchin monkeys are capable of individual recognition of outgroup members through signals encoded in their vocalizations, using acoustic analysis and playback experiments. Vocal recognition is particularly beneficial for wide-ranging arboreal species such as capuchins, as this allows them to maintain the socio-spatial makeup of their social systems (Briseño-Jaramillo et al., 2015). Acoustic analysis determined that capuchin food call vocalizations are individually distinct, providing potential for individual recognition. The preliminary results of the playback experiment did not reveal significant evidence towards varying behavior that could imply individual recognition. Further exploration with a greater number of trials and condition types will be needed for conclusive results.

Vocal Recognition and Social Knowledge in captive Tufted Capuchin Monkeys (*Sapajus apella*)

By

Atlas S. Moss

Marcela E. Benítez

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Anthropology

2023

Acknowledgements

I would like to thank the following people, without whom I would not have been able to complete this project:

Marcela Benitez, who oversaw my thesis as my advisor, for her support and understanding every step of the way in my thesis and over the past three years I've spent in her research team.

Elizabeth Lonsdorf, Ben Wilson, and Aubrey Kelly for serving on my thesis committee and for their insight and support along the way.

Isabelle Wagoner for her assistance in my call collection and playback experiments over the course of the study.

Sierra Simmons for her assistance with my playback experiment. Her assistance in my project allowed me to include more trials and integrate IR camera data into the playback experimental results.

The Language Research Center at Georgia State University for allowing me to conduct my research at their facility.

Table of Contents

Atlas Spencer Moss

Vocal Recognition and Social Knowledge in captive Tufted Capuchin Monkeys (*Sapajus apella*)

ABSTRACT

Individual recognition is a key ability for social animals, particularly primates. Recognizing individuals, tracking changing social relationships, deceiving others, and theory of mind may confer a fitness advantage to individuals living in complex groups and is thus considered an important factor in the development of the large brains in primates (Emery et al., 2007). In this study, I examine whether captive tufted capuchin monkeys are capable of individual recognition of outgroup members through signals encoded in their vocalizations, using acoustic analysis and playback experiments. Vocal recognition is particularly beneficial for wide-ranging arboreal species such as capuchins, as this allows them to maintain the socio-spatial makeup of their social systems (Briseño-Jaramillo et al., 2015). Acoustic analysis determined that capuchin food call vocalizations are individually distinct, providing potential for individual recognition. The preliminary results of the playback experiment did not reveal significant evidence towards varying behavior that could imply individual recognition. Further exploration with a greater number of trials and condition types will be needed for conclusive results.

INTRODUCTION

Like humans, other social animals benefit from identifying and maintaining information about others within their social world. The social complexity hypothesis (Jolly, 1966; Humphrey, 1976) argues that sociality, especially at more complex levels involving large groups and varying ranks and social standings, is cognitively challenging and correlates with the evolution of large brains and complex cognitive abilities (Bergman et al., 2015). Recognizing individuals, tracking changing social relationships, deceiving others, and theory of mind (Herrmann et al., 2007), all critical components of social knowledge, may confer a fitness advantage to individuals living in complex groups and is thus considered an important part of the large primate brain (Emery et al., 2007). There is, at least on the level of statistical correlation, a strong link between size of a social group and the relative size of the neocortex (Dunbar, 1995).

At the core of social knowledge is individual recognition. Individual recognition, whether through visual, auditory, or olfactory cues, is present in a variety of species and is highly advantageous socially as it allows individuals to monitor relationships within and between social groups (Tibbetts and Dale, 2007). Because many primate species have large and/or complex social structures, individual recognition not only of identity, but of status in relation to an individual is key for competitive and cooperative relationships that help individuals gain access to resources (Emery et al., 2007).

In order to benefit individuals in these complex social systems, individual recognition must be consistently accurate to avoid potentially costly misidentification, and it must adapt with changing social structures to remain relevant for the future (Bergman & Sheehan, 2013). While frequently shifting social structures and the risk of misidentification can limit the value and utility of individual recognition, it is still a highly valuable trait that social animals, especially humans and other primates, rely on in maintaining social relationships. Because of this, recognition not only of ingroup conspecifics, but outgroup individuals as well, can greatly benefit individuals in complex social systems. Identifying and recognizing ingroup and outgroup conspecifics allows for an individual to track of the numerous cooperative and competitive relationships they are a part of, as well as the territories in which those interactions take place through socio-spatial awareness (Seyfarth and Cheney, 2009). Awareness of these varying cooperative and competitive relationships is a key measure for social complexity (Bergman & Beehner, 2015).

While there are many ways that individuals recognize each other, recognition based on vocalizations is particularly useful for and prevalent in many primate species (Briseño-Jaramillo et al., 2015. Fischer et al., 2013). Acquisition of social knowledge through vocalizations is widely documented in primates (Benítez, 2017, Radick, 2005.). For example, baboon mothers are known to respond to the calls on their own infants and ignore calls from other infants, ensuring that they do not unintentionally expend energy searching for or tending to an infant that is not their own, especially useful with the energetically-costly nature of primate infant care (Rendall et al., 2007).

In this study, I examine whether captive tufted capuchin monkeys, *Sapajus apella*, are able to distinguish between individuals based on identifying information encoded into their vocalizations (Pokorny et al., 2009).

Capuchin monkeys are good candidates for understanding the evolution of social intelligence because of their relatively advanced cognitive and socio-cognitive capacities. Various capuchin species have been shown to exhibit tool use (Westergaard, 2010), a strong working memory (Travares & Tomaz, 2002), social tolerance (particularly around feeding) (Perry, 2011), and prosociality (Lakshminarayanan & Santos, 2008). They have, through convergent evolution, developed many of the same socio-cognitively significant traits as humans, such as large brain size, extractive foraging, cooperative behavior, and social learning (Perry, 2011).

While humans are seen as going beyond the complex sociality of nonhuman primates and behaving as “ultra-social,” using socio-cognitive skills of such as social learning and communication in the development of a cultural world (Herrmann et al., 2007), capuchins exhibit many of the characteristics which are foundational for the emergence of culture. Capuchins have relatively long lifespans and slow development, the largest brain to body ratio of any non-human primate, and they excel in cognitive tasks as mentioned previously (Perry, 2011). Capuchins also exhibit gaze-following, which is considered foundational to the development of the complex social systems that humans have formed (Perry, 2011).

In the wild, tufted capuchins live in relatively large multi-male, multi-female groups. They are arboreal, making visual access to conspecifics difficult in many circumstances. Vocal recognition is particularly beneficial for wide-ranging, arboreal species such as capuchins, as it allows them to maintain an understanding of the socio-spatial makeup of their social systems, such as locating members of their ingroup in low-visibility environments and recognizing the territorial boundaries with neighboring groups (Briseño-Jaramillo et al., 2015). Generally, if an individual hears an outgroup conspecific calling consistently from a specific location, that would inform them of that individual’s general territory if they are capable of vocal individual recognition.

Visual individual recognition of conspecifics has been well documented in capuchins. Using pictures of group mates, one study determined that tufted capuchins are capable of distinguishing between images of in-group and out-group members (Pokorny et al., 2009). Tufted capuchins are also capable of recognizing the relationships they have with individuals and sex identification based on images of familiar conspecifics faces (Talbot, 2016). Capuchins’ ability to recognize individuals based on visual cues and facial recognition shows that they are capable of recognition and have an understanding of their relationships with these individuals.

If capuchin vocalizations contain identifying information, it would incur a fitness advantage for individuals to be able to recognize callers based on their calls. Playback studies are a key method for examining the ability of primates, and animals in general, to recognize individuals based on their vocalizations. Playback studies involve playing recorded calls for individuals and examining their behaviors following the playback to determine whether they are responding based on specific information they have gained from the call. A previous playback study on wild tufted capuchins has shown that their vocalizations contain communicable information about the context of the call (Di Bitetti, 2003), though few studies have been done regarding the identity of callers through playbacks with capuchins. This is in part due to the fact that capuchins are arboreal, making playback studies in a wild setting less feasible.

One playback study with chimpanzees has shown that they are able to distinguish between and vary their responses to calls played based on whether the call was taken from an in-group, out-group, or unfamiliar individual to the focal subject, which found that the chimpanzees did respond differently based on the caller’s identity for all three categories, showing individual voice recognition of social relationship to other individuals (Herbinger et al., 2009). My study is building on this idea by comparing responses to different outgroups played from locations either congruent or incongruent with the outgroup’s territory, examining not whether they can distinguish between “familiar versus unfamiliar” and “ingroup versus outgroup,” but whether they recognize the distinct territories of different outgroups. The goal is to understand whether capuchins can recognize and distinguish the calls of different outgroup members, showcasing an understanding of the composition of other groups and their territories. In other words, does their comprehension of outgroup members and territories extend beyond "mine" versus "not mine.”

I aim to address two main questions in this study. First, I examined how the acoustic properties of affiliative whistle vocalizations may encode information about the identity of the caller. In my previous research, I examined two types of whistle calls, food calls and lost calls, and found that they differed along a few key temporal and acoustic parameters including call duration, number of elements, maximum pitch, pitch range, and mean F1 (Moss, Unpublished Data, 2020). Tufted capuchin food calls are a whistle series call consisting of 2 or more elements (syllables). These calls are acoustically very similar to the tufted capuchin lost call, a whistle series call produced when an individual is lost or in some way separated from the group. Black-horned capuchins, which have a similar vocal repertoire to that of tufted capuchins, also produce whistle series calls in a variety of contexts, including ones in a food call (food-associated whistle series) and a lost call context (long distance whistle series), so it is unsurprising that the tufted capuchins have a whistle-series call associated with both contexts (Di Bitetti et al., 2018). In another playback study, food calls and lost calls have been found to elicit differing responses in tufted capuchins, acting as functionally referential signals to the context of the call (Di Bitetti, 2003).

Here, I focus on how the acoustic features of food calls convey important information about the identity of a caller. Food calls are widely given during feeding contexts and as such there is ample opportunity for individuals to recognize the identity of the caller from these calls, if they are acoustically distinct by individual. I predict that several acoustic features that relate to call production, such as pitch and call duration, will differ between individuals.

Second, I examined whether capuchins are able to recognize individuals based on these calls. To test this, I use a playback study with a violation-of-expectations paradigm, directed at the socio-spatial cognitive abilities of the capuchins (Briseño-Jaramillo et al., 2015). The violation-of-expectations paradigms predicts that individuals will respond more strongly and attend longer to unexpected stimuli than to expected (Sankey et al., 2007). If capuchins recognize outside group members based on their calls, they should respond more strongly to food calls that are inconsistent with their expectations, meaning an individual is calling from a group that they don't belong to. I record audiovisual data for each individual's reactions in the time after the playback and analyze the videos for factors such as time spend looking towards the speaker and latency of response as measures for surprise. Other behaviors that may indicate surprise that were recorded include lost call responses and other vocalizations, pacing, and moving towards or away from the direction of vocalization (Fischer et al., 2013).

This study serves to build on previous research in the social cognition and communication of capuchin monkeys, while also broadening our understanding of nonhuman primate communication and socio-spatial cognition. The results of the playback study also have implications for between-group relations and familiarity in capuchins and other non-human primates and their cognitive ability to produce and interpret individually unique calls.

METHODS

Study Site and Subjects

The study subjects were 23 capuchin monkeys (5 males and 18 females) living in four multi-male, multi-female social groups at the Language Research Center of Georgia State University. Subjects lived in large indoor-outdoor enclosures. During the day, subjects were given the opportunity to voluntarily enter individual testing chambers affixed to the indoor section of the group enclosure. Water was available ad libitum, including in the testing chambers. Food, including primate chow and fresh fruits and vegetables, was provided multiple times per day, regardless of the testing schedule. Call recording and testing a conducted between February 2021 to March 2023. All protocols were approved by the Georgia State University & Emory IACUCs (A16031) and adhered to all applicable legal requirements of the United States; in addition, all research was conducted in adherence with the ASP Principles for the Ethical Treatment of Primates.

Acoustic Analysis

Recording Vocalizations

Calls were recorded on a battery-powered Marantz professional digital solid-state recorder PMD661 MK II with a Sennheiser directional shotgun microphone with an attached modular windscreen and shock mount to preserve recording quality from interference from wind and hand movement. The Marantz recorder was set to PCM-24 channel system, with a 24-bit Mono setting, and a max recording frequency of 44.1kHz. This is a standard recorder type and setting for the recording of high frequency primate calls (Fischer et al., 2013).

Call collection consisted of a combination of ~30-min focal recordings on individual groups and opportunistic recordings of calls from all groups. I collected calls distances of 1m-10m from the callers, in keeping with the maximum distance recommended for replicable measurements of high-amplitude calls (Fischer et al., 2013). For each recording session, the time of day, weather conditions, and groups analyzed were noted at the time of recording. For each individual call, identity of the caller, and call type, as well as any additional necessary context (accompanying behaviors, identity of individual call was directed towards, and any background noise, when applicable) were also noted at the time of recording. I collected calls outdoors between 12:00pm-3:00pm when the capuchins were released from their indoor enclosures to their outdoor enclosures, where they receive food access. When possible, groups were let out successively for recordings to reduce call overlap, though calls were taken when multiple/all groups had access to their outdoor enclosures as well. I prioritized the collection of food calls, which were analyzed for this paper, but I collected all call types indiscriminately during recording sessions.

Call Analyses

Following collection, I inspected calls using Praat acoustic analysis software (Boersma & Weenink, 2023) to view spectrograms and labelled and cut each individual food call. Only calls for which the caller had been identified were isolated. The caller ID, call type, and all other context provided were recorded for each extracted call. Based on the spectrogram in Praat, the calls were also assigned a call quality of either excellent, good, fair, or poor based on pre-established definitions. (see Table 1).

Table 1. Definitions of Call Quality Categorizations

Category	**Definition**
Excellent	The complete call is present with sufficient energy that does not exceed amplitude limits of the recorder, with no background noise or interruption/overlap from other callers.
Good	The complete call is present with sufficient energy that does not exceed amplitude limits of the recorder, with no overlapping calls, but may contain some background noise at a lower frequency and/or amplitude.
Fair	The complete call is present, but exceeds amplitude limits of recorder, or contains excessive background noise/interruptions from other callers. The start and end of the call is still clearly discernable.

Poor	The call all is not discernible from background noise/interruptions from other callers, or does not contain complete call.

Using RAVEN pro 1.6 software (Cornell Lab of Ornithology, 2023), I conducted acoustic analyses on the entire food call (*call analysis*) and on each individual element of each call (*element analysis*). This software was used to measure the spectral and temporal acoustic features of the calls. I generated spectrograms in Raven Hamming window, 3dB Filter Bandwidth of 112 Hz, Overlap of 87.5%, DFT size of 512 samples, and Grid Spacing of 86.1Hz, averaging 1 spectra. For this analysis, I selected only complete recordings that were previously categorized as Excellent or Good. In total, I accumulated 325 food call recordings of sufficient quality from 23 subjects for acoustic analysis. I examined 8 call parameters (Table 2) related to the acoustic and temporal features of the calls.

Table 2. Acoustic and temporal features collected for Capuchin food calls and elements

Variable	**Variable Definition**	**Units**
Duration	Time lapsed from the start to end of each call and element.	s
Low frequency (Pitch)	Lowest frequency where frequency bands occur.	Hz
Peak frequency	Peak of dominant frequency band.	Hz

Frequency 5%	Minimum frequency ignoring 5% below total energy of selected call/element	Hz
First quartile frequency	Mean frequency (Hz) under which 25 % of the energy is located in the selection.	Hz
Third quartile frequency	Mean frequency (Hz) under which 75 % of the energy is located in the selection.	Hz
Frequency 95%	Maximum frequency of call, ignoring 5% above selected energy in the selection.	Hz
# of elements per call	The number of discrete syllables or elements within a call	(count)

Data Analysis

Factor Analysis. Before all analyses, I checked the distribution of each parameter and, where necessary, I transformed the data to approximate a normal distribution. Next, to remove redundancy between the acoustic parameters (i.e., many of the acoustic properties were highly correlated), I ran a factor analysis (FA) using all 8 call parameters. The FA was run in SPSS using a varimax rotation. To examine how elements change throughout a call, I constructed a Mixed Model in R with the factors scores as the dependent variable, element position as a fixed effect, and ID of caller as the random effect to control for pseudo-replication.

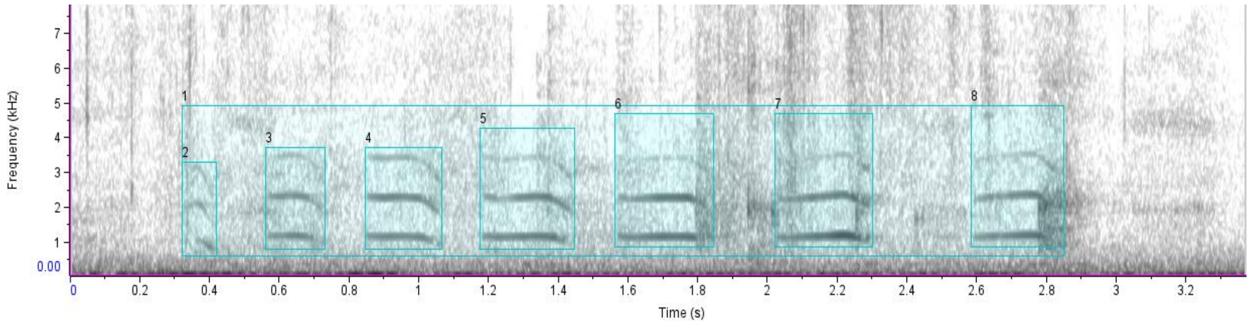


Figure 1. Spectrogram of tufted capuchin food call, made in Raven Pro with the following Spectrographic parameters: Hamming window, 3dB Filter Bandwidth of 112 Hz, 87.5% Overlap, DFT size of 512 samples, and Grid Spacing of 86.1Hz, averaging 1 spectra. Labeled selections were defined as follows: 1. “Full food call”, 2. “Element 1”, 3. “Element 2”, 4. “Element 3”, 5. “Element 4”, 6. “Element 5”, 7. “Element 6”, 8. “Element 7”. Measures for each of the acoustic and temporal features outlined in Table 2 were collected in Raven Pro for each selection.

Individual Differences. To test for differences between calls due to individual identity, I used discriminant function analysis (DFA). DFA is commonly used in behavioral and bioacoustics research to determine whether calls differ between subjects, groups, and context (Mundry & Sommer, 2007). The benefit of DFA is that it addresses how reliably groups can be distinguished and which acoustic parameters are important in those classifications. In addition, discriminate function analyses are used to discriminate between the individuals or groups (Mundry & Sommer, 2007). To balance the variance in the number of calls from each individual in the dataset, I used a subset of the data for each DFA analysis. In the DFA data sets, I randomly selected 14 calls from the 10 individuals with the largest number of collected calls. I compared the entire call, rather than each element, to account for differences in total call duration and the number of elements per call.

Playback Experiment

Playback Setup

To examine if capuchin monkeys monitor the identity of outer group members, I conducted a playback experiment on 6 individuals (2 males, 4 females) across two neighboring groups. The playback stimuli were selected from the calls collected in the acoustic analysis portion of the study. Only high-quality calls were used for the playbacks. To remove low-frequency wind noise from the calls, a bandpass filter was applied over all stimuli with a lower limit of 700hz, and no upper limit. As tufted capuchin food calls are high frequency, this did not alter the call itself as the bandpass filter did not interfere with the lowest frequency band of any of the calls. Before each playback, calls were played and I auditorily matched calls for consistency and controlled speaker volume using the application “Volume Pro”. For this experiment, I only used playback calls from the dominant males in each group (“Logan”, “Mason”, “Nkima" and “Griffin”).

The stimuli for the playback experiment were played over a Bluetooth JBL FLIP6 speaker, an effective speaker type for small-bodied primates with high calls, played from short distances (5-10m) (Fischer et al., 2013). Test subjects were filmed using a 4K Plus handheld digital video camera with 48megapixels, set up on a tripod directly facing the subject’s faceplate in order to have a clear view of the subject’s head turn.

Playback Design

Playbacks took place in the capuchins’ indoor enclosures around 1pm before the test group was given outdoor access. As shown in Fig 2, the four groups’ indoor enclosures are arranged beside each other, with group 1 in room 1, the two in-most groups, 2a and 2b, both in room 2, and group 3 in room 3 (see Fig. 2). Playbacks were conducted on individuals from groups 2a and 2b (for a total of 9 test subjects), as they each have neighboring groups on two adjacent sides. My design was a “violation of expectation paradigm” which is commonly used to examine what subjects know about the social world around them (Sankey et al., 2007. Fischer et al., 2013).

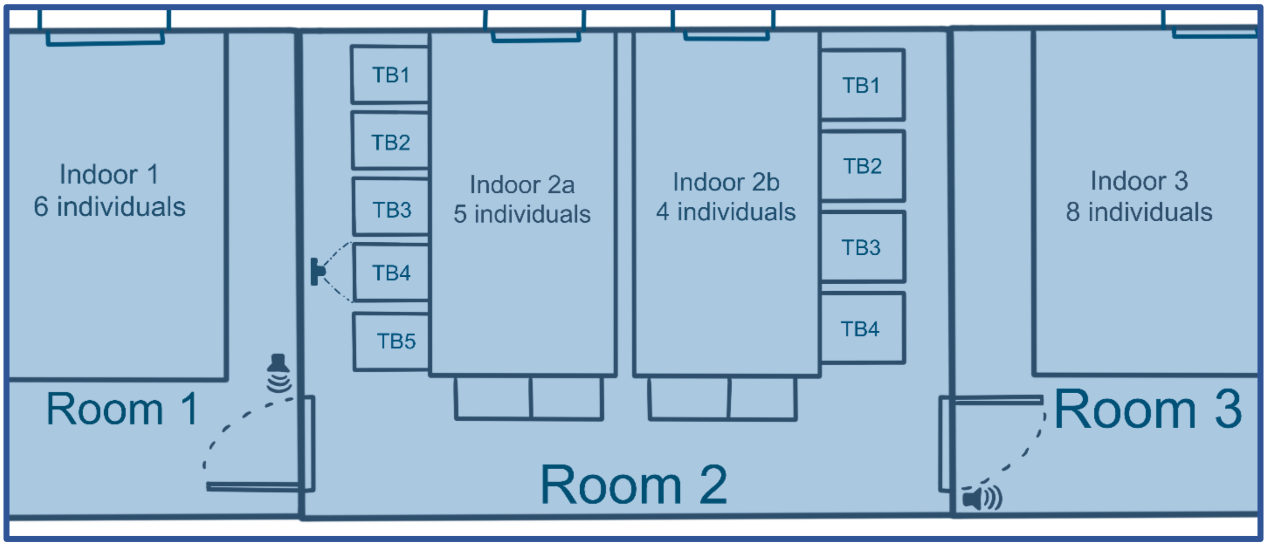


Figure 2. Depiction of Playback setup showing the layout of each group's indoor enclosures, labeled as "Indoor [Group #]." Group 1 dominant male is referred to in this paper as "Logan," Group 2a dominant male is "Griffin," Group 2b dominant male is "Nkima," Group 3 dominant male is "Mason." This example shows the camera setup for a playback on an individual in Group 2a, in box testing box 4 (labeled "TB4" connecting to "Indoor 2a"). The speaker symbol in Room 1 shows where calls would play from in all Group 2a congruent and incongruent trials. The speaker symbol in Room 3 shows where calls would play from in all Group 2b congruent and incongruent trials.

There were two conditions in the playback study:

1) *Congruent Condition* where the dominant male's call was played from the location consistent with that male's enclosure (*e.g.*, a Logan call was played from Group 1's enclosure, to an individual in Group 2a) (see Fig. 2).

2) *Incongruent Condition* where a male's call from the group's closest neighboring group was played from a location inconsistent with the caller's enclosure (*e.g.*, a Nkima call was played from Group 1's enclosure, to an individual in Group 2a) (see Fig. 2).

Each subject was tested in both a Congruent and an Incongruent condition, allowing for a within-subject comparison. I tested subjects while they were in their individual testing boxes.

The capuchins have voluntary testing each day in which they enter individual testing boxes (labeled in Fig. 2 as TB#) to complete computer-based testing for other research projects, after which they are given fruit and/or peanut butter treat rewards in their testing boxes. The testing boxes are wire mesh on all sides except for an opaque barrier between the testing box and the full enclosure, allowing them to see individuals in the testing boxes to the left and right of them, but not behind them.

Playback trials occurred after all the individuals had completed testing and received their fruit. Each playback trial tested one individual, with no individuals in testing boxes in between the test subject and the playback stimulus, to avoid potential for the test subject's response being influenced by other group members' responses to the stimulus. The testing box used for playback trials varied between subjects, but was kept consistent for all three trials within subject design.

During playbacks, all groups other than the test group were moved into their outdoor enclosures with no indoor access. This was done in order to assure that 1) there were no other calls from out-group members that could affect the test subject's response to the stimulus 2) individuals from outgroups could not hear their own calls being played, and (in incongruent trials) the caller is not in the test subject's line of sight, in a different location from where the call is being played.

The second point is especially crucial for the effectiveness of the playback, as nonhuman primates are likely to no longer respond to playback stimuli once they understand that the call is coming from an artificial sound source (Fischer et al., 2013). For this same reason, the speaker playing the call was hidden from the test subject's view (fig. 2). To ensure that the test subjects did not become familiar with the playback stimuli through over-exposure, each playback stimulus was used a maximum of two times for each test group. The test group and subject were both alternated as much as possible during the playback period to allow for the longer durations of time between playback trials for each subject.

During playbacks, I set up a camera directly in front of the test subject's box (fig. 2). Calls were only played when the test subject was facing the camera, to assure that any movement they make towards the stimulus (*e.g.*, head turn) was a clear, approximately 90° angle. At the start of filming, a small amount of peanut butter was applied to the front cage in each trial to encourage the test subject's orientation towards the camera. When the test subject was oriented appropriately, one researcher would make a gesture, and another researcher who was out of the test subject's line of vision would initiate the call. Each test subject's behavior was filmed for one full minute beginning at the onset of the call, at which point the trial would be completed and the recording would end.

Playback Response

All videos were scored using Behavioral Observation Video-Coding software (BORIS; Friard, O. and Gamba, M., 2016). All files were then renamed and randomized such that observers were blind to the condition (i.e., control or incongruent). I generated spectrograms of each trial's audio in BORIS in order to set the trial start directly at the initiation of the call.

To standardize responses, I coded 8 variables for 60 seconds beginning at the onset of the stimulus for each trial. All playback analyses were performed using R version 4.2.2 (R core team, 2021). Specifically, I measured: (1) duration of first look towards stimulus, (2) total duration of looking time towards stimulus in 60 secs, (3) latency of first look towards stimulus, (4) vocal response (and type: food call, stress trill, scream, or "other"), (5) visual threats/stress responses (i.e. fear grimace), (6) grooming (as a possible stress response), (7) manipulating enclosure (rattling enclosure door, turning on water source), and (8) foraging/eating (to account for the time individuals spent attending to the peanut butter applied to the enclosure faceplate).

Playback Analysis

To examine if capuchins responded differently to the congruent and incongruent condition, I conducted a paired-t test analyses on the following behavioral measures: duration of first look towards stimulus, latency of first look towards stimulus from the onset of the call, and total duration of time spent looking towards the stimulus in the 60s window. To account for the small sample size (6 individuals, 12 trials), I tested the distribution of the congruent and incongruent data sets for each of the above variables with a Shapiro-Wilk normality test. For variables that exhibited normal distribution, I ran a paired parametric t-test to assess whether there was a significant difference in the results between the congruent and incongruent trials. For variables that did not exhibit normal distribution, I used a Wilcoxon signed rank test (paired non-parametric t-test) in order to test whether the congruent and incongruent trials yielded significantly different results.

RESULTS

Do the acoustic properties of capuchin food calls convey information about caller identity?

Call Analysis

To analyze the components of capuchin food calls broadly, I ran a factor analysis on all food calls labeled “good” or “excellent,” a total of 303 calls from 22 individuals. The factor analysis resulted in two factors with eigenvalues >1 together explaining 66.8 % of the total variance. Factor 1 explained 38.1% of the variance with high positive loading (> 0.4) on pitch (low frequency), minimum frequency, and first quartile. Factor 2 explained 28.8% of the variance and showed a high positive loading on peak frequency, max frequency, and third quartile. Based on the loadings of these parameters on the rotated components, I characterized and labeled factor 1 as lower frequency measures and factor 2 as higher frequency measures. In other words, a high score along factor 1 indicates a higher pitch, a higher minimum frequency, and more energy concentrated at higher frequencies in the first quartile of the call. A high score in factor 2 indicates a high peak frequency, high max frequency, and more energy concentrated at higher frequency in the 3rd quartile of the call.

Throughout a call, I found that Factor 1 increases and Factor 2 decreases (Fig. 3). The first element in the call is lower in pitch, has a lower minimum frequency, 1st quartile energy concentrated at a lower frequency and a higher maximum frequency, 3rd quartile energy concentrated a higher frequency, and a higher overall peak frequency. The first call element exhibits a wider range of frequency measures, with lower lows and higher highs, indicating a lower pitch but wider range of vocal frequency within that element.

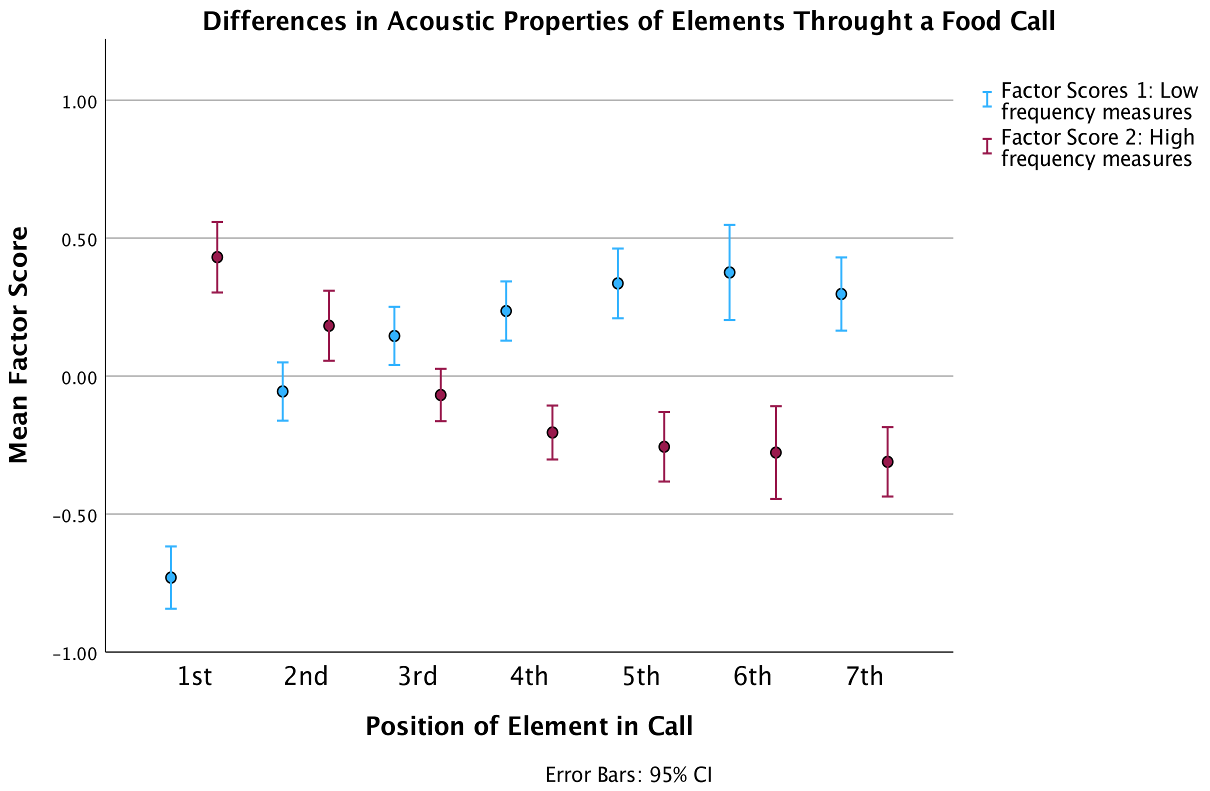
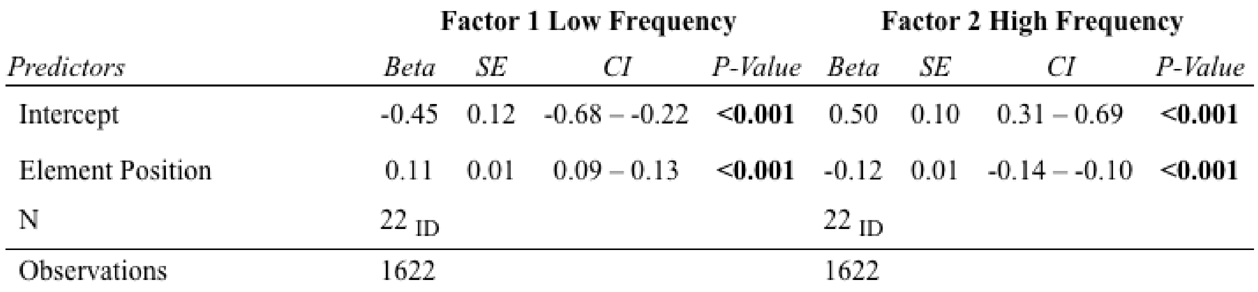


Figure 3. Differences in Acoustic Properties of Elements Throughout a Food Call. This dot plot shows the Mean Factor Score (y axis) for the two factors (described below) for each element over the course of the call (x axis). Factor 1 (in blue) indicates a higher pitch, a higher minimum frequency, and more energy concentrated at higher frequencies in the first quartile of the call at higher mean Factor Score values. Factor 2 (in red) indicates a high peak frequency, high max frequency, and more energy concentrated at higher frequency in the 3rd quartile of the call at higher mean Factor Score values.



	Factor 1 Low Frequency				**Factor 2 High Frequency**			
Predictors	*Beta*	*SE*	*CI*	*P-Value*	*Beta*	*SE*	*CI*	*P-Value*
Intercept	-0.45	0.12	-0.68 – -0.22	**<0.001**	0.50	0.10	0.31 – 0.69	**<0.001**
Element Position	0.11	0.01	0.09 – 0.13	**<0.001**	-0.12	0.01	-0.14 – -0.10	**<0.001**
N	22 ID				22 ID			
Observations	1622				1622			

Table 3. Predictors of Factor 1 (positively loading for Low Frequency), and Factor 2 (positively loading for High Frequency) based on the Element Position in the call.

Individual Differences.

The DFA revealed two factors with eigenvalues >1. This first factor (eigenvalue=0.88) accounted for 36 % of the variance and exhibited a high positive loading (>0.4) on minimum frequency and first quartile energy and a high negative loading on maximum frequency. A high loading Factor 1 suggested that calls were higher in frequency and exhibited less range of frequency within the call (higher min and lower max). The second function (eigenvalue = 0.7) accounted for 29.3% of the variance and exhibited a high absolute loading on number of elements and call duration. A high loading on Factor 2 signifies longer calls with more elements. The results from the DFA indicated that calls were significantly different between individuals, with 60% of original calls and 45% of cross-classified calls (chances= 1/10 or 10%) assigned correctly (p<0.001)

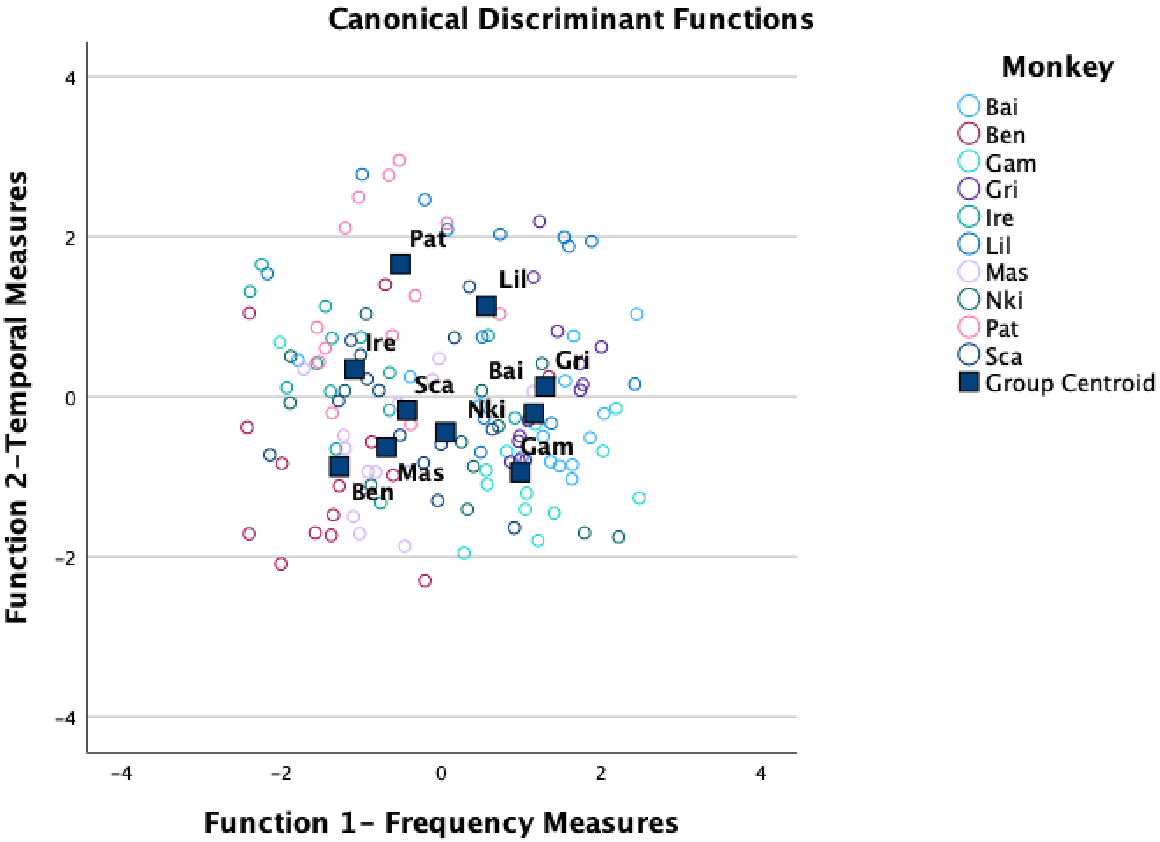


Figure 4. Canonical Discriminant Functions. This scatter plot shows distribution of callers by identity (unfilled circles) in relation to the two functions calculated through the discriminant function analysis. Function 1 (x axis) relates to frequency measures of the call, and Function 2 (y axis) relates to temporal measures of the call. The average alignment for each individual is shown as a group centroid (filled square), labeled by individual on the figure.

Do capuchins attend to acoustic information regarding the composition of neighboring groups?

The results from the playback study are currently preliminary. Of 8 subjects and 16 trials, I have run and analyzed 12 trials from 6 individuals. 16 trials in total were run for the 6 individuals, and 4 of these trials were aborted. Two trials were aborted due to speaker malfunctions during the playback. Two trials were aborted due to the individual not facing the camera at the onset of the call, making it impossible to test the onset of their turn towards the call. The aborted trials were rerun at later dates, and the reruns for these trial conditions were analyzed here.

Duration of first look towards the stimulus

On average, individuals appear to attend longer to the incongruent stimulus than to the congruent in their first turn towards the stimulus (Fig. 5) (Con Avg = 2.85 +/- 2.88; Inc Avg = 3.00 +/- 1.51), though these results were not significant. The data was normally distributed (Shapiro-Wilk normality test, W = 0.94346, p-value = 0.6871), so a paired parametric t-test was run. The paired parametric t-test yielded a p-value >0.05 (p-value = 0.8695), so the null-hypothesis of no significant difference between the two conditions was accepted.

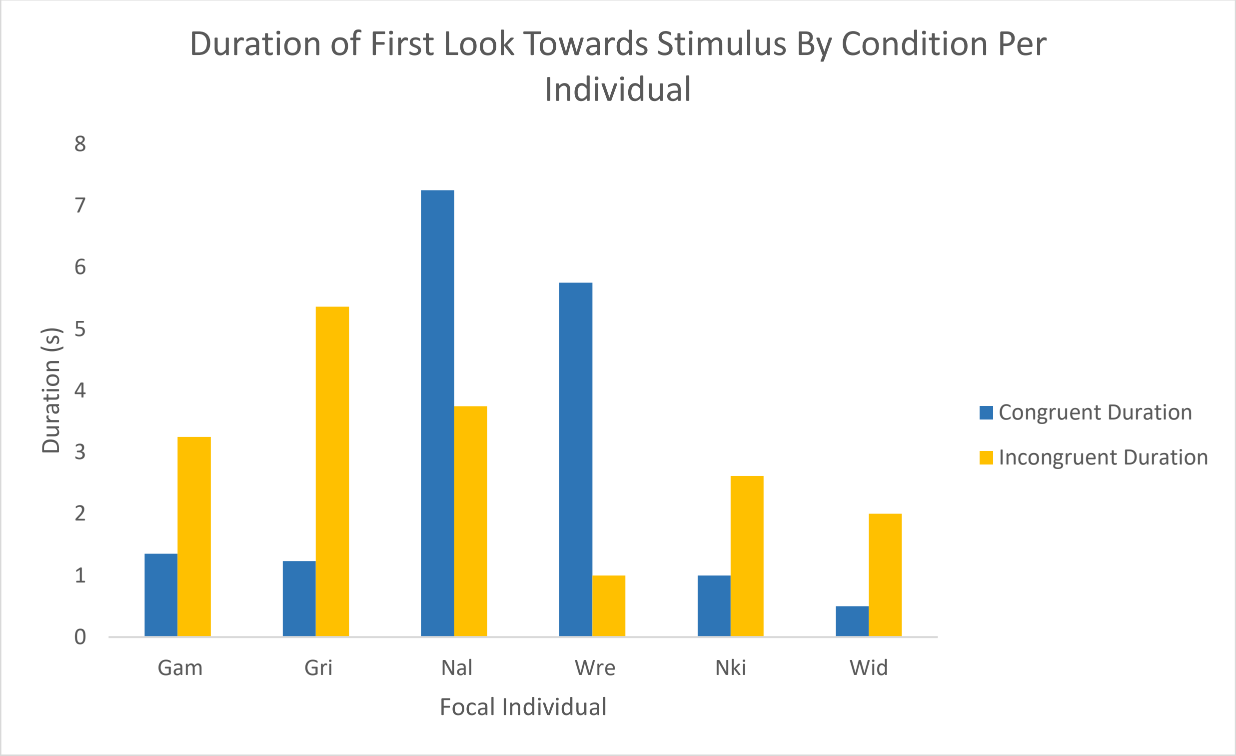


Figure 5. Duration of First Look Towards the Stimulus by Condition Per Individual. This bar plot shows the duration in seconds (y axis) of each individual's first look towards the stimulus by condition. The congruent condition is depicted in blue, and the incongruent condition is depicted in yellow. Based on the results of the paired parametric t-test (p-value = 0.8965), these results show no consistent trend across individuals.

Latency of first turn towards stimulus

No significant difference was found in the latency of an individual's first turn towards the stimulus following the onset of the call depending on the condition. The results for each condition were skewed, based on the Shapiro-Wilk normality test (W = 0.77204, p-value = 0.03245), so a Wilcoxon Signed Rank Test (or paired non-parametric t-test) was run and yielded a p-value > 0.05 (p-value = 0.4375). Based on these results, the null-hypothesis of no significant difference in latency of first look between the two conditions was accepted.

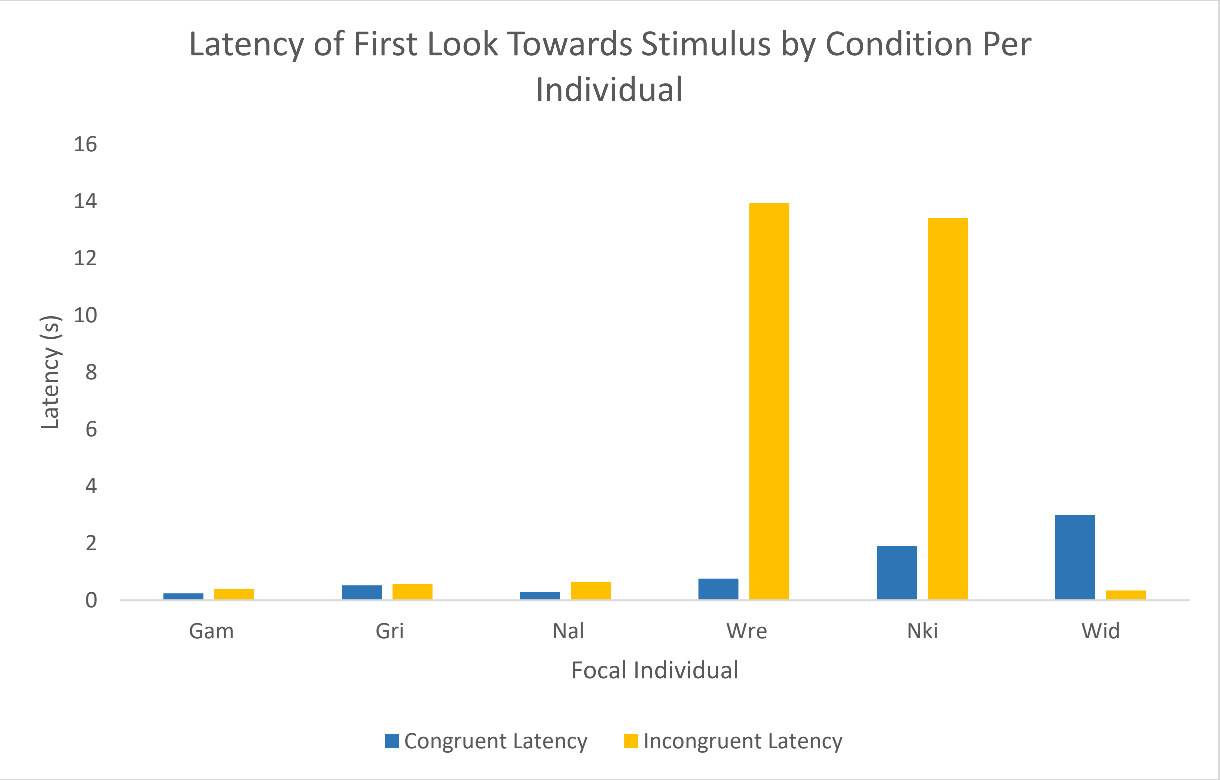


Figure 6. Latency of First Look Towards Stimulus by Condition Per Individual. This bar plot shows the latency in seconds (y axis) of each individual's first look towards the stimulus by condition. The congruent condition is depicted in blue, and the incongruent condition is depicted in yellow. Based on the results of the Wilcoxon Signed Rank Test (p-value = 0.4375), these results show no consistent trend across individuals.

Total Duration Looking Towards Stimulus

The total duration of time spent looking towards the stimulus over the 60s trial interval also showed no significant difference between the two conditions. The results for the two conditions showed a normal distribution according to the Shapiro-Wilk normality test (p-value = 0.6565), so a paired parametric t-test was run. The paired t-test yielded a p-value > 0.05 (p-value = 0.9984), so the null-hypothesis that no significant difference existed between the two conditions was accepted.

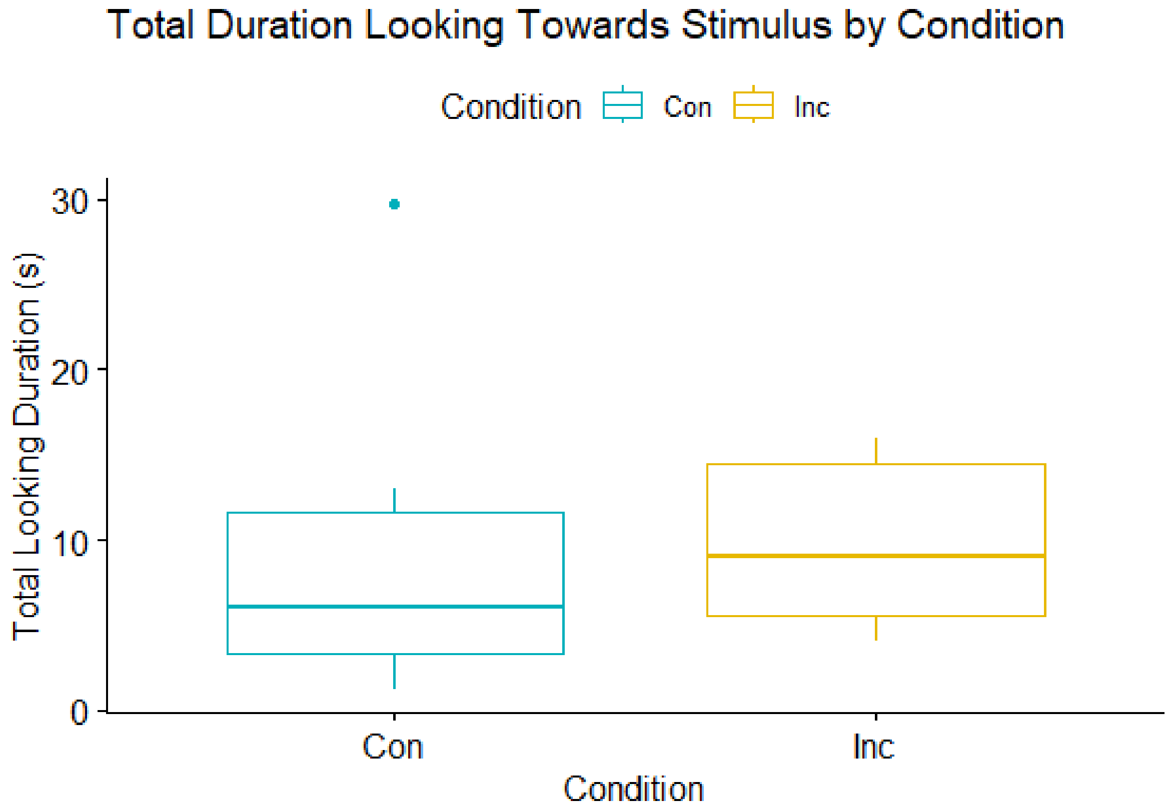


Figure 7. Total Duration Looking Towards Stimulus by Condition. This Ggplot shows the duration in seconds (y axis), averaging the total time each individual spent looking towards the stimulus over the course of the trial, by condition (x axis). The congruent condition is depicted in blue, and the incongruent condition is depicted in yellow. Based on the results of the paired parametric t-test (p-value = 0.9984), these results show no consistent trend across individuals.

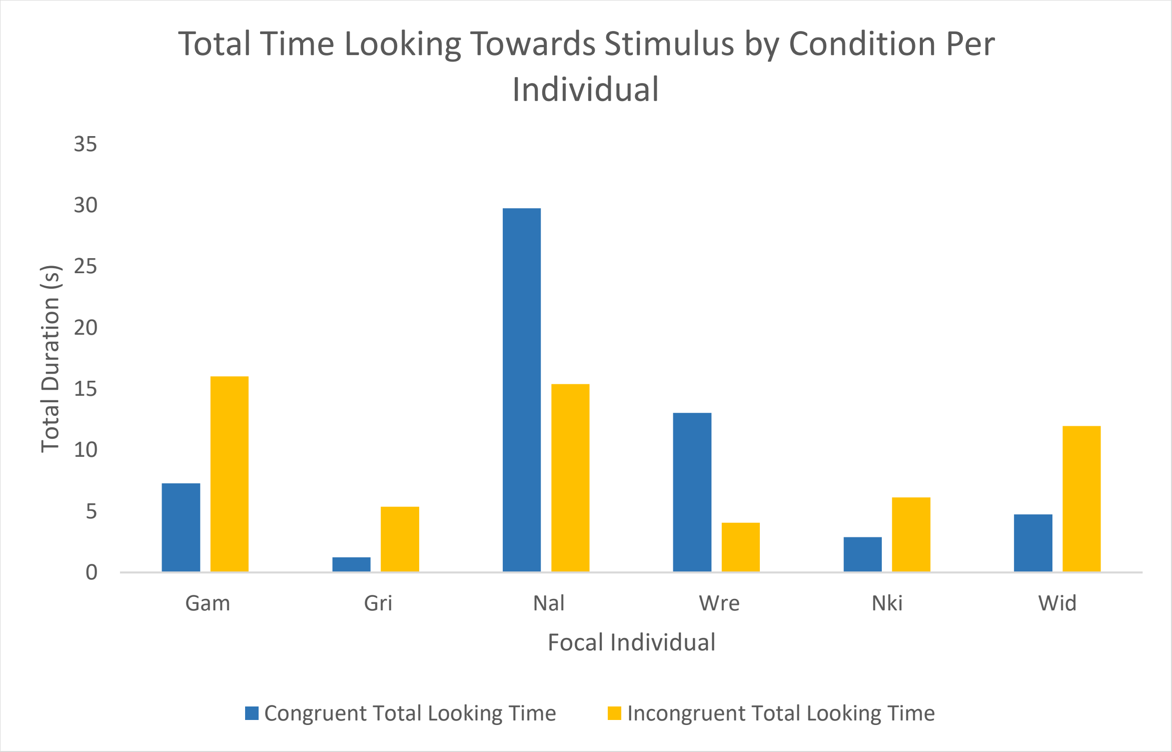


Figure 8. Total Time Looking Towards Stimulus by Condition Per Individual. This bar plot shows the duration in seconds (y axis) of each individual's cumulative time spent looking towards the stimulus over the course of the trial by condition. The congruent condition is depicted in blue, and the incongruent condition is depicted in yellow. Based on the results of the paired parametric t-test (p-value = 0.9984), these results show no consistent trend across individuals.

DISCUSSION

In this paper I set out to explore whether tufted capuchin vocalizations are individually distinct and whether individuals can distinguish the identity of callers from their neighboring groups based on these vocalizations. In line with my prediction, I found that tufted capuchin food calls do contain identifying information in their acoustic composition. The results from the DFA indicated that calls were significantly different between individuals, allowing the potential for individual recognition based on these differences. Preliminary results from the playback experiment did not show a statistically significant difference in how individuals responded between calls congruent with their social knowledge, that is calls from a male within his groups enclosure, and calls from male in another's group enclosure.

I found that acoustic features of the food call encode information about the identity of the caller. Specifically, the acoustic properties of the food call varied most between individual in terms of two factors, one encompassing minimum frequency and maximum frequency (Factor 1), and the other involving the number of elements and call duration (Factor 2). These findings are in line with studies of vocalizations in other primates. For example, in a study with free ranging black howler monkeys (Briseño-Jaramillo et al., 2015), minimum frequency was also a key feature in distinguishing the identity of the caller. In gelada monkeys, both acoustic and temporal features, including call duration and number of elements, were important in individual variation among gelada male loud calls (Benítez, 2017). While food whistles are quieter and more affiliative calls than howler and gelada loud calls, these vocalizations can still carry important information about caller identity. As such, individuals may attend to these calls to know which individuals are feeding around them. Future studies could examine how these, and other acoustic and temporal factors differ between individuals in other call types, such as the lost call whistle series tufted capuchins also display.

While food calls convey information about the identity of the caller, my preliminary results suggest that capuchins may not be using information from these calls to monitor identity of out-group individuals. I found no significant results between congruent and incongruent trials in terms of the duration and latency of their first look towards the stimulus, and the total time they spent looking towards the stimulus in the 60s analysis window. These results are surprising as playbacks in the wild with white-faced capuchin monkeys have shown that capuchins respond differently to food calls of within group and between group members (Bergman & Benítez, in prep). It is too soon, however, to draw any conclusions based on these results as the sample size is too small. I have two final playback trials to run which will allow us to compare 8 individuals and given the general trends (4/6 look more at incongruent stimulus), it is possible that this will alter the findings.

There are several reasons why I may have found the results I did in this playback study. The first may be that capuchins do not recognize the composition of other groups. This would be surprising, especially in a captive population, where capuchins have ample opportunity to attend to this information. The fact that all of the subjects responded to the stimuli suggests that they are interested in attending to food calls from the direction of other groups. I await the final two playback trials to make more conclusive remarks on whether capuchins gleam information about other groups from these calls.

Two individuals responded in ways that I did not predict but are interesting to further discuss. During the incongruent playback trial, one individual immediately turned away from the stimulus, towards the direction of the enclosure of the individual whose call was being played, and consistently looked in that direction throughout the playback, which the individual did not do during the congruent playback. This could be less a result of a lack of interest or surprise regarding the unexpected stimulus (e.g., “why is Griffin over there?”), and instead a result of an underlying awareness of the territory of the individual whose call was played (“wait isn’t Griffin back there?”). The latency of the individual’s head turn towards the back enclosures during the incongruent trial was lower than the head turn towards the stimulus in the congruent condition (Latency in Incongruent trial = 0.726 (s), Latency in Congruent Trial = 0.997 (s)). I did not code for this in my original analyses, however, I believe this is an interesting anecdote and may be something I code for in the final analysis.

The congruent trial for another individual in the group was one of the pilot sessions for the experiment. During the congruent trial, the individual did not receive peanut butter on the face of their enclosure, and did receive it in their incongruent trial at a later date. It is possible that they attended more to the stimulus in the congruent trial than the incongruent trial due to the lack of another stimulus to attend to, as individuals had in all other trials.

In addition to behavioral measures, the playback experiments in this study were also filmed using an Infrared Thermographic Camera (FLIR T540) that was recently validated for use assessing physiological changes to arousing stimuli within both a positive and negative context in the focal species *Sapajus apella*, and is also in the process of being validated for assessing comparative cortisol levels. Once this data is analyzed, it will be possible to more objectively examine the physiological responses individuals have to the various stimuli.

Finally, I recognize that this playback study has many limitations, including a relatively small samples size. I have three additional conditions that I plan to run on this population that will increase the number of trials per individual and allow me to piece apart what exactly is surprising from the stimulus being played. First, I plan to include a less invocatory stimuli consisting of radio calls recorded with the same recorder and recorder settings, and played at the same volume as the vocalizations were. This will provide a better “baseline” for the capuchins’ responses to familiar stimuli other than an individual’s call. Second, I will conduct a second incongruent condition where calls from the two neighboring groups are played from the opposite group. This differs from the first incongruent trial as it involves two groups that are further away than the adjacent group. Lastly, I will play a food call collected from an unknown individual. I predict that this condition will illicit very strong responses if the individuals are able to identify callers, as an unknown individual may be a larger threat than any of the individuals in their neighboring groups.

It is not possible to determine how well these playback results reflect capuchins in the wild. As these are captive capuchins, their territories are very discrete and unchanging. Future studies with this playback setup in a wild will be valuable for our understanding of how capuchins may acoustically map their socio-spatial relationships with neighboring groups when outgroup interactions and territories are more variable.

References

Bergman, Thore J. & Beehner, Jacinta C. (2015). Measuring social complexity, Animal Behaviour, Volume 103, Pages 203-209, https://doi.org/10.1016/j.anbehav.2015.02.018.

Bergman, T., & Sheehan, M. (2012, December 10). Social Knowledge and Signals in Primates. Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1002/ajp.22103.

Benítez, M. E., le Roux, A., Fischer, J., Beehner, J. C., & Bergman, T. J. (2016, October 27). Acoustic and temporal variation in Gelada (Theropithecus gelada) loud calls advertise male quality - International Journal of Primatology. SpringerLink. https://link.springer.com/article/10.1007/s10764-016-9922-0.

Boersma, Paul & Weenink, David (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.09. http://www.praat.org/

Briseño-Jaramillo, M., Estrada, A., & Lemasson, A. (2015). Individual voice recognition and an auditory map of neighbours in free-ranging black howler monkeys (Alouatta pigra). American Psychological Association. https://psycnet.apa.org/record/2014-51416-001.

Calcutt S. E., Rubin T. L., Pokorny J. J., de Waal F. B. Discrimination of emotional facial expressions by tufted capuchin monkeys (Sapajus apella). J Comp Psychol. 2017 Feb; Volume 131, pages:40-49. doi: 10.1037/com0000055.

Champely, Stephane. (2018). PairedData: Paired Data Analysis. R package version 1.1.1. https://CRAN.R-project.org/package=PairedData

Di Bitetti, M. S. (2003). Food-Associated Calls of Tufted Capuchin Monkeys (Cebus apella nigritus) Are Functionally Referential Signals. Behaviour, 140(5), 565–592. http://www.jstor.org/stable/4536046.

Di Bitetti, Mario & Wheeler, Brandon. (2018). The vocal repertoire of the black-horned capuchin (Cebus [Sapajus] nigritus): an acoustic and contextual analysis.

Emery, N. J., Clayton, N. S., & Frith, C. D. (2007, April 29). Introduction. social intelligence: From brain to culture. Philosophical Transactions of the Royal Society B: Biological Sciences. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2042522.

Fischer, J., Noser, R., & Hammerschmidt, K. (2013). Bioacoustic field research: a primer to acoustic analyses and playback experiments with primates. American journal of primatology, Volume 75, Issue 7, pages 643–663. https://doi.org/10.1002/ajp.22153.

Friard, O., & Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. Methods in Ecology and Evolution, 7(11), 1325–1330.

Herbinger, I., Papworth, S., Boesch, C., & Zuberbühler, K. (2009, October 27). Vocal, gestural and locomotor responses of wild chimpanzees to familiar and unfamiliar intruders: A playback study. Animal Behaviour. https://www.sciencedirect.com/science/article/abs/pii/S0003347209004151?via%3Dihub

Herrmann, E., Call, J., Hernandez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007, September 7). Humans have evolved specialized skills of social cognition The Cultural Intelligence Hypothesis. Science.org. https://www.science.org/doi/10.1126/science.1146282.

IBM Corp. (2020). IBM SPSS Statistics for Windows (Version 27.0) [Computer software]. IBM Corp.

Kassambara, Alboukadel. (2023). “ggpubr: 'ggplot2' Based Publication Ready Plots.” R package version 0.6.0. https://CRAN.R-project.org/package=ggpubr.

K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology. (2023). Raven Pro: Interactive Sound Analysis Software (Version 1.6.4) [Computer software]. Ithaca, NY: The Cornell Lab of Ornithology. https://ravensoundsoftware.com/.

Lakshminarayanan, V. R., & Santos, L. R. (2008, November 10). Capuchin monkeys are sensitive to others' welfare. Current Biology. https://www.sciencedirect.com/science/article/pii/S0960982208011354.

Mundry, R., & Sommer, C. (2007, October). Discriminant function analysis with nonindependent data: consequences and an alternative. ScienceDirect. Animal Behaviour, Volume 74, Issue 4, Pages 965-976. https://www.sciencedirect.com/science/article/abs/pii/S0003347207002618?via%3Dihub.

Perry, S. (2011, April 12). Social traditions and social learning in Capuchin Monkeys (cebus). Philosophical transactions of the Royal Society of London. Series B, Biological sciences. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3049088/.

Pokorny, J. J., & de Waal, F. B. M. (2009, December 22). Monkeys recognize the faces of group mates in photographs. PNAS. Retrieved from https://doi.org/10.1073/pnas.0912174106

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Seyfarth, R. M., & Cheney, D. L. (2016, June 22). The origin of meaning in animal signals. Animal Behaviour. https://www.sciencedirect.com/science/article/abs/pii/S0003347216300768.

Talbot, C. (2016). Discrimination of Faces, Sex, and Relationships by Capuchin Monkeys. Semantic Scholar. https://scholarworks.gsu.edu/psych_diss/156/.

Tavares, M. C. H., & Tomaz, C. (2002, February 6). Working memory in Capuchin Monkeys (Cebus apella). Behavioural Brain Research. https://www.sciencedirect.com/science/article/abs/pii/S0166432801003680.

Westergaard, G. C. (2010, July 15). The stone-tool technology of Capuchin Monkeys: Possible Implications for the evolution of symbolic communication in Hominids. Taylor & Francis. https://www.tandfonline.com/doi/abs/10.1080/00438243.1995.9980289.