

Lack of referential vocal learning from LCD video by Grey parrots (*Psittacus erithacus*)*

Irene M. Pepperberg and Steven R. Wilkes

Massachusetts Institute of Technology, Brandeis University /
Brandeis University

Grey parrots (*Psittacus erithacus*) do not acquire referential English labels when tutored with videotapes displayed on CRT screens if (a) socially isolated, (b) reward for attempted labels is possible, (c) trainers direct birds' attention to the monitor, (d) live video feed avoids habituation or (e) one trainer repeats labels produced on video and rewards label attempts. Because birds learned referential labels from live tutor pairs in concurrent sessions, we concluded that video failed because input lacked live social interaction and modeling (Pepperberg, 1999). Recent studies (e.g. Ikebuchi & Okanoya, 1999), however, suggest that standard CRT monitor flickering could instead have prevented learning. Using an LCD monitor, we found that eliminating flickering did not enable birds to learn from video under conditions of limited social interaction. Results emphasize the role of social interaction in referential label learning and may generalize to other systems (e.g. disabled children, or possibly software and robotic agents).

Keywords: video learning, social interaction, animal language

Grey parrots (*Psittacus erithacus*), unlike normal young children, seem unable to learn referential English labels from video, even if such videos portray modeling of referential, functional use of targeted labels by two trainers and another parrot, and a co-viewer directs the birds' attention appropriately (Pepperberg, 1999; Pepperberg, Gardiner & Luttrell, 1999). The birds' failures have been attributed to an absence of live, interactive modeling that adjusts to their behavior (Pepperberg et al., 1999), but recent songbird studies (e.g. Ikebuchi & Okanoya, 1999) suggest that failure may instead be due to the birds' inability to process the physical input when standard cathode-ray tubes (CRTs), with

screen refresh rates set to match the flicker-fusion of human vision, are used as monitors. We thus tested whether CRT flickering indeed prevented learning from video; the information may generalize to other systems (e.g., teaching disabled children, possibly software and robotic agents; Pepperberg, 2003).

To clarify the issues involved, we briefly review earlier work on Grey parrot vocal learning. In previous studies, Greys learned referential English speech (e.g. to comprehend and produce labels for objects, colors, shapes; to answer questions about concepts of number, category, relative size, same/different) through Model/Rival (M/R) training (Pepperberg, 1999). This procedure, introduced by Todt (1975) and adapted by Pepperberg (1981), involves three-way interactions between two human speakers and an avian student. The two humans demonstrate a label's referentiality (i.e. the connection between the label and the object to which it refers), model a label's functionality (i.e. demonstrate label use, as in a request), attend to both bird and object during training, exchange roles of trainer and model, involve the parrot in interactions, and adjust training to the bird's level (Pepperberg, 1981, 1994; Pepperberg & McLaughlin, 1996; Pepperberg, Naughton & Banta, 1998; Pepperberg et al., 1999, 2000). Video input lacks these combined qualities, and birds at most learn to produce nonreferential labels when exposed to videos of M/R training, in which functionality and reference are demonstrated but modeled social interaction is not live. Learning fails even when a trainer is present to direct birds' attention to the video, when reward is possible, and when likelihood of habituation to the tape is avoided (Pepperberg et al., 1998, 1999). Moreover, a bird taught by one trainer who demonstrates reference and functionality (i.e. only omitting human-human modeling) stores the label but does not understand when or how to use it (Pepperberg et al., 2000). Such data suggested that modeled, live, referential social interactions are crucial for Grey parrots' referential vocal learning of another species' communication code (i.e. allo-specific learning). If, however, video tutoring fails merely because of CRT flicker, then parrots might, like normal children, learn human speech under less stringent social conditions.

Video input for avian subjects has had varying success depending on an investigation's focus and aims (review in Pepperberg et al., 1999) and flicker-fusion is only one potential issue of concern (Oliveira et al., 2000). At least four other possible factors might explain why the birds did not learn referential labels in CRT video sessions (Pepperberg et al., 1999). First, standard video monitors might distort audio output, thus providing inadequate information to process (Adret, 1997); using appropriate speakers easily corrects this problem

(Pepperberg, 1994). Second, differences in avian and primate visual systems — for example, that birds likely perceive colors differently than primates because most birds have oil droplets in their cones and see in the ultraviolet spectrum (e.g. Bennett & Cuthill, 1994; Bowmaker et al., 1996; Varela, Palacios, & Goldsmith, 1993) — might affect a parrot's ability to perceive video images as valid representations of reality; however, parrots, like humans, respond appropriately to video despite possible color distortion (Rutledge & Pepperberg, 1988). Other perceptual issues (e.g. focal length and image size) might also be important (Adret, 1997; Dawkins & Woodington, 1997; D'Eath & Dawkins, 1996; Patterson-Kane et al., 1997; review in Pepperberg et al., 1999) but, again, at least one parrot responded appropriately to questions posed via video about physical objects (Rutledge & Pepperberg, 1988), suggesting that some correlation was made between the video and the physical images. Third, for birds, as for children, visual aspects of video may be more salient and memorable than audio aspects, thereby distracting attention from sounds to be learned (Pezdek & Stevens, 1984; Watkins et al., 1980). This issue might explain why birds, like some children, appropriately respond to, but fail to learn, auditory information provided via video (Pepperberg, 1999). Nevertheless, Grey parrots occasionally attempt nonreferential production of labels introduced via video if a trainer, without uttering labels in sessions, directs the birds' attention to the video (Pepperberg et al., 1998). Such nonreferential learning must be due to video exposure, suggesting that some information processing occurs, but not enough for reference. Fourth, birds may also be confused by the physical separation between the presentation of visual and audio signals on video (i.e. sound emanating from boxes abutting the screen rather than from speakers' mouths; Radeau, 1994). This confusion might cause birds to ignore and thus fail to learn connections between auditory signals and the object or action (Pepperberg et al., 1999). Whereas issues two through four can be somewhat discounted, although not completely resolved at present, that of flicker-fusion can be examined fairly easily and potentially be eliminated.

Rather than viewed as an integrated whole, CRT output may be seen as an interrupted collection of flashing images that might distract the bird's attention, render difficult the observation of physical objects to be labeled, or disrupt formation of a connection between audio and visual input; male songbirds responded to videos of females only when liquid crystal display (LCD) monitors were used (Ikebuchi & Okanoya, 1999). Although Greys, as noted above, respond to video (i.e. correctly answer questions about objects presented via video, Rutledge & Pepperberg, 1988), the overall effect could prevent learning

of targeted labels. According to Adret (1997), simply testing the Grey parrot flicker-fusion frequency is insufficient because the decay of the video scan lines is not the same as that for the high-contrast patterns used to measure flicker-fusion frequency, and measurements of the latter also depend on background conditions during testing, which in turn may differ from what is experienced during playback. Although examination of flicker-fusion using standard procedures for one psittacid, the budgerigar (*Melopsittacus undulatus*), suggested its rate was similar to that of humans (Ginsburg & Nilsson, 1971), Ikebuchi & Okanoya's (1999) data on what birds likely see on CRT screens demonstrate the validity of Adret's arguments. Thus, to test whether elimination of flickering enables Greys to learn from video under conditions of limited social interaction, we used an LCD monitor to replicate a previous study (Pepperberg et al., 1998).

Methods

Subjects and Housing

The subjects were two male Grey parrots (*Psittacus erithacus*). Griffin, 7½ years old when the study began, had been obtained from a breeder at 7½ weeks, was hand-raised in our lab, and had participated in several previous CRT video studies (Pepperberg et al., 1998, 1999). Arthur (a.k.a. Wart), approximately 3 years old when the study began, had been obtained when a year old from a breeder. He had had no prior video training on labels, but previous projects had required him to view an LCD monitor (Pepperberg, 2000). Training and testing occurred with birds on "T" stands. The birds' living conditions outside of testing and training were described previously (e.g. Pepperberg et al., 1998), except now the birds shared a room at all times.

A 25-year-old Grey, Alex, acted as a subject in the videos. He already routinely labeled numerous objects and categories referentially (Pepperberg, 1999). His housing conditions matched those of the other birds, including sharing their room.

Training

Videotape presentations

The procedure matched that described in Pepperberg et al., 1998. Each tape lasted approximately 25 minutes, closely resembling that of live M/R sessions

(~15–30 min/45-min session; birds are allowed to interrupt for food, water, etc.) and included multiple interactions among Alex and trainers. A video for each targeted label was shown to each bird 50 times on the LCD; the number corresponded to that used in earlier experiments and was approximately twice that needed for learning labels via live M/R sessions (Pepperberg, 1999). Also as in live M/R training, video sessions occurred 2–6 days/week, alternating between mornings and afternoons, when possible, to balance potential circadian effects. A “T” stand was placed in front of the LCD so as to replicate the average distance between bird and object during live M/R sessions; zoomed images (~ $\frac{1}{3}$ of the video), but not the full scenario, were roughly life-sized at this distance. Alex previously responded to objects presented via a live video link at this distance (Rutledge & Pepperberg, 1988). Audio speakers were placed as close as physically possible to the LCD screen. One trainer pointed to the screen during and after taped interactions involving targeted objects, making comments like, “Look what Alex has!” (Pepperberg, 1996), but neither repeated targeted labels nor related content to other training sessions. A bird’s attempts at the targeted label would be rewarded with vocal praise, not the object (Pepperberg, 1996). We intentionally limited trainer-bird interactions and eliminated live reference and functionality. Our goal was not to establish optimal conditions for video training, but instead to determine whether a bird could learn labels from video without live referential, functional input if CRT flickering were eliminated. The trainer recorded all relevant behavior in a notebook.

The two birds’ training periods did not overlap. Onset date for each set of trials varied because of trainers’ schedules and our decision to limit the number of trainers. For the duration of all trials and testing a sign was posted outside the laboratory requesting persons entering not to use *pen* or *chalk* (the targeted labels) while inside.

The videotapes provided input that closely resembled M/R format but avoided live social interaction. We videotaped Alex during a standard M/R review session for *chalk* and a novel M/R session for *pen* (i.e. Alex knew the label *chalk* but not *pen*; rationale below). Thus tapes proceed as follows: As in standard M/R sessions, Alex observes two humans handling a targeted object. One human (the trainer) queries the other about the item; this second human is a model for the behavior and the parrot’s rival for the trainer’s attention. The trainer rewards the model/rival’s correct identification with the object itself, providing one-to-one correspondence between label and object, thereby demonstrating referential, functional label use. The model/rival also illustrates

aversive consequences of errors. When s/he intentionally gives an incorrect or unclear answer, the trainer responds by chiding and temporarily hiding the object from view; the model/rival is told to speak more clearly or try again, thereby allowing Alex to observe “corrective feedback” (Goldstein, 1984; Vanayan, Robertson, & Biederman, 1985). Unlike other modeling procedures (e.g. Todt, 1975; review in Pepperberg & Sherman, 2000), the M/R protocol then requires repeating an interaction while reversing the roles of human trainer and model/rival to demonstrate that the communicative process can be used by either party to request information or effect environmental change. Alex is included in these interactions; he is rewarded for correct responses or approximations and chided for errors. Occasionally, after correctly answering, Alex discards the targeted object and requests something else (“Want X”). In such cases, we positively reinforced this behavior with X because such requests are referential. During taping, Alex never queried the humans, so his role was at most that of another model/rival; previous experiments showed his addition in this capacity, in live M/R training sessions, did not affect Griffin’s speed or accuracy of learning (Pepperberg et al., 2000).

Pen video

This video was created in Spring, 2001 at the MIT Media Lab. Pen exemplars were Bic™ ball-point pens with ink and ball point removed. Alex was trained exactly as in non-taped M/R sessions. He never produced “pen” during this single session and thus was never rewarded. Two student trainers behaved as in standard M/R sessions (Pepperberg, 1981); that is, by demonstrating correct responses and receiving the object, or erring and being told “No, you’re wrong”, “Say better”, “Talk clearly”, or “Try again” to retain an approximate 80% correct/20% error mix overall. Such procedures allowed birds viewing the video to observe which utterances generated rewards versus chiding. This tape began with an exact correlation between trainers’ vocal actions and produced sounds, but midway through started to exhibit a minor delay that, by the end of the session, was about one second. This delay allowed us to determine whether birds were influenced by a separation between physical movements of trainers’ faces and sounds heard. Because half the video did not involve this delay, birds still had the same amount of input, approximately that received in 25 sessions (half of 50 sessions), that enabled learning in standard M/R training; learning, however, might be less efficient because of the delay.

Chalk video

The video was created at Brandeis University in January, 2003. Again, Alex was trained as in non-taped M/R sessions. Exemplars were variously colored pieces of chalkboard chalk, 2–5 cm long. For *chalk*, Alex's accuracy was > 80%, but he occasionally erred or requested other items and was treated as described above. Trainers responded as in the *pen* video, creating an 80%/20% mix overall of correct responses and errors. Now sounds produced by trainers exactly matched corresponding vocal actions throughout.

M/R sessions

Griffin had acquired several M/R-trained labels prior to this study (Pepperberg et al., 1999); Arthur had received the same live, interactive M/R training on *spool*, *wool*, *wood*, and *spoon*. He had not yet been tested on those labels when video training began, and thus formal testing of those labels was included in this experiment.

Choice of labels trained

We chose English labels *pen* and *chalk* for six reasons. First, we normally use labels in video sessions that Alex produces with good fidelity to ensure that sounds involved are within a Grey's vocal capacity. We thus chose *chalk*. We also chose *pen*, which Alex had not yet learned, to determine if Alex's correct responses on video might somehow inhibit vocal learning. Although this possibility may seem odd, its roots are both in some earlier data and in parrots' natural behavior. In a previous study, Griffin had learned more slowly when Alex substituted for a human in M/R sessions (Pepperberg et al., 2000). We proposed several reasons for such results, including those involving Alex's error rate and his inability to query Griffin, but did not consider a possibility derived from parrots' natural behavior. Wild, mated Grey pairs engage in specific duets with one another (May, personal communication, August, 1999); other birds in the flock do not often use the same vocalization patterns in their interactions. We wondered if Griffin viewed Alex's interactions with us as part of a pairbond duet, and thus had to overcome natural inhibitions to learn an utterance pattern already in use by Alex with a trainer. This possibility seemed unlikely because Griffin learned referential labels equally well whether Alex was or was not an *additional* participant in live M/R training, and had acquired labels for some preferred foods by observing Alex's interactions with students outside of sessions (Pepperberg et al., 1999); nevertheless Alex's competence might be an

issue in video training. Hence the choice of *pen*. If birds learned *pen* but not *chalk*, Alex's competence could be a factor. Alex's incompetence would likely be unimportant, because trainers would provide correct responses exactly as in regular M/R sessions. If neither bird learned *pen*, but either learned *chalk*, the evolving disassociation between sound/facial movement in the video might have interfered with learning.¹ Second, Griffin had never received *pen* training but had failed to learn *chalk* in other video conditions; if he learned *chalk* but not *pen*, and Arthur failed to learn either label, we could suggest that Griffin acquired some latent learning in previous sessions but, as in single-trainer sessions, he did not understand what to do with the label (Pepperberg et al., 2000) — that is, that LCD video could not train referential labeling *de novo*, but could enable latent learning to emerge. Third, to ensure that Alex's casual speech would not affect video training, we chose labels Alex either did not produce outside his own sessions (*chalk*) or that he didn't know (*pen*). Fourth, we wanted to test Arthur on one label previously given to Griffin via video to ensure that nothing specific about the label or object to which it referred existed that might either enhance or deter video learning, and that only LCD use differed between studies. Fifth, if Arthur, who had no video training experience, but who had already acquired referential English vocalizations via M/R sessions, learned any labels and Griffin failed, we could argue that Griffin's previously unsuccessful video experiences may have blocked his ability to learn even with an LCD. Sixth, if neither bird learned any label referentially, we could argue that eliminating flickering did not enable birds to learn from video under conditions of limited social interaction.

Note that because Alex more frequently practiced and readily acquired labels for favored items in contexts in which a label was initially used as a request (Pepperberg, Brese, & Harris, 1991), we eliminated motivational factors by training labels that referred to items with which birds had interacted in free-play with approximately equal frequency (e.g., for 3–5 consecutive min in sessions over several wks); moreover, once training began, these objects were available only during sessions to maintain motivation to obtain them. This pre-exposure also ruled out the possibility that birds would not recognize test items as those on the video because of the items' novelty.

Equipment

The *pen* session was recorded on a Sony Digital camera (model DCR-TRV820 NTSC) onto a Hi8 ME 8mm cassette; the *chalk* session was recorded on a Sony

Digital camera (model DCR-TRV18) with the VHS setting onto a TDK VHS tape. We used a zoom lens to include life-size images of Alex and the targeted object in addition to somewhat smaller images of an entire training scenario (the object, Alex, and two humans). For consistency with previous studies (e.g., Pepperberg et al., 1998), life-sized images constituted roughly one-third of each video. Viewing quality did not differ between tapes.

Playbacks were on an 18" ViewSonic model VP181 high resolution LCD monitor, set to display at 1024 by 768 pixels. This system had, at purchase, one of the highest screen performance ratings (e.g., lowest amount of ghosting) out of the commercially available monitors. Video height-width were adjusted to set screen images as close to life-size as possible. Again, for consistency with previous studies, we wished to replicate input used in human studies; thus we did not purchase a monitor that would show the entire training scenario as life-sized.

Testing procedures

To evaluate what parrots learned from video, Irene Pepperberg, assisted by trainers who had not participated in video sessions involving targeted objects, administered tests one-to-four times/week for each label. Specific details of testing protocol, including descriptions of precautions against inadvertent and expectation cuing, have previously been published (Pepperberg 1981, 1990, 1994; Pepperberg et al., 1998, 1999, 2000).

Criterion prior to testing

The criterion for initiating testing after training is generally based on a bird's ability to emit a label, not the effect of our procedures on the bird's ability to associate label and referent (Pepperberg, 1981). For testing to begin, targeted vocalizations merely must be recognized by trainers in blind trials with ~90% interobserver agreement. As in previous video experiments, however, the criterion for initiating testing was completing 50 training sessions.

Correction procedure

During testing, if a bird's initial response is incorrect or indistinct, examiners present a misnamed object a total of three times or until an identification is correct; and errors are recorded. Immediate re-presentations occur only if birds err; the protocol thus penalizes a "win-stay" strategy.

Scoring procedure

We report “first” and “all trials” scores for targeted labels. First trial data are the percentage of correct first responses (the initial time a question is posed in a trial; that is, before any correction procedures are allowed) and are used for statistical analyses. For comparison, we report all trials’ scores: total number of correct identifications divided by total number of presentations required. If a bird cannot produce a correct response, we report details of only first trial data (i.e. the misidentifications produced).

Results

Neither Griffin nor Arthur learned referential use of *pen* or *chalk* in LCD sessions; they did, however, referentially label objects presented in prior M/R training. In video sessions, they most frequently produced non-targeted human utterances or parrot sounds (e.g. clucks, squawks, chirps). Griffin made at least one clear attempt at each targeted label; Arthur did not. Trainers’ notes provided details of session activities.

Pen video

Griffin

LCD video training was unsuccessful. In most sessions, Griffin did not attempt *pen* but produced numerous human vocalizations (Figure 1), most often *cork*, *shower* and *go back*. He frequently ignored the video, opting to preen, whistle, flap his wings, or stare at the human trainer. He often turned around in mid-session, stretching and flapping his wings in the direction of his cage, signaling a desire to return there; the human trainer would reposition Griffin and draw his attention back to the monitor. In sessions 6, 7 and 39, video was prematurely stopped because of Griffin’s excessive inattention and movement on the perch. Overall vocalizing decreased nonsignificantly ($p=0.09$, $r=-0.24$) as sessions progressed and, except for session 41, vocalizations were infrequent during the last 19 sessions. On four separate occasions (sessions 17, 25, 30, 35), Griffin produced utterances directly after hearing Alex speak in the video or responded with previously acquired labels when Alex was asked “What toy?” or “What matter?” Such behavior showed he correctly processed at least the audio portion of the video. Griffin possibly made one *pen* attempt during session 24, voicing something “en”-like.

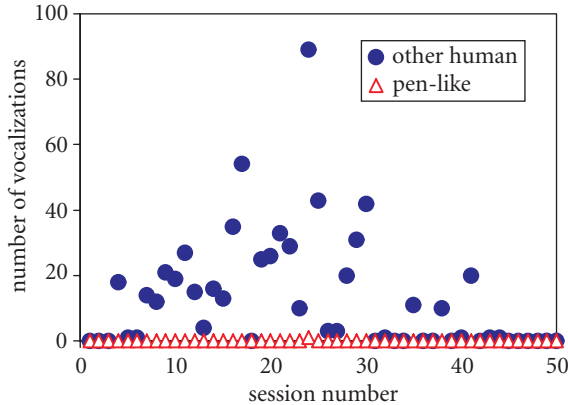


Figure 1. Griffin’s vocalizations during “pen” sessions.

Griffin’s identification tests confirmed his failure to learn *pen* from LCD sessions. He scored 0%; 0/40 first-trial scores, 0/120 overall. As a control within the same tests, he was briefly queried on previously acquired M/R-trained labels (Pepperberg et al., 1999), *paper*, *key*, *wool*, *nail*, *cork*, *rock*, *ring*, and *wood*; on these he scored >96% on first trials overall ($\geq 75\%$ for each M/R label, binomial test, $p < .01$, probability 1/8 considering eight possible answers). Table 1 gives a breakdown of errors.

Arthur

Arthur also failed to learn *pen* in LCD sessions (Figure 2). During sessions, his most frequent human utterances were *wool*, *sss* (/s/), and *ooo* (/u/); M/R training had begun on phonemes /s/ and /u/, for which he received plastic letters he would chew (thus establishing some level of reference). He showed the same attentional problems as had Griffin, only to a greater degree. Rather than watch the screen, he preened excessively or requested tickles from the human trainer; this human co-viewer constantly repositioned Arthur, attempting to focus his attention on the screen. Arthur emitted parrot calls throughout many sessions; overall, he produced far fewer human vocalizations than Griffin. During sessions 33, 34, 40, and 49, he produced well above his average numbers of human utterances ($n=8, 15, 11, 12$, respectively; overall $M=2.3$, $SD=3.5$), but most were requests for tickles and irrelevant to the video. These utterances caused his overall vocalization rate to increase slightly but significantly ($p=0.04$, $r=0.29$) over the course of training.

Once each in sessions 5 and 30, Arthur responded to “What toy?”, asked of

Table 1. Pen Identification Tests

Object	Test Score	Erroneous identifications (no. errors)
<i>Griffin</i>		
Pen	0/40	ring (18), key (6), cork (4), sss (4), nail (2), paper (2), rock (2), orange (1), green (1)
Paper	8/8	
Nail	8/8	
Wood	8/8	
Wool	8/8	
Rock	8/8	
Key	6/8	nail (2)
Cork	8/8	
Ring	8/8	
<i>Arthur</i>		
Pen	0/40	wood (21), wool (11), spool (7), /u/ (1)
Spool	35/40	wool (4), wood (1)
Wood	39/40	spool (1)
Wool	35/40	wood (4), spool (1)

NB: Griffin was tested on non-*pen* labels only as a control; he previously demonstrated competence on these labels (Pepperberg et al., 1999); these were Arthur's first tests on M/R labels and thus he had the full 40 trials standard for our procedure (Pepperberg, 1999).

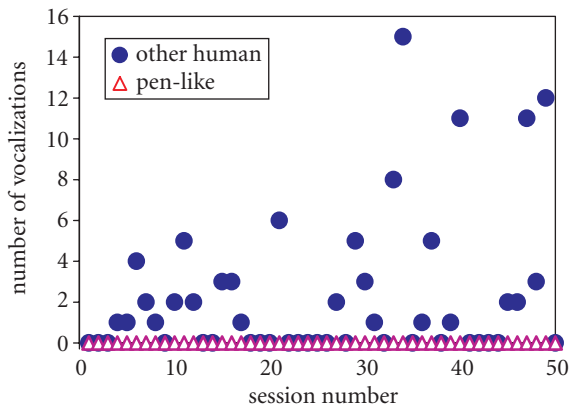


Figure 2. Arthur's vocalizations during "pen" sessions (note difference in scale from Griffin).

Alex by a human trainer in the video. Arthur answered "sss" and "wool", respectively. Although these responses were incorrect, his behavior revealed that, despite lack of interest in the video, he, like Griffin, demonstrated some

understanding of the observed procedure.

Arthur's tests confirmed that he, too, failed to learn *pen* from LCD training. For *pen*, he scored 0%; 0/40 first-trial scores, 0/120 overall. As a control within the same tests, he was queried on M/R-trained labels, *wool*, *wood*, and *spool*; on these he scored >90% on first trials overall (>80% on first trials for each M/R label, binomial test, $p < .01$, probability $\frac{1}{3}$ considering three possible answers); Table 1 gives a breakdown of errors.

Chalk video

Griffin

Griffin did not learn referential use of *chalk*, but did produce the label once outside of sessions. Attempts during sessions were initially mumbled utterances starting with *ch-*, but developed into something resembling *chalk* in sessions 44 and 48 (Figure 3). This behavior contrasted with his attempts in previous video studies, where he began producing *alk*-like ("awk") sounds toward the end of training (Pepperberg et al., 1999). His attention span and actions during this video were similar to those in *pen* sessions. Griffin generally attended more to the live human trainer than the video, or ignored the video and whistled or preened; he would grind his beak, request random items, or lunge toward his cage. Griffin seemed to attempt *chalk* in sessions where trainers subjectively felt he attended more closely to the LCD monitor (numbers 3, 11, 14, 15, 17, 24, 32, 37, 40, 44, 48). The frequency of these vocalizations, however, was very low and stayed constant across sessions ($p=0.6$, $r=0.08$). Griffin's utterances overall increased non-significantly during the study ($p=0.17$, $r=0.20$). We separately plotted (as *chalk*-related, Figure 3) vocalizations like *cork* and *chair* that have phonemes in common with *chalk*; few were produced (14.7% of total utterances) and their numbers did not increase as the experiment progressed ($p=0.45$, $r=0.11$). In a previous video study, *chalk*-related utterances (*cork*, *rock*) comprised 41.6% of Griffin's vocalizations in sessions (Pepperberg et al., 1999).

Results of Griffin's tests showed that despite some *chalk*-like utterances, he failed to use *chalk* referentially after LCD training. For *chalk*, he scored 0%; 0/40 first-trials, 0/120 overall. Again, as a control within the same tests, he was briefly queried on M/R-trained labels (Pepperberg et al., 1999), *paper*, *key*, *wool*, *nail*, *cork*, *rock*, *ring*, and *wood*; on these he scored 98% on first trials overall ($\geq 85\%$ for each label, binomial test, $p < .01$, probability $\frac{1}{8}$ considering eight possible answers). Table 2 gives a breakdown of errors.

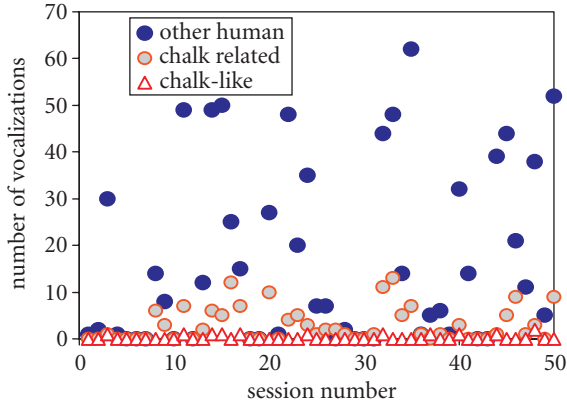


Figure 3. Griffin’s vocalizations during “chalk” sessions (note difference in scale from “pen”).

Table 2. Chalk Identification Test

Object	Test Score	Erroneous identifications (no. errors)
<i>Griffin</i>		
Chalk	0/40	rock (20), cork (19), ring (1)
Paper	8/8	
Nail	8/8	
Wood	8/8	
Wool	8/8	
Rock	8/8	
Key	8/8	
Cork	8/8	
Ring	7/8	ss (1)
<i>Arthur</i>		
Chalk	0/40	wood (28), wool (8), spool (4)
Spool	14/16	wood (2)
Wood	16/16	
Wool	14/16	wood (2)
Spoon	35/40	wood (3), wool (2)

NB: Arthur was retested more often than Griffin on previously acquired labels to control for expectation cuing.

Arthur

Arthur also failed to learn *chalk*. His behavior was the same as during the *pen* video. His attention strayed from the monitor; he continually turned around, preened, and requested tickles. He did not attempt anything *chalk-like*; overall

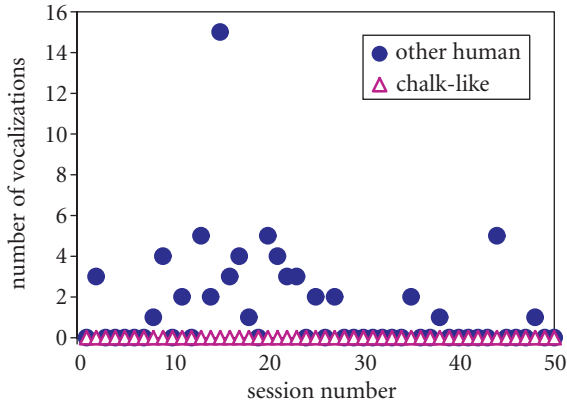


Figure 4. Arthur’s vocalizations during “chalk” sessions (note differences in scale from Griffin).

his vocalizations decreased nonsignificantly, $p=0.11$, $r=-0.23$ (Figure 4).

Arthur’s *chalk* tests confirmed his failure to learn the label. For *chalk*, he scored 0%; 0/40 first-trials, 0/120 overall. As a control within these tests, he was again queried on M/R-trained labels, *wool*, *wood*, *spool* and now *spoon*. Fewer trials were given on the first three labels because he had demonstrated their referential use in *pen* tests, but more trials were given to Arthur than Griffin on each familiar label so as to avoid expectation cuing for *spoon* and *chalk*. He scored >89% on first trials for all M/R labels ($\geq 85\%$ on first trials for each M/R label, binomial test, $p < .01$ for all M/R labels, probability $\frac{1}{4}$ given four possible answers); Table 2 gives a breakdown of errors.

Discussion

We tested whether a parrot could learn labels from video without live, referential, functional input if CRT flickering was eliminated. Flicker may contribute to parrots’ failures to learn referential labeling from video (Ikebuchi & Okanoya, 1999), but is not the critical factor. Arthur did not even attempt targeted labels; although Griffin produced “chalk”-like utterances during training and “chalk” once after training, he did not transfer labeling behavior to formal testing. In contrast, M/R-trained labels were used referentially in formal tests. Although we have not eliminated all nonsocial reasons for failure to learn from video, we maintain that the video’s lack of live, interactive modeling is

likely the most important factor. Detailed discussions of the social effects on learning have been presented previously (e.g. Pepperberg, 1985, 1998; Pepperberg et al., 1998, 1999, 2000); we briefly review that material and present an additional reason to support our contention.

The need for social interaction for vocal learning is likely based in parrots' natural behavior. Greys are social animals; they probably learn not only vocalizations but also other critical behavior patterns from one another (Todt, 1975). For wild parrots, social interaction is likely crucial not only for determining what should be learned (e.g. allospecific alarm calls versus sounds of a stream) and for assessing the nature of interactions (e.g. aggressive versus affiliative), but also for targeting input as worthy of processing (e.g. what is gained by learning; Pepperberg, 1998). In the laboratory, they adapt these behavior patterns, learning through interaction with humans. Social interaction thus highlights various components of input, that is, signals which environmental components should be noted. Social interaction may also emphasize common attributes, and thus possible underlying rules, of diverse actions and, possibly most importantly, allows input to be continually adjusted to the learner's level (Pepperberg, 1985). Even with live human trainers to designate which video aspects should be noted, birds often do not attend to video, preferring to interact directly with humans (e.g. requesting tickles, Pepperberg et al., 1998, 1999). Removing human interaction to avoid such distraction does not, however, improve video learning (Pepperberg, 1994), nor does presentation of live video to avoid habituation to a single tape (Pepperberg et al., 1999). Given that material is learned best if input is finely tuned to a learner's level (and for allospecific learning, the more attuned must be the input; e.g. Bandura, 1977), and that video alone cannot make such adjustment (Schramm, 1961; Stoneman & Brody, 1982), birds may, like some children (Lorch, Anderson, & Levin, 1979), withdraw attention and thus fail to learn because they simply do not process video to learn vocal labels. Griffin's more frequent human vocalizations during sessions compared to Arthur were likely due to Griffin's larger vocabulary and more years of training, not individual differences between birds. The *pen* video disassociation of timing for speech and speakers' utterances likely had no effect over and above that of being part of the video presentation, given that neither Griffin nor Arthur learned from the *chalk* video without such dissociation. Nor did Alex's labeling competence or lack thereof apparently affect outcomes. Interestingly, recent studies (Goldstein, King & West, 2003) show that for both young birds and humans, even nonvocal social interaction (e.g. attention to changing behavior patterns) is necessary for normal communicative competence

to develop, and previous studies in our laboratory (Pepperberg, 1981, 1994, 1999) have demonstrated the effectiveness of full social interaction training.

Although birds did not always attend to the video, concentration was not completely absent. The birds occasionally responded to questions posed to Alex on video, demonstrating that they could interact with the video. Alex's prior work (Rutledge & Pepperberg, 1988) has provided additional data that birds perceive objects presented via video as representations of reality. Griffin's *chalk* attempts were made when he seemingly attended more closely to the video; even so, he didn't transfer knowledge from training to testing, which suggests that amount of attention is not a crucial factor.

Interestingly, Griffin failed to learn *chalk* referentially despite previous video training on *chalk* (both taped and live video with a co-viewer, the latter to see if he had habituated to single tapes; Pepperberg et al., 1999). Possibly, nonreferential acquisition of *chalk*, unlike his complete failure for *pen*, was a consequence of cumulative training. Nevertheless, his acquisition was non-referential, reinforcing our conclusion that live M/R training is necessary to teach referential allospecific communication.

Arguably, Griffin might have acquired nonreferential "chalk" from Alex, caged nearby. We had, however chosen *chalk* specifically because we had not heard Alex produce it outside his sessions (see above). Moreover, even if Alex did utter the label during evening babbling after trainers departed (note Pepperberg et al., 1991), the label would have been produced out-of-context (no chalk being present) and Griffin's nonreferential acquisition would again reinforce our conclusion of the need for all aspects of M/R training for referential acquisition. Moreover, Arthur, also caged near Alex, would have had the same exposure; he, however, failed to acquire "chalk", even nonreferentially. In contrast, Arthur did referentially learn *spool* and *spoon* in M/R sessions, neither of which were in Alex's repertoire at the time. Finally, Griffin and Arthur acquired their M/R labels before birds were housed together so acquisition could not have been from Alex.

An additional issue, mostly correlational but intriguing, involves "mirror neurons" (MNs, e.g. Fadiga et al., 1995). MNs are part of the nervous system that responds to an observed action just as though the action were actively being executed by the observer (see Arbib & Rizzolatti, 1996). Human MNs may be involved in imitative learning, possibly including speech (Fadiga et al., 2002). The idea is that observation of an act (one that is goal-directed in monkeys, not necessarily so in humans) entrains the nervous system for its production, and execution then becomes a matter of maturation and practice

(Arbib, 2002). Interestingly, Jaervelaeinen and colleagues (2001) found stronger reactivity of MN-related brain areas during observation of live rather than video motor acts. Jellema et al. (2000) show that the agent of the observed action must attend to the action for the observer's MNs to respond appropriately, suggesting a role for joint attention (Pepperberg & McLaughlin, 1996) and maybe a need for live demonstrations; note possible connections with data on nonvocal social interaction (Goldstein et al., 2003). Data for avian MN systems are circumstantial (review in Pepperberg, in press), but Jaervelaeinen et al.'s data are consistent with our findings that video modeling is less successful than that presented live.

Our work may have implications for intervention programs for children with various learning disabilities and in robotics and computer learning. Non-interactive video is likely adequate for issues involving nonexceptional behavior, but not exceptional behavior (learning unlikely during normal development, but possible under certain conditions; e.g. allospecific vocal learning, Pepperberg, 1985, 1993, 1997). In conjunction with previous studies on the acquisition of communicative competence, particularly exceptional forms, our current data support Bandura's (1977) correlation in humans between the extent of inhibition toward learning (whatever the cause) and the extent to which input must be referential, functional, and involve full social interaction. Subjects may respond appropriately to, but not learn exceptional communicative behavior from, video. Some exceptional learning can, of course, occur when input is not maximally referential, functional, or interactive. Children handicapped in a variety of ways acquire some language competence under impoverished conditions (e.g. Goldin-Meadow, 1997; Rice, 1991; Watkins, Sprafkin, & Krolikowski, 1990), humans learn some aspects of second languages in non-interactive settings (e.g. Secules, Herron, & Tomasello, 1992), and parrots (and other animals) acquire allospecific production without comprehension under such conditions (Pepperberg, 1997, 1998), but subjects learn best when training is fully interactive (see Pepperberg et al., 1999; Pepperberg & Sherman, 2000, 2002). Previous work along these lines has engendered interest from researchers involved in computer and robotics learning (e.g. could MNs be mathematically modeled for computers and trained via M/R procedures? See Pepperberg, 2001, 2003); our data might encourage more studies on interactive learning in these fields.

In sum, although LCDs eliminated one potential source of distraction that may have prevented video learning by parrots, the birds still did not learn labels through video presentations. As noted previously (e.g. Pepperberg et al., 1999), we still may find that parrots could learn conspecific rather than allospecific vocalizations via video or learn referential English labels if videos were fully

interactive (i.e. changed appropriately in response to birds' actions such as in a conference video situation), or if birds controlled the extent and amount of what they see on video (Pepperberg, 2000). Our present data, however, suggest that psittacids learn referential English labels more readily, more extensively, and with full comprehension as well as production when input is not impoverished.

Notes

* This study was supported by the MIT Media Lab, the American Foundation, and donors to *The Alex Foundation* (particularly the Pearl Family Foundation, Eleanor Lloyd Dees Foundation, South Bay Bird Society, Rainbow Feathers Bird Club, Midwest Avian Seminar, Midwest Avian Research Expo, Susan and Michael Harmon, the Fowler family, Corrine and Rick Unger, Birds&More, Lauren Mitchell, Bill Broach and Nancy Clark, University of Illinois Veterinary School, the Baker family, Kenosha Exotic Bird Club, the Dioletis family, Feathers Bird Store, Nalani Toys, West Suburban Cage Bird Club, River City Bird Club, Imperial Bird Club, the Ingram family, Lee and Nancy Wright, the Simmons family, the Medwick Foundation, the Andrew De Mar Family Foundation, the Dulski family, Diane Grindol, and Arlene Levin). We thank Spencer Lynn and unidentified reviewers for comments on the manuscript, Terry Clyne of Apalachee River Aviary for donating Griffin, Kim Gaudette for handraising Arthur, the Harrison family for Harrison's Bird Diet, Fowl Play and Corky's Organics for treats, Carol D'Arezzo for Griffin's and Arthur's perches, and Avian Adventures for cages. Betsy Lindsay, Vincent Capanelli, and Susan Abrams assisted with testing. Kerry Jordan, Arlene Levin, and undergraduates at both MIT and Brandeis assisted with video training.

1. Alex's lack of interest in the pen, unlike Griffin's and Arthur's playful interactions in pre-session tests, meant we could not train him in M/R sessions to determine if the label might be difficult to acquire. All phonemes involved were, however, already in Alex's repertoire, so pronunciation problems would not likely be the case (Pepperberg, 1999).

References

- Adret, P. (1997). Discrimination of video images by zebra finches (*Taeniopygia guttata*): Direct evidence from song performance. *Journal of Comparative Psychology*, 111, 115–125.
- Arbib, M.A. (2002). The mirror system, imitation, and the evolution of language. In K. Dautenhahn and C. L. Nehaniv (Eds.), *Imitation in animals and artifacts* (pp. 229–279). Cambridge, MA: MIT Press.
- Arbib, M.A., & Rizzolatti, G. (1996). Neural expectations: A possible evolutionary path from manual skills to language. *Communication and Cognition*, 29, 393–424.
- Bandura, A. (1977). *Social modeling theory*. Chicago: Aldine-Atherton.

- Bennett, A. T. D., & Cuthill, I. C. (1994). Ultraviolet vision in birds: What is its function? *Vision Research*, 34, 1471–1478.
- Bowmaker, J. K., Heath, L. A., Wilkie, S. E., Das, D., & Hunt, D. M. (1996). Middle-wave cone and rod visual pigments in birds: Spectral sensitivity and opsin structure. *Investigative Ophthalmology & Visual Science*, 37, S804.
- Dawkins, M. S., & Woodington, A. (1997). Distance and the presentation of visual stimuli to birds. *Animal Behaviour*, 54, 1019–1025.
- D’Eath, R. B., & Dawkins, M. S. (1996). Laying hens do not discriminate between video images of conspecifics. *Animal Behaviour*, 52, 903–912.
- Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic simulation study. *Journal of Neurophysiology*, 73, 2608–2611.
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, 15, 399–402.
- Ginsburg, N., & Nilsson, V. (1971). Measuring flicker thresholds in the budgerigar. *Journal of the Experimental Analysis of Behavior*, 15, 189–192.
- Goldin-Meadow, S. (1997). The resilience of language in humans. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development* (pp. 293–311). Cambridge: Cambridge University Press.
- Goldstein, H. (1984). The effects of modeling and corrected practice on generative language and learning of preschool children. *Journal of Speech and Hearing Disorders*, 49, 389–398.
- Goldstein, M. H., King, A. P., & West, M. J. 2003. Social interaction shapes babbling: testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, 100, 8050–8055.
- Ikebuchi, M., & Okanoya, K. (1999). Male zebra finches and Bengalese finches emit directed songs to the video images of conspecific females projected onto a TFT display. *Zoological Science*, 16, 63–70.
- Jaerveläinen, J., Schürmann, M., Avikainen, S., & Hari, R. (2001). Stronger reactivity of the human primary motor cortex during observation of live rather than video motor acts. *Neuroreport*, 12, 3493–3495.
- Jellema, T., Baker, C. I., Wicker, B., & Perrett, D. I. (2000). Neural representation for the perception of the intentionality of actions. *Brain and Cognition*, 44, 280–302.
- Lorch, C. P., Anderson, D. R., & Levin, S. R. (1979). The relationship of visual attention to children’s comprehension of television. *Child Development*, 50, 722–727.
- Oliveira, R. F., Rosenthal, G. G., Schlupp, I., McGregor, P. K., Cuthill, I. C., Endler, J. A., Fleishman, L. J., Zeil, J., Barata, E., Burford, F., Gonçalves, D., Haley, M., Jakobsson, S., Jennions, M. D., Körner, K. E., Lindström, L., Peake, T., Pilastro, A., Pope, D. S., Roberts, S. G. B., Rowe, C., Smith, J., & Waas, J. R. (2000). Considerations on the use of video playbacks as visual stimuli: The Lisbon Workshop consensus. *Acta Ethologica*, 3, 61–65.
- Patterson-Kane, E., Nicol, C. J., Foster, T. M., & Temple, W. (1997). Limited perception of video images by domestic hens. *Animal Behaviour*, 53, 951–963.
- Pepperberg, I. M. (1981). Functional vocalizations by an African Grey Parrot (*Psittacus erithacus*). *Zeitschrift für Tierpsychologie*, 55, 139–160.

- Pepperberg, I. M. (1985). Social modeling theory: A possible framework for understanding avian vocal learning. *Auk*, 102, 854–864.
- Pepperberg, I. M. (1990). 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. (1993). A review of effects of social interaction on vocal learning in African Grey Parrots (*Psittacus erithacus*). *Netherlands Journal of Zoology*, 43, 104–124.
- Pepperberg, I. M. (1994). Vocal learning in Grey Parrots (*Psittacus erithacus*): Effects of social interaction, reference, and context. *Auk*, 111, 300–313.
- Pepperberg, I. M. (1996). Categorical class formation by an African Grey parrot. In T. R. Zentall & P. R. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 71–90). Amsterdam: Elsevier.
- 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. (1998). The African Grey parrot: How cognitive processing might affect allospecific vocal learning. In R. P. Balda, I. M. Pepperberg, & A. C. Kamil (Eds.), *Animal cognition in nature* (pp. 383–409). London: Academic Press.
- Pepperberg, I. M. (1999). *The Alex studies*. Cambridge, MA: Harvard University Press.
- Pepperberg, I. M. (2000, April). “Pet Projects” Paper presented at the *Wired Kingdom* conference, MIT Media Lab, Cambridge, MA.
- Pepperberg, I. M. (2001). Lessons from cognitive ethology: Animal models for ethological computing. In *Proceedings of the 1st Conference on Epigenetic Robotics*, Lund, Sweden.
- Pepperberg, I. M. (2003). Training behavior by imitation: from parrots to people...to robots. In *Proceedings of the AISB '03 Second International Symposium on Imitation in Animals and Artifacts*, University of Wales, Aberystwyth.
- Pepperberg, I. M. (in press). Insights into vocal imitation in Grey Parrots (*Psittacus erithacus*). In S. L. Hurley & N. Chater (Eds.), *Perspectives on imitation*. Cambridge, MA: MIT Press.
- 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., & 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., 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, 151–178.
- Pepperberg, I. M., Gardiner, L. I., & Luttrell, L. J. (1999). Limited contextual vocal learning in the Grey parrot (*Psittacus erithacus*): effect of co-viewers on videotaped instruction. *Journal of Comparative Psychology*, 113, 158–172.

- 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 (*Psittacus erithacus*): Effect of species identity and number of trainers. *Journal of Comparative Psychology*, *114*, 371–380.
- Pezdek, K., & Stevens, E. (1984). Children's memory for auditory and visual information on television. *Developmental Psychology*, *20*, 212–218.
- Radeau, M. (1994). Auditory-visual spatial interaction and modularity. *Current Psychology of Cognition*, *13*, 3–51.
- Rice, M. L. (1991). Children with specific language impairment: Toward a model of teachability. In N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Biological and behavioral determinants of language development* (pp. 447–480). Hillsdale, NJ: Erlbaum.
- Rutledge, D., & Pepperberg, I. M. (1988). [Video studies of same/different.] Unpublished raw data.
- Schramm, W. (1962). What we know about learning from television. In L. Asheim (Ed.), *Educational television: the next ten years* (pp. 52–76). Stanford, CA: Institute for Communication Research.
- Secules, T., Herron, C., & Tomasello, M. (1992). Effect of video context on foreign language learning. *The Modern Language Journal*, *76*, 480–490.
- Stoneman, A., & Brody, G. H. (1982). An in-home investigation of maternal teaching strategies during *Sesame Street* and a popular situation comedy. *Journal of Applied Developmental Psychology*, *3*, 275–284.
- Todt, D. (1975). Social learning of vocal patterns and modes of their applications in Grey Parrots. *Zeitschrift für Tierpsychologie*, *39*, 178–188.
- Vanayan, M. H., Robertson, A., & Biederman, G. B. (1985). Observational learning in pigeons: The effects of model proficiency on observer performance. *Journal of General Psychology*, *112*, 349–357.
- Varela, F. J., Palacios, A. G., & Goldsmith, T. H. (1993). Color vision of birds. In H. P. Zeigler & H. J. Bischof (Eds.), *Vision, brain, and behavior in birds* (pp. 77–98). Cambridge, MA: MIT Press.
- Watkins, B., Calvert, S., Huston-Stein, A., & Wright, J. C. (1980). Children's recall of television material: Effects of presentation mode and adult labeling. *Developmental Psychology*, *16*, 672–674.
- Watkins, L. T., Sprafkin, J. N., & Krolkowski, D. M. (1990). Effects of video based training on spoken and signed language acquisition by students with mental retardation. *Research in Developmental Disabilities*, *11*, 273–288.

Authors' addresses

Dr. Irene M. Pepperberg
MIT School of Architecture and Planning
7–231
77 Massachusetts Ave

Cambridge, MA 02139
(617) 253-0364 (ph)
(617) 324-0040 (fax)
impepper@media.mit.edu

About the authors

Irene M. Pepperberg is a Research Associate Professor in the Psychology Department at Brandeis University and a Research Scientist at the MIT School of Architecture and Planning.

Steven R. Wilkes was a Biology Department undergraduate at Brandeis University and portions of this manuscript constituted his senior thesis. Wilkes is now a research technician in the Department of Pathology, University of Virginia HSC, Charlottesville, VA.