

10-2010

# Vocal Learning in Grey Parrots: A Brief Review of Perception, Production, and Cross-Species Comparisons

Irene M. Pepperberg  
*Harvard University*, [impepper@media.mit.edu](mailto:impepper@media.mit.edu)

Follow this and additional works at: [http://animalstudiesrepository.org/acwp\\_asie](http://animalstudiesrepository.org/acwp_asie)

 Part of the [Animal Studies Commons](#), [Comparative Psychology Commons](#), and the [Other Animal Sciences Commons](#)

---

## Recommended Citation

Pepperberg, I. M. (2010). Vocal learning in Grey parrots: A brief review of perception, production, and cross-species comparisons. *Brain and language*, 115(1), 81-91.

This Article is brought to you for free and open access by the Humane Society Institute for Science and Policy. It has been accepted for inclusion by an authorized administrator of the Animal Studies Repository. For more information, please contact [eyahner@humanesociety.org](mailto:eyahner@humanesociety.org).

# Vocal Learning in Grey Parrots: A Brief Review of Perception, Production, and Cross-Species Comparisons

Irene M. Pepperberg<sup>1,2</sup>

<sup>1</sup> *Harvard University*

<sup>2</sup> *Brandeis University*

## KEYWORDS

Grey parrots, allospecific vocal learning, Psittacine auditory perception, Psittacine model

## ABSTRACT

*This chapter briefly reviews what is known—and what remains to be understood—about Grey parrot vocal learning. I review Greys' physical capacities—issues of auditory perception and production—then discuss how these capacities are used in vocal learning and can be recruited for referential communication with humans. I discuss cross-species comparisons where applicable and conclude with a description of recent research that integrates issues of reference, production and perception.*

## **1. Introduction**

Grey parrot (*Psittacus erithacus*) imitative abilities have been lauded at least since Aristotle. Only fairly recently, however, have we learned how Greys may achieve such feats, how such vocal ability connects to cognitive capacities, and how their abilities compare to other common vocal learners—songbirds, cetaceans, humans, and possibly other mammals. This chapter briefly reviews what is known—and what remains to be understood—about Grey parrot vocal learning. I review Greys' physical capacities—issues of auditory perception and production—then discuss how these capacities are used in vocal learning and can be recruited for referential communication with humans. I discuss cross-species comparisons where applicable and conclude with a description of recent research that integrates issues of reference, production and perception.

## 2. Physical capacities

### 2.1. Perception

No studies directly test Grey parrot vocal perception; we infer that Greys perceive sounds at least as well, and in similar ranges, as humans in order to reproduce human vocalizations accurately (see below). Perceptual data do exist for oscine songbirds (e.g., Okanoya & Dooling, 1987) and for psittacids distantly related to Greys: budgerigars (*Melopsittacus undulatus*) and a cockatiel (*Nymphicus hollandicus*; Okanoya & Dooling, 1987), and orange-fronted conures (*Aratinga canicularis*; Wright, Cortopassi, Bradbury, & Dooling, 2003). These studies suggest both possible parallels and differences among parrots, other avian species, and humans.

For most psittacids and oscines that were tested, regions of both greatest vocal energy and best auditory sensitivity were between 2 and 5 kHz, also the optimal range for human auditory sensitivity (Dooling, 1980; Wright et al., 2003). Such data imply that basic hearing capabilities and the responsible auditory structures are evolutionarily conserved across many species (Dooling, Lohr, & Dent, 2000); however, sensitivities of some hummingbirds, also vocal learners, are at higher ranges (see Ferreira, Smulders, Sameshima, Mello, & Jarvis, 2006; Ficken, Rusch, Taylor, & Powers, 2000). Interestingly, the conure and budgerigar, but neither the cockatiel nor most other avian species so tested (see review, Wright et al., 2003), exhibit critical ratios (signal-to-noise ratio at masked threshold) that differ considerably from those of humans; for most vertebrates, the critical ratio increases by 3 dB/octave (Dooling et al., 2000), whereas for the budgerigar, the ratio is lowest in that same 2–4 kHz window. Such findings suggest, given the phylogenetic and habitat distance of conures and budgerigars (de Kloet & de Kloet, 2005), and the closer relationships between budgerigars and cockatiels, that local effects, such as habitat and sympatric species' vocalizations, may have influenced acoustical niche diversification and evolved perceptual capacities (note Ryan & Brenowitz, 1985).

Other research (e.g., Weisman, Njegovan, Williams, Cohen, & Sturdy, 2004) demonstrates that birds in general, including the budgerigar, discriminate more finely than humans in some domains such as absolute pitch perception. Here, pigeons, a species without significant vocal learning (e.g., ten Cate, Slabbekoorn, & Ballintijn, 2002 data for doves), fall somewhere between humans and songbirds (Friedrich, Zentall, & Weisman, 2007). Playback of birdsong at slower speeds than normal reveals considerable complexity missed by human ears; conceivably such detail exists for its usefulness to the birds themselves (Kroodsma, 2005).

Of interest in terms of the relationship of perception to production, including allospecific production—especially with respect to human speech—are studies demonstrating budgerigar perception of human vowel tokens /i/, /a/, /e/ and /u/ (Dooling, 1992; Dooling & Brown, 1990) and their ability to distinguish consonants /ba-/ pa/ (bilabial), /da-/ta/ (alveolar), /ga-/ka/ (velar) (Dooling, Okanoya, & Brown, 1989), /ra-/la/ (Dooling, Best, & Brown, 1995), and /ba-/wa/ consistent with the stimulus-length effect shown in humans (Dent, Pierce, Brittan-Powell, & Dooling, 1997). Clearly, birds perceive sound differences that are critical to production distinctions. Scanlon's (1999) interpretation of similarities in budgerigar and human acoustic features is that budgerigars perceive human speech as modified budgerigar sound. He argues that “. . . shared neural mechanisms which correspond to these shared acoustic patterns could constitute a phylogenetically conservative level of auditory perception which is communication-sound-specific but not species-specific”.

Despite phylogenetic and spatial distances of Greys from the above-mentioned psittacids (i.e., possible separation from Australian parrots 76 MYA and from South American parrots ~35 MYA; de Kloet & de Kloet, 2005), Greys likely have, given the need for perception to underlie production, perceptual abilities

similar to birds discussed above. These perceptual abilities are, however, likely a general vertebrate trait: Even animals lacking vocal learning (e.g., chinchillas) perceive some auditory distinctions of human speech (e.g., Kuhl & Miller, 1975).

## 2.2. Production

Few studies have examined natural Grey parrot vocalizations (imitative: Cruickshank, Gautier, & Chappuis, 1993; conspecific: May, 2004); considerably more is known with respect to their learning human speech, particularly for the Grey, Alex (Pepperberg, 1999). Using standard phonetic and statistical techniques, students and I analyzed spectrograms and videotapes of Alex's speech samples and compared his and my vocalizations (I was his principal trainer). Although Alex's anatomy prevented him from exactly reproducing human articulatory acts, he (though maybe not all parrots) used a two-tube system and frequency modulation, and employed his tongue, glottis, and larynx in some of the same ways used by humans to produce vowels and stop consonants (Patterson & Pepperberg, 1994, 1998; Warren, Patterson, & Pepperberg, 1996). Alex's vowels each exhibited formant structures similar to those of humans (see Fig. 1); his stop consonants exhibited voiced/ voiceless, labial, alveolar, and velar groupings (/p-b/, /d-t/, /g-k/). I summarize only our most relevant findings (details in Patterson & Pepperberg, 1994, 1998; Pepperberg, 2002).

### 2.2.1. $F_0$ analyses

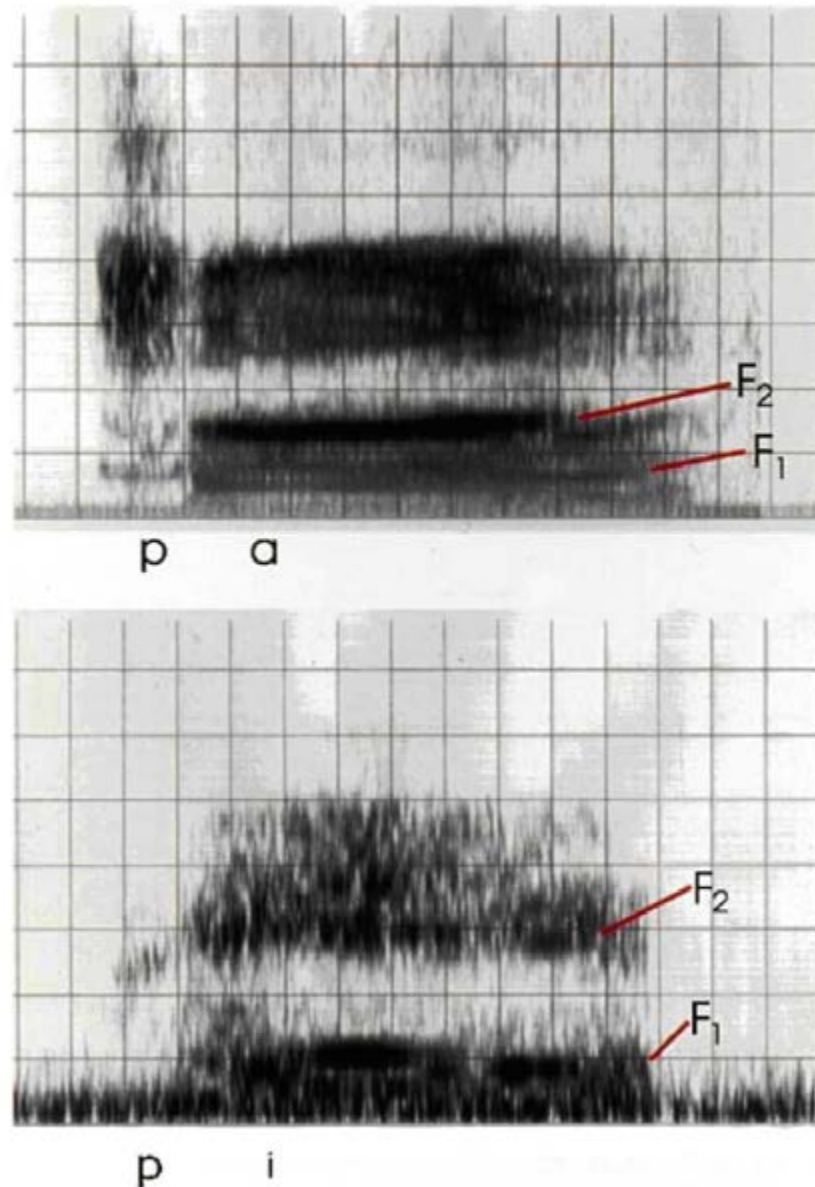
We measured fundamental frequencies,  $F_0$  (source vibration at the avian syrinx; syringeal constriction at the tracheal base functionally resembles human vocal folds in phonation; Gaunt & Gaunt, 1985; Scanlan, 1988; Suthers & Zollinger, 2004).  $F_0$  analyses show that Alex, like humans and unlike most songbirds (Greenewalt, 1968; Miller, 1977), used only one set of articulators to produce speech (Larsen & Goller, 1999; Patterson & Pepperberg, 1994)—not by using two halves of his syrinx independently to produce formant-like structure from  $F_0$ 's via sine wave interference as Lieberman (1984) suggests. His  $F_0$ s approximated those of an adult human (124–276 Hz, Peterson & Barney, 1952). Even mynahs (*Gracula religiosa*), an oscine, do not use sine wave interference to produce human speech, but their mechanism differs from that of Greys (Klatt & Stefanski, 1974). Budgerigars produce human vowels by amplitude modulation, but not quite as Lieberman suggests. Because their syrinx cannot produce two independent sounds, they produce speech via a different, as yet undetected, mechanism (Banta Lavenex, 1999).

### 2.2.2. Vowels— $F_1$ and $F_2$ analyses

Accurate production of human speech involves formants, resonant frequencies of the human mouth opening ( $F_1$ ) and oral cavity ( $F_2$ ) that could also represent specific parrot vocal tract areas; sometimes  $F_3$  is necessary (Borden & Harris, 1984; Lieberman, 1984). A formant is not a harmonic of  $F_0$ ; harmonics, if present, appear as odd multiples of  $F_0$ .  $F_1$  and  $F_2$  variations may differentiate vowels for classes of human speakers. We compared Alex's and my  $F_x$ 's and tested whether Alex's formants predict his vowels (Patterson & Pepperberg, 1994). For both  $F_1$  and  $F_2$ , his range was smaller than mine and lacked my low frequency values. His  $F_1$ s, which in humans roughly corresponds to tongue height, differed considerably from mine and, unlike mine, varied little across most vowels (Warren et al., 1996). Our  $F_2$ s, which in humans corresponds to front/back tongue position, were similar: Like mine, his vowels were somewhat distinguishable by  $F_2$ ; his front/back position varied considerably. Thus, as for humans, his formants correlated with specific vocal tract area resonances (Beckers, Nelson, & Suthers, 2004; Remez, Rubin, Nygaard, & Howell, 1987; Warren et al., 1996). His  $F_3$ s, which in humans indicate whether speech-producing constrictions occur toward the front or back of the vocal tract, were rarely observed or nearly indistinguishable from  $F_2$ . All his vowels clustered in the high/low midrange but his /i, I, E, ae/ and /U, ə, a/ respectively cluster as 'front' and 'back' vowels. /u/ is a back vowel for most Standard American English

speakers; for Alex and me it was a diphthong—had two parts, /u<sub>1</sub>/ and /u<sub>2</sub>/. Mine are front vowels; Alex's were front and back vowels, respectively (Patterson & Pepperberg, 1994).

**Fig. 1. Alex's production of "pah" /pa/, his label for pasta and for "pea" /pi/, his label for a green pea.**



### 2.2.3. Stop consonants

Traditional acoustic analyses for stops (consonants with blocked air flow) generally applied to Alex (Patterson & Pepperberg, 1998). We examined voice-onset timing (time for humans between lip opening to release air and start of sound source vibration), number of bursts (bursts result from release of air pressure built up by lip closure), and stop loci (high energy peaks associated with voiceless aspirated portions of stops). We measured formant frequency in the first 10 ms of voicing following a stop, average formant frequency during the steady state portion of a vowel following a stop, and time from beginning of voicing to when formants assume a steady state pattern. Similarities and differences exist for our stops

(Patterson & Pepperberg, 1998). Similarities involve statistically distinct acoustic characteristics and separation of stops into voicing and place of constriction subsets. Differences involve specific subset divisions and mechanisms for /b,p/ production. Compared to my subsets, Alex's /b/ was an outlier; his /p/ clustered nearer /d,g/ than /t,k/. Alex produced bilabial /b,p/ without lips; unlike normal humans, he likely used esophageal involvement (Patterson & Pepperberg, 1998).

#### 2.2.4. *Related data*

X-ray video data (Warren et al., 1996) suggest that Alex used anticipatory co-articulation, setting up his vocal tract for the next speech sound as the previous one was completed (e.g., his /k/ in “key” differs from that in “cork”, Patterson & Pepperberg, 1998). Thus he appeared to separate specific phonemes from the speech flow and produce them so as to facilitate production of upcoming phonemes. In humans, these abilities are taken as evidence for top-down processing (Ladefoged, 1982), and are necessary for segmentation and phonological awareness, discussed below.

Alex's reproduction of human speech was likely constrained only by human–avian vocal tract differences (e.g., no lips). His vocalizations all resemble human samples; characteristics that describe human speech and processes used to produce human speech apply to him (Patterson & Pepperberg, 1994, 1998).  $F_0$  data specifically show that Alex, like humans, produces true resonances, not sinusoids at appropriate formant frequencies. His brain areas devoted to vocal learning (Jarvis, 2004) control numerous vocal tract areas in order to produce human speech, a form of gestural learning (Arbib, 2005). Such data relate to gestural theories of song and language evolution, to mirror neurons (MNs, neurons that respond the same way whether one performs or views an action), and to mirror neuron systems (MNS; e.g., Arbib, 2005).

#### 2.3. *Mirror neurons*

Gestural (or motor) theories of language and song evolution suggest how vocal communication evolved in, respectively, primate (Hewes, 1973) and avian lines (Williams & Nottebohm, 1985). A very brief summary of rather complex ideas is that (a) manual gestures—possibly accompanied and modified by grunts or calls—shifted to vocal tract—articulatory—gestures that supported finer-grained vocal communication and (b) an individual understands and acquires vocal communication by representing the speech that is heard as motor articulatory behavior (Liberman & Mattingly, 1985; Vihman, 1993), which can then be both interpreted and reproduced. Such representation and acquisition is assisted if, for example, your articulatory system responded to my voice as if you were talking. Interestingly, some MNs involve exactly that parity (e.g., Arbib, 2005; Craighero, Buccino, & Rizzolatti, 2002). And, because MNs have an inhibitory component, allowing you not to repeat my utterance if you so chose (Baldissera, Cavallari, Craighero, & Fadiga, 2001), they do not preclude voluntary action. The specific connection to language is that MNs are found in Broca's area—one of the human ‘language centers’—and in the monkey analogue of Broca's area, F5; the monkey's system and part of the human system both react to manual grasping and related actions (e.g., Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995), including ‘mouthing’ actions in monkeys and articulatory actions in humans (Craighero et al., 2002). As hypothesized by Rizzolatti and Arbib (1998), the development of vocal from manual communication involves evolution of Broca's area from F5: no longer would speech require a major brain reorganization, only reorganization of the MN system from voluntary control of manual gestures to include articulatory ones (e.g., Corballis, 2003), so as to handle the evolution of learning to communicate. Various types and levels of MN systems likely exist, both across and within species. Notably, monkeys, unlike humans, fail to engage in vocal learning or even true manual imitation (Visalberghi & Fragaszy, 1990) but have intact MN systems. Even in humans, MN systems possibly vary developmentally or with respect to control of various levels of behavioral complexity (e.g., Pepperberg, 2006). Similarly, brain structures responsible

for avian unlearned calls could have evolved to control suboscines' unlearned songs, then to control the complex learned oscine system and parrots' open-ended learning (Farries, 2001; Pepperberg, 2006). But until recently, mammalian–avian parallels were unclear.

Certain avian and mammalian brain structures, including those responsible for vocal learning, are now thought to have evolved from the same reptilian pallial structures (Jarvis et al., 2005); thus it does not seem unreasonable to propose some form of avian MN system. One cannot, however, argue for direct correlations to Broca's area in birds as one can for F5 in monkeys. Possibly the bits of brain corresponding to specific mammalian language/articulatory gesture centers are apportioned across several song centers in the avian brain (Reiner, pers. comm., April 19, 2005; see Jarvis, 2004; also Kirn, this volume, for discussions and definitions of songbird brain circuitry).<sup>1</sup> Interestingly, for the oscine song sparrow, Prather, Peters, Nowicki, and Mooney (2008) found HVC<sub>x</sub> neurons (a population of neurons in the songbird HVC, the higher vocal center—necessary for the learning and production of song—that innervates Area X, another brain area important to song learning and perception) that display nearly identical patterns of activity when a bird sings and hears the same sequence of notes. Although the authors are careful not to claim that these neurons are indeed MNs, and suggest that the exact MN system could lie downstream (e.g., as part of a loop involving other song areas), the involved brain areas do correlate with those involved in human vocal behavior. Given our knowledge of parallels among songbird, human, and parrot vocal centers (Jarvis et al., 2005), a comparable parrot MN system is likely. Because parrots lack a specific correlate to HVC<sub>x</sub>, quite possibly the MN-related area is indeed downstream for both parrots and songbirds, or also involves connections between what are parrot correlates of songbird areas: i.e., connections between AAC (nucleus centralis arcopallii anterioris) and MMSt (magnocellular nucleus of the medial striatum), the parrot versions, respectively, of RA (robust nucleus of the arcopallium, part of the song production pathway) and Area X (the perception area noted above). The parrot system also likely differs somewhat from that of most songbirds so as to allow for more diverse forms of imitation. Specifically, given studies demonstrating how, in adult songbirds, LMAN (lateral magnocellular nucleus of the nidopallium) inactivation reduces singing variability (Kao, Doupe, &

Brainard, 2005; Őlveczky, Andalman, & Fee, 2005) and the connections in songbirds between X, LMAN, and RA for perception and production, quite likely in parrots the MNS for imitative behavior will involve NAOc (the oval nucleus of the anterior nidopallium complex)—which is the parrot equivalent of LMAN—and its connections to MMSt and AAC. Interestingly, cross-species MN responses are already documented: Monkey *F5* MNs respond to human hand gestures; might some form of parrot MNs involved in articulatory gestures respond to those of humans and assist with psittacine learning of human speech?

### **3. Cognitive capacities and vocal learning**

Studies of perception, production, and even MNs address parrots' physical abilities, but not the extent and processes of their vocal learning. Parrots, unlike most songbirds, are open-ended learners; moreover, many parrots master allospecific acquisition. This trait, observed in nature (Cruickshank et al., 1993), is most obvious in captivity. And, although many captive parrots may mindlessly mimic, Greys, at least, can learn to use human speech referentially (Pepperberg, 1981, 1999). What appears to enable reference is the form of training.

Parrots do not readily acquire any novel sound. Pet birds may indeed mimic random utterances and environmental noises without apparent instruction, but fail to learn English speech via psychological conditioning techniques (e.g., Grosslight & Zaynor, 1967; Mowrer, 1950; Pepperberg, 1999). Thus, for parrots, different types of vocal learning likely require different types of input: Simple mimicry used only to attract attention differs from acquisition of vocalizations having communicative content (Pepperberg, 2002). Like young children (e.g., Hollich, Hirsh-Pasek, & Golinkoff, 2000), Greys acquire communication

skills most effectively when teaching is referential, functional, and socially rich (Pepperberg, 1997, 2004). Reference is an utterance's meaning—the relationship between labels and objects to which they refer—and is exemplified by presenting to the bird, during training, the specific objects being named, as the primary reward for vocalizing appropriately. Functionality involves the context and effects of an utterance's use; opportunities during training for using labels to request objects provide a rationale for learning unique, unfamiliar sets of sounds constituting human speech. Social interaction between subject and trainer focuses attention on certain aspects of the environment, emphasizes common attributes of objects or actions (and possible underlying rules for their use) and allows trainers to adjust procedures to a learner's level. Interaction with trainers engages subjects directly, provides a contextual explanation for actions, and demonstrates the consequences of those actions (Pepperberg, 2004). These elements of training define what has become known as the model/rival technique.

### *3.1. The model/rival (M/R) technique*

In the 1970s, after reviewing what was then known about parrots' vocal learning, child language acquisition, and human social learning, I devised the M/R procedure to teach parrots referential use of human speech (Pepperberg, 1981). M/R training, based on methods developed by Todt (1975) and Bandura (1971), uses three-way social interactions among two humans and a parrot to demonstrate the targeted vocal behavior (Pepperberg, 1981). The parrot watches and listens as one trainer presents objects and queries the other trainer about them (e.g., "What's here?", "What color?"), giving praise and transferring the named object to the human partner to reward correct answers. Incorrect responses are punished by scolding and temporarily removing items from sight. Thus the second human is both a model for the parrot's responses and its rival for the trainer's attention, and illustrates consequences of errors. The model must try again or talk more clearly if the response was deliberately incorrect or garbled; that is, the model is subject to corrective feedback, which the bird observes. The parrot is included in interactions, being queried and rewarded for successive approximations to correct responses; training is adjusted to its performance level. If a bird is inattentive or accuracy regresses, trainers threaten to leave.

Unlike M/R procedures others have used (see Pepperberg & Sherman, 2000), ours exchanges roles of trainer and model. The parrot thus sees how questioner and respondent interchange roles, and how the procedure causes environmental change. Role reversal also counteracts an earlier methodological problem: birds whose trainers always maintained their respective roles responded only to the human questioner (Todt, 1975). With our system, birds respond to, interact with, and learn from any human.

To ensure the closest possible link between labels or concepts to be learned and their appropriate referent, M/R training uses only intrinsic reinforcers: reward for uttering "X" is X, the object to which the label or concept refers. Earlier unsuccessful attempts to teach birds to communicate with humans used extrinsic rewards: a single food that neither related to, nor varied with, the label or concept being taught (Pepperberg, 1999). This procedure delayed label and concept acquisition by confounding the label of the targeted exemplar or concept with that of the food reward. Initial use of labels as requests also demonstrates that uttering labels has functionality; later, birds learned "I want X", to separate requesting and labeling (Pepperberg, 1988) and to enable them to request preferred rewards while learning labels for items of little interest to them.

M/R training successfully enabled label and concept acquisition (for Alex's repertoire, see Pepperberg, 1999), but did not specify which training elements were necessary and sufficient. Earlier unsuccessful attempts to teach birds referential speech lacked at least some of these elements but neither deconstructed the learning environment nor tested the significance of different input elements individually and in various combinations. To learn why M/R training was so successful, and whether some subset of input elements would suffice, I began testing how changes might affect learning not only with respect to



sound reproduction, but also with respect to comprehension and appropriate use, focusing on actions that require cognitive processing.

### *3.2. Eliminating aspects of M/R training*

To provide training input that varied in amounts of reference, functionality, and social interaction, I contrasted M/R tutoring with sessions using audiotapes, videos, and different levels of human interaction; I also involved additional subjects. Had I significantly changed Alex's training protocols, he might have ceased learning simply because I had made a change, not because of the quality of the change. New subjects, lacking training, would not be influenced by prior experience. I performed one M/R variant with Alex and more extensive permutations with four naïve Greys: Kyaaro, Alo, Griffin, and Arthur.

The juveniles Alo and Kyaaro experienced sessions in social isolation (Pepperberg, 1994b): Each heard audiotapes of Alex's sessions (input now nonreferential, not contextually applicable, and non-interactive) and saw videotapes showing how Alex and trainers received referential rewards for labeling (input now referential, non-interactive, minimally functional and contextually applicable). Nothing the juveniles uttered could enable reward. In all cases, no new labels were acquired; however, both birds learned referential labels in concomitant M/R sessions, showing that the birds' abilities were not the issue.

Additional experiments explored video-training variations (Pepperberg, Gardiner, & Luttrell, 1999; Pepperberg, Naughton, & Banta, 1998; Pepperberg & Wilkes, 2004). First, a co-viewer provided social approbation for viewing and pointed to the screen while making comments like "Look what Alex has!" but did not repeat targeted words, ask questions, or relate content to other training. Birds' attempts to utter a label could garner only vocal praise; social interaction was limited; referentiality and functionality matched earlier video sessions. Second, the co-viewer was more interactive, uttering targeted labels and querying birds. Third, a socially isolated parrot watched videos while a student in another room monitored its utterances via headphones and could deliver intrinsic rewards remotely. Fourth, we used live video from Alex's sessions to minimize habituation. Fifth, in a study with Arthur, using an LCD monitor, we tested whether CRT screen flicker could affect learning. Birds did not learn labels from any of these procedures, but, again, learned labels presented in concurrent, traditional M/R sessions.

Based on studies with children (Baldwin, 1991), I examined the role of joint focusing of attention between bird and trainer. A single trainer faced away from a bird, who was within reach of, for example, a key; the trainer talked about the object, emphasizing its label, "Look, a shiny key!" "Do you want key?" etc. These sentences framed the label, allowing for repeated label use while minimizing possible habituation (Pepperberg, 1981). The trainer had no visual or physical contact with the parrot or object. If a bird attempted to utter the targeted label, it would receive vocal praise. Overall, some functionality and considerable social interaction were eliminated (Pepperberg & McLaughlin, 1996). Again, no new label learning occurred.

An experiment with Griffin eliminated the second trainer. One student labeled objects to which she and Griffin jointly attended (Pepperberg, Sandefer, Noel, & Ellsworth, 2000). Griffin did not utter any targeted label during or immediately after 50 such sessions, but did so after two or three subsequent M/R sessions involving the same items. Because birds switched to M/R training after 50 video sessions needed about 20 sessions before beginning to acquire targeted labels (Pepperberg, 1999)—about the same number of sessions needed when labels are introduced initially in M/R training—Griffin had likely acquired labels during the solo-trainer sessions, but did not understand how to use them until he observed their usage modeled. Importantly, he learned new labels and their use in concurrent M/R sessions with other objects.

Learning that occurs with impoverished input may produce mimetic speech, as in Alex's M/R-variant study, where I eliminated as much reference and functionality as possible while retaining a two-trainer system (Pepperberg, 1994b). Alex viewed a string of pictured Arabic numerals as two humans interacted: the model/rival stating "Say number", the other uttering a vocal sequence based on Korean count labels, but without reference either to specific numerals or to Alex's previously acquired English number labels (Pepperberg, 1994a, 1997). The Korean-based labels, required for a simultaneous study on serial learning (Silverstone, 1989), provided maximal contrast with English. Training lacked functional meaning and all but minimal referentiality—the page with numbers was simply a focus for joint attention. As in the basic M/R procedure, roles were reversed and Alex was included in interactions. Reward was the opportunity to request a favored item ("I want X") and vocal praise; errors elicited scolding and time-outs.

Alex learned the string of modeled vocal labels—but results differed from those of M/R studies (Pepperberg, 1981, 1994b; Silverstone, 1989). Strikingly, he did not use these sounds as labels, nor could he subsequently learn to use them referentially—that is, with respect to either serial labeling or quantity, even after we paired them with objects as one would state "first, second, . . .", etc. Alex had thus learned to produce, but apparently could not learn to comprehend, these sounds (Pepperberg, 1994b), like a child who, having learned the 'alphabet song' thinks 'LMNOP' is one letter (Pepperberg, 2004). Given his previous success on label production and comprehension after M/R training (e.g., Pepperberg, 1990a, 1990b, 1992, 1994a), Alex's failure was likely a consequence of the training protocol (Pepperberg, 1994b). Thus referential imitation differs from 'mere mimicry', and training that engenders the latter is usually incapable of engendering the former, with the implication that different cognitive processes—and possibly different neural areas (e.g., see below: separate or different levels of MN systems; Pepperberg, 2006)—are involved in the two situations.

In sum, Grey parrots apparently acquire referential speech only via original M/R training. Training must involve two humans who demonstrate referentiality and functionality of a label to be learned, who socially interact with each other and the bird, exchange roles of questioner and respondent, portray the effects of labeling errors, provide corrective feedback, and adjust the level of training as the subject begins to learn. When training lacks some or all of these elements, birds typically fail to learn referential speech.

Similarities exist with respect to the importance of social interaction for vocal learning in many species. Without going into detail, I note that songbirds given impoverished input—audiotapes in sound isolation boxes (Marler, 1970)—demonstrate limitations not observed in nature, for example, on the period during which they can learn song and their capacity for allospecific and conspecific (multiple dialect) acquisition (see Baptista & Gaunt, 1997). Birds given live tutors, in contrast, may extend the song acquisition period and learn songs of allospecific, sympatric species (Baptista & Gaunt, 1997). Dolphins also exhibit more rapid and flexible vocal learning when exposed to procedures that are referential and functional (Reiss & McCowan, 1993) compared to operant techniques (Richards, Wolz, & Herman, 1984), and preliminary evidence exists for some level of social allospecific learning in elephants (Poole, Tyack, Stoeger-Horwath, & Watwood, 2005). Children who experience labeling that lacks joint reference fail to acquire targeted labels (Baldwin, 1991) and children who watch videos in isolation learn less than children with interactive trainers (e.g., Rice, Huston, Truglio, & Wright, 1990). Children who have developmental communicative disabilities also learn more in model/rival situations than with procedures that lack various elements of M/R training (see Pepperberg & Sherman, 2000, 2002).

Specifically, Grey parrots' simultaneous acquisition of sound patterns and their association with specific objects during M/R training argue for a mechanism different from simple mimicry, as do data showing the difficulty of training label-object associations after mimetic label acquisition for Greys, apes, and dolphins (Pepperberg, 1999). Note, too, data (below) suggesting that, as Alex matured, he began constructing

novel labels not by directly reproducing his human trainers' speech but by composition from existent labels (and segmented parts of existent labels) in his repertoire.

#### **4. Developmental similarities with children and songbirds: untrained vocal practice**

Like children (e.g., Kuczaj, 1983; Nelson, 1989) and songbirds (e.g., Baptista, 1983; Marler, 1970), Alex engaged in active vocal practice before acquiring targeted vocalizations. Using sound play (phonetic 'babbling'/recombination; Pepperberg, Brese, & Harris, 1991) he derived new speech patterns spontaneously from existent ones. After acquiring "grey", for example, he produced sound variants that we eventually mapped to appropriate referents (Pepperberg, 1990c; see below). Notably, these recombinations suggest he abstracted rules for utterances' beginnings and endings. In over 22,000 vocalizations, he never made backwards combinations (e.g., "percup" instead of "cupper/copper"; Pepperberg et al., 1991). As noted above, he seemed to exhibit anticipatory co-articulation; such behavior observed along with sound play is often considered evidence for top-down processing (Ladefoged, 1982). Finally, he engaged in label segmentation in a manner suggesting some form of phonological awareness (Pepperberg, 2007, 2009). I discuss all but co-articulation in some detail.

##### *4.1. Vocal play*

Sound play—repetitions and variations on vocal patterns—has many forms, including "monologue speech": active practice of a communication code in situations without overt social stimulation (Kuczaj, 1983). Monologue speech further divides into private speech produced in solitude and social-context speech produced in the presence of potential receivers but lacking obvious communicative purpose. When alone in his cage, Alex engaged in the former, which explained some of his acquisition data: Although most of his labels appeared initially as rudimentary patterns that improved during training—first as a vocal contour, then with vowels, finally with consonants (Patterson & Pepperberg, 1994, 1998)—sometimes completely-formed new labels or entire phrases materialized after minimal training and without practice in his trainers' presence. By eavesdropping, we found he was practicing these utterances privately (Pepperberg et al., 1991).

Such behavior patterns are presumably integral to development; their occurrence across species (e.g., songbirds (Marler, 1970); children (Nelson, 1989; Weir, 1962); cetaceans (McCowan, Hanser, & Doyle, 1999)), suggests an evolutionary theory of language play (Kuczaj, 1998). Like most play, monologue speech facilitates learning by allowing experimentation with adult systems: full freedom to choose topics and contexts, to attempt novel words, and compare familiar and novel forms to extract regularities and discover differences (Kuczaj, 1983). Lack of negative feedback or the need to use correct forms may encourage practice and even accelerate learning.

Equally important for repertoire expansion, Alex's social-context speech appeared when, with trainers nearby, he produced new vocalizations by recombining existing label parts in their corresponding orders (Pepperberg, 1990c). Such spontaneous novel phonemic combinations often occurred outside of testing and training, and juvenile Greys behave similarly (e.g., Neal, 1996; Pepperberg, 2004), with combinations appearing in contexts and forms reminiscent of children's play. These utterances were rarely if ever used by trainers, but sometimes resembled both existing labels and separate human vocalizations (e.g., "grain" from "grey"). When we referentially mapped these spontaneous utterances—providing relevant objects to which they could refer—Alex rapidly integrated these labels into his repertoire. For "grain", trainers gave Alex seed, something not normally available; later sprouted legumes were substituted. He received a paper clip ring for "chain", appropriate fruit for "grape", and a nutmeg grater for trimming his beak after uttering "grate". "Cup" (from "up") was rewarded with metal cups and plastic mugs, "copper" (first produced as "cupper") with pennies, and "block" with cubical wooden beads. "Chalk" (from "talk")

mapped to blackboard chalk; “truck” to toy cars and trucks. We ignored “shane” and “cheenut”; he abandoned these sounds (Pepperberg, 1990c). Thus Alex’s spontaneous utterances that initially lacked communicative value could, as they do for children, acquire such value if caretakers interpreted them as meaningful and intentional (Pepperberg, 1990c); he behaved as though our interactions “conventionalized” both the sound patterns and the sound-meaning connections in the direction of standard communication.

Somewhat like children engaged in testing meanings of newly-acquired labels (e.g., Brown, 1958, 1973), my birds also use humans to provide additional referential information for relatively novel labels (Pepperberg, 2002). Our youngest subject, Arthur, uttered a label used in a very specific context—“wool”, for a woolen pompon—while pulling at a trainer’s sweater. He seemed to be testing the situation, and our responses—given with high positive affect—stimulated him further, revealing the potential power of an utterance and encouraging his early categorization attempts. Even if birds err in initial categorizations, they are reinforced when we provide a correct, new label for the item: replying that an almond isn’t “cork”, but “cork nut”—a label they will subsequently use (Pepperberg, 1999, 2002).

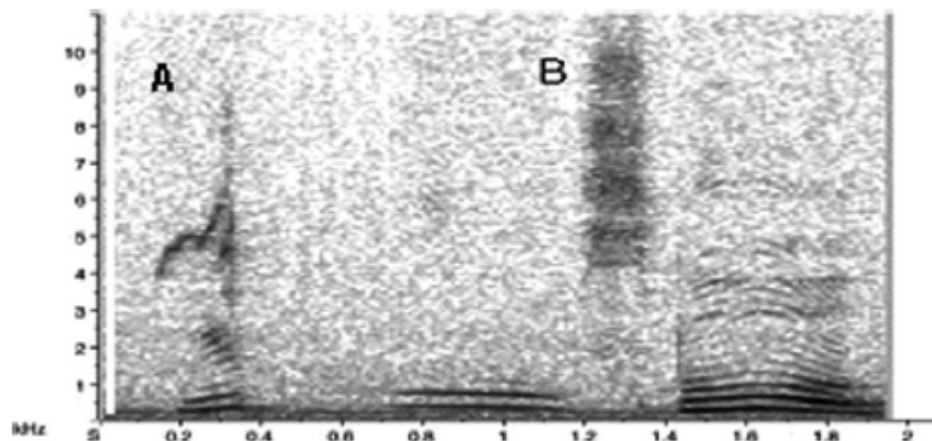
#### *4.2. Segmentation*

Of even greater interest is that Alex appeared capable of vocal segmentation—a special form of vocal combinatory behavior—that (a) implied that he understood that his existent labels were formed of individual morphemes or phonemes that could be combined in novel ways to create what were for him novel vocalizations (e.g., Greenfield, 1991; Peperkamp, 2003), and (b) would also demonstrate phonological awareness (Pepperberg, 2007). Such behavior is not only considered basic to human language development (Carroll, Snowling, Stevenson, & Hulme, 2003), but also a uniquely human trait: most animals, lacking speech, are never exposed to, nor trained nor tested on, phonological awareness, nor are they expected to have internal representations of phonemes that would engender such combinatorial behavior (Pepperberg, 2007). Such behavior is not innate even for humans: Children apparently shift from recognizing and producing words holistically (Studdert-Kennedy, 2002) to recognizing words as being constructed via a rule-based phonology when about three years old (Carroll et al., 2003; Vihman, 1996); furthermore, manipulation of individual parts of words is presumed to require development of an internal representation of phonological structure (Byrne & Liberman, 1999). That is, in order to sound out—rather than mimic—a novel label, children must segment the sound stream into discrete elements, recognize a match between those elements and elements (or close approximations) in their own repertoires, and then recombine these elements into an appropriate sequence (see Gathercole & Baddeley, 1990; Treiman, 1995). Moreover, children’s ability to focus on the sounds of words and sound elements of words rather than solely on word meaning may be assisted by training in sound–letter associations (Carroll et al., 2003; Mann & Foy, 2003). Little evidence exists for segmentation in animals, even a less advanced form involving combination and/or recombination of whole labels to describe novel situations. The few existent incidents—apes’ “water bird” for a swan, “cry hurt food” for a radish (Fouts & Rigby, 1977), dolphins’ “ring–ball” during simultaneous play with two items (Reiss & McCowan, 1993)—have been considered as descriptors of the entire situation rather than as specific combinations to denote one element. The following data present evidence for segmentation by Alex (Pepperberg, 2009).

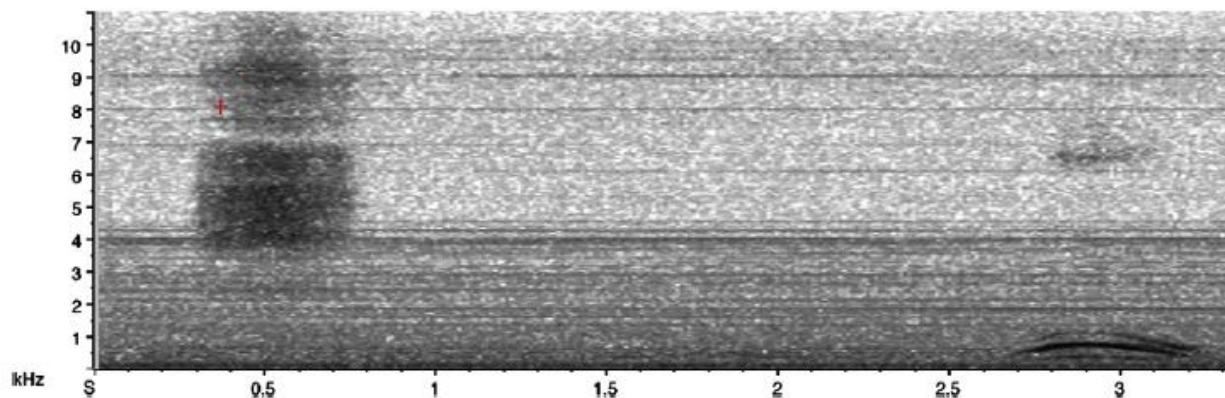
After observing the attention that Arthur received for labeling a wooden bobbin as “spool” (Fig. 2A), Alex evinced interest in the item, and received M/R training to produce the label. For spool, unlike his usual form of acquisition, Alex began using a combination of existing phonemes and labels to identify the object: /s/ (unvoiced, trained in conjunction with the alphabet letter, S) and wool, to form “s” (pause) “wool” (“s-wool”; /s-pause-wUl/; see Fig. 3; Pepperberg, 2007). The pause seemed to provide space for the absent (and difficult) /p/ (possibly to preserve the number of syllables or prosodic rhythm of the targeted vocalization; see Leonard, 2001; Peters, 2001). None of Alex’s prior labels contained /sp/, nor

did his repertoire include “pull” or “pool”, nor any label including /u/. He knew “paper”, “peach”, “parrot”, “pick”, etc. and “shape” and “sich” (six); technically, /p/, /sh/ and /s/ but not /sp/ were available. He knew /u/ from labels such as “two” and “blue” (Pepperberg, 2007). Note that both Alex’s and my /p/, when analyzed for VOT (voice onset time), fall solidly into the voiceless category, distinct from the voiced /b/ (Patterson & Pepperberg, 1998).

**Fig. 2. (A) Arthur’s “spool” compared to (B) Pepperberg’s “spool” (from Pepperberg, 2007).**



**Fig. 3. Alex’s “s-wool” (/s-pause-wU/) (from Pepperberg, 2007).**

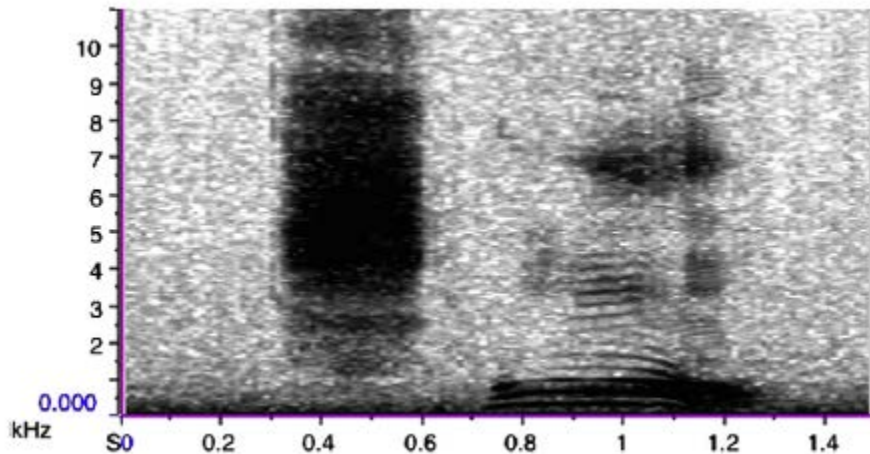


Alex retained “s-wool” for almost a year, even though usually only several weeks of training enable learning of a new label with existent phonemes (Pepperberg, 1999). After this year-long period, he spontaneously produced a perfectly formed “spool” (/spul/), replacing the space with /p/ and shifting the vowel to /u/ from /U/ (see Figs. 4 and 5; Pepperberg, 2007). His utterance sounded distinctly human, differed from Arthur’s whistled version, and clearly resembled mine (Figs. 2B and 5C), even though students had performed 90% of the training.

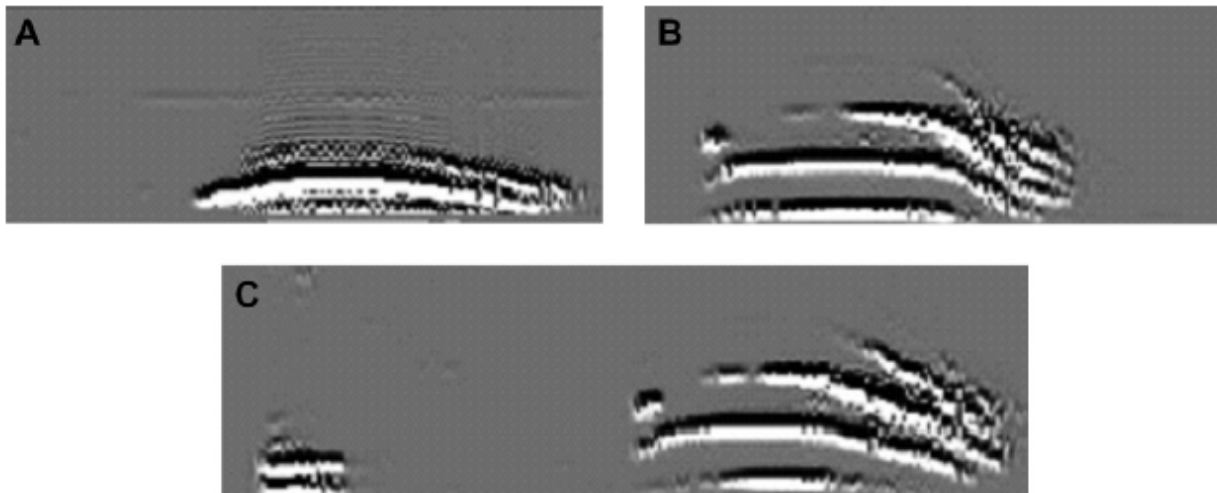
Whether Alex’s shift from /U/ to /u/ was gradual is unknown. Unlike a previous laboratory situation in which Alex was alone for specific periods each day to enable eavesdropping on his solitary practice

(Pepperberg et al., 1991), three birds (Alex, Griffin, Arthur) were now together 24/7. A gradual shift was unlikely if Alex had maintained his previous pattern of vocalizing in private: Significant portions of Alex's solitary practice involved what in humans would be considered rhymes (e.g. "green, cheen, bean"; "mail, banail") in which ends of labels were stable (Pepperberg et al., 1991); that is, he seems to have (or have acquired) categorical distinctions and minimal pairs similar to those of his human models (Patterson & Pepperberg, 1994, 1998). An abrupt shift could indicate some level of self-monitoring and even some additional awareness that the appropriate vowel for "spool" derived from yet another label such as "two" (/tu/); such information was unavailable to Arthur (Pepperberg, 2009)

**Fig. 4. Alex's "spool" (/spul/) (from Pepperberg, 2007).**



**Fig. 5. (A) Alex's /U/, (B) Alex's /u/, (C) part of Pepperberg's /spu/ (from Pepperberg, 2007).**

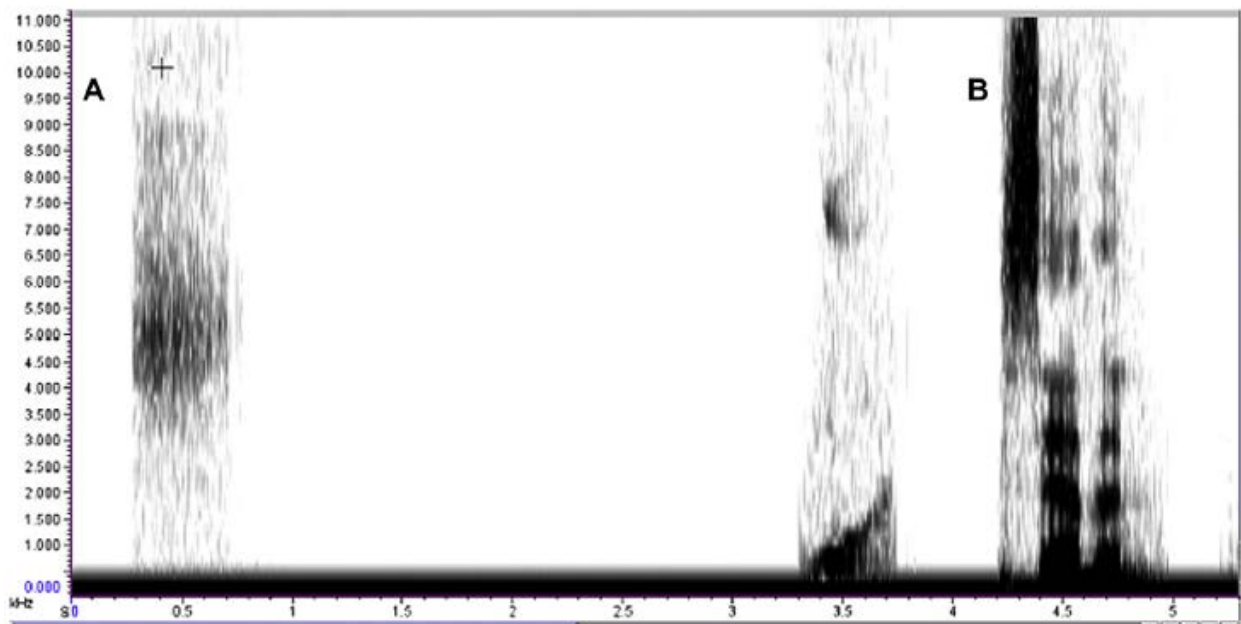


The pattern of acquisition is not unique to "spool"; Alex exhibited a similar pattern for the vocalization "seven" (first in reference to the Arabic numeral, then to an object set). Alex's first production of the label could best be described as "s . .n", a bracketing using the phonemes /s/ and /n/; he then quickly progressed to "s-one" (see Fig. 6; /s/-pause-/wEn/) which looked quite different from my "seven", but followed the form of "s-pause-wool".

Eventually, he replaced “s-one” with something sounding to the human ear like “sebun”, much closer to my “seven” (Fig. 7; sonagraph expanded for reference).

Alex’s training likely enabled him to use phonological awareness (in the sense defined in Anthony & Francis, 2005) to create a difficult new label from existing bits of sounds already in his repertoire (i.e., via segmentation) and to carefully produce the appropriate phonemes. As presented above, Alex’s data demonstrate that he had a functional understanding that his existent labels were comprised of individual units that could intentionally be recombined in novel ways to create referential, novel vocalizations (Pepperberg, 2007). Although Alex seemingly generated novel meaningful labels from a finite set of elements, the rule system he demonstrated was relatively limited. Nevertheless, the data add another intriguing parallel between Alex’s and young children’s early label acquisition.

**Fig. 6. (A) Alex’s “s-one” /s-pause-wən/ followed by (B) Pepperberg’s “seven” /sEvIn/ (from Pepperberg, 2009).**



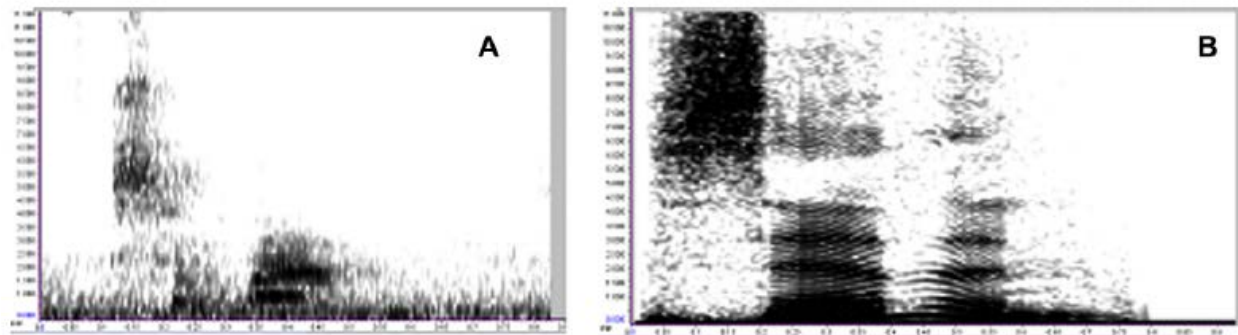
#### 4.3. Neural correlates

Notably, the different vocal behaviors discussed above might involve different neural mechanisms or combinations thereof. That is, Alex’s social-context speech, in which he generated vocal variability—without necessarily having a specific targeted outcome—might be neurally different from practice or segmentation in which he appeared to actively compare auditory production with a stored template in an attempt to produce a specific label.

Ólveczky et al. (2005) and Aronov, Andalman, and Fee (2008) have shown that juvenile songbirds’ LMAN is involved in generating extreme vocal variability across song renditions, a behavior that enables exploration of a range of possible patterns from which birds will home in on those that match a specified type of input. Lesioning LMAN prevents juvenile birds from expressing variable patterns, resulting in aberrant, stereotypic final vocalizations. Lesioning LMAN in adult songbirds, in contrast, has little effect on song production other than reducing some stereotypy (Doupe, Solis, Kimpo, & Boettiger, 2004; Kao et al., 2005). As noted earlier, in open-ended learning parrots, their area NAOc—corresponding to LMAN—might remain active throughout life and be responsible for Alex’s ability to generate a vocal series such as

“grey” → “grate” → “grain” → “chain” → “cane”. Here, Alex did not appear to have specific goals; he vocalized spontaneously and the vocalizations were mapped to meaning through actions of his human trainers (rewards).

**Fig. 7. (A) Alex’s “seben” [sEbln/] compared to (B) Pepperberg’s “seven” [sEvln/] (from Pepperberg, 2009).**



But what happens when there was a goal, as in Alex’s production of “spool” or “seven”? Here, Fiete, Fee, and Seung (2007) propose an intriguing model in songbirds that could similarly be adapted for parrots, and could also tie into the MN system suggested earlier. A very simplistic description of the Fiete et al. model is that if a particular coincident input (out of many possibilities) to RA from LMAN and  $HVC_{RA}$  is followed by some form of reinforcement (unspecified as to origin), the output from RA leads to what corresponds to crystallized song. Interestingly, however,  $HVC_x$  neurons (those proposed as possibly connected to an MN system) but not  $HVC_{RA}$  neurons respond robustly to bird’s own song (Prather et al., 2008), and  $HVC_x$  neurons may trigger LMAN activity through disinhibitory mechanisms (see Mooney, 2004; note Doupe et al., 2004). Although considerable work remains to determine specific details of which avian brain area determines when the output is ‘correct’ and how this mechanism functions, some integration of pathways including MN systems must exist for iterative learning in songbirds along with a corresponding parrot loop involving NAOc, AAC (the RA correlate in parrots), and NLC (the nidopallial vocal nucleus correlate of HVC in parrots). A caveat is that no equivalent  $HVC_x$  connection exists in parrots (Jarvis, 2004), but rather AAC to NAOc (what would be RA to LMAN in songbirds), and NAOc to NLC instead. Likely Alex’s slow, deliberative matching of vocal output to specific human labels involves such an iterative neural pathway, in contrast to the rapid variations produced in his vocal play.

## 5. Summary

Vocal learning is rare, but exists in species that are phylogenetically distant, even within the avian (songbirds, parrots) and mammalian (humans, cetaceans, elephants) lines. Similarities exist among vocal learners, not only in their ability to learn a vocal code, but also often in the manner in which the code is acquired. Striking parallels exist particularly within the human and Grey parrot system, suggesting that birds, and particularly parrots, are appropriate models for studying the evolution not only of vocal communication, but also of vocal learning (Pepperberg, 2009).

## Notes

<sup>1</sup> Arguments also exist, of course, that what some researchers consider specific language modules are actually more like bottlenecks in distributed networks connecting various populations of neurons necessary for communicative behavior; see review in Scharff & White (2004); note also Liberman and Mattingly (1985).



## References

- Anthony, J. L., & Francis, D. (2005). Development of phonological awareness. *Current Directions in Psychological Science*, 14, 255–259.
- Arbib, M. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28, 105–167.
- Aronov, D., Andalman, A. S., & Fee, M. S. (2008). A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science*, 320, 630–634.
- Baldissera, F., Cavallari, P., Craighero, L., & Fadiga, L. (2001). Modulation of spinal excitability during observation of hand actions in humans. *European Journal of Neuroscience*, 13, 190–194.
- Baldwin, D. A. (1991). Infants' contributions to the achievement of joint reference. *Child Development*, 62, 875–890.
- Bandura, A. (1971). Analysis of social modeling processes. In A. Bandura (Ed.), *Psychological modeling* (pp. 1–62). Chicago: Aldine-Atherton.
- Banta Lavenex, P. A. (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.
- Baptista, L. (1983). Song learning. In A. H. Brush & G. A. Clark, Jr. (Eds.), *Perspectives in ornithology* (pp. 500–506). Cambridge: Cambridge University Press.
- Baptista, L. F., & Gaunt, S. L. L. (1997). Social interaction and vocal development in birds. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development* (pp. 23–40). New York: Cambridge University Press.
- Beckers, G. J. L., Nelson, B. S., & Suthers, R. A. (2004). Vocal-tract filtering by lingual articulation in a parrot. *Current Biology*, 14, 1592–1597.
- Borden, G. J., & Harris, K. S. (1984). *Speech science primer: Physiology, acoustics, and perception of speech*. Baltimore: Williams and Wilkins.
- Brown, R. (1958). *Words and things*. New York: Free Press.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard.
- Byrne, B., & Liberman, A. M. (1999). Meaninglessness, productivity and reading: Some observations about the relation between the alphabet and speech. In J. Oakhill & R. Beard (Eds.), *Reading development and the teaching of reading* (pp. 157–174). Oxford: Blackwell.
- Carroll, J. M., Snowling, M. J., Stevenson, J., & Hulme, C. (2003). The development of phonological awareness in preschool children. *Developmental Psychology*, 39, 913–923.
- Corballis, M. C. (2003). From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral & Brain Sciences*, 26, 199–260.
- Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, 15, 99–402.
- Cruickshank, A. J., Gautier, J. P., & Chappuis, C. (1993). Vocal mimicry in wild African Grey parrots *Psittacus erithacus*. *Ibis*, 135, 293–299.
- de Kloet, R. S., & de Kloet, S. R. (2005). The evolution of the spindlin gene in birds: Sequence analysis of an intron of the spindlin W and Z gene reveals four major divisions of the Psittaciformes. *Molecular Phylogenetics and Evolution*, 36, 706–721.
- Dent, M. L., Pierce, A., Brittan-Powell, E. F., & Dooling, R. J. (1997). Perception of synthetic /ba-/wa/ speech continuum by budgerigars (*Melopsittacus undulatus*). *Journal of the Acoustical Society of America*, 102, 1891–1897.
- Dooling, R. J. (1980). Behavior and psychophysics of hearing in birds. In A. N. Popper & R. R. Fay (Eds.), *Comparative studies of hearing in vertebrates* (pp. 261–288). New York: Springer-Verlag.

- Dooling, R. J. (1992). Perception of speech sounds by birds. In Y. Cazals, L. Demany, & K. Horner (Eds.), *Advances in biosciences: Auditory physiology and perception* (pp. 407–413). UK: Pergamon Press.
- Dooling, R. J., Best, C. T., & Brown, S. D. (1995). Discrimination of synthetic full-formant and sinewave /ra/-/la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*). *Journal of the Acoustical Society of America*, 97, 1839–1846.
- Dooling, R. J., & Brown, S. D. (1990). Speech perception by budgerigars (*Melopsittacus undulatus*): Spoken vowels. *Perception & Psychophysics*, 47, 568–574.
- Dooling, R. J., Lohr, B., & Dent, M. L. (2000). Hearing in birds and reptiles. In R. J. Dooling, R. R. Fay, & A. N. Popper (Eds.), *Comparative hearing in birds and reptiles* (Vol. 13, pp. 308–359). New York: Springer.
- Dooling, R. J., Okanoya, K., & Brown, S. D. (1989). Speech perception by budgerigars (*Melopsittacus undulatus*): The voiced-voiceless distinction. *Perception & Psychophysics*, 46, 65–71.
- Doupe, A., Solis, M. M., Kimpo, R., & Boettiger, C. A. (2004). Cellular, circuit, and synaptic mechanisms in song learning. *Annals of the New York Academy of Sciences*, 116, 495–523.
- Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, 73, 2608–2611.
- Farries, M. A. (2001). The oscine song system considered in the context of the avian brain: Lessons learned from comparative neurobiology. *Brain, Behavior, and Evolution*, 58, 80–100.
- Ferreira, A. R. J., Smulders, T. V., Sameshima, K., Mello, C. V., & Jarvis, E. D. (2006). Vocalizations and associated behaviors of the sombre hummingbird (*Aphantochroa cirrhochloris*) and the rufous-breasted hermit (*Glaucis hirsutus*). *Auk*, 123, 1129–1148.
- Ficken, M. S., Rusch, K. M., Taylor, S. J., & Powers, D. R. (2000). Blue-throated hummingbird song: A pinnacle of nonoscine vocalizations. *Auk*, 117, 120–128.
- Fiete, I. R., Fee, M. S., & Seung, H. S. (2007). Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. *Journal of Neurophysiology*, 98, 2038–2057.
- Fouts, R. S., & Rigby, R. L. (1977). Man–chimpanzee communication. In T. Sebeok (Ed.), *How animals communicate* (pp. 1034–1054). Bloomington: Indiana University Press.
- Friedrich, A., Zentall, T. R., & Weisman, R. J. (2007). Absolute pitch: Frequency-range discriminations in pigeons (*Columba livia*)—comparisons with zebra finches (*Taeniopygia guttata*) and humans (*Homo sapiens*). *Journal of Comparative Psychology*, 121, 95–105.
- Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, 81, 439–454.
- Gaunt, A. S., & Gaunt, S. L. L. (1985). Electromyographic studies of the syrinx in parrots (Aves: Psittacidae). *Zoomorphology*, 105, 1–11.
- Greenewalt, C. H. (1968). *Bird song: Acoustics and physiology*. Smithsonian Institution Press: Washington, DC.
- Greenfield, P. M. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, 14, 531–595.
- Grosslight, J. H., & Zaynor, W. C. (1967). Vocal behavior of the mynah bird. In K. Salzinger & S. Salzinger (Eds.), *Research in verbal behavior and some neurophysiological implications* (pp. 5–9). New York: Academic Press.
- Hewes, G. W. (1973). Primate communication and the gestural origin of language. *Current Anthropology*, 33, 65–84.
- Hollich, G. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, 262, 1–138.
- Jarvis, E. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, 116, 749–777.

- Jarvis, E., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., et al. (2005). Avian brains and a new understanding of vertebrate evolution. *Nature Reviews Neuroscience*, 6, 151–159.
- Kao, M. H., Doupe, A. J., & Brainard, M. S. (2005). Contributions of the avian basal ganglia–forebrain circuit to real-time modulation of song. *Nature*, 433, 638–643.
- Klatt, D. H., & Stefanski, R. A. (1974). How does a mynah bird imitate human speech? *Journal of the Acoustical Society of America*, 55, 822–832.
- Kroodsma, D. E. (2005). *The singing life of birds*. New York: Houghton-Mifflin.
- Kuczaj, S. A. (1983). *Crib speech and language play*. New York: Springer-Verlag.
- Kuczaj, S. A. (1998). Is an evolutionary theory of language play possible? *Cahiers Psychologie Cognitive*, 17, 135–154.
- Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190, 69–72.
- Ladefoged, P. (1982). *A course in phonetics*. San Diego, CA: Harcourt Brace Javonovitch.
- Larsen, O. N., & Goller, F. (1999). Role of syringeal vibrations in bird vocalizations. *Proceedings of the Royal Society London B: Biological Sciences*, 266, 1609–1615.
- Leonard, L. B. (2001). Fillers across languages and language abilities. *Journal of Child Language*, 28, 257–261.
- Lieberman, A., & Mattingly, I. (1985). The motor theory of speech perception revisited. *Cognition*, 21, 1–36.
- Lieberman, P. (1984). *The biology and evolution of language*. Cambridge, MA: Harvard.
- Mann, V. A., & Foy, J. G. (2003). Phonological awareness, speech development, and letter knowledge in preschool children. *Annals of Dyslexia*, 53, 149–173.
- Marler, P. (1970). A comparative approach to vocal learning: Song development in white-crowned sparrows. *Journal of Comparative & Physiological Psychology*, 71, 1–25.
- May, D. L. (2004). *The vocal repertoire of Grey Parrots (*Psittacus erithacus*) living in the Congo Basin (Central African Republic, Cameroon)*. PhD Dissertation, University of Arizona, Tucson.
- McCowan, B., Hanser, S. F., & Doyle, L. R. (1999). Quantitative tools for comparing animal communication systems: Information theory applied to bottlenose dolphin whistle repertoires. *Animal Behaviour*, 57, 409–419.
- Miller, D. B. (1977). Two-voiced phenomenon in birds: Further evidence. *Auk*, 94, 567–572.
- Mooney, R. (2004). Synaptic mechanisms for auditory-vocal integration and the correction of vocal errors. *Annals of the New York Academy of Sciences*, 116, 476–494.
- Mowrer, O. H. (1950). *Learning theory and personality dynamics*. New York: Ronald.
- Neal, K.B. (1996). *The development of a vocalization in an African Grey parrot (*Psittacus erithacus*)*. Unpublished senior thesis, University of Arizona, Tucson.
- Nelson, K. (1989). *Narratives from the crib*. Cambridge, MA: Harvard.
- Okanoya, K., & Dooling, R. J. (1987). Hearing in passerine and psittacine birds: A comparative study of absolute and masked auditory thresholds. *Journal of Comparative Psychology*, 101, 7–15.
- Ólveczky, B., Andalman, A. S., & Fee, M. S. (2005). Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *Public Library of Science, Biology*, 3, 0902–0908.
- Patterson, D. K., & Pepperberg, I. M. (1994). A comparative study of human and parrot phonation: Acoustic and articulatory correlates of vowels. *Journal of the Acoustical Society of America*, 96, 634–648.
- Patterson, D. K., & Pepperberg, I. M. (1998). Acoustic and articulatory correlates of stop consonants in a parrot and a human subject. *Journal of the Acoustical Society of America*, 103, 2197–2215.
- Peperkamp, S. (2003). Phonological acquisition: Recent attainments and new challenges. *Language and Speech*, 46, 87–113.
- Pepperberg, I. M. (1981). Functional vocalizations by an African Grey parrot (*Psittacus erithacus*). *Zeitschrift für Tierpsychologie*, 55, 139–160.

- Pepperberg, I. M. (1988). An interactive modeling technique for acquisition of communication skills: Separation of 'labeling' and 'requesting' in a psittacine subject. *Applied Psycholinguistics*, 9, 59–76.
- Pepperberg, I. M. (1990a). An investigation into the cognitive capacities of an African Grey Parrot. In P. J. B. Slater, J. S. Rosenblatt, & C. Beer (Eds.), *Advances in the study of behavior* (pp. 357–409). New York: Academic Press.
- Pepperberg, I. M. (1990b). Cognition in an African Grey parrot (*Psittacus erithacus*): Further evidence for comprehension of categories and labels. *Journal of Comparative Psychology*, 104, 41–52.
- Pepperberg, I. M. (1990c). Referential mapping: Attaching functional significance to the innovative utterances of an African Grey parrot. *Applied Psycholinguistics*, 11, 23–44.
- Pepperberg, I. M. (1992). Proficient performance of a conjunctive, recursive task by an African Grey parrot (*Psittacus erithacus*). *Journal of Comparative Psychology*, 106, 295–305.
- Pepperberg, I. M. (1994a). Evidence for numerical competence in an African Grey parrot (*Psittacus erithacus*). *Journal of Comparative Psychology*, 108, 36–44.
- Pepperberg, I. M. (1994b). Vocal learning in African Grey parrots: Effects of social interaction. *Auk*, 111, 300–313.
- Pepperberg, I. M. (1997). Social influences on the acquisition of human-based codes in parrots and nonhuman primates. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development* (pp. 157–177). Cambridge: Cambridge University Press.
- Pepperberg, I. M. (1999). *The Alex studies*. Cambridge, MA: Harvard University Press.
- Pepperberg, I. M. (2002). Allospecific referential speech acquisition in Grey parrots (*Psittacus erithacus*): Evidence for multiple levels of avian vocal imitation. In K. Dautenhahn & C. Nehaniv (Eds.), *Imitation in animals and artifacts* (pp. 109–131). Cambridge, MA: MIT Press.
- Pepperberg, I. M. (2004). Human speech: Its learning and use by Grey parrots. In P. Marler & H. Slabbekoorn (Eds.), *Nature's music* (pp. 363–373). London: Elsevier.
- Pepperberg, I. M. (2006). Emergence of linguistic communication: An avian perspective. In C. Lyon, C. L. Nehaniv, & A. Cangelosi (Eds.), *Emergence and evolution of linguistic communication* (pp. 355–386). London: Springer-Verlag.
- Pepperberg, I. M. (2007). Grey parrots do not always “parrot”: Roles of imitation and phonological awareness in the creation of new labels from existing vocalizations. *Language Science*, 29, 1–13.
- Pepperberg, I. M. (2009). Grey parrot (*Psittacus erithacus*) vocal learning: Creation of new labels from existing vocalizations and issues of imitation. *LACUS Forum*, 34, 21–30.
- Pepperberg, I. M., Brese, K. J., & Harris, B. J. (1991). Solitary sound play during acquisition of English vocalizations by an African Grey parrot (*Psittacus erithacus*): Possible parallels with children's monologue speech. *Applied Psycholinguistics*, 12, 151–177.
- Pepperberg, I. M., Gardiner, L. I., & Luttrell, L. J. (1999). Limited contextual vocal learning in the Grey parrot: The effect of co-viewers on videotaped instruction. *Journal of Comparative Psychology*, 113, 158–172.
- Pepperberg, I. M., & McLaughlin, M. A. (1996). Effect of avian–human joint attention on allospecific vocal learning by Grey parrots. *Journal of Comparative Psychology*, 110, 286–297.
- Pepperberg, I. M., Naughton, J. R., & Banta, P. A. (1998). Allospecific vocal learning by Grey parrots (*Psittacus erithacus*): A failure of videotaped instruction under certain conditions. *Behavioural Processes*, 42, 139–158.
- Pepperberg, I. M., Sandefer, R. M., Noel, D., & Ellsworth, C. P. (2000). Vocal learning in the Grey Parrot: Effect of species identity and number of trainers. *Journal of Comparative Psychology*, 114, 371–380.
- Pepperberg, I. M., & Sherman, D. V. (2000). Proposed use of two-part interactive modeling as a means to increase functional skills in children with a variety of disabilities. *Teaching and Learning in Medicine*, 12, 213–220.

- Pepperberg, I. M., & Sherman, D. V. (2002). A two-trainer modeling system to engender social skills in children with disabilities. *International Journal of Comparative Psychology*, 15, 138–153.
- Pepperberg, I. M., & Wilkes, S. R. (2004). Lack of referential vocal learning from LCD video by Grey Parrots (*Psittacus erithacus*). *Interaction Studies*, 5, 75–97.
- Peters, A. N. (2001). Filler syllables: What is their status in emerging grammar? *Journal of Child Language*, 28, 229–242.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the identification of vowels. *Journal of the Acoustical Society of America*, 24, 175–184.
- Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Elephants are capable of vocal learning. *Nature*, 434, 455–456.
- Prather, J. F., Peters, S., Nowicki, S., & Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature*, 451, 305–310.
- Reiss, D., & McCowan, B. (1993). Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): Evidence for vocal learning. *Journal of Comparative Psychology*, 107, 301–312.
- Remez, R., Rubin, P., Nygaard, L., & Howell, W. (1987). Perceptual normalization of vowels produced by sinusoidal voices. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 40–61.
- Rice, M. L., Huston, A. C., Truglio, R., & Wright, J. (1990). Words from 'Sesame Street': Learning vocabulary while viewing. *Developmental Psychology*, 26, 421–428.
- Richards, D. G., Wolz, J. P., & Herman, L. (1984). Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin. *Journal of Comparative Psychology*, 98, 10–28.
- Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neuroscience*, 21, 188–194.
- Ryan, M. J., & Brenowitz, E. A. (1985). The role of body size, phylogeny, and ambient noise in the evolution of bird song. *American Naturalist*, 126, 87–100.
- Scanlan, J. (1988). Analysis of avian "speech": Patterns and production. PhD Dissertation, University College, London.
- Scanlan, J. (1999). The function and significance of inter-species acoustic cues in the transformation of budgerigar (*Melopsittacus undulatus*) sounds into "speech". *International Journal of Comparative Psychology*, 3, 111–152.
- Scharff, C., & White, S. A. (2004). Genetic components of vocal learning. *Annals of the New York Academy of Sciences*, 1016, 325–347.
- Silverstone, J.L. (1989). Numerical abilities in the African Grey Parrot: Sequential numerical tags. Unpublished senior honors thesis, Northwestern University, Evanston, IL.
- Studdert-Kennedy, M. (2002). Mirror neurons, vocal imitation, and the evolution of particulate speech. In M. Stamenov & V. Gallese (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 207–227). Amsterdam: John Benjamins.
- Suthers, R. A., & Zollinger, S. A. (2004). Producing song: The vocal apparatus. *Annals of the New York Academy of Sciences*, 1016, 109–129.
- ten Cate, C., Slabbekoorn, H., & Ballintijn, M. R. (2002). Bird song and male–male competition: Causes and consequences of vocal variability in the collared dove (*Streptopelia decaocto*). In P. J. B. Slater, J. S. Rosenblatt, C. T. Snowdon, & T. J. Roper (Eds.), *Advances in the study of behavior* (pp. 31–75). San Diego, CA: Academic Press.
- Todt, D. (1975). Social learning of vocal patterns and models of their applications in Grey parrots. *Zeitschrift für Tierpsychologie*, 39, 178–188.
- Treiman, R. (1995). Errors in short-term memory for speech: A developmental study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1197–1208.

- Vihman, M. (1993). Vocal motor schemes, variation, and the production–perception link. *Journal of Phonetics*, 21, 163–169.
- Vihman, M. (1996). *Phonological development: The origins of language in the child*. Malden, MA: Blackwell.
- Visalberghi, E., & Fragaszy, D. M. (1990). Do monkeys ape? In S. T. Parker & K. R. Gibson (Eds.), “Language” and intelligence in monkeys and apes (pp. 247–273). Cambridge, UK: Cambridge University Press.
- Warren, D. K., Patterson, D. K., & Pepperberg, I. M. (1996). Mechanisms of American English vowel production in a Grey parrot (*Psittacus erithacus*). *Auk*, 113, 41–58.
- Weir, R. (1962). *Language in the crib*. The Hague: Mouton.
- Weisman, R. G., Njegovan, M. G., Williams, M. T., Cohen, J. S., & Sturdy, C. B. (2004). A behavior analysis of absolute pitch: Sex, experience, and species. *Behavioural Processes*, 66, 289–307.
- Williams, H., & Nottebohm, F. (1985). Auditory responses in avian vocal motor neurons: A motor theory for song perception in birds. *Science*, 229, 279–282.
- Wright, T. F., Cortopassi, K. A., Bradbury, J. W., & Dooling, R. J. (2003). Hearing and vocalizations in the orange-fronted conure (*Aratinga canicularis*). *Journal of Comparative Psychology*, 117, 87–95.