

This PDF includes a chapter from the following book:

The Cognitive Animal

Empirical and Theoretical Perspectives on Animal Cognition

© 2002 Massachusetts Institute of Technology

License Terms:

Made available under a Creative Commons
Attribution-NonCommercial-NoDerivatives 4.0 International Public License
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

OA Funding Provided By:

The open access edition of this book was made possible by generous funding from Arcadia—a charitable fund of Lisbet Rausing and Peter Baldwin.

The title-level DOI for this work is:

[doi:10.7551/mitpress/1885.001.0001](https://doi.org/10.7551/mitpress/1885.001.0001)

Cory T. Miller and Asif A. Ghazanfar

An experience common to all of us who travel to foreign countries is trying to make sense of the confusing sounds being uttered by native speakers. Unless one can find a translator or at least a good bilingual dictionary, one often will have an extraordinarily difficult time expressing one's basic needs for food and shelter. In such situations, we lack not only the knowledge of what the different words mean but also an understanding of where the boundaries are for the different acoustic units within a foreign stream of speech. The extreme version of this Quinean problem of translation (Quine 1973) applies not only to linguists and foreign travelers but also to those of us who wish to shed light on the vocal communication systems of other species (Hauser 1996).

Ethologists studying nonhuman animal communication systems are faced with the daunting task of dividing the vocal repertoire into different types of acoustic units (e.g., bouts, vocalizations, syllables). Specifically, how can one determine whether a sequence of temporally distinct units emitted by an animal represents a single functional unit (such as meaningless syllables put together to form a word in speech), a string of functionally independent units (such as words forming a sentence in speech), or something simpler, such as the repetition of one small unit? A true understanding of how vocal signals are parsed must be derived from the animal's perspective. Vocal signals must be parsed into the acoustic units that are meaningful in terms of eliciting specific behaviors from the intended receivers (Green and Marler 1979; Hauser 1996). Thus, an animal's behavior serves as a "translator" for ethologists entering a species' perceptual world. Using this approach, we have learned much about the meaningful acoustic units in many avian (e.g., Podos et al. 1992; Searcy et al. 1999) and anuran vocal repertoires (e.g., Narins and Capranica 1978; Ryan and Rand 1990).

Studies in these taxa have given us significant insights into how vocal behavior relates to brain design.

Like birds and anurans, many nonhuman primate (hereafter, primate) species produce bouts of vocalizations containing sequences of similar acoustic units and/or different-sounding acoustic units (figure 33.1), but we know very little about the meaningful units in primate vocal signals. An understanding of how primates perceive and produce such vocalizations is important for several reasons. First, the evolution of speech and language may have involved selection for capabilities that existed in extant primates (Lieberman 1984; Ghazanfar and Hauser 1999; Fitch 2000). One such capability may be to produce vocal signals that mean one thing when produced individually, but something different when recombined into sequences of sounds serve an entirely different function. Second, from a more general perspective, an understanding of the constraints on the perceptual and motor domains of primates' vocal behavior may provide us with insights into the species-specific perceptual world and thus their cognitive abilities and limitations. In this essay we review our understanding of the meaningful acoustic units of production and perception in primate vocal communication. Given the space constraints, a complete literature review cannot be provided here. For a list of studies relevant to this issue, see table 33.1.

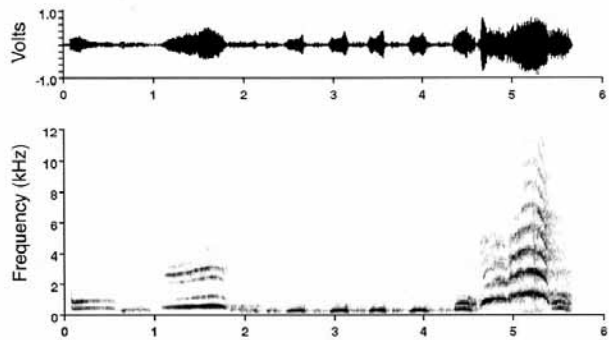
To begin this discussion, we would like to clarify our use of the term *acoustic unit*. For the purposes of this review, we refer to all temporally distinct acoustic pulses as syllables. This criterion is arbitrary in the sense that it is based solely on an acoustic measure, not a behavioral one. Nevertheless, it serves as a good starting point. Our own research is aimed at refining this definition as we gain a better understanding of the functional and perceptual significance of all

A

Chimpanzee



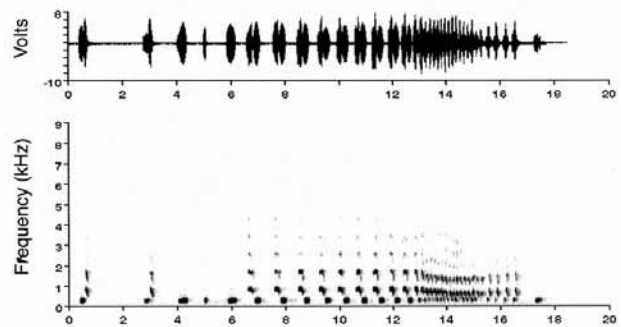
Pant-hoot

**B**

Gibbon/Siamang



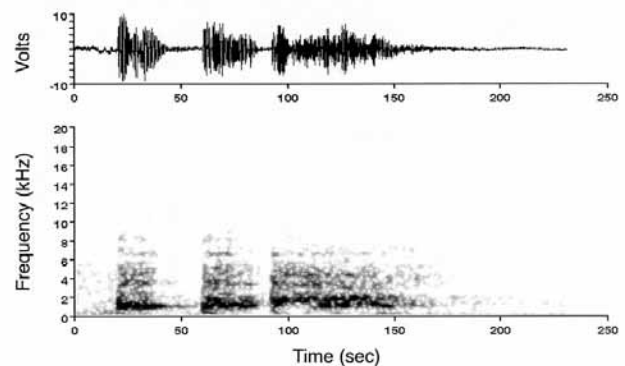
'Song'

**C**

Rhesus monkey



Shrill Bark

**Figure 33.1**

Examples of multisyllabic calls in primates. Time-amplitude waveforms (top right) and spectrograms (bottom right). (A) Chimpanzee pant-hoot. (Photograph courtesy of P. Marler; vocalization courtesy M. Wilson and M. Hauser.) (B) Siamang (gibbon) song. (Photograph courtesy of Roy Fontaine/Monkey Jungle; vocalization courtesy of Thomas Geissmann.) (C) Rhesus monkey shrill bark. (Photograph courtesy of Cayo Santiago photo archives; vocalization courtesy of M. Hauser.)

Table 33.1
Investigations of meaningful acoustic units in primate communication

Vocal Production	Vocal Perception
<i>Observations</i>	
Titi monkeys	Titi Monkeys
Robinson (1979)	Robinson (1979)
Capuchins	Capuchins
Robinson (1984)	Robinson (1984)
Gibbons	Gibbons
Mitani and Marler (1989)	Mitani and Marler (1989)
Goustard (1976)	
Chimpanzees	
Marshall et al. (1999)	
<i>Experiments</i>	
Tamarins	Titi monkeys
Miller and Hauser (in prep)	Robinson (1979)
	Gibbons
	Mitani and Marler (1989)
	Tamarins
	Ghazanfar et al. (2001)
	Rhesus monkeys
	Hauser et al. (1998)

acoustic units within primate vocal signals; i.e., determining how acoustic units are organized to form units that are “meaningful” from the specie’s perspective.

Observations and Experiments in the Field

Insights into the units of production and perception are possible from studies of the orderly arrangement of syllables within primate long calls. Long calls serve as localization cues for conspecifics and are often produced in the context of territorial encounters, mate attraction, and isolation or group cohesion (Marler 1968; Waser 1982). These multisyllabic calls are produced by a range of different primate species, such as the “whoop-gobble” of mangabeys (*Cercocebus*

albigena; Waser 1977), the “chirrup-pant-bellow-pump-honk” call sequences of titi monkeys (*Callicebus moloch*; Robinson 1979), the “chirp-trills-squaws” of wedge-capped capuchins (*Cebus olivaceus*; Robinson 1984), the “twitters” of squirrel monkeys (*Saimiri sciureus*; Newman et al. 1978), the “whinny” of spider monkeys (*Ateles geoffroyi*; Chapman and Weary 1990; Teixidor and Byrne 1999), the “pant-hoot” of chimpanzees (figure 33.1A) (*Pan troglodytes*; Clark Arcadi 1996; Mitani and Nishida 1993), and the “songs” of gibbons (figure 33.1B) (*Hylobates* spp.; Mitani 1985; Mitani and Marler 1989). Given the similarities in the acoustic structure of long calls throughout the primate order, data on meaningful units in one species may be generalized to other species’ communication systems.

In vocal signals consisting of multiple syllables, both the individual syllables and combinations of syllables may mediate specific aspects of behavior. As a result, for any vocalization, information may be encoded at multiple levels. For example, the chimpanzee pant-hoot consists of a series of “hoo” calls followed by a series of screams (figure 33.1A). Since both hoos and screams are produced individually in other contexts, the pant-hoot could be either a single vocalization or a bout of several vocalizations. To reframe the question from a production standpoint, is the vocal control during pant-hoot production akin to a tape recorder that once activated must run to completion, or can control be exerted for each syllable? Recently, Marshall and colleagues (1999) reported that two different, genetically unrelated groups of captive chimpanzees incorporated a novel vocal element (dubbed the “Bronx cheer”) at either the beginning or end of pant-hoots, but never in the middle of the call. This suggests that although each syllable can be produced independently, perhaps the pant-hoot is encoded as a single vocal signal that can be added to, but cannot be internally modified.

Not only is the syntactic arrangement of syllables within long calls relevant to our understanding of vocal production, it can also provide insights into the units of perception. Robinson (1979) and Mitani and Marler (1989) observed that the long calls of titi monkeys and the songs of gibbons contain sequences of syllables that vary in their order. The males of these species recombine different syllable types to create unique phrases or songs, and certain orders of syllables have a higher probability of occurring than others. Based on these observations, a series of field playback experiments were conducted to determine whether the different syllable arrangements serve different functions in these primates. Playbacks in which the syllables were arranged in a low probability sequence revealed that both gibbons and titi monkeys recognize when the syllables in conspecific vocalizations occur in an

atypical order. Gibbons produced significantly more “squeak” calls (given during intergroup encounters) upon hearing a song with experimentally rearranged notes (Mitani and Marler 1989). Similarly, titi monkeys produced significantly more moaning responses (given in response to interspecies and intergroup encounters) following playbacks of conspecific vocalizations with abnormal arrangements of syllables (Robinson 1979).

The first step in assessing the meaningful units of a vocal signal is to determine the smallest invariable acoustic unit that can influence behavior. For example, Hauser et al. (1998) recently conducted a field playback study on rhesus macaques (*Macaca mulatta*) demonstrating that the intersyllable interval plays an important role in conspecific vocal recognition for some vocalization types. Using the orienting preference experimental assay developed by Hauser and Andersson (1994), they increased or decreased the intersyllable interval beyond the species’ typical range for three call types: grunts, shrill barks (figure 33.1C), and copulation screams. Each of these three call types consists of repeated syllables with similar acoustic structures. Playbacks of these manipulated calls revealed that rhesus monkeys recognize grunts and shrill barks based on the overall, multisyllabic call structure, while copulation screams are recognized at the level of the individual syllable. In other words, for grunts and shrill barks, the smallest meaningful unit of perception is the whole, multisyllabic call, while for copulation screams, a single syllable is the smallest meaningful unit of perception.

These data provide an important foundation for primate communication studies, but there are still many unanswered questions concerning the meaningful units of primate vocal signals. Do each of the individual syllables represent distinct units that can be recombined under the caller’s volition, or are calls with different syllable orders each representative of completely different vocalizations? Such questions are directly relevant to understanding the constraints on vocal control.

In terms of perception, no studies of primate calls have systematically varied syllable order or presented subjects with isolated syllables to determine the functionality of syllables within a vocal signal. Such an approach is necessary if we are to fully understand the organization of complex design features in primate vocal signals.

Unit of Production in Cotton-Top Tamarins

Our research focuses on a small New World primate, the cotton-top tamarin (*Saguinus oedipus*). This species has a complex vocal repertoire in which similar-sounding syllables can be used in several different vocalizations (Cleveland and Snowdon 1981). One vocal signal of particular interest is the combination long call, a contact call emitted when individuals are isolated from members of their colony (Cleveland and Snowdon 1981; Weiss et al. in press) (figure 33.2A). The vocalization is multisyllabic, composed of 1–3 chirps followed by 1–5 whistles, and thus is ideal for addressing questions about meaningful units of production and perception.

We first set out to examine the organization of syllables within the cotton-top tamarin long call from the perspective of vocal production. Specifically, is the entire long call produced as a single unit, or is each syllable its own encapsulated unit that can be produced in isolation? To address this question, we borrowed an elegant experimental technique originally used by Cynx (1990) to test a similar question in a songbird, the zebra finch (*Taeniopygia guttata*).

In this experiment, a bright light was flashed at subjects while they produced their multisyllabic song. Cynx predicted that if the song was its own unit of production, then subjects should be unaffected by the light flash and continue singing until the song was complete. If the smaller acoustic units within the song (i.e., notes, syllables, or motifs) represented the unit of production, then subjects should be able to stop singing in the middle of the song following the

completion of one of these units. He found that zebra finches interrupted their song approximately 60 percent of the time following a light flash. Interestingly, when the song was interrupted, subjects consistently completed the syllable already being produced before ending the vocalization. Based on these results, Cynx argued that the minimal unit of production in zebra finch song is the syllable.

We used the same logic to assess the units of production in tamarin long call production and attempted to interrupt vocal production using both light flashes and white noise bursts (Miller and Hauser, in preparation). Overall, only 14 percent of all calls were interrupted. Of those trials in which an interruption occurred, the auditory stimulus was more effective (22 percent of occurrences) than the light flash (5 percent) at causing interruptions. During instances when calls were interrupted, the subjects always completed the syllable already in production, suggesting that the minimal unit of production in the tamarin long call is the syllable. The low rate of interruptions in comparison with birds, however, suggests that although the syllable is the smallest unit of production, the organization of the syllables is dramatically different than in bird song. While zebra finches appear to have good control over the acoustic units within their song, tamarins exhibit a more limited degree of vocal control over syllables within the long call.

Unit of Perception in Cotton-Top Tamarins

In a variety of vertebrate species, individuals often respond to territorial and/or contact calls by producing an identical or similar-sounding call. Such “antiphonal calling” can be used as a robust behavioral assay for investigating which acoustic features are important in a given vocal signal. For example, in an elegant study of the two-note mating call of the Coqui frog (*Eleutherodactylus coqui*), Narins and Capranica (1978) used the antiphonal calling of male frogs to de-

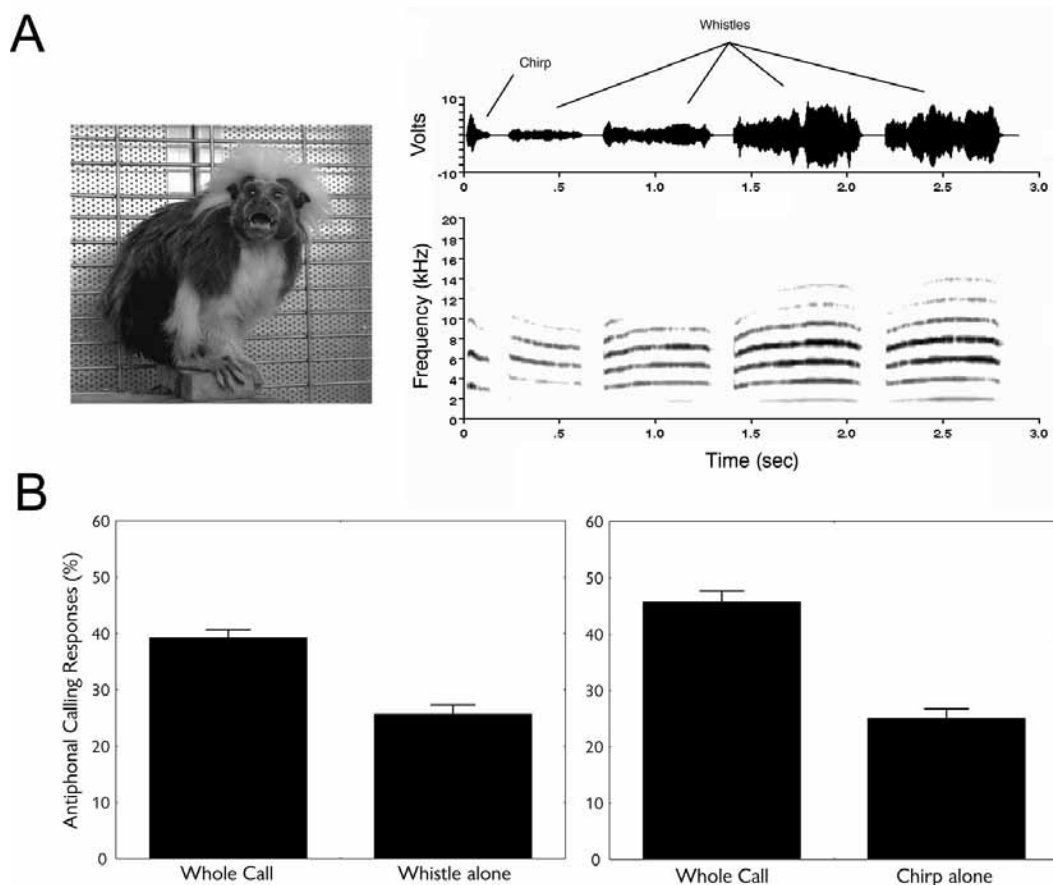


Figure 33.2

(A) Image of a cotton-top tamarin (left). Cotton-top tamarin combination long call: Time-amplitude waveforms (top right) and spectrogram (bottom right) of a representative combination long call recorded in captivity. Chirps and whistles are labeled. (B) Antiphonal calling responses to playbacks of whole long calls versus single syllables (whistles and chirps). Error bars show 1 standard error of mean.

termine which of the two notes in the call elicited the antiphonal response. Using field playbacks, they found that the first note elicits antiphonal calling responses from conspecific males, but the second note does not.

Recently, we used a similar antiphonal calling assay to examine the unit of perception for the cotton-top tamarin long call (Ghazanfar et al. 2001). We measured the antiphonal long calling responses of isolated tamarin subjects to playbacks of (1) whole long calls, (2) isolated whistles, (3) isolated chirps, (4) white noise bursts, and (5) artificial sequences of chirps. Our goal was to determine the acoustic cues necessary to elicit antiphonal long calls from isolated individuals.

In some situations, cotton-top tamarins produce chirps and whistles independently, but combination long calls consist of a concatenation of both syllable types occurring sequentially in a single vocalization (Cleveland and Snowdon 1981). We hypothesized that the combination of both syllable types in the form of a combination long call would be a more effective stimulus for eliciting antiphonal long calls from our subjects than either unit alone (see figure 33.2A). This was indeed the case. Whole calls were much more effective in eliciting antiphonal long calls than either single whistles or single chirps (figure 33.2B). It is important to note that the level of antiphonal calling to single whistles or chirps was not significantly different from the level of calling back to white noise, suggesting that those response levels represent a baseline. In terms of overall long call production rate, whole call playbacks were also able to elicit more calls than single whistles or chirps. Response levels to whole calls versus sequences of chirps were equivalent, providing clues to the acoustic cues necessary in eliciting antiphonal calls by isolated tamarins. Overall, our data suggest that the most meaningful unit from the perspective of socially isolated receivers is the whole call. This represents an important first step in deciphering the perceptually salient features of the cotton-top tamarins' combination long call.

Future Directions

One of the central questions in the study of animal vocal communication concerns how the structure of acoustic signals relates to the behavior of receivers. Like learning any foreign communication system, one of the first problems ethologists confront is the difficulty of parsing acoustic signals into the relevant units of analysis—that is, units that are meaningful in terms of eliciting specific behaviors from the intended receivers (Chomsky and Halle 1968; Green and Marler 1979). For primate vocal behaviors, such investigations are only in their infancy, yet the implications for understanding the cognitive capacities and brain design of primates are tremendous (Ghazanfar and Hauser 1999).

Our research on cotton-top tamarin long calls suggests that the smallest unit of production is the syllable, but that the smallest unit of perception is the whole call. While our understanding of the units of perception and production is relatively limited at this point, new data will provide more insights into the meaningful units in tamarin vocal communication and the neural substrates for vocal behavior. At the perceptual level, for example, there are many behavioral experiments in both humans and other taxa demonstrating that a common strategy for signal identification and localization is sensitivity to two or more spectral or temporal components. Detection of such sound combinations often leads to individuals producing species-specific behaviors. In our case, tamarins antiphonally call more readily to combination long calls containing both the chirp and multiple whistles than to either syllable type alone.

One candidate neuronal mechanism that may underlie this behavioral selectivity is the nonlinear response of neurons to the combined elements of a signal compared with responses to the individual elements alone. Neurons that are combination sensitive to multisyllabic species-specific vocalizations are described extensively

in the auditory systems of bats, frogs, and songbirds (Fuzessery and Feng 1983; Margoliash and Fortune 1992; O'Neill and Suga 1979). We predict that, like other organisms, primates with long, multisyllabic vocalizations will have neurons that are combination sensitive, particularly when the meaningful unit of perception is the whole call, not the individual syllables. In the present case, we predict that the central auditory system of tamarins should contain many neurons that are combination sensitive to both temporal and spectral components of the chirp-whistle sequence.

Acknowledgments

We would like to thank Marc Hauser for his support, guidance, and unfaltering enthusiasm in all aspects of our investigations. We would also like to thank R. Delgado, M. Hauser, V. Janik, D. Katz, and L. Santos for their excellent comments on earlier versions of this manuscript. This work was funded by Harvard University and by a grant from the National Institutes of Health (grant DC00377-02) to A.A.Ghazanfar and National Science Foundation grants (SBR-9602858 and 9357976) to M. Hauser.

References

Chapman, C. A. and Weary, D. M. (1990). Variability in spider monkey's vocalizations may provide basis for individual recognition. *American Journal of Primatology* 22: 279–284.

Chomsky, N. and Halle, M. (1968). *The Sound Patterns of English*. Cambridge, Mass.: MIT Press.

Clark Arcadi, A. (1996). Phrase structure of wild chimpanzee pant hoots: Patterns of production and interpopulation variability. *American Journal of Primatology* 29: 159–178.

Cleveland, J. and Snowdon, C. T. (1981). The complex vocal repertoire of the adult cotton-top tamarin, *Saguinus oedipus oedipus*. *Zeitschrift für Tierpsychologie* 58: 231–270.

Cynx, J. (1990). Experimental determination of a unit of song production in zebra finch (*Taeniopygia guttata*). *Journal of Comparative Psychology* 104: 3–10.

Fitch, W. T. (2000). The evolution of speech. *Trends in Cognitive Science* 4: 258–267.

Fuzessery, Z. M. and Feng, A. S. (1983). Mating call selectivity in the thalamus and midbrain of the leopard frog (*Rana p. pipiens*): Single and multiunit analyses. *Journal of Comparative Physiology A* 150: 333–344.

Ghazanfar, A. A. and Hauser, M. D. (1999). The neuroethology of primate vocal communication: Substrates for the evolution of speech. *Trends in Cognitive Sciences* 3: 377–384.

Ghazanfar, A. A., Flombaum, J. I., Miller, C. T., and Hauser, M. D. (2001). The units of perception in the antiphonal calling behavior of cotton-top tamarin (*Saguinus oedipus*): Playback experiments with long calls. *Journal of Comparative Physiology A* 187: 27–35.

Goustard, M. (1976). The vocalizations of hylobates. In *Gibbon and Siamang*, D. M. Rumbaugh, ed., pp. 135–166. Basel, Switzerland: S. Karger.

Green, S. and Marler, P. (1979). The analysis of animal communication. In *Social Behavior and Communication, Handbook of Behavioral Neurobiology*. Vol. 3, P. Marler and J. Vandenbergh, eds., pp. 73–158. New York: Plenum.

Hauser, M. D. (1996). *The Evolution of Communication*. Cambridge, Mass.: MIT press.

Hauser, M. D. and Andersson, K. (1994). Left hemisphere dominance for processing vocalizations in adult, but not infant rhesus monkeys. *Proceedings of the National Academy of Sciences, U.S.A.* 91: 3946–3948.

Hauser, M. D., Agnetta, B., and Perez, C. (1998). Orienting asymmetries in rhesus monkeys: Effect of time-domain changes on acoustic perception. *Animal Behaviour* 56: 41–47.

Lieberman, P. (1984). *The Biology and Evolution of Language*. Cambridge, Mass.: Harvard University Press.

Margoliash, D. and Fortune, E. S. (1992). Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *Journal of Neuroscience* 12: 4309–4326.

Marler, P. (1968). Aggregation and dispersal: Two functions in primate communication. In *Primates*:

- Studies in Adaptation and Variability*, P. C. Jay, ed., pp. 420–438. New York: Holt, Rinehart, and Winston.
- Marshall, A. J., Wrangham, R. W., and Arcadi, A. C. (1999). Does learning affect the structure of vocalizations in chimpanzees? *Animal Behaviour* 58: 825–830.
- Miller, C. T., and Hauser, M. D. (in prep). What is the unit of production of cotton-top tamarin (*Saguinus oedipus*) long calls?
- Mitani, J. C. (1985). Responses of gibbons (*Hylobates muelleri*) to self, neighbor, and stranger duets. *International Journal of Primatology* 6: 193–200.
- Mitani, J. C. and Marler, P. (1989). A phonological analysis of male gibbon singing behavior. *Behaviour* 109: 20–45.
- Mitani, J. and Nishida, T. (1993). Contexts and social correlates of long distance calling by male chimpanzees. *Animal Behaviour* 45: 735–746.
- Narins, P. M. and Capranica, R. R. (1978). Communicative significance of the two-note call of the treefrog *Eleutherodactylus coqui*. *Journal of Comparative Physiology, A* 127: 1–9.
- Newman, J. D., Liebllich, A., Talmage-Riggs, G., and Symmes, D. (1978). Syllable classification and sequencing in twitter calls of squirrel monkeys (*Saimiri sciureus*). *Zeitschrift für Tierpsychologie* 47: 77–88.
- O'Neill, W. E. and Suga, N. (1979). Target range-sensitive neurons in the auditory cortex of the mustache bat. *Science* 251: 565–568.
- Podos, J., Peters, S., Rudnicki, T., Marler, P., and Nowicki, S. (1992). The organization of song repertoires in song sparrows: Themes and variations. *Ethology* 90: 89–106.
- Quine, W. V. (1973). On the reasons for the indeterminacy of translation. *Journal of Philosophy* 12: 178–183.
- Robinson, J. G. (1979). An analysis of the organization of vocal communication in the titi monkey *Callicebus moloch*. *Zeitschrift für Tierpsychologie* 49: 381–405.
- Robinson, J. G. (1984). Syntactic structure in the vocalizations of wedge-capped capuchin monkeys, *Cebus olivaceus*. *Behaviour* 90: 46–79.
- Ryan, M. J. and Rand, A. S. (1990). The sensory basis of sexual selection for complex calls in the Tungara frog, *Physalaemus pustulosus* (sexual selection for sensory exploitation). *Evolution* 44: 305–314.
- Searcy, W. A., Nowicki, S., and Peters, S. (1999). Song types as fundamental units in vocal repertoires. *Animal Behaviour* 58: 37–44.
- Teixidor, P. and Byrne, R. W. (1999). The “whinny” of spider monkeys: Individual recognition before situational meaning. *Behaviour* 136: 279–308.
- Waser, P. M. (1977). Individual recognition, intra-group cohesion, and intergroup spacing: Evidence from sound playback to forest monkeys. *Behaviour* 60: 28–74.
- Waser, P. M. (1982). The evolution of male loud calls among mangabeys and baboons. In *Primate Communication*, C. T. Snowdon, C. H. Brown, and M. R. Petersen, eds., pp. 117–143. New York: Cambridge University Press.
- Weiss, D., Garibaldi, B., and Hauser, M. D. (in press). Individual recognition of cotton-top tamarin (*Saguinus oedipus oedipus*) long calls. *Journal of Comparative Psychology*.

