



---

# Failure of Operant Control of Vocal Learning in Budgerigars

Yoshimasa Seki<sup>1,2\*</sup>, Michael S. Osanski<sup>1,3</sup>, and Robert J. Dooling<sup>1</sup>

<sup>1</sup>Department of Psychology, University of Maryland, College Park, MD

<sup>2</sup>Faculty of Letters, Aichi University, Toyohashi, Japan

<sup>3</sup>Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD

\*Corresponding author (Email: [yoshimasa.seki@gmail.com](mailto:yoshimasa.seki@gmail.com))

**Citation** – Seki, Y., Osanski, M. S., & Dooling, R. J. (2018). Failure of operant control of vocal learning in budgerigars. *Animal Behavior and Cognition*, 5(1), 154-168. <https://doi.org/10.26451/abc.05.01.11.2018>

**Abstract** – Budgerigars were trained by operant conditioning to produce contact calls immediately after hearing a stimulus contact call. In Experiments 1 and 2, playback stimuli were chosen from two different contact call classes from the bird's repertoire. Once this task was learned, the birds were then tested with other probe stimulus calls from its repertoire, which differed from the original calls drawn from the two classes. Birds failed to mimic the probe stimuli but instead produced one of the two call classes as in the training sessions, showing that birds learned that each stimulus call served as a discriminative stimulus but not as a vocal template for imitation. In Experiment 3, birds were then trained with stimulus calls falling along a 24-step acoustic gradient which varied between the two sounds representing the two contact call categories. As before, birds obtained a reward when the bird's vocalization matched that of the stimulus above a criterion level. Since the first step and the last step in the gradient were the birds' original contact calls, these two patterns were easily matched. Intermediate contact calls in the gradient were much harder for the birds to match. After extensive training, one bird learned to produce contact calls that had only a modest similarity to the intermediate contact calls along the gradient. In spite of remarkable vocal plasticity under natural conditions, operant conditioning methods with budgerigars, even after extensive training and rigorous control of vocal discriminative stimuli, failed to show vocal learning.

**Keywords** – Vocal conditioning, Parrot, Auditory-vocal association, Synthetic sound

---

Budgerigars (*Melopsittacus undulatus*) have become a highly successful model of animal vocal production learning (for vocal production learning, see Tyack (2008)). Budgerigars are capable of acquiring new conspecific vocalizations both as juveniles (Brittan-Powell, Dooling, & Farabaugh, 1997) and adults (e.g., Farabaugh, Linzenbold, & Dooling, 1994). They readily imitate other bird vocalizations as well as human speech (Banta Lavenex, 1999, 2000; Eda-Fujiwara & Okumura, 1992; Gramza, 1970). The contact call of budgerigars is about 100-200ms in duration with complex frequency/amplitude modulations and has been particularly well studied (e.g., Dooling, 1986). Adult males and females engage in call convergence readily when housed together (Farabaugh et al., 1994; Hile, Plummer, & Striedter, 2000; Hile & Striedter, 2000; Moravec, Striedter, & Burley, 2006). A similar phenomenon has been observed in at least two parrot species (orange-fronted conures, *Aratinga canicularis*, and the galah, *Eolophus roseicapillus*; Balsby & Scarl, 2008; Scarl & Bradbury, 2009; Vehrencamp, Ritter, Keever, & Bradbury, 2003). Taken together, these studies suggest a high degree of vocal plasticity in parrots and a suspicion that frequent exposure to acoustic models might facilitate vocal imitation. Some studies even

report spontaneous imitation of body movements by budgerigars in operant tasks, which could be related to the capacity for vocal learning (e.g., Ikkatai, Okanoya, & Seki, 2016; Miller, Gallup, Vogel, Vicario, & Clark, 2012; Mui, Hazelgrove, Pearce, & Heyes, 2008).

Previous work on operant control of vocal behavior in budgerigars developed a method for vocal operant training (Manabe & Dooling, 1997; Manabe, Dooling, & Brittan-Powell, 2008; Manabe, Kawashima, & Staddon, 1995; Manabe, Staddon, & Cleaveland, 1997). Specifically, these studies showed that budgerigars could be trained to produce two different contact call types in response to two different visual stimuli (Manabe & Dooling, 1997; Manabe et al., 1995) but failed to produce two different call types to two different contact calls produced from the same location (Osmanski, Seki, & Dooling, submitted).

Bringing vocal learning more fully under operant control could be extremely valuable because budgerigars have become a popular model for investigating the neuroscience of vocal motor control (e.g., Banta Lavenex, 2000; Heaton & Brauth, 2000; Paton, Manogue, & Nottebohm, 1981). The present experiments were undertaken to investigate the potential and constraints in operant vocal production learning in budgerigars.

The question addressed here is whether birds could learn a strategy that involved “vocal template matching” wherein any sound heard in an operant context would be imitated, or whether birds instead would adopt a discrimination strategy, in which the sounds played to the birds were used merely as discriminative stimuli to indicate to the bird which vocalization was to be produced. In other words, audible templates simply served as discriminative stimuli and vocal responses were essentially the equivalent of key pecking response.

Actually, in this respect, the training of the vocal template matching under the operant context in budgerigars in the present study ended in failure. However, the results do provide strong evidence that communicative social situations are critical factors in the vocal learning in this species. A less likely possibility is that food reward was not a salient enough reward for budgerigars to learn to imitate vocal sounds, and/or process of vocal imitation is sufficiently different from the process of association learning normally observed operant paradigms. This failure, though, has important implications for future studies in animal vocal learning especially for the systematic vocal training procedures attempted with this species.

### **General Methods**

Each bird’s calls were recorded in the same experimental system as that used for conduct of the experiment since earlier work suggests that budgerigars sometimes produce particular vocalizations in particular contexts (Brockway, 1969). This helped to ensure that the specific vocalizations we recorded were not affected by changing conditions or environments and that the bird could produce only the required vocalizations during testing.

#### **Subjects**

Four male and three female budgerigars were used. Three birds were trained and tested in a previous experiment on operant control of vocal production and four were naïve birds. Birds were kept in each separate cage in an aviary in the Biology/Psychology building of the University of Maryland. They could access water ad libitum but their food was restricted outside of the test apparatus so as to maintain an 80-85% free-feeding body weight.

#### **Apparatus**

A metal wire cage (19 cm X 17 cm X 22 cm) was mounted in a sound attenuation chamber (AC-1, Industrial Acoustic Company, NY). The cage was equipped with a perch, and a food hopper that contained millet. Subjects could access food from a hole within the floor of the cage when the food

hopper was in the up position. A response panel with 3 red LEDs (left, center and right) and a microphone (ECM-77B, Sony, Japan) was mounted on the side of the cage about 2 cm above from the food hopper opening. Vocalizations picked up by the microphone were sent to a signal processor (RP2.1, Tucker-Davis Technology (TDT), FL) after band pass filtering (450-10kHz) by a frequency filter (Model 3550, Krohn-Hite, MA) and an amplifier (MA-3, TDT, FL). A small loud speaker (40-245, Archer, Korea) mounted on the cage wall was directed at the bird when it faced the response panel. Another loudspeaker (40-1289, Realistic, Japan), placed in the corner of the IAC booth, was used for delivering flock sounds on the background to induce the bird to vocalize during the initial stages of training. The TDT RP2.1 was also used to control the sound stimuli, LED illumination and the food hopper. The overall background noise level was 33dB(A) SPL in the test booth.

### Sound Analysis and Configuration of Stimulus Sounds

Contact calls produced by the birds were analyzed with the same method as Osmanski and Dooling (2009). Briefly, similarities among vocalizations were determined by the MATLAB functions “spectrogram” and “xcorr2” (2-D cross-correlation). The first function provided spectrograms of the vocalizations and made matrices of these spectrograms. The second function found the best shift-point between two spectrogram matrices to calculate a similarity index using correlation values of the matrices. A function “cmdscale” (Classical Multidimensional Scaling) was used to categorize the contact calls and to view their location in a similarity space. Vocalizations with similar patterns formed a cluster in the similarity space. Vocalizations located near the center of these clusters were selected to serve as contact call templates for each vocal production pattern and were also used to create synthetic vocalizations along the acoustic dimension spanning these two vocalization categories. Extraneous noise was filtered from these selected calls and the duration of the calls were adjusted to 250 ms by Audition 2.0 (Adobe Systems Inc., CA). We used also Praat acoustic software (Boersma, 2001) to analyze the recorded sounds. The sound pressure level of each stimulus was adjusted to 69 dB SPL at the position of the birds’ head.

### Design and Procedure

After a period of adaptation to the apparatus, the 4 naïve birds were trained to produce contact calls in the apparatus. In the first phase of training, flock sounds from their aviary were played in the test booth at the level of about 40-45 dB. This facilitated calling by the bird and every vocalization the bird produced was reinforced by food access. Once the bird reliably produced contact calls under these conditions, the flock sound was muted and the birds were trained to vocalize only when the center LED was lit. The inter trial interval was randomized between 3-10 s and the LED was lit and stayed on until the bird vocalized. Vocalizations produced when the LED was not lit, were not reinforced and the random interval before the next trial occurred was extended.

The birds were trained to produce at least three different vocal patterns in this paradigm, using a previously described 1- or 2-back procedure (see Manabe et al., 1997). Briefly, birds were reinforced only for producing a contact call that was significantly different than the previous contact call it produced. In the 1-back procedure, the bird was only reinforced with food when it produced a call different from the last call. In the 2-back procedure the bird was reinforced only when the bird produced a call different from the last two calls it had produced. Whether a call was similar or different from a previously produced call was determined by calculating a similarity index in real time between the vocalizations as described above. The criterion for similarity between any two calls was initially set low and then gradually increased over sessions until the birds were reliably producing distinctly different calls. This phase of training ended when a large number of contact calls produced by the bird fell into two distinct acoustic categories. All vocalizations were recorded at 24 kHz sampling rate.

## Data Analysis

Data were analyzed with a Mann-Whitney U-test for comparison between two groups. For more than two groups, a Kruskal-wallis method and a Steel-Dwass multiple comparison test were used.

### Experiment 1: Acoustic Control of Two Vocalizations

#### Stimulus

For each bird, a single representative call was selected from each of the two categories of vocalizations and designated as the template call representing that category to be used in subsequent training phases. These two calls, one from each category, were labeled as template-A and template-B for convenience. The birds' other vocal patterns produced during the 1- or 2- back procedure or training trials were used as probe stimuli in Phase 2 of Experiment 1.

#### Design and Procedure

**Phase 1: Training sessions.** A trial began with a presentation of a stimulus call, which was either Template-A or Template-B followed by lighting of the center LED. If no response occurred within 20 s the call was played again and again until the bird vocalized. Once a sound was detected, the system calculated a similarity index between the sound and template-A and template-B in real time (this value was referred to as *rvalue-A* or *rvalue-B*). In template-A trials, if the *rvalue-A* was greater than both the *rvalue-B* and a pre-set criterion, the vocal response was reinforced. A separate criterion was chosen for each template because of the variability of the birds' vocal productions for the two categories. This criterion procedure resulted in calls that were highly consistent with those made by visual inspection of spectrograms by the experimenter (> 99%). If a vocalization was not reinforced, the same template was played back repeatedly until criterion was met and the vocalization was reinforced.

In initial training sessions, we also added an additional visual cue by using left and right LEDs as additional discriminative stimuli because past work has shown that it is relatively easy to train for budgerigars to make a vocal response to two different visual stimuli (Osmanski et al., submitted). So, following playback of template-A, both left and center LEDs were turned on as response cues, while both the right and center LEDs were turned on after a template-B playback. As training progressed, left and right LEDs were gradually dimmed as the number of correct responses increased. Eventually a high level of responding was reached with only the center LED used as a response cue.

The correct percent was calculated by # of reinforced call / # of total call X 100. One session consisted of 100 reinforced vocalizations. The session was divided into two parts for morning and afternoon sessions (50 correct responses each). When the number of correct response was shorter than 50 and birds did not respond within 10 min, the session was terminated.

When a bird showed over 80% correct response in a session, the reinforcement probability was gradually reduced to 86%. Then, the discrimination criterion was set to 80% correct or above for two consecutive sessions.

**Phase 2: Probe test.** When a bird met the criterion, it was tested in a probe session on the next day. In the probe session, ten probe trials were interspersed randomly among the daily training trials. Three of the bird's own vocalizations, which differed from the training sounds by visual inspection of the spectrogram, and two vocalizations from other budgerigars, were used as the probe stimuli. Thus, the number of probe trials was 50, or 5 probe stimuli (3 own and 2 other birds' vocal patterns) X 10 trials each, for each bird. Vocal responses to the probe stimuli resulted in neither food reinforcement nor punishment.

One probe stimulus set (5 probe stimuli) was used in each probe session, thus each bird received 10 probe test sessions in total. Two normal training sessions (i.e., without probes) with at least an 80% correct response were required before running a probe test session.

## Results and Discussion

All birds eventually acquired the conditioned vocal behavior just in response to the illumination of the center LED. However, two of the birds failed to produce two or more vocal patterns that were sufficiently and consistently distinguishable from one another. Therefore, these two birds were excluded from further training.

**Sampling template calls.** For the remaining birds, two training template calls were chosen from the bird's own vocal repertoire as described above for Phase 2. The mean durations of these stimulus sounds were 143.7 ms (maximum 203 ms, minimum 108 ms, SD 23.6 ms).

**Call dissociation and template matching training.** During Phase 2, no birds repeated the playback calls spontaneously. It was very clear that it was difficult for the bird to produce its own vocalization played to it during test sessions as a discriminative stimulus. This resulted in extremely long training sessions. Moreover, as the number of training sessions increased, the birds' vocal patterns began to drift, which is a common occurrence even under non-operant conditions (see Hile et al., 2000). This reduced the similarity index between the vocalization and the template sound. As a consequence, some of the bird's initial vocalizations that were "correct" and therefore reinforced, were no longer reinforced using the original criterion. The consequence of this was that each template had to be occasionally updated and the birds had to be trained with these new templates.

In the end, five birds met the discriminative criterion successfully but only after exceedingly long training sessions. The total number of training trials (vocalizations) in this phase for each bird was: Eros: 5,857, Inko: 11,325, Tori: 19,286, Artemis: 23,869, Sora: 42,778. These birds needed far more trials to meet the discriminative criterion using the acoustic template references in these experiments than was the case for budgerigars used in other vocal operant studies which used visual discriminative stimuli (e.g., Manabe et al., 1995). This difficulty in training birds using two different acoustic references emanating from a single sound source is closely in line with that found in earlier studies (Osmanski et al., submitted).

**Analysis of response latencies.** Each bird's vocal response provided two measures. One was the degree of acoustic matching and the other was the response latency. Comparing the response latencies of reinforced responses to the two vocal patterns for each subject for the last two sessions just before the probe tests showed that the latencies to one vocal pattern were typically much shorter than the response latencies to the other. This difference within subjects was statistically significant (Table 1). Thus, here, SLStim refers to the template stimulus pattern which resulted in the bird typically producing Short-Latency calls and LLStim refers to the template stimulus pattern which resulted in Long-Latency calls. A call pattern typically produced in response to the SLStim is called SLProd and a call pattern produced in response to the LLstim is referred to as LLProd.

By a computed similarity index, as well as a visual inspection of the spectrograms, the birds did not produce any vocalizations that matched the probe stimuli. Instead, four birds had a bias for producing the LLProd responses to the probe stimuli (Eros 48/50; Inko 44/50; Artemis 43/50; Sora 44/50), while one bird (Tori) produced equal numbers of the SLProd and the LLProd responses. Overall, the probability of the LLProd response was higher than that of the SLProd response ( $\chi^2 = 104.98$ ,  $p < .001$  in pooled data). These results also show that the birds took more time to vocalize in response to the probe stimuli than to the training stimuli (i.e., template A/B). On average, the response latency for probe stimuli was significantly longer than that for the SLStim in all birds and longer also for the LLStim in 4 of 5 birds (Table 2).

Because 4 birds tended to use the LLProd vocalization in response to the probe stimuli, two natural variations of the SLProd, previously produced by the subjects themselves, were selected as stimuli for an additional generalization test. In this test, the four birds continued to produce the LLProd vocalization to the natural variations of the new SLStim (Eros: 20/20; Inko: 19/20; Artemis: 14/20; Sora: 14/20). In other words, birds failed to generalize their vocal response even to the natural variations of their own template vocalization ( $\chi^2 = 0.04$ ,  $p = .841$ ).

Table 1

*Difference of Response Latency Between Two Playback Sounds*

Subjects	mean latency (ms)		rank sum		Mann-Whitney U	Significance
	SL	LL	SL	LL		
Eros	488	2,536	4,998	14,308	147	***
Inko	409	1,308	6,533	13,170	1,583	***
Tori	700	801	8,272	10,256	3,616	*
Artemis	1,063	1,976	5,700	13,606	849	***
Sora	658	670	8,301	11,200	3,450	***

Note: \*\*\* $p < .001$ , \* $p < .05$ . SL; Short Latency, LL; Long Latency

Table 2

*Difference of Response Latency Between Training and Probe Stimuli in Two Template Task*

Subjects	mean latency (ms)			KW $\chi^2$	df	$p$	$t$		$p$	
	SL	LL	Probe				SL vs. Probe	LL vs. Probe	SL	LL
Eros	747	2,854	5,585	326.75	2	***	10.69	4.34	***	***
Inko	238	998	6,468	177.57	2	***	10.32	7.09	***	***
Tori	845	940	2,748	64.4	2	***	7.3	6.24	***	***
Artemis	911	1,651	2,594	272.36	2	***	9.6	4.79	***	***
Sora	676	749	757	23.54	2	***	2.05	3.88	n.s.	***

Note: \*\*\* $p < .001$

Taken together, the results of Experiment 1 showed that the response latency was significantly different between the training stimuli and the probe stimuli and further suggested that birds failed to perceive the template stimuli as their own call. Moreover, the birds appeared to attend, not to the global properties of the probe stimuli, but to a particular acoustic feature in only one of the probe stimuli when discriminating between the stimulus sounds.

## Experiment 2: Training with Multiple Natural Call Variations

In this experiment, the 5 birds were retrained using multiple natural variations of the calls used in Experiment 1.

### Design and Procedure

The overall method was the same as Experiment 1 except that, instead of one exemplar of the template-A and B being used as a stimulus on each trial, five natural variations of the template-A (A1-A5) or the template-B (B1-B5) were used as training stimuli. During a test session, a single template, chosen at random from each of the two sets, was played back on each trial. After a subject learned the task (over 80% reinforced vocalization twice under reinforce probability of 0.86), the birds were trained on another stimulus set (A6-A10, B6-B10) by the same method. When the bird met the same criterion for these new stimulus sets, the bird was tested in a probe session.

In these probe tests, the probe stimuli were chosen from the birds' own vocalizations as in Experiment 1. In addition, two natural variations of SL-type template were used as stimuli for a

“generalization test” in 4 of the birds. Because one bird (Tori) predominately used the SL-type vocalizations, natural variations of the LL-type vocalization were used for the generalization test for this bird instead of the SL-type vocalization.

## Results and Discussion

All birds learned the task in significantly fewer sessions (Eros: 16; Inko: 11; Tori: 6; Artemis: 24; Sora: 4) than they did in Experiment 1.

In the probe test, birds produced the same vocal patterns as in the training sessions in response to the probe stimuli in this experiment just as they did in Experiment 1. The birds also showed a bias in producing one particular dominant call for the probe test stimuli ( $\chi^2 = 51.8, p < .001$ ). Four birds preferred to use the LL-type vocalization for those probe stimuli (Eros: 20/20; Inko: 19/20; Artemis: 14/20; Sora: 14/20), while Tori used the SL-type vocalization (19/20).

In the generalization test, however, all birds showed a bias for using the non-dominant call in response to the playback of the natural variations of the corresponding (i.e., non-dominant vocal type) pattern (Eros: 17/20; Inko: 15/20; Artemis: 15/20; Sora: 12/20; Tori: 16/20;  $\chi^2 = 25.0, p < .001$ ). In contrast to Experiment 1, this pattern of results does show a generalization of the vocal response to natural variations of the stimulus patterns. Nevertheless, there is no direct evidence that the birds could repeat the playback sounds they heard, or that they used these template sounds as a “vocal reference.” The birds still persisted in producing a dominant pattern to the probe stimuli.

### Experiment 3: Step-by-Step Transition Between Two Templates

Because the acoustic properties of the probe stimuli were quite different from the training stimuli and the probe stimuli were selected at random in the daily test sessions, the possibility exists that birds were uncertain as to which kind of call to produce in response to the probe calls. An obvious next question is whether the bird could track small variations in the probe stimuli. In other words, if the playback sound was slightly and gradually changed so as to guide the subjects’ response, would the birds then begin to use the stimulus sounds as the vocal reference?

In this experiment, the acoustic characteristics of template-A was slightly shifted toward that of template-B in a step by step fashion. To ensure the intermediate sounds were like the template, we used a spectral “contour” of the template sounds (see below). We know from earlier work that budgerigars perceive the similarity among sounds with a similar spectral contour (Dooling & Okanoya, 1995).

The original sounds were analyzed by a 256 point FFT. A value of frequency which had the strongest intensity and a value of sound intensity was extracted at each time point. Those frequency and intensity data were concatenated to create a “frequency contour” which kept the original sound intensity at each point. By this method, we had “contour-A” and “contour-B”. Then, 24 intermediate sounds between those two sounds were synthesized by shifting 4% at each step. For example, at the 1<sup>st</sup> step, the frequency value was calculated by (frequency of contour-A $\times$ 0.96 + frequency of contour-B $\times$ 0.04) at each time point. This procedure was likewise followed with the amplitude values. These values were then concatenated and a synthetic sound for each of the 24 steps was synthesized from those frequency and amplitude contours. All calculations were executed by custom-made MATLAB programs

### Design and Procedure

In advance of this experiment, the birds were tested in a probe session using the synthetic contour sound of its “non-dominant” call using the same test method of Experiment 2. Because there was no evidence that the birds generalize their vocal response to these contour sounds (see Result section) despite the perceptual similarity between original sound and contour sound, it was necessary to train the birds using variations of “contour sounds” by the same method used in Experiment 2. In this training, two of natural variations (A1 and A2, B1 and B2) in a stimulus set were converted to the contour sounds (as

contour-A1 and -A2, contour -B1 and -B2). When a bird met the criterion (over 80% reinforced vocalizations for those contour sounds excluding correction trials in a session), the bird went to a “step-by-step experiment” on the next day.

The step-by-step experiment began from a playback of a contour-A. The same stimulus was used until the bird was reinforced (i.e., responded correctly) in 3 successive trials at each step. Once this criterion was met, the playback sound on the next trial shifted to the next step in the synthetic series. Finally, for the last 3 trials, a contour-B stimulus was presented. In all, one session consisted then of 78 trials (26 steps X 3). The criterion for reinforcement was decided prior to each session. As the sessions progressed, the criterion was gradually raised.

## Results and Discussion

In the initial stage of training in this experiment, three of the five birds showed large fluctuations in their original vocal patterns and drifted further away during the course of the experiment. As a consequence, vocalizations produced by the birds failed to match the pattern of the synthetic sound contour. Therefore, for this step-wise experiment, only two of five birds were used. At the beginning, these two birds produced their preferred or dominant calls for both types of the contour sounds as in Experiment 1 and 2 which showed that these birds did not readily generalize their call responses matched to the original template sounds to the new synthetic contour sounds. With additional training, these two birds quickly reached the discriminative criterion on this task in relatively few sessions (Eros 14; Artemis 7).

Initially, birds responded with one of the two vocal patterns (i.e., call-A or call-B) in response to every synthetic sound and then the birds switched the vocal pattern to the other one in the middle of the synthetic series through trial and error. This means that when examined over the course of a test session, the similarity indices between the birds' vocalizations and playback sounds within a session produced a U-curve function (Figure 1a) because the playback sound for the first three trials was contour-A which was similar to call-A, while the playback sound for the last three trials was contour-B which was quite similar to call-B.

As the matching criterion was raised in successive sessions, the task became increasingly difficult for the birds. As shown in Figure 3, for the first several trials, calls produced in response to Step 2 and Step 3 stimuli typically led to reinforcement. However, as the step stimuli became more dissimilar from Call-A and moved into the middle of the continuum, the birds' calls failed to reach the similarity criterion required for reinforcements (i.e., vocal productions for the stimuli such as Step 8, 9). As the result, Eros struggled (Figure 1b) and occasionally produced different call types while moving around in the cage. Those calls made two novel clusters in the MDS space (Figure 2). Our visual inspection of the spectrograms was completely consistent with the MDS of call differences which were statistically significant (Wilks' Lambda < .0001,  $df = 12$ ,  $\chi^2 = 1.340$ ,  $p < .0001$ ). Figure 2 showed that the vocalizations Call A, B and other novel calls produced by the bird, did not form a continuous sound gradation but rather four discrete clusters suggesting a physical limitation in the vocal production apparatus of the bird.

Spectrograms of two of three novel call patterns produced by the bird were similar to the intermediate synthetic sounds (Figure 3). These had not been seen before or during the training sessions. This observation was supported by not only visual inspection of the spectrograms but also by higher similarity indices between one of novel vocal sounds and an intermediate synthetic sound (118<sup>th</sup> trial, Figure 3; Figure 4). The novel pattern 2 could be created as a result of combination of some parts from contour-A and contour-B (145<sup>th</sup> trial, Figure 3). This recombination of call components has been reported previously in vocal learning in freely behaving budgerigars (Farabaugh et al., 1994). Interestingly, the novel calls also had less power at the harmonic frequencies which was entirely consistent with the lack of harmonics in the synthetic stimuli (Figure 3). The peak in the power spectrum of these novel sounds occurred roughly intermediate to the frequency content between contour-A and -B (Figure 5).



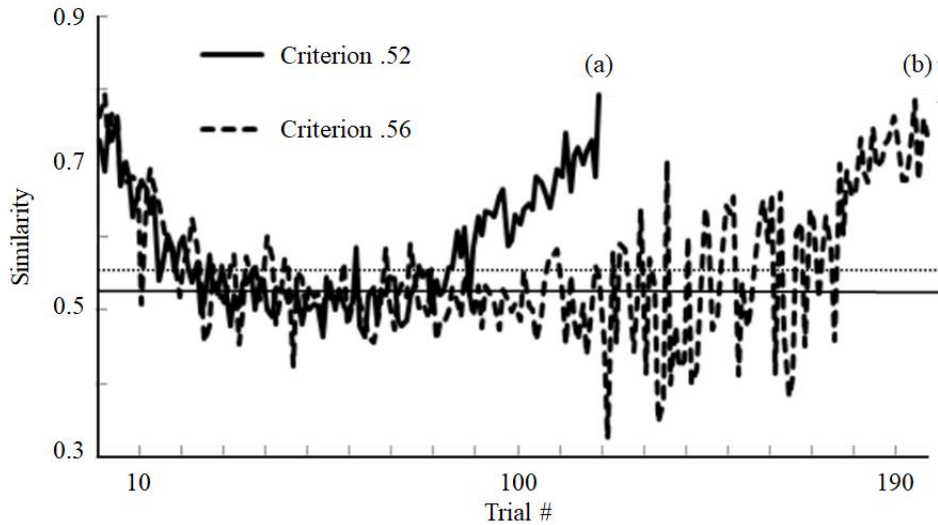


Figure 1. Example U-curves of similarity index between vocalization and playback sound in two sessions of Experiment 3 (Subject; Eros). (a) With lower criterion, the bird could pass the bottom of the curve at easy with production of either vocal-A or B. (b) With higher criterion, the bird began to struggle around the bottom of the curve. In such condition, the bird vocalized novel patterns with moving around in the cage.

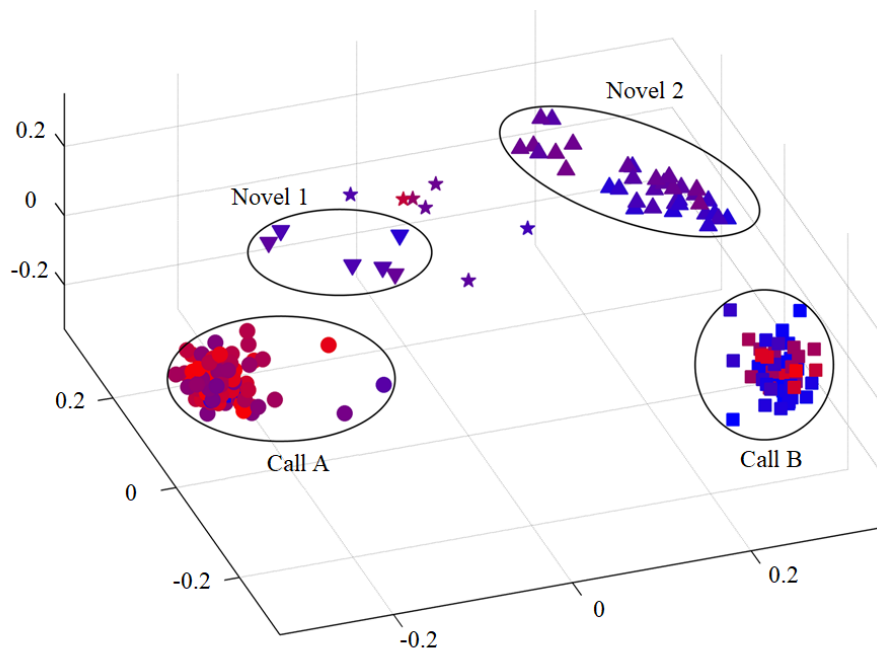


Figure 2. A multidimensional scaling (MDS) space built from similarity of sound spectrograms acquired in a session of Experiment 3 (corresponding to the dotted line of Figure 1). The bird created several vocal patterns in addition to call A and B, when the bird was struggling. The marker represents different pattern of the spectrogram categorized by visual inspection (circle; call-A, square; call-B, downward-triangle; novel call-1, upward-triangle; novel call-2, star; others). The plot color indicates progress of the trials; purely red means the first trial and purely blue means the last trial.

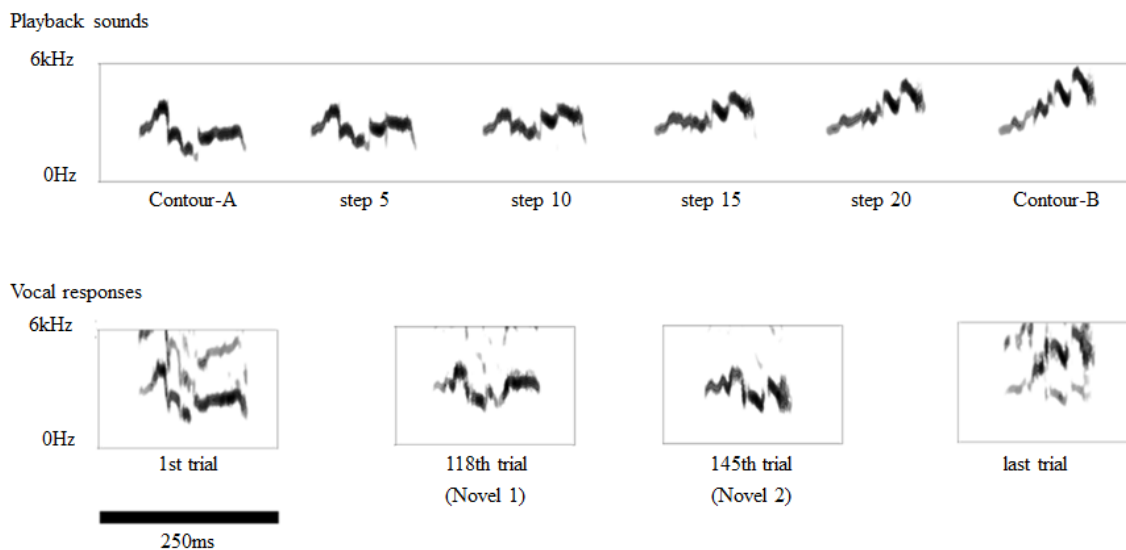


Figure 3. Example of the sound spectrograms of playback sounds and vocal responses in a session of Experiment 3 (corresponding to Figures 2). Novel pattern 1 is quite similar to an intermediate sound (the 5<sup>th</sup> step) on the spectrogram. Novel pattern 2 looks like a combination of several parts of the contour-A and B.

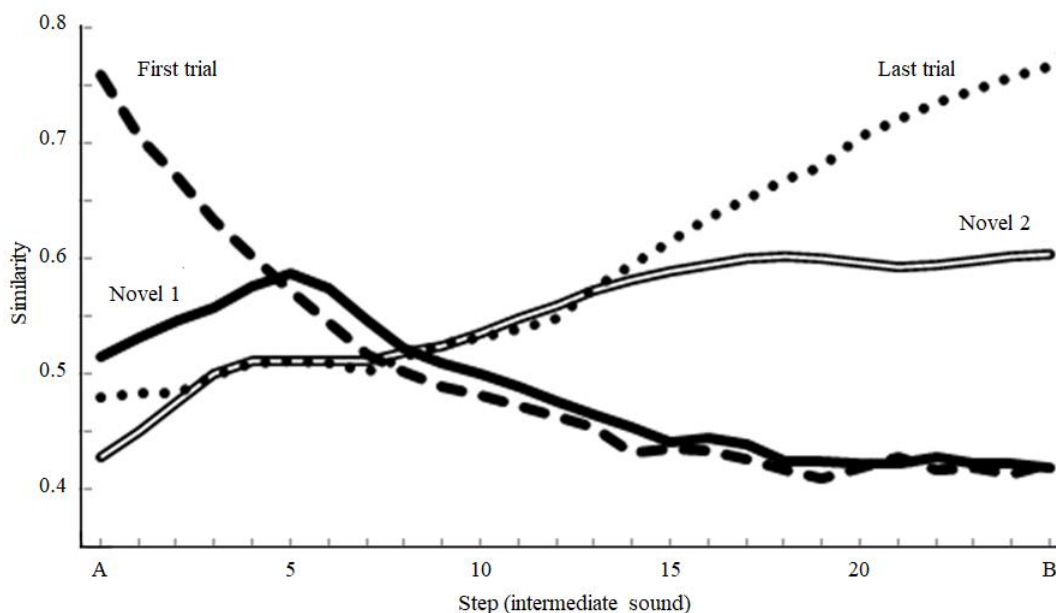


Figure 4. Transition of the similarity between the playback sounds and each vocal pattern through a session in Experiment 3 (corresponding to Figures 2 and 3). Because the playback sounds were shifting from the contour-A to contour-B, similarity between vocal pattern A (produced at the 1<sup>st</sup> trial) and playback sound was going down as the trial was going. Likewise, similarity between vocal pattern B (produced at the last trial) and playback sound was going up as trial was going. However, novel vocal pattern 1 was more similar to intermediate sounds than A and B. Values of novel vocal repertoire 1 and 2 were average of each type.

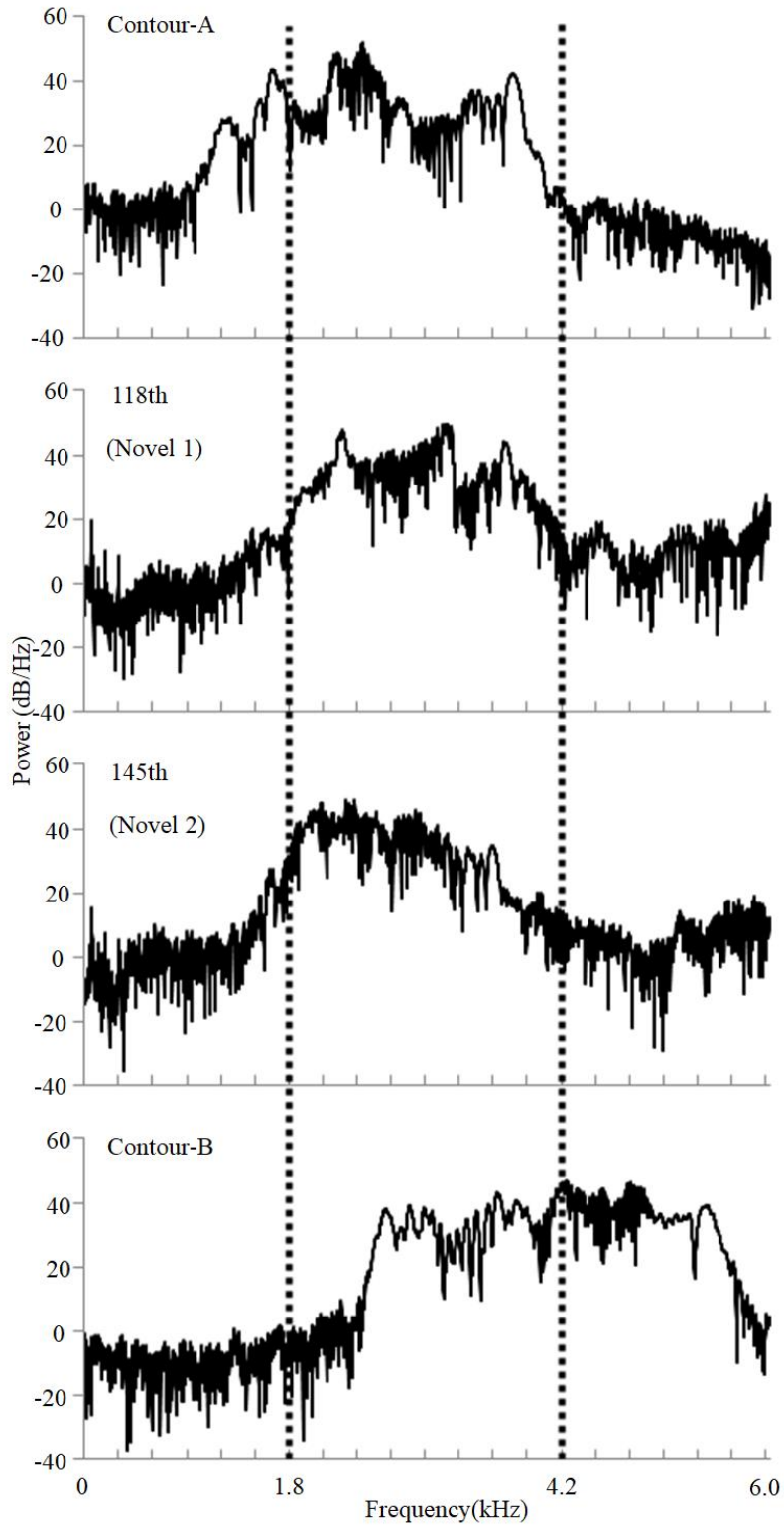


Figure 5. Power spectrums of the playback sounds (contour-A and B) and vocalizations (novel patterns 1 and 2) the vocal patterns shown in Figure 3 (after normalizing in the total energy). Peak power of the novel patterns appeared in the middle frequency range between contour-A and B.

For Artemis, as the similarity criterion was raised, the number of calls produced in a single test session reached as high as over 650. This bird did not produce variable calls like Eros and also did not give up on the task. Although the bird was sometimes moving around, the bird returned to the original position to vocalize. Eventually, the bird stopped responding.

### General Discussion

There is a plethora of evidence that budgerigars learn contact calls from one another, that budgerigars in small groups share similar contact call types, and that they can learn to mimic other environmental sounds (Farabaugh et al., 1994; Hile et al., 2000; Hile & Striedter, 2000). Here birds failed to produce calls that were similar to the probe stimuli in an operant task. Birds failed to use the playback sounds as a model of what was to be learned. Instead, birds were learning to choose between one of two previously learned vocal patterns to use as a response to sound stimuli. The acoustic pattern of the birds' calls during a session drifted despite hearing the template sounds on every trial. Compared to experiments with visual discriminative stimuli, this task also required a long time for the birds to learn. Birds also failed to use other call patterns in their own vocal repertoire that may have been more similar to the stimulus pattern. Taken together, these results show that vocal imitation is not a simple, automatic process moving from sound perception to vocal production. Rather, the kind of call convergence (i.e., learning) that has been described in more natural contexts (e.g., Farabaugh et al., 1994) gives some insight that interaction with other conspecifics is a mandatory requirement. Present results show these elements may be critical.

Another possibility is the operant food-reward contingencies may strongly and tightly constrain the birds' behavior so that some essential components of vocal learning cannot occur. An indication that something like this might be operating is seen in Experiment 3. This bird did produce novel calls which were more similar to the intermediate calls than original sounds but it did so only when it was moving around in the test cage away from the food hopper. One possibility is that the bird listened to the synthetic intermediates, stored them as acoustic references, but then could only produce these calls under more unconstrained or 'natural' conditions with freedom of movement than under operant conditions.

### Task Difficulty

Budgerigars can be easily trained in conspecific-call discrimination tasks with key-pecking operant-conditioning within small number of trials (e.g., Dooling, 1986; Park & Dooling, 1985). In a vocal production task, using visual discriminative stimuli, they can learn to produce one of two calls selectively depending on the stimulus color or spatial location of either visual or acoustic stimuli within relatively small number of trials (Manabe et al., 1997; Osmanski et al., submitted). Clearly it is not a difficult task for them to learn simple visual or spatial -vocal contingencies. Here, however, the birds either failed to learn the task or required many more trials to partially learn such a response. This is consistent with Osmanski et al. (submitted). This argues that the association between auditory information processing and vocal motor control are tightly constrained probably by social factors and innate predispositions. Farabaugh et al. (1994) showed that budgerigars housed together shared multiple calls and the convergence generally started with modifications of one particular call rather than parallel, multiple call learning (also see Hile et al., 2000; Hile & Streidter, 2000). Perhaps budgerigars cannot learn multiple sound-vocal associations in parallel but rather require serial adjustments in an existing call type through repeated occurrences and social contexts. This interpretation is consistent with the finding that juvenile songbirds re-learned syllable transitions (i.e., from an original A-B-C-A-B-C-... pattern to a novel A-C-B-A-C-B-... pattern) in their singing behavior by a stepwise manner during song learning (Lipkind et al., 2013). This means that the birds learned novel syllable transition patterns serially rather than in parallel or simultaneously. It is worth keeping in mind that while tens of thousands of trials and weeks of daily test sessions to learn a task in an operant paradigm may be considered a lot, the contact

call of budgerigars is the most frequent call in their repertoire and is easily produced thousands of times a day by a bird in social situations.

### **Importance of Social Interaction and Contexts**

Clearly, songbirds copy the songs by only hearing playback sounds. Several operant experiments showed that social relationship did not affect to song learning in juvenile songbirds (Bolhuis, Van Mil, & Houx, 1999; Houx & tenCate, 1999) and male songbirds (vocal learners) did not use visual information more than females (non-vocal learners) in discrimination of singing conspecifics (Seki & Okanoya, 2008). While results from the present experiments are not directly comparable to these earlier findings, they do suggest that social interactions, including visual and physical contact, might be particularly important in vocal behavior of budgerigars (also see Farabough et al., 1994). These birds may simply need greater motivation in order to be able to vocalize what they heard in spite of instances that they can readily mimic even instrumental sounds or artificial noises occasionally (Gramza, 1970). Several playback experiments in other two parrot species (Balsby & Scarl, 2008; Scarl & Bradbury, 2009; Vehrencamp et al., 2003) suggests that not only direct social interaction but other contexts such as like a territorial protection, may also play an important role for vocal imitation. Perhaps constructing some of these more “natural contexts” in the experimental domain may be able to identify the key components that facilitate vocal imitation in budgerigars.

### **Possible Neural Mechanisms**

Since the stimulus templates were the bird’s own vocalizations, these should be encoded as a special auditory signals at a part of the vocal control nervous systems (i.e., NLC: Plummer & Striedter, 2000). Work on songbirds revealed some neurons are activated by both singing and hearing the same part of song (Prather, Peters, Nowicki, & Mooney, 2008) suggesting the existence of mirror (or mirror-like) neurons in a part of the song nervous system (i.e., HVC parallel of NLC: Jarvis, 2006). Perhaps both auditory processing and vocal motor control mechanisms involving calls are shared also in budgerigars. When they are hearing their own call, it would drive the vocal motor nervous system with a special type of input. The result is that the same neurons are immediately recruited to produce the same sound. In the present experiment, birds were repeatedly exposed to the sound stimuli during the task and the birds were required to focus on the stimuli and then to vocalize to obtain the food rewards. Perhaps this resulted in an overload of the auditory-vocal nervous system. Consistent with this, it was occasionally observed that once the subjects made an error, the same errors were repeated in consecutive trials, although the birds had already learned the task.

We again review the possible explanations of the cause of the failure and implications for the cognitive control of vocal learning. First, as shown in the initial training, the birds did not positively utilize the sound stimuli, but a visual stimulus, in the auditory-vocal association task using the food-rewarded operant procedure. Although dominance of visual stimuli might be universally observed in the stimulus control under food-reward operant tasks in birds (e.g., Foree & Lolordo, 1973), a part of the reason why they did not prefer to associate the sounds with the task might be involved in the fact that their own calls are neutrally and behaviorally special sounds for each individual in nature.

Second, at the beginning of the study, we had assumed that it should be obvious that they would easily recognize the exemplars as their own calls, because it is well known that neural responses of the vocal control system are sensitive to their own vocalization. So that we thought just two exemplars were enough for the training, but finally we found the idea was wrong. As shown in Experiment 2, before giving up to let the birds to learn the task, we had to use, not single exemplar of A and B, but a wider set of calls as playback stimuli. This might suggest that the birds rejected to hear the sounds as their own calls in Experiment 1 even though they should be capable to do it, instead, they focused on the acoustic differences between the two exemplars as mere discriminative stimuli.

Third, by way of summary, we could not exclude a possibility that we saw something different in the results if we use vocal sounds of other conspecifics as the training stimuli, as we did not examine it (also see Osmanski et al., submitted). However, even if they learn to produce a similar vocalization as such sound stimuli, it means the sound stimuli have already become one of their own vocal repertoire at that time point. So, we might also obtain the similar results also in this alternative method using the food-reward operant task.

And finally, we can imagine an operant conditioning experiment using social reinforcement such as a contact with cage mate (via wire mesh) or a visual presentation of a cage mate (via video), not food rewards, in socially isolated birds (e.g., a production of similar sound to the call of Bird A results in brief contact with Bird A). This situation might make sense for budgerigars rather than association only of sound stimuli with vocal responses via food rewards.

In summary, in spite of concerted efforts, budgerigars failed to learn to mimic a sound that they just heard when trained under operant conditions with food reward. These results highlight the complexity of vocal learning and mimicry in budgerigars and argue, as with humans, these phenomena cannot be fully understood without gaining control over other social, environmental and motivational variables that typically accompany vocal learning in natural situations.

### Acknowledgments

We thank Drs. Peter Marvit, Ed Smith and Elizabeth Brittan-Powell for technical support. All experimental procedures were in agreement with the Animal Care and Use Committee at the University of Maryland. This work was supported by NIH/NIDCD R01-DC 000198 to RJD.

### References

- Balsby, T. J. S., & Scarl, J. C. (2008). Sex-specific responses to vocal convergence and divergence of contact calls in orange-fronted conures (*Aratinga canicularis*). *Proceedings of the Royal Society of London, Series B*, 275, 2147–2154.
- Banta Lavenex, P. (1999). Vocal production mechanisms in the budgerigar (*Melopsittacus undulatus*): The presence and implications of amplitude modulation. *The Journal of the Acoustical Society of America*, 106, 491–505.
- Banta Lavenex, P. (2000). Lesions in the budgerigar vocal control nucleus NLc affect production, but not memory, of English words and natural vocalizations. *Journal of Comparative Neurology*, 421, 437–460.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5, 341–345.
- Bolhuis, J. J., Van Mil, D. P., & Houx, B. B. (1999). Song learning with audiovisual compound stimuli in zebra finches. *Animal Behaviour*, 58, 1285–1292.
- Brittan-Powell, E. F., Dooling, R. J., & Farabaugh, S. M. (1997). Vocal development in the budgerigar (*Melopsittacus undulatus*): Contact call learning. *Journal of Comparative Psychology*, 111, 226–241.
- Brockway, B. F. (1969). *Roles of budgerigar vocalization in the integration of breeding behavior*. In R. A. Hind (Ed.), *Bird vocalizations* (pp. 131–158). London: Cambridge University Press.
- Dooling, R. J. (1986). Perception of vocal signals by budgerigars (*Melopsittacus undulatus*). *Experimental Biology*, 45, 195–218.
- Dooling, R. J., & Okanoya, K. (1995). *Psychophysical methods for assessing perceptual categories*. In G. M. Klump, R. J. Dooling, R. R. Fay, & W. C. Stebbins (Eds.), *Methods in comparative psychoacoustics* (pp. 307–318). Basel, Switzerland: Birkhauser.
- Eda-Fujiwara, H., & Okumura, H. (1992). The temporal pattern of vocalizations in the budgerigar. *Journal of the Yamashina Institute for Ornithology*, 24, 19–31.
- Farabaugh, S. M., Linzenbold, A., & Dooling, R. J. (1994). Vocal plasticity in budgerigars (*Melopsittacus undulatus*): Evidence for social factors in the learning of contact calls. *Journal of Comparative Psychology*, 108, 81–92.
- Foree, D. D., & LoLordo, V. M. (1973). Attention in the pigeon: The differential effects of food-getting vs. shock avoidance procedures. *Journal of Comparative and Physiological Psychology*, 85, 551–558.
- Gramza, A. F. (1970). Vocal mimicry in captive budgerigars (*Melopsittacus undulatus*). *Ethology*, 27, 971–983.

- Heaton, J. T., & Brauth, S. E. (2000). Effects of lesions of the central nucleus of the anterior archistriatum on contact call and warble song production in the budgerigar (*Melopsittacus undulatus*). *Neurobiology of Learning and Memory*, *73*, 207–242.
- Hile, A. G., Plummer, T. K., & Striedter, G. F. (2000). Male vocal imitation produces call convergence during pairbonding in budgerigars. *Animal Behaviour*, *59*, 1209–1218.
- Hile, A. G., & Striedter, G. F. (2000). Call convergence within groups of female budgerigars (*Melopsittacus undulatus*). *Ethology*, *106*, 1105–1114.
- Houx, B. B., & tenCate, C. (1999). Do stimulus–stimulus contingencies affect song learning in zebra finches (*Taeniopygia guttata*)? *Journal of Comparative Psychology*, *113*, 235–242.
- Ikkatai, Y., Okanoya, K., & Seki, Y. (2016). Observing real-time social interaction via telecommunication methods in budgerigars (*Melopsittacus undulatus*). *Behavioural processes*, *128*, 29–36.
- Jarvis, E. D. (2006). Selection for and against vocal learning in birds and mammals. *Ornithological Science*, *5*, 5–14.
- Lavenex, P. B. (1999). Vocal production mechanisms in the budgerigar (*Melopsittacus undulatus*): The presence and implications of amplitude modulation. *Journal of the Acoustical Society of America*, *106*, 491–505.
- Lavenex, P. B. (2000). Lesions in the budgerigar vocal control nucleus NLC affect production, but not memory, of English words and natural vocalizations. *Journal of Comparative Neurology*, *421*, 437–460.
- Lipkind, D., Marcus, G. F., Bemis, D., Sasahara, K., Jacoby, N., Takahashi, M., ...Tchernichovski, O. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature*, *498*, 104–108.
- Manabe, K., & Dooling, R. J. (1997). Control of vocal production in budgerigars (*Melopsittacus undulatus*): Selective reinforcement, call differentiation, and stimulus control. *Behavioural Processes*, *41*, 117–132.
- Manabe, K., Dooling, R. J., & Brittan-Powell, E. F. (2008). Vocal learning in budgerigars (*Melopsittacus undulatus*): Effects of an acoustic reference on vocal matching. *Journal of the Acoustical Society of America*, *123*, 1729–1736.
- Manabe, K., Kawashima, T., & Staddon, J. E. R. (1995). Differential vocalizations in budgerigars: Toward an experimental analysis of naming. *Journal of the Experimental Analysis of Behavior*, *63*, 111–126.
- Manabe, K., Staddon, J. E. R. & Cleaveland, J. M. (1997). Control of vocal repertoire by reward in budgerigars. *Journal of Comparative Psychology*, *111*, 50–62.
- Miller, M. L., Gallup, A. C., Vogel, A. R., Vicario, S. M., & Clark, A. B. (2012). Evidence for contagious behaviors in budgerigars (*Melopsittacus undulatus*): An observational study of yawning and stretching. *Behavioural Processes*, *89*, 264–270.
- Moravec, M. L., Striedter, G. F., & Burley, N. T. (2006). Assortative pairing based on contact call similarity in budgerigars, *Melopsittacus undulatus*. *Ethology*, *121*, 1108–1116.
- Mui, R., Hazelgrove, M., Pearce, J. M. & Heyes, C. M. (2008). Automatic imitation in budgerigars. *Proceedings of the Royal Society of London: B*, *275*, 2547–2553.
- Osmanski, M. S., & Dooling, R. J. (2009). The effect of altered auditory feedback on control of vocal production in budgerigars (*Melopsittacus undulatus*). *Journal of the Acoustical Society of America*, *126*, 911–919.
- Osmanski, M. S., Seki, Y., & Dooling, R. J. (submitted). Control of vocal production in budgerigars (*Melopsittacus undulatus*): Effectiveness of acoustic vs. visual stimuli in cueing vocal responses.
- Paton, J. A., Manogue, K. R., & Nottebohm, F. (1981). Bilateral organization of the vocal control pathway in the budgerigar, *Melopsittacus undulatus*. *Journal of Neuroscience*, *1*, 1279–1288.
- Park, T. J., & Dooling, R. J. (1985). Perception of species-specific contact calls by the budgerigar (*Melopsittacus undulatus*). *Journal of Comparative Psychology*, *99*, 391–402.
- Plummer, T. K., & Striedter, G. F. (2000). Auditory responses in the vocal motor system of budgerigars. *Developmental Neurobiology*, *42*, 79–94.
- Prather, J. F., Peters, S., Nowicki, S., & Mooney, R. (2008). Precise auditory–vocal mirroring in neurons for learned vocal communication. *Nature*, *451*, 305–310.
- Scarl, J. C., & Bradbury, J. W. (2009). Rapid vocal convergence in an Australian cockatoo, the galah *Eolophus roseicapillus*. *Animal Behaviour*, *77*, 1019–1026.
- Seki, Y., & Okanoya, K. (2008). Sex difference of audio-visual discrimination task of Bengalese finches (*Lonchura var domestica*). *Journal of Comparative Psychology*, *122*, 26–34.
- Tyack, P. L. (2008). Convergence of calls as animals form social bonds, active compensation for noisy communication channels, and the evolution of vocal learning in mammals. *Journal of Comparative Psychology*, *122*, 319–331.
- Vehrencamp, S. L., Ritter, A. F., Keever, M., & Bradbury, J. W. (2003). Responses to playback of local vs. distant contact calls in the orange-fronted conure, *Aratinga canicularis*. *Ethology*, *109*, 37–54.