Cognitive Control of Distinct Vocalizations in Rhesus Monkeys

Steffen R. Hage, Natalja Gavrilov, and Andreas Nieder

Abstract

■ Whether nonhuman primates can decouple their innate vocalizations from accompanied levels of arousal or specific events in the environment to achieve cognitive control over their vocal utterances has been a matter of debate for decades. We show that rhesus monkeys can be trained to elicit different call types on command in response to arbitrary visual cues. Furthermore, we report that a monkey learned to switch between two distinct call types from trial to trial in response to different visual cues. A controlled behavioral protocol and data analysis based on signal detection theory showed that noncognitive factors as a cause for the monkeys' vocalizations could be excluded. Our findings also suggest that monkeys also have rudimentary control over acoustic call parameters. These findings indicate that monkeys are able to volitionally initiate their vocal production and, therefore, are able to instrumentalize their vocal behavior to perform a behavioral task successfully. ■

INTRODUCTION

Speech is one of the key defining features of humans, allowing us most sophisticated audio-vocal communication (Balter, 2010; Ghazanfar, 2008). A necessary criterion for language production is volition. Speech sounds that we learn throughout our lives can be uttered or withheld on command. In contrast, vocal utterances of our closest relatives in the animal kingdom, nonhuman primates, are innate and genetically predetermined. Unlike arm or hand movements, the vocal motor network of nonhuman primates is not controlled by the primary motor cortex, premotor cortex, cerebellum, and corresponding structures, which usually enable volitional motor control. Rather, vocal utterances in monkeys are produced by an extrapyramidal vocal motor network that includes the ACC and several subcortical structures such as the periaqueductal gray and the ventrolateral reticular formation (Hage, 2009; Jürgens, 2002, 2009).

Because of these differences, the degree to which nonhuman primates are capable of volitional call initiation and modulation has been discussed controversially for decades: On the one hand, monkey vocalizations can both be linked to different levels of arousal while simultaneously providing listeners with information about specific events in the environment (Seyfarth & Cheney, 2003; Manser, Seyfarth, & Cheney, 2002). Such vocal utterances show only involuntary modulation of specific call patterns in response to external stimuli. Here, the most prominent example is the Lombard effect, which is an involuntary rise in call amplitude in response to masking ambient noise (Brumm & Zollinger, 2011; Zollinger & Brumm, 2011; Brumm & Slabbekoorn, 2005).

On the other hand, several behavioral studies report that monkeys are able to control—at least rudimentarily—vocal initiation and to "decide" which call type to utter. Nonhuman primates are able to produce a vocalization or remain silent when submitted to operant conditioning tasks (Coude et al., 2011; Koda, Oyakawa, Kato, & Masataka, 2007; Hihara, Yamada, Iriki, & Okanoya, 2003; Aitken & Wilson, 1979; Sutton, Larson, Taylor, & Lindeman, 1973). These studies support field studies that show that nonhuman primates vocalize in different ways when addressing different individuals (for a review, see Cheney & Seyfarth, 2007) and produce or withhold alarm calls depending on the social context (Rendall, Seyfarth, Cheney, & Owren, 1999; Seyfarth, Cheney, & Marler, 1980). Hereby, calls might encode the vocalizing individuals' information about the presence of a predator (Zuberbühler, Cheney, & Seyfarth, 1999), other individuals' behavior (Wich & de Vries, 2006), or specific external events (Ouattara, Lemasson, & Zuberbühler, 2009). Furthermore, nonhuman primates modulate vocal timing in response to external ambient noise (Roy, Miller, Gottsch, & Wang, 2011; Egnor, Wickelgren, & Hauser, 2007) or conspecific calls (Hage, 2013; Miller, Beck, Meade, & Wang, 2009) and are able to make fine-scale modifications in the acoustic structure of their vocalizations (for a review, see Seyfarth & Cheney, 2010). Despite this rudimentary control of calls, several studies suggest that nonhuman primates are unable to use different calls interchangeably in different contexts or in different conditions (for reviews, Seyfarth & Cheney, 2010; Hammerschmidt & Fischer, 2008).

University of Tübingen

Journal of Cognitive Neuroscience 25:10, pp. 1692–1701
doi:10.1162/jocn_a_00428

A major concern of most behavioral studies is that potential motivational effects that might affect vocal behavior cannot be excluded. Specific call types that are produced in response to distinct external events might just be the result of different motivational states bound to specific vocalizations that are associated with these external events. Changes in vocal timing in response to external auditory stimuli might simply arise from threshold effects of audio-vocal integration mechanisms. In most conditioning experiments, nonhuman primates were trained to utter a vocalization in response to visually presented food rewards that trigger calls motivationally (Coude et al., 2011; Koda et al., 2007; Hihara et al., 2003). Moreover, success rates seems to be highly variable in such conditioning experiments (Pierce, 1985). Therefore, it cannot be excluded that these studies, instead of demonstrating the capability to volitionally vocalize on command, rather confirm that nonhuman primates produce adequate motivationally based responses to hedonistic stimuli. A strong argument for cognitive control of onset and type of vocal output would require evidence that nonhuman primates are capable to reliably vocalize in response to arbitrary (i.e., nonhedonic and nonsocial) cues in a highly controlled experimental design.

In the current study, we therefore first trained two rhesus monkeys (*Macaca mulatta*) to perform a computer-controlled go/no-go visual detection task by using their vocalizations as a response. We show that they are able to achieve volitional control on their vocal output and use it as immediate response to an abstract, learned cue, thus demonstrating the ability to instrumentalize their vocal output to perform a task successfully. Second, we trained one of these monkeys to selectively emit two different call types in response to distinct visual cues that were presented in random order. Thus, we demonstrate that the monkey is capable to switch call types from trial to trial. These behavioral results open the door to later investigate the neuronal precursors of the cognitive control of vocalizations in the monkey brain.

METHODS

Experimental Animals

We used two male rhesus monkeys (*Macaca mulatta*) weighing 4.2 and 4.5 kg for this study. All procedures were authorized by the Regierungspräsidium Tübingen, Germany.

Data Acquisition

Stimulus presentation and behavioral monitoring was automated on PCs running the CORTEX program (NIH) and recorded by a Plexon multiacquisition system. Vocalizations were recorded by the same system with a sampling rate of 40,000 Hz via an A/D converter. A custom-written MATLAB program running on another PC monitored the vocal behavior in real time. Vocal on- and offset times were detected off-line by a custom-written MATLAB program to assure precise timing for data analysis.

Behavioral Protocol

During the first part of the study, we trained both monkeys to perform a visual go/no-go detection task using their vocalizations as response (detection task). A trial began when the monkey initiated a "ready" response by grasping a bar (see Figure 1A). A visual cue, indicating the "no-go" signal ("precue"; white square, diameter = 0.5° of visual angle) appeared for a randomized time of 1–5 sec for monkey 1 (M1) and 0.5–5 sec for monkey 2 (M2). During this period, vocal output had to be withheld. Next, in 80% of the trials, the visual cue was changing to a colored "go" signal (red or blue square; diameter = 0.5° of visual angle) lasting for 3000 msec. During this time, the monkey had to emit a vocalization to receive a reward. To demonstrate that the motivational state, which might be associated with a call type, did not influence the cognitive control of vocalizations, each monkey was trained to utter a different call type. M1 was trained to utter "coo" vocalizations (harmonic vocalization used for intraspecific long distance communication), M2 was taught to emit "grunts" (noisy call; used for intraspecific short distance communication and as indicator of low quality food; Hauser & Marler, 1993; see Figure 2). Both colors appeared with equal probability ($p = .5$). Our results show that cue color had no significant influence on call probability (Wilcoxon sign rank test, $p > .1$ for both monkeys). In 20% of the trials, the cue remained unchanged for another 3000 msec ("catch" trial). During this period, the monkey had to withhold calls. "Catch" trials were not rewarded. "False alarms" were indicated by visual feedback (blue screen) and by trial abortion. To demonstrate its readiness to work, the monkey had to grab the bar throughout the "precue" as well as the "go" phases. Bar releases aborted the trials instantaneously, followed by visual feedback (red screen).

In the second part of the study, we trained monkey M1 to switch between two vocalizations on command. Here, the animal had to produce two different vocalizations in response to distinct visual cues (discrimination task). As previously, the monkey initiated a trial by grasping a bar and the "no-go" signal appeared for a randomized time of 1–5 sec (see Figure 1B). Next, in 80% of the trials, the visual cue was changed to either a colored (red or blue square) or a shaped "go" signal (cross or ring). All "go" signals appeared pseudorandomly with equal probability ($p = .2$). The monkey was trained to utter a "coo" vocalization in response to the red square and the cross and to emit a "grunt" vocalization in response to the blue square and the ring. Our results show that the type of cue had no significant effect on call probability [two-way ANOVA, $F(1, 39) = 0.33$ ("grunt" calls) and $F(1, 39) = 0.02$ ("coo" calls), $p > .1$ for both call types]. "Catch" trials were

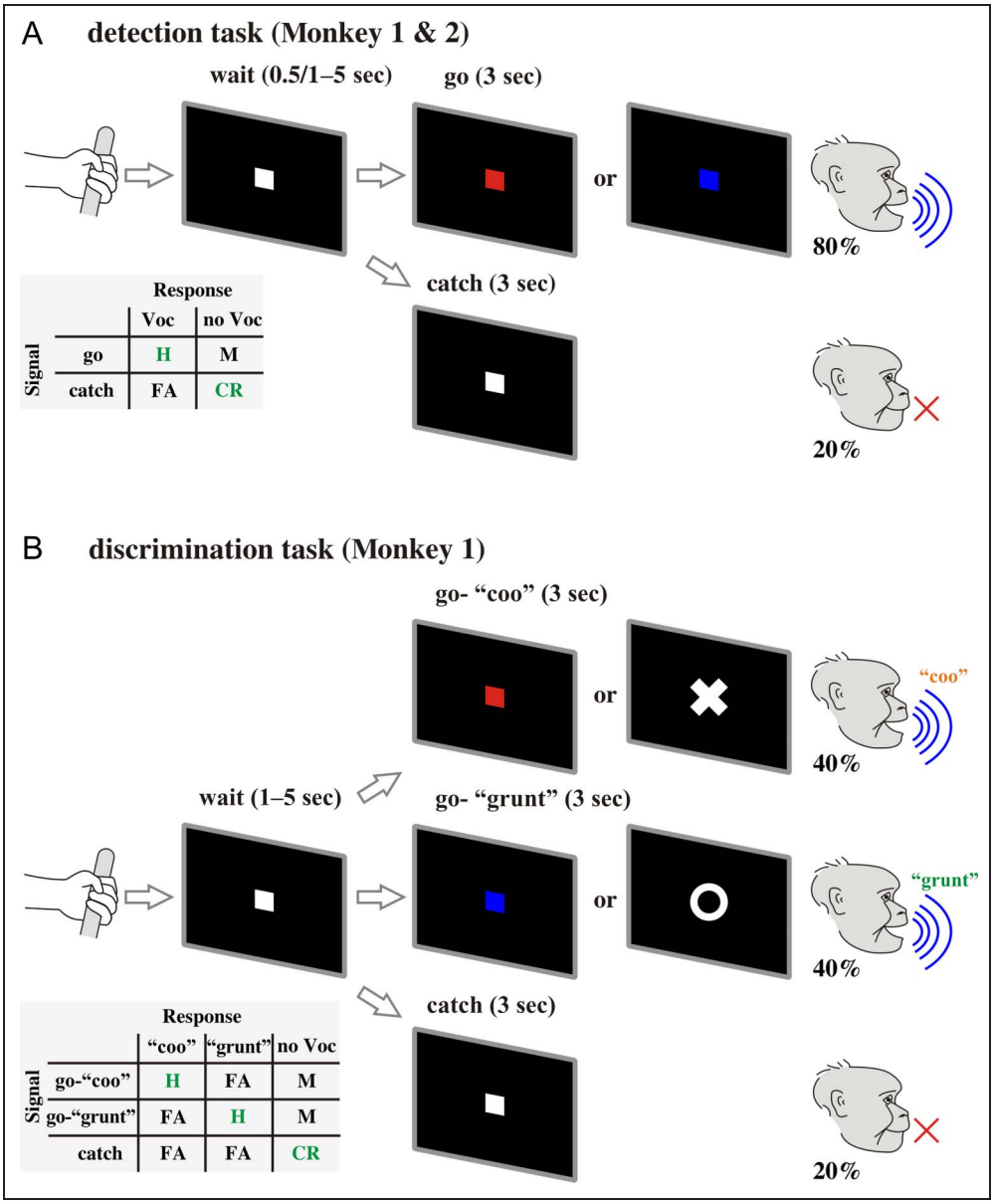

Figure 1. Experimental design. (A) Both monkeys were trained in a go/no-go protocol to vocalize whenever a visual cue appeared (detection task). (B) One monkey was trained in a successive training period to utter distinct vocalizations in response to specific visual cues. H = hit; M = miss; FA = false alarm; CR = correct rejection.

again presented in 20% of the trials. Also, as previously, the monkey had to grab the bar throughout the trials to report its readiness and bar releases aborted the trials. More than 15 bar releases yielding less than five vocal utterances within a sliding window of the last 50 trials defined the end of a daily session. One session was recorded per individual per day. Animals were head-fixated during all experiments, maintaining a constant distance of 5 cm between the animal's head and the microphone.

Data Analysis

Fifteen consecutive sessions per individual during the detection task and 10 consecutive sessions during the discrimination task were used for data analysis in which more than 50 vocalizations were uttered. Behavioral

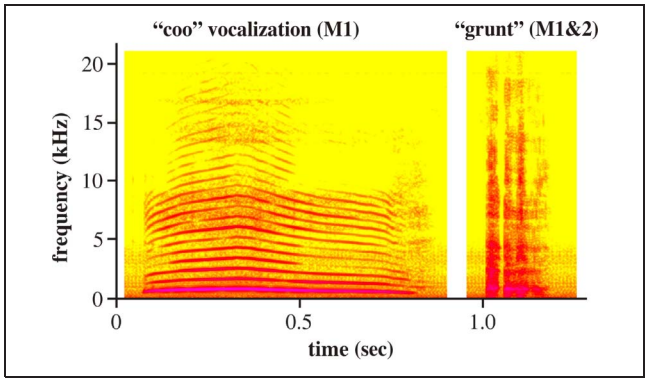

Figure 2. Spectrograms of representative "coo" and "grunt" vocalizations uttered by the experimental animals M1 and M2. Intensity is represented by different shades of color.

sessions with less than 50 vocalizations were excluded from data analysis. In accordance to the go/no-go detection protocol, successful "go" trials were defined as "hits," unsuccessful "catch" trials as "false alarms" in the detection paradigm. For the discrimination protocol, the utterance of the correct vocalization in response to a specific visual cue was defined as "hits," a vocal response with the wrong call type as "false alarms."

In M1 ("coo" vocalizations), we compared differences in call parameters, such as call duration, amplitude and frequency, of "hit" vocalizations and adjacent spontaneously uttered vocalization with differences between two adjacent vocalizations uttered directly before or after the latter mentioned call pair during the performance of the detection protocol. A fast Fourier transformation, with 2048 points resulting in a frequency resolution of 19.5 Hz, was performed to analyze peak frequencies of single calls. Call parameters were calculated automatically by a custom-written MATLAB program.

Statistical Analysis

Statistical analysis was performed with MATLAB (Math Works Statistics Toolbox). We computed d' sensitivity values derived from signal detection theory (Green & Swets, 1966) by subtracting z scores (normal deviates) of median "hit" rates from z scores of median "false alarm" rates. Detection threshold for d' values was set to 1.8. We performed a one-way MANOVA with post hoc Wilcoxon sign rank tests to test for significant differences in call parameters between volitionally and spontaneously calls and volitional call pairs that were uttered in direct succession. A one-way ANOVA (Kruskal–Wallis test) was performed to test for significant differences in call response latency according to the duration of the "no-go" signal (cue delay). A Friedman test with post hoc Wilcoxon sign rank test was calculated to test for significant differences in the hit and false alarm rates during the discrimination task.

RESULTS

We recorded 5040 vocalizations (3312 "coo" calls and 1728 "grunt" vocalizations) of two rhesus monkeys that where uttered in response to abstract, learned visual cues in 35 daily sessions while performing either a detection or discrimination task.

Vocal Performance during the Detection Task

The current data basis consists of 15 daily sessions per monkey for the detection task. Both monkeys showed consistent vocal performances with mean call rates of 174 ± 11 (*SEM*) for M1 and 96 ± 10 vocalizations for M2 per session. Figure 3A shows a representative session by M1 with 202 uttered "coo" vocalizations. Call frequency was high during go signals with 188 calls (93.7%), resulting in a "hit" rate of 76.1% (188 of 245 go trials). One single call was vocalized during catch phases, resulting in a "false alarm" rate of 1.6% (1/64 trials). The remaining 13 vocalizations were uttered during the wait periods (nine calls; 4.5%) and after cue offset of go trials (four calls; 2%). The obtained "hit" and "false alarm" values led to a mean

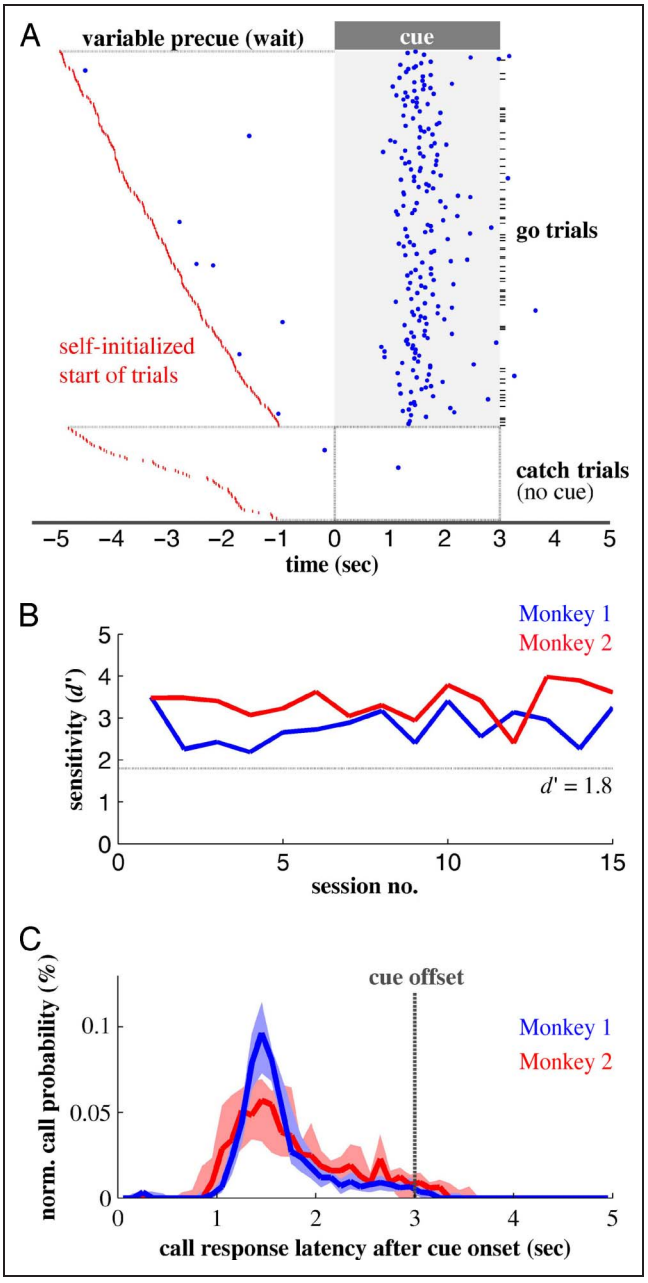

Figure 3. Vocal performance in the detection task. (A) Example of a single session of M1. Responses to "go" and "catch" trials are sorted according to the length of the "pre cue" signal. Each line represents a single trial; blue circles indicate vocal onsets. "Go" trials ignored by the monkey ("misses") are marked with a horizontal black bar at trial end. (B) Sensitivity of signal detection of 15 sessions for both monkeys indexed by the d' value. The dotted line indicates the border for successful signal detection. (C) Call probability of M1 and M2 during "go" trials (normalized for 15 sessions each; bin width = 100 msec, shaded areas indicate first and third quartiles).

d' sensitivity value of 2.9. This shows that the monkey produced calls reliably and almost exclusively in response to the visual go cues. Throughout the sessions, mean values were high in both monkeys for "hit" (M1: 57.3 ± 3.2%, M2: 63.3 ± 2.8%) and low for "false alarm" rates (M1: 0.9 ± 0.2%, M2: 0.2 ± 0.1%). In both monkeys, d' values were above detection threshold in all sessions (mean $d' = 2.8 \pm 0.1$ in M1 and 3.4 ± 0.1 in M2; see Figure 3B). Both monkeys showed similar response patterns with median latencies of 1.53 sec (M1) and 1.64 sec (M2; Figure 3C).

M1 occasionally elicited spontaneous "coo" vocalizations between trials. A comparison of acoustic call features (call duration, peak frequency and amplitude) of spontaneous (unconditioned) and volitional (conditioned) vocalizations uttered in direct succession showed interesting differences ($p < .01$, MANOVA). These differences resulted from differences in call duration ($p < .05$, post hoc Wilcoxon rank test), and peak frequency ($p < .01$, see Figure 4). Conditioned "coo" vocalizations were characterized by longer durations and lower frequencies when compared with unconditioned calls.

Vocal Performance during the Discrimination Task

We recorded 10 additional sessions of M1 while performing the discrimination task. A representative session by M1 with 128 uttered vocalizations (92 "coo" and 36 "grunt" calls) is depicted in Figure 5A. Eighty-six "coo" calls and 36 "grunt" calls were produced during the corresponding "go" signals, resulting in a "hit" rate of 54. 8% (86 of 157 "coo" trials) for "coo" utterances and a "hit" rate of 16.3% (36/160 trials) for "grunt" vocalizations. Only six "coo" calls and no "grunt" calls were uttered during the wrong "go" signals resulting in "false alarm" rates of 3.8% (6/160 trials) and 0% (0/157 trials), respectively. The obtained "hit" and

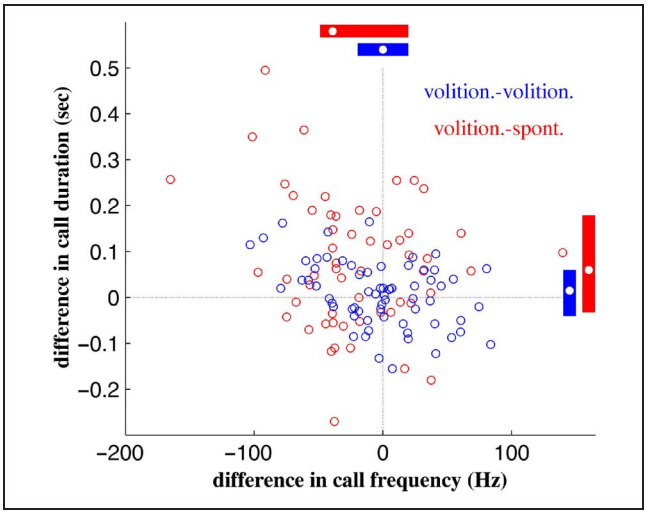

Figure 4. Distribution of differences in call duration and peak frequency of consecutive volitional–volitional ($n = 66$) and volitional–spontaneous "coo" pairs ($n = 56$) of M1 in the detection task. Bars indicate medians, first and third quartiles for each parameter in each group.

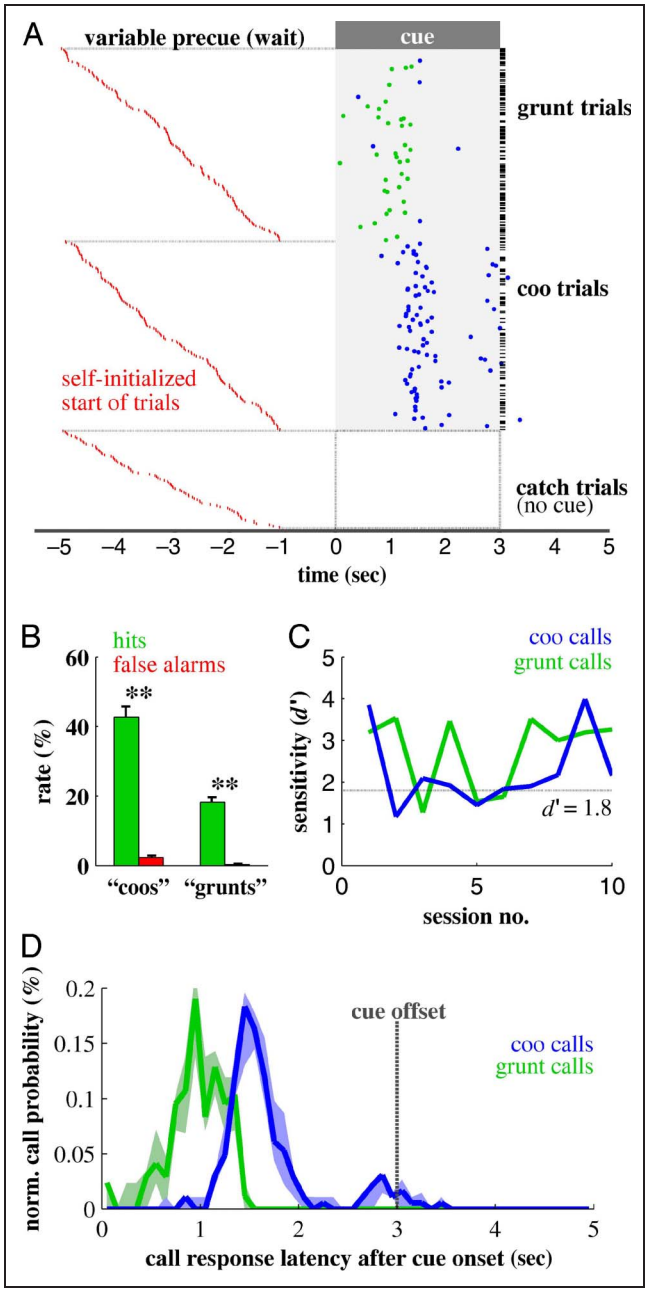

Figure 5. Vocal performance of Monkey 1 in the discrimination task. (A) Example of a single session. Responses to "grunt," "coo," and "catch" trials are sorted according to the length of the "pre cue" signal. Each line represents a single trial, blue circles indicate "coo" call onsets, and green circles indicate "grunt" call onsets. Trials ignored by the monkey ("misses") are marked with a horizontal black bar at trial end. (B) Distribution of hit and false alarm rates of 10 sessions for "coo" and "grunt" vocalizations. Both call types were uttered with significantly higher probabilities during the corresponding "go" trials (hits) than during the other "go" trials (false alarms); $^{**}p < .01$, Friedman test with post hoc sign rank test). (C) Sensitivity of signal detection of 10 sessions (same sessions as in B) for "coo" and "grunt" calls indexed by the d' value. The dotted line indicates the border for successful signal detection. (D) Call probability of "coo" and "grunt" calls during "coo" and "grunt" trials, respectively (normalized for 10 sessions each; same sessions as in B; bin width = 100 msec, shaded areas indicate first and third quartiles).

"false alarm" values led to a mean d' sensitivity value of 1.9 for "coo" performance and 3.5 for "grunt" performance. Throughout the sessions, the vocal performance was consistent although lower than during the detection task with mean call rates of 103 ± 6. Interestingly, mean call rates were higher for "coo" calls (73 ± 5) than "grunt" vocalizations (30 ± 2), resulting in a higher mean "hit" rate for "coo" calls (42.6 ± 3.2%) than "grunt" calls (18.2 ± 1.4%). Mean "false alarm" rates were low for both call types ("coo": 2.3 ± 0.6%, "grunt": 0.3 ± 0.2%). Statistical analysis revealed a significantly non-homogeneous distribution for call probabilities for both call types in response to the corresponding visual cues (hits) and the visual cues accompanied with the other call type (false alarms; $p < .001$, $df = 3$, $\chi^2 = 29.51$, Friedman test). These differences were mainly because of the considerably higher probabilities for the utterances of a vocalization in response to its accompanied visual cues than in response to the other cues for both call types ($p < .01$ post hoc Wilcoxon sign rank test). These findings are also reflected in the d' values. For both call types, d' values were above detection threshold in most sessions (8 of 10 sessions for coo calls, 7/10 for grunt calls) with mean d' values of 2.3 ± 0.3 for "coo" calls and 2.8 ± 0.3 for "grunt" calls see (Figure 5C). These data show that the monkey was able to produce the specific call type reliably in response to the corresponding visual cues.

The monkey showed similar response latencies for "coo" vocalizations than during the detection task with a median latency of 1.55 sec. Response latencies for "grunt" vocalizations, however, where significantly shorter with a median latency of 0.97 (Wilcoxon rank sum test, $p < .001$; see Figure 5D).

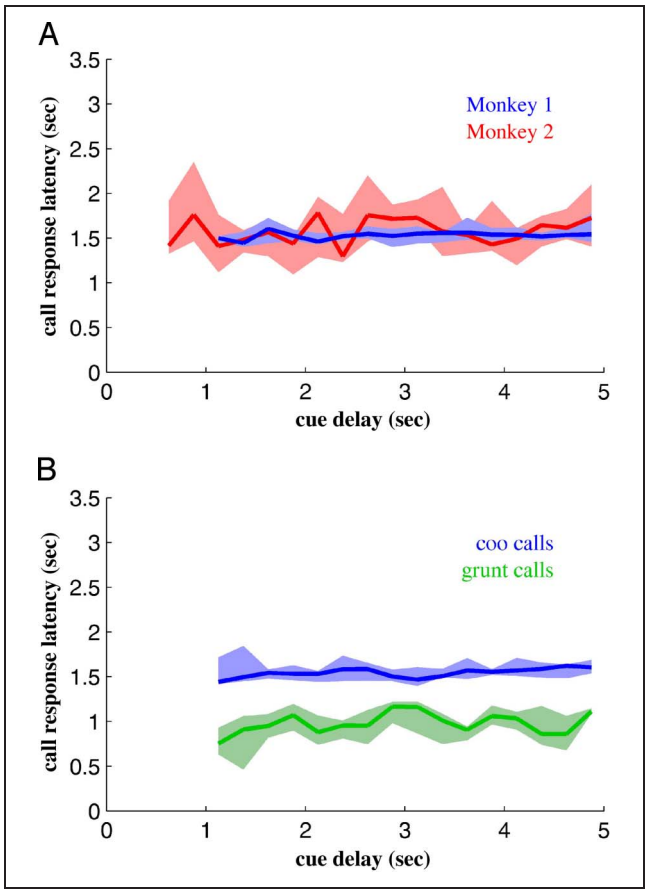

Figure 6. Relationship between the median call response latencies after cue onset and the preceding waiting period (cue delay). (A) Call response latencies show no significant relation between the call response latency and the duration of the preceding cue delay in both monkeys during the detection task (15 sessions, $p > .1$ for both monkeys, Kruskal–Wallis test). (B) No significant correlation was observed between call response latency and the duration of the preceding cue delay for both vocalizations that were uttered by Monkey 1 in the discrimination task (10 sessions, $p > .1$ for both call types, Kruskal–Wallis test). Bin size = 250 msec.

No Correlation between Call Response Latencies and the Preceding Duration of the Waiting Period

Finally, we investigated whether the vocal response latency was dependent of the cue delay, that is, the wait period between self-induced trial initiation and "go" cue onset. Therefore, we tested the relationship between the vocal response latency and the duration of the corresponding cue delay for both monkeys in the detection task and both call types of M1 in the discrimination task. We did not find any significant changes of the vocal response latency dependent of the cue delay for both monkeys in the vocal detection task (M1: $p > .1$, $df = 15$, $\chi^2 = 9.64$; M2: $p > .1$, $df = 17$, $\chi^2 = 18.22$, Kruskal–Wallis test; see Figure 6A) and both call types of M1 in the discrimination task (coo calls: $p > .1$, $df = 15$, $\chi^2 = 9.35$; grunt calls: $p > .1$, $df = 15$, $\chi^2 = 18.61$, Kruskal–Wallis test; see Figure 6B).

DISCUSSION

We demonstrate that rhesus monkeys are capable to volitionally initiate vocal output in a highly controlled experimental design. By applying abstract cue stimuli and rigorous psychophysical measurements, our study complements earlier investigations of volitional call initiations. We show, first, that monkeys can be trained to vocalize on command in response to arbitrary visual cues in a go/no-go detection task. Second, we report that a monkey learned to switch between two distinct call types from trial to trial in response to different visual cues in a discrimination task. Third, our findings also suggest that monkeys have rudimentary control over acoustic call parameters.

Evidences for Volitional Call Initiation in Both Detection and Discrimination Tasks

The results of the detection experiment show that monkeys were able to instrumentalize their vocal utterances, irrespective of call type, in response to arbitrary visual

cues to receive a reward. This indicates that monkeys were able to volitionally initiate their vocal output. In both monkeys, vocal response latencies in relation to the preceding waiting period (white cue) were comparable. These findings indicate that both monkeys initiated vocal output in response to the onset of the abstract learned cues. During the discrimination task, we show that a rhesus monkey was able to produce two distinct vocalizations in response to different visual cues. To our knowledge, this is the first evidence that a nonhuman primate was capable to switch call types from trial to trial and use them in a goal-directed way to perform a behavioral task successfully.

Differences to Earlier Studies on Volitional Call Initiation

Over the last decades, several studies provided suggestive evidence that monkeys could be conditioned to produce vocalizations in response to visual stimuli (Coude et al., 2011; Koda et al., 2007; Hihara et al., 2003; Pierce, 1985; Aitken & Wilson, 1979; Sutton et al., 1973). After careful examination of methodological approaches that were used in these studies, however, it is unclear to what extent the uttered vocalizations were the result of the nonhuman primates' ability to volitionally initiate its vocal output. Alternative explanations, like motivationally triggered calls in response to stimuli with hedonic value, cannot be excluded because nonhuman primates were trained to vocalize in response to the presentation of food items (Coude et al., 2011; Koda et al., 2007; Hihara et al., 2003). Gemba, Kyuhou, Matsuzaki, and Amino (1999) presented auditory playbacks of species-specific vocalizations to Japanese macaques and defined vocal responses of the subjects as volitional utterances; here, purely motivational responses are the most parsimonious explanation for the calls. Furthermore, a problem in interpreting the data in two of these studies (Coude et al., 2011; Hihara et al., 2003) relates to the subjects being trained to produce vocalizations that are naturally used as "food indicators" by these monkey species (Hauser & Marler, 1993). It is not surprising that monkeys utter food calls at the sight of food or contact calls in response to species-specific vocalizations.

In other conditioning studies, monkeys were trained in rigid protocols to vocalize in response to visual cues in rather long time windows of up to 5 min in which the monkeys were rewarded for every single vocalization (Aitken & Wilson, 1979; Sutton et al., 1973). Two aspects are noteworthy. First, monkeys were rewarded for every single vocalization that they produced in these rather long time windows. Therefore, it is not clear whether the monkeys were vocalizing as a response to the visual cue or rather produced vocalizations in response to the preceding food reward or self-produced vocalization ("motivational self-enhancement"). Second, using rigid temporal protocols might allow monkeys to develop other strategies to solve the tasks, such as vocalizing after a specific period rather than paying attention to a visual cue. In our study, we eliminated this possibility by introducing pseudorandomized "pre cue" phases in which the monkeys have to omit vocal output or controls such as "catch" trials in which the monkeys are not rewarded.

In yet another study, Japanese monkeys were trained to produce "coo" vocalizations in response to a tool. After successful vocalization, the tool could be used to reach a food reward (Hihara et al., 2003). However, because no time limit was set for the monkey to vocalize, every single vocalization, whenever produced, resulted in the presentation of a food reward, which then could be reached with the tool. It is therefore difficult to exclude that the monkeys indeed learned to vocalize in response to the arbitrary tool to reach the reward, rather than randomly producing vocalizations without any learning effect. The observed increase in call performance from day to day might also be explained by motivational changes in the monkey subjects, which learned that they will be rewarded during the daily sessions.

To summarize, these studies confirmed that specific motivational states of monkeys are accompanied by specific vocal utterances (Jürgens, 1979) rather than showing that monkeys are capable to call on command in response to an arbitrary visual stimulus. In contrast, our results indicate that monkeys are able to volitionally initiate their vocal output in response to an arbitrary visual cue. Especially our finding that a rhesus monkey was able to switch between two distinct call types from trial to trial in response to different visual cues shows that nonhuman primates are capable to volitionally control which vocalization to utter. Our findings are based on data that were collected with a highly controlled experimental approach. Therefore, we are able to exclude possibilities other than cognitive control that might cause the initiation of vocal behavior.

Differences in Call Parameters in Conditioned and Unconditioned Vocalizations

Our data show that conditioned vocalizations were significantly longer in duration and lower in frequency than unconditioned vocalizations in the detection task. A recent study revealed similar results by reporting spontaneous differentiation of "coo" vocalizations with respect to changes of the fundamental frequency in Japanese macaques during a tool use training (Hihara et al., 2003). Furthermore, a few other studies showed that monkeys are capable of acquiring volitional control on call amplitude and duration (Trachy, Sutton, & Lindeman, 1981; Larson, Sutton, & Lindeman, 1978; Sutton et al., 1973). The experimental findings of these studies, including the present one, suggest that monkeys might have at least some control over modulating their call parameters within natural constraints, thus indicating a rather complex

mechanism underlying volitional control of vocal production. However, the interpretation of these findings is quite more complex. Call parameters such as call frequency and amplitude are directly correlated with the level of hedonic or aversive state in monkeys (Fichtel & Hammerschmidt, 2003). Therefore, the observed changes in call parameters might be simply because of a change in the motivational state of the monkey subject. Additionally, minor changes in call parameters such as call amplitude, fundamental frequency, or duration can be explained by the adjustments of respiratory functions rather than articulatory functions and do not conclusively imply operant control over spectro-temporal call features in monkeys (Janik & Slater, 2000). Other involuntary mechanisms can also cause changes in call features without cognitive control. The most prominent example for such involuntary changes in call parameters is an involuntary rise in call amplitude and frequency in response to masking ambient noise. This so-called Lombard effect is present in several mammals, including monkeys and man (e.g., Hage, Jiang, Berquist, Feng, & Metzner, 2013; Brumm, Voss, Kollmer, & Todt, 2004; Sinnott, Stebbins, & Moody, 1975; Lombard, 1911) and has been shown to be most likely be controlled by the pontine brainstem (Hage et al., 2013; Hage, Jürgens, & Ehret, 2006; Nonaka, Takahashi, Enomoto, Katada, & Unno, 1997).

On the basis of the presented data, it would thus be premature to conclude that monkeys are capable to volitionally modulate their call parameters. Further studies are needed to find out if and to what extent monkeys are able to volitionally control spectro-temporal compositions of their calls, for example, by training them to modulate their call parameters in a controlled discrimination protocol.

Neurobiological Implications for Cognitive Control on Vocal Output

Primate vocalization is a complex behavioral pattern that is generated by a complex neuronal network in the brainstem (Hage, 2009; Jürgens, 2002). The periaqueductal gray in the midbrain and the vocal pattern generator in the ventrolateral pontine brainstem have been shown to be crucial components within this network (Hage & Jürgens, 2006). Both structures are directly involved in triggering call onset. This brainstem network receives facilitating input from several sensory and limbic structures such as the ACC, amygdala, hypothalamus, and the septum (Hage, 2009) underpinning the strong motivational character of primate vocalization.

This study provides evidences for the ability of rhesus monkeys to decouple their vocal production from the accompanied motivational state and instrumentalize distinct call types to perform a specific task successfully. This indicates that monkeys are able to volitionally initiate their vocal output. Of course, the relationship between vocal control in monkeys and a speech-and-language system in humans are still very superficial. First, a voluntary call does not yet constitute some kind of "conversation" in which a listener receives information and responds appropriately, thus establishing a communicative feedback loop. Second, even if animals are trained to transmit information to another, unlike humans they do so exclusively to receive some sort of reward (Epstein, Lanza, & Skinner, 1980; Savage-Rumbaugh, Rumbaugh, & Boysen, 1978).

The underlying neuronal networks that are responsible for cognitive control of vocal production by modulating the vocal motor system are not well understood .A recent study demonstrates vocalization-correlated activity in the premotor cortex of rhesus monkeys (Coude et al., 2011). On the basis of shared anatomical and physiological features and comparable cytoarchitectonics, recent studies suggested monkey homologues of human brain structures essential for human speech, such as Broca's and Wernicke's area (Gil-da-Costa et al., 2006; Petrides, Cadoret, & Mackey, 2005) as well as specific voice areas in the auditory cortex (Petkov et al., 2008). In particular, neurons in Brodmann's areas 44 and 45 of the ventral pFC in monkeys are known to represent sign–object associations (Diester & Nieder, 2007) and rules guiding the structuring of conceptual information (Bongard & Nieder, 2010), putative precursors for semantic and syntactical processing in sign systems (Nieder, 2009). The present conditioning approach paves the way for further electrophysiological investigations of how forebrain homologues of these brain structures that are essential for human speech control are involved in the control of monkey vocalization.

Reprint requests should be sent to Steffen R. Hage, Animal Physiology, University of Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany, or via e-mail: steffen.hage@uni-tuebingen.de.

REFERENCES

Aitken, P. G., & Wilson, D. A. (1979). Discriminative vocal conditioning in rhesus monkeys: Evidence for volitional control? *Brain and Language, 8,* 227–240.

Balter, M. (2010). Animal communication helps reveal roots of language. *Science, 328,* 969–971.

Bongard, S., & Nieder, A. (2010). Basic mathematical rules are encoded by primate prefrontal cortex neurons. *Proceedings of the National Academy of Sciences, U.S.A., 107,* 2277–2282.

Brumm, H., & Slabbekoorn, H. (2005). Acoustic communication in noise. *Advances in the Study of Behaviour, 35,* 151–209.

Brumm, H., Voss, K., Kollmer, I., & Todt, D. (2004). Acoustic communication in noise: Regulation of call characteristics in a New World monkey. *Journal of Experimental Biology, 207,* 443–448.

Brumm, H., & Zollinger, S. A. (2011). The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour, 148,* 1173–1198.

Cheney, D. L., & Seyfarth, R. M. (2007). *Baboon metaphysics: The evolution of a social mind*. Chicago, IL: University Chicago Press.
Coude, G., Ferrari, P. F., Rodà, F., Maranesi, M., Borelli, E., Veroni, V., et al. (2011). Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PloS One, 6,* e26822.
Diester, I., & Nieder, A. (2007). Semantic associations between signs and numerical categories in the prefrontal cortex. *PLoS Biology, 5,* 2684–2695.
Egnor, R. S. E., Wickelgren, J., & Hauser, M. D. (2007). Tracking silence: Adjusting vocal production to avoid acoustic interference. *Journal of Comparative Physiology A, 193,* 477–483.
Epstein, R., Lanza, R. P., & Skinner, B. F. (1980). Symbolic communication between two pigeons (*Columba livia domestica*). *Science, 207,* 543–545.
Fichtel, C., & Hammerschmidt, K. (2003). Responses of squirrel monkeys to their experimentally modified mobbing calls. *Journal of the Acoustical Society of America, 113,* 2927–2932.
Gemba, H., Kyuhou, S., Matsuzaki, R., & Amino, Y. (1999). Cortical field potentials associated with audio-initiated vocalization in monkeys. *Neuroscience Letters, 272,* 49–52.
Ghazanfar, A. A. (2008). Language evolution: Neural differences that make the difference. *Nature Neuroscience, 11,* 382–384.
Gil-da-Costa, R., Martin, A., Lopes, M. A., Munoz, M., Fritz, J. B., & Braun, A. R. (2006). Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque. *Nature Neuroscience, 9,* 1064–1070.
Green, D. M., & Swets, J. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
Hage, S. R. (2009). Neuronal networks involved in the generation of vocalization. In S. M. Brudzynski (Ed.), *Handbook of mammalian vocalization* (pp. 329–338). Oxford: Academic Press.
Hage, S. R. (2013). Audio-vocal interactions in vocal communication of squirrel monkeys and their neurobiological implications. *Journal of Comparative Physiology A, 7,* 663–668.
Hage, S. R., Jiang, T., Berquist, S., Feng, J., & Metzner, W. (2013). Ambient noise induces independent shifts in call frequency and amplitude within the Lombard effect in echolocating bats. *Proceedings of the National Academy of Sciences, U.S.A., 110,* 4063–4068.
Hage, S. R., & Jürgens, U. (2006). On the role of the pontine brainstem in vocal pattern generation. A telemetric single-unit recording study in the squirrel monkey. *Journal of Neuroscience, 26,* 7105–7115.
Hage, S. R., Jürgens, U., & Ehret, G. (2006). Audio-vocal interaction in the pontine brainstem during self-initiated vocalization in the squirrel monkey. *European Journal of Neuroscience, 23,* 3297–3307.
Hammerschmidt, K., & Fischer, J. (2008). Constraints in primate vocal production. In U. Griebel & K. Oller (Eds.), *The evolution of communicative creativity: From fixed signals to contextual flexibility* (pp. 93–119). Cambridge, MA: MIT Press.
Hauser, M. D., & Marler, P. (1993). Food-associated calls in rhesus macaques (Macaca mulatta): I. Socioecological factors. *Behavioral Ecology, 4,* 194–205.
Hihara, S., Yamada, H., Iriki, A., & Okanoya, K. (2003). Spontaneous vocal differentiation of coo-calls for tools and food in Japanese monkeys. *Neuroscience Research, 45,* 383–389.
Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour, 60,* 1–11.
Jürgens, U. (1979). Vocalization as an emotional indicator a neuroethological study in the squirrel monkey. *Behaviour, 69,* 88–117.
Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews, 26,* 235–258.
Jürgens, U. (2009). The neural control of vocalization in mammals: A review. *Journal of Voice, 23,* 1–10.
Koda, H., Oyakawa, C., Kato, A., & Masataka, N. (2007). Experimental evidence for the volitional control of vocal production in an immature gibbon. *Behaviour, 144,* 681–692.
Larson, C. R., Sutton, D., & Lindeman, R. C. (1978). Cerebellar regulation of phonation in rhesus monkey (Macaca mulatta). *Experimental Brain Research, 33,* 1–18.
Lombard, E. (1911). Le Signe de l'Elévation de la Voix. *Annales des Maladies de l'Oreille Larynx, 37,* 101–119.
Manser, M. B., Seyfarth, R. M., & Cheney, D. L. (2002). Suricate alarm calls signal predator class and urgency. *Trends in Cognitive Sciences, 6,* 55–57.
Miller, C. T., Beck, K., Meade, B., & Wang, X. (2009). Antiphonal call timing in marmosets is behaviorally significant: Interactive playback experiments. *Journal of Comparative Physiology A, 195,* 783–789.
Nieder, A. (2009). Prefrontal cortex and the evolution of symbolic reference. *Current Opinion in Neurobiology, 19,* 99–108.
Nonaka, S., Takahashi, R., Enomoto, K., Katada, A., & Unno, T. (1997). Lombard reflex during PAG-induced vocalization in decerebrate cats. *Neuroscience Research, 29,* 283–289.
Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences, U.S.A., 106,* 22026–22031.
Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature Neuroscience, 11,* 367–374.
Petrides, M., Cadoret, G., & Mackey, S. (2005). Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature, 435,* 1235–1238.
Pierce, J. D., Jr. (1985). A review of attempts to condition operantly alloprimate vocalizations. *Primates, 26,* 202–213.
Rendall, D., Seyfarth, R. M., Cheney, D. L., & Owren, M. J. (1999). The meaning and function of grunt variants in baboons. *Animal Behaviour, 57,* 583–592.
Roy, S., Miller, C. T., Gottsch, D., & Wang, X. (2011). Vocal control by the common marmoset in the presence of interfering noise. *Journal of Experimental Biology, 214,* 3619–3629.
Savage-Rumbaugh, E. S., Rumbaugh, D. M., & Boysen, S. (1978). Symbolic communication between two chimpanzees (*Pan troglodytes*). *Science, 201,* 641–644.
Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual Review of Psychology, 54,* 145–173.
Seyfarth, R. M., & Cheney, D. L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain and Language, 115,* 92–100.
Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence for

predator classification and semantic communication. *Science, 210,* 801–803.
Sinnott, J. M., Stebbins, W. C., & Moody, D. B. (1975). Regulation of voice amplitude by the monkey. *Journal of the Acoustical Society of America, 58,* 412–414.
Sutton, D., Larson, C., Taylor, E. M., & Lindeman, R. C. (1973). Vocalization in rhesus monkeys: Conditionability. *Brain Research, 52,* 225–231.
Trachy, R. E., Sutton, D., & Lindeman, R. C. (1981). Primate phonation: Anterior cingulate lesion effects on response rate and acoustical structure. *American Journal of Primatology, 1,* 43–55.
Wich, S. A., & de Vries, H. (2006). Male monkeys remember which group members have given alarm calls. *Proceedings of the Royal Society, Series B, 273,* 735–740.
Zollinger, S. A., & Brumm, H. (2011). The Lombard effect. *Current Biology, 21,* R614–R615.
Zuberbühler, K., Cheney, D. L., & Seyfarth, R. M. (1999). Conceptual semantics in a nonhuman primate. *Journal of Comparative Psychology, 113,* 33–42.