

## 10 Insights into Vocal Imitation in African Grey Parrots (*Psittacus erithacus*)

Irene M. Pepperberg

### 10.1 Introduction

The study of animal imitation, like that of animal cognition, animal-human communication, or numerical competence, generates considerable debate and as many definitions of the behavior as there are researchers in the discipline, and generally focuses on primates. Outside of the ornithological literature on song acquisition, avian vocal imitation is often ignored or considered a special, irrelevant case of imitative behavior (e.g., Shettleworth, 1998) because the imitated actions are perceived and produced within the auditory, not the visual, channel and their rationale may be difficult to determine. Avian allospecific vocal learning, however, is one of the few cases in which the copied behavior is, in accord with Thorpe's (1963) definition, completely novel—replication of an otherwise improbable act—and thus less likely to be confounded with social facilitation, stimulus enhancement, or other forms of nonimitative socially mediated learning (Fritz & Kotrschal, 1999). Here I examine replication of human speech by African Grey parrots, explain why I believe it is both referential and true imitation, show how various forms of input (e.g., my model/rival training) affect such learning, use my data to support the suggestion that imitation most likely occurs at multiple levels of complexity (see R. Byrne & Russon, 1998), and examine neurological issues involving imitation in a nonprimate, nonmammalian species.

### 10.2 Vocal Learning as an Imitative Act

Vocal imitation, even in a sense less strict than Thorpe's, is uncommon. Even vocal *learning* occurs in few nonhuman species (Lachlan & Slater, 1999) and its occurrence is rarely truly imitative. The acquisition of conspecific songs or calls by oscine birds, for example, does not necessarily

qualify as imitation. Its development in the absence of a model is abnormal (e.g., Marler, 1970), but oscines are generally predisposed toward conspecific learning (Petrinovich, 1985); thus the modeled act is not improbable (Thorpe, 1963). Also, the resultant behavior is not usually identical to the model's behavior (Kroodsma, 1996). True vocal imitation can be claimed only when subjects engage in goal-directed learning and use of allospecific utterances, a task that involves building some representation of an alien communication code (e.g., bottlenosed dolphins, *Tursiops truncatus*, reproducing computer-generated whistles for specific rewards, McCowan & Reiss, 1997; white-crowned sparrows, *Zonotrichia leucophrys*, learning song sparrow, *Melospiza melodia*, song, Baptista & Catchpole, 1989; African Grey parrots, *Psittacus erithacus*, using English labels as intentional requests, Pepperberg, 1988a).

Even when avian vocal learning is considered imitation, few researchers examine levels of behavioral complexity or data suggesting that successful imitation depends on the form of input (cf. Fragaszy & Visalberghi, 2001). Considerable data exist on input and levels of avian allospecific acquisition (see Baptista & Gaunt, 1994; Pepperberg, 1985, 1991b) and on oscine neural learning pathways (e.g., deVoogd et al., 1993; Jarvis & Mello, 2000), but not with respect to true imitation. If, as I maintain, vocal imitation involves acquiring *exceptional* vocalizations (Pepperberg, 1985, 1986a,b), that is, communication characterized by vocal learning unlikely to occur during normal development and that is used functionally after learning, such as acquisition of referential allospecific forms, then issues of complexity, input and output, and neural bases (e.g., mirror neurons) are critical for understanding avian imitation. With a somewhat unusual species—one known for its mimicry—the African Grey parrot (*Psittacus erithacus*), I study how input affects learning, how these parrots precisely reproduce human speech, and their referential use of such sounds. Because their vocalizations are often requests, such imitation, although involving auditory and vocal channels, is as goal directed as that in studies involving physical actions (see Akins & Zentall, 1996; Fritz et al., 1999). Not being a neurobiologist, I merely speculate on the bases for such behavior.

### 10.3 Reproduction of Human Speech by African Grey Parrots

To the human ear, an African Grey parrot's speech closely matches that of its trainers. Some researchers (e.g., Greenewalt, 1968; Lieberman, 1991), however, argue that birds cannot inherently produce human speech sounds and that birds' "speech" is an artifact of human perception (see following discussion). But might physical characteristics of avian speech match those

of humans? If so, and if these characteristics are those that scientists use to describe human speech, such data would support the contention that vocal learning in Grey parrots is an imitative act. Consequently, a student and I compared acoustic and articulatory data for human and psittacine vowels and consonants.

### 10.3.1 Phonological Analyses

Using standard phonetic and statistical techniques, we analyzed spectrograms and videotapes of speech samples from a human-raised, male Grey parrot, Alex, who had spent the last 14 of his 15 years interacting with humans, including training on cognitive concepts and English labels. We compared Alex's and my vocalizations because I was Alex's principal trainer, and observers claim his utterances closely resemble mine. The analyses were complicated and here I summarize only our most relevant findings (for details, see Patterson & Pepperberg, 1994, 1998; Pepperberg, 2002).

**F<sub>0</sub> Analyses** We measured fundamental frequencies, F<sub>0</sub> (source vibration at the avian syrinx; syringeal constriction at the tracheal base functionally resembles that of human vocal folds in phonation; Gaunt & Gaunt, 1985; Scanlan, 1988). Some researchers propose that an oscine bird (e.g., a mynah, *Gracula religiosa*) uses each half of the syrinx independently to produce two different sinusoidal pure tones whose interaction produces "formant frequencies of the original human speech sound that the bird is mimicking. . . . We perceive these nonspeech signals as speech because they have energy at the formant frequencies" (Lieberman, 1984, p. 156).<sup>1</sup> F<sub>0</sub> analyses, however, show that Alex, like humans and unlike most songbirds (Greenewalt, 1968; D. Miller, 1977), uses one set of articulators to produce speech (Patterson & Pepperberg, 1994). He has one F<sub>0</sub> and his absolute values are in the general range of those of an adult human (124–276 Hz; Peterson & Barney, 1952).

**F<sub>1</sub> and F<sub>2</sub> Analyses** Accurate perception of human speech involves processing formants (F<sub>x</sub>): resonant frequencies of the human mouth opening

1. Mynahs do not independently use two halves of their syrinx to produce formants from F<sub>0</sub>s as Lieberman (1984) suggests, but their mechanism differs from that of Grey parrots (Klatt & Stefanski, 1974). Budgerigars produce human vowels by amplitude modulation (as Lieberman suggests), but because their syrinx cannot produce two independent sounds, they produce the modulation via a different, as yet undetected, mechanism (Banta Lavenex, 1999).

( $F_1$ ) 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. Differences and similarities exist between our data. For both  $F_1$  and  $F_2$ , his range is less than mine and lacks my low-frequency values. For most vowels, his  $F_1$ , which in humans roughly corresponds to tongue height, differs considerably from mine and, unlike mine, varies little across vowels; his tongue height varies little. Our  $F_2$ s, which in humans correspond to front-to-back tongue position, are similar. Like mine, his vowels are somewhat distinguishable by  $F_2$ ; his front-to-back position varies considerably. Thus, as for humans, his formants correlate with specific resonances of the vocal tract area (Remez et al., 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 were nearly indistinguishable from  $F_2$ . All his vowels cluster in the high to 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 is a diphthong. It has two parts, /u<sub>1</sub>/ and /u<sub>2</sub>/. Mine are front vowels; Alex's are front and back vowels, respectively.

**Comparisons of Psittacine and Human Stop Consonants** Traditional acoustic analyses for stops (consonants with blocked air flow) generally applied to Alex. We examined voice-onset timing (time, in humans, between the lips opening to release air and the start of vibration by the sound source), number of bursts (bursts result from the 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 the beginning of voicing to when formants assume a steady-state pattern.

Similarities and differences exist for our stops (Patterson & Pepperberg, 1998). The similarities involve statistically distinct acoustic characteristics and separation of stops into voicing and place of constriction subsets. The differences involve specific subset divisions and likely mechanisms for /b,p/ production. Compared with my subsets, his /b/ is an outlier; his /p/ clusters nearer /d,g/ than /t,k/. Alex must produce bilabial /b,p/ without lips; he seems to compensate with mechanisms (esophageal involvement) not normally used by humans (Patterson & Pepperberg, 1998).

### 10.3.2 Evidence for Physical Imitation

Our research suggests that Alex's fidelity of vowel and stop imitation is most likely limited only by differences in human and avian vocal tracts (e.g., lack of lips). His vocalizations all resemble human samples; characteristics that describe human speech and processes used to produce human speech generally work for him (Patterson & Pepperberg, 1994, 1998).  $F_0$  data specifically show that Alex, like humans, produces true resonances, not sinusoids at appropriate formant frequencies. Such data support an interpretation of Alex's speech acts as physical imitation. Also, he must learn to manipulate various vocal tract areas to produce human speech; some researchers call such behavior gestural learning. The process may take considerable time and practice; Alex's patterns evolve slowly toward human speech unless new labels involve previously acquired phonemes (Pepperberg, 1999).

Three other aspects of Alex's speech resemble young children's speech and suggest purposeful imitation. First, Greys also use sound play (phonetic "babbling" and recombination; Pepperberg et al., 1991) to derive new speech patterns from existing ones. After acquiring "gray," for example, Alex produced "grape," "grate," "grain," "chain," and "cane," which we eventually mapped to appropriate referents (Pepperberg, 1990b). Second, his recombinations suggest that he abstracts rules pertaining to the beginnings and endings of utterances. In over 22,000 vocalizations, we never observed backward combinations (e.g., "percup" instead of "cupper" or "copper"; Pepperberg et al., 1991). His behavior suggests that he parses human sound streams and acoustically represents labels in humanlike ways, and has similar phonetic categories (Patterson & Pepperberg, 1994, 1998). Third, he may exhibit anticipatory co-articulation. He may separate specific phonemes from speech flow *and* produce these sounds to facilitate production of upcoming phonemes (e.g., /k/ varies significantly between "key" and "corn"; Patterson & Pepperberg, 1998). Such behavior observed in conjunction with sound play is often used as evidence for top-down processing (Ladefoged, 1982).

### 10.4 Referential Use of Learned Vocalizations

Crucial to my argument for Alex's imitation is that his speech is *referential*. Purposeless duplication of an act (e.g., rote reproduction of human speech) lacks cognitive complexity and is "mere" mimicry. But if an act is performed because the imitator understands its purpose—to reach a goal, be it an object or intentional communication, that is otherwise impossible to obtain—then the imitation is intentional and complex, most likely

indicating cognitive processing. Moreover, to use imitated speech referentially, Alex must discriminate among, appropriately categorize, and treat as meaningful even minor human sound differences; he must know that "want pea" and "want tea" elicit different responses in trainers.

Alex exhibits such processing. He can label more than 50 exemplars, 7 colors, 5 shapes, quantities to 6, 3 categories (color, shape, material), and uses "no," "come here," "wanna go  $x$ " and "want  $y$ " ( $x$  and  $y$  are appropriate location or item labels). He combines labels to identify, request, comment upon, or refuse about a hundred items and to alter his environment (Pepperberg, 1990a). He processes queries about category, relative size, quantity, and the presence or absence of similarity or difference in attributes; shows label comprehension; and semantically separates labeling and requesting (Pepperberg, 1983, 1987a,b,c, 1988a,b, 1990b, 1992; Pepperberg & Brezinsky, 1991). Given an array of, for example, different numbers of intermingled red and blue balls and blocks, he can quantify any subarray (Pepperberg, 1994a). His responses are not paired associations; he answers multiple questions about the same objects. He easily adds new exemplars to his categories of color, shape, material or object labels, and new categories to his repertoire (e.g., number). Previous studies on mimetic birds, using standard conditioning techniques, failed to produce such learning (Pepperberg, 1999). Why did Alex—a creature with a walnut-sized brain, and one so different from that of humans, nonhuman primates, mammals, and even songbirds—succeed?

### 10.5 Effects of Input on Vocal Learning in Grey Parrots

Alex was taught via the model/rival (M/R) procedure, which differs considerably from standard animal training techniques (Pepperberg, 1999). Was the technique responsible for his success? To find out, my students and I compared this technique with other procedures. Here I describe this system, contrast it with less successful forms, and discuss our findings.

#### 10.5.1 Description of the Different Training Techniques

**M/R Technique** Basic model/rival (M/R) training, adapted from Todt (1975), involves three-way interactions among two humans and one bird. M/R training introduces labels and concepts and helps shape pronunciation; it demonstrates targeted vocal behavior through *social interaction*. A parrot observes two humans handling one or more objects in which it has shown interest, then watches the humans interact. The trainer presents

an object or objects to the other human and queries the other human about the selection (e.g., "What's here?" "What color?"), rewarding correct answers with praise and the object (or objects). The second human is a model for the parrot's responses and the bird's rival for the trainer's attention. This person is also used to illustrate the consequences of errors; incorrect responses (such as those the bird might make) are punished by scolding and temporarily removing the object (or objects). The model is asked to try again or to talk more clearly if its response was (deliberately) incorrect or garbled, thereby showing the bird "corrective feedback" (see Goldstein, 1984; Vanayan et al., 1985). The human pair demonstrate *referential* and *functional* use of labels, respectively, by providing a one-to-one correspondence between label and object and modeling label use as a means to obtain the object (Pepperberg, 1981, 1990a,c). The bird is also queried and initially rewarded for approximating correct responses. As training progresses, the reward criteria tighten, adjusting training to the parrot's level. The model/rival and the trainer also reverse roles, showing how either can use the communicative process to request information or effect environmental change. Without experiencing this role reversal, birds exhibit two behavior patterns that are inconsistent with interactive, referential communication. They neither transfer responses to anyone other than the human who posed questions during training, nor learn both parts of the interaction (Todt, 1975).

M/R training intentionally uses *intrinsic* reinforcers. To ensure the correlation of the labels or concepts to be learned with their appropriate referents, the birds' rewards are the specific object to which the uttered label or concept refers. Earlier unsuccessful avian-human communication attempts used *extrinsic* rewards (Pepperberg, 1999); for example, one food that neither related to, nor varied with, the label or concept being taught. Extrinsic rewards delay acquisition by confounding the label of the targeted exemplar or concept with the food. My birds never receive extrinsic rewards.

After using M/R input successfully, my students and I tested how changing the system might affect learning, not only with respect to sound reproduction, but also comprehension and appropriate use (i.e., actions requiring cognitive processing). We varied the major aspects of input in M/R training—reference, context or function, and interaction—and a related aspect, modeling. Reference is considered to be an utterance's meaning—the relationship between labels and the objects to which they refer—and is exemplified by our use of referential rewards. Determining the referent—and whether one true referent exists—often requires cognitive processing by the receiver (W. Smith, 1991). To determine the referent for

signal  $a$ , the receiver cannot simply process the signal and remember and interpret it as being associated with situation  $x$ , but rather must choose, based on additional information, among different aspects of situation  $x$ . The more explicit the referent, the more easily the signal is learned. Context or function involves the particular situation and effects of an utterance's use; our initial use of a label as a request gives the bird a reason to learn the unique and unfamiliar set of sounds. Both cognitive processes and motivation are important for extracting a signal's function from many possibilities in a given situation; explicit functionality assists learning. Social interaction indicates which environmental components of the modeled act should be noted, assists in determining their relative importance and order, emphasizes common attributes—and thus possible underlying rules—of diverse actions, and allows input to be continuously adjusted to the learner's level. Interaction and modeling engage a subject directly, provide contextual explanations of reasons for, and demonstrate consequences of, actions. All such aspects are crucial to successful imitation and require cognitive abilities.

To vary training with respect to these factors and to test their relative importance, I contrasted M/R tutoring with sessions using different forms of input involving human interaction, videotapes, and audiotapes. Most of these studies had to await additional subjects. Alex might have failed to learn simply because protocol changes had been made, not because of the quality of the changes; new subjects would not be influenced by prior experience. We thus performed one M/R variant experiment with Alex and seven sets of experiments with three naive Greys—Kyaaro, Alo, and Griffin (Pepperberg, 1999; Pepperberg et al., 2000). In each study, the labels were counterbalanced across training conditions and were always labels that could be produced by Grey parrots.

**Eliminating Aspects of Training** In Alex's "M/R-variant 1" (Pepperberg, 1994b), two humans modeled a sequence derived from Korean counting labels, *il ee bam ba oo yuk chil gal*, but without reference to either specific items or Alex's existing English number labels (Pepperberg, 1994a, 1997). He saw only a line of pictured numerals; no sound was attached to a particular numeral. The labels were chosen as part of another study on serial learning (Silverstone, 1989) and to differ maximally from English. The training lacked functional meaning and all but minimal referentiality, but maintained joint attention among bird, humans, and pictured numbers (known to facilitate label acquisition by children; Baldwin, 1991). The trainers maintained role reversal, included Alex in interactions, and re-



warded him with vocal praise and the opportunity to request favored items ("I want *x*"; Pepperberg, 1987a); errors elicited scolding and time-outs.

The first set of experiments with the juvenile parrots compared simultaneous exposure of Alo and Kyaaro to labels in three conditions. Socially isolated in condition I, they heard audiotapes of Alex's sessions; input was nonreferential, not contextually applicable, and noninteractive. Socially isolated in condition II, they saw videos of Alex's sessions; input was referential, minimally contextually applicable, and noninteractive. Condition III involved standard M/R training. Condition I paralleled early allospecific song-acquisition studies (e.g., Marler, 1970); condition II involved still-unresolved issues about avian vision and video (i.e., whether the flickering of a cathode ray tube and its lack of ultraviolet, which birds can see, affect their ability to attend to or recognize video as representing reality; see Ikebuchi & Okanoya, 1999; Pepperberg & Wilkes, 2004). Each bird received training in all three conditions.

Experiments two through five tested video learning in more detail. Because interactive co-viewers can increase video learning in children (M. Rice et al., 1990), in experiment two 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 labels, ask questions, or relate content to other training. The birds' attempts at a label could receive only vocal praise. Social interaction was limited; referentiality and functionality matched earlier video sessions (Pepperberg et al., 1998). Because the extent of co-viewer interaction might affect video learning in children (St. Peters et al., 1989), in experiment three our co-viewer uttered targeted labels and asked questions (Pepperberg et al., 1999). Because the absence of a reward might deter video learning, in experiment four an isolated parrot watched videos while a student in another room monitored its utterances through headphones and could deliver rewards remotely (Pepperberg et al., 1998). Experiment five used live videos from Alex's sessions because the birds might have habituated to the single video previously used per label (although each tape depicted many different responses and interactions among Alex and the trainers; Pepperberg et al., 1999). We are replicating video studies with a liquid crystal monitor to see if the flickering of the cathode ray tube rather than our experimental manipulations affects learning (Pepperberg & Wilkes, 2004).

Experiments six and seven examined the role of bird-trainer joint attention. For children, label acquisition occurs when adult-child duos focus jointly on objects being labeled (e.g., D. Baldwin, 1995). In experiment six, a single trainer had no visual or physical contact with parrot or object.

She faced away from the bird and talked about an object placed within reach of the bird, emphasizing its label (“Look, a shiny *key!*”, “Do you want *key?*” etc.—phrases framing the label, allowing repeated label use without causing habituation; Pepperberg, 1981). A bird’s attempts to utter the targeted label could receive only vocal praise, thus eliminating considerable functionality and social interaction (Pepperberg & McLaughlin, 1996). In experiment seven, we retained bird–trainer–object joint attention, but eliminated aspects of modeling by having only a single student label objects and query the bird (Pepperberg et al., 2000).

### 10.5.2 Results

In sum, the birds failed to acquire referential use of targeted labels in any non-M/R condition, but succeeded in concurrent M/R sessions on other labels (Pepperberg, 1999). For Grey parrots, at least, purposeful imitation requires input involving two humans who demonstrate the referentiality and functionality of a targeted label, socially interact with each other and the bird, exchange roles of questioner and respondent, portray effects of errors, provide corrective feedback, and adjust training as a subject learns. Success also requires cognitive processing by the parrot to evaluate the input. Without such input and processing, even supposedly mimetic birds fail to learn entirely or acquire limited associations rather than full referential use of labels; that is, they cannot transfer label use from training to testing and from training exemplars to other instances of a relevant object or concept; or they reproduce a label but do not comprehend what they produce (Pepperberg, 1994b; Pepperberg et al., 1998, 1999; Todt, 1975).

Thus three birds—Kyaaro, Alo, and Griffin—failed to learn referential labels in any video session; the first two failed to learn from audiotapes (Griffin did not receive audio training), and Kyo and Griffin did not utter labels in experiments six and seven. In contrast, all birds appropriately identified items (e.g., cork, paper, wood, wool) targeted in basic M/R sessions. Tapes of solitary sounds produced by the birds revealed that they practiced only M/R-trained labels (Pepperberg et al., 1991; Pepperberg, 1994b; Pepperberg & McLaughlin, 1996). Of interest is that Griffin clearly produced targeted labels after two or three M/R sessions subsequent to experiment seven. We suspect he actually acquired labels in that experiment, but did not use them until he observed their use modeled. (Note: birds that were switched to M/R training after fifty video sessions needed about twenty sessions before they produced labels.)

In M/R variant-1, Alex learned the modeled label string, but the results differed from those of M/R studies in two important ways. First, acquisition

took an unusual 9 months (Pepperberg, 1981, 1994b; Silverstone, 1989). Second, and most striking, was that he could not immediately use, nor subsequently learn to use, these labels referentially, that is, for either serial labeling or quantity. Even after we modeled one-to-one correspondences between eight objects and the label string, he could not refer to smaller quantities using elements in the string (e.g., say “il ee bam ba” when shown four items and told “say number”); he had learned an unanalyzed phrase (Pepperberg, 1994b). Given his previous M/R-based success (e.g., Pepperberg, 1990b, 1992, 1994a), the current failure was most likely a consequence of training, not a cognitive deficit (Pepperberg, 1994b).

The M/R technique also encompasses other elements of input (e.g., quality of input, identity of model, trainer role reversal; Pepperberg, 1991a,b) that we did not initially vary. Subsequent work (Pepperberg et al., 2000; Pepperberg & Wilcox, 2000) studied these forms and their cognitive correlates, and reaffirmed that standard M/R training, with all its elements of input, is the more effective protocol for referential learning.

#### 10.6 Implications for Studies of Imitation

The results of the work reported here highlight issues relevant to imitation studies in all creatures, with implications for general learning theory. Training that produces mere mimicry does not result in the acquisition of purposeful “improbable acts” (Thorpe, 1963), implying that mimicry and imitation involve different learning mechanisms and possibly cognitive processes. Imitation requires specialized input; generating imitation and determining the level of imitation in a given species with a known range of cognitive capacity and learning ability may depend upon environmental and tutoring conditions. Japanese quail (*Coturnix japonica*), for example, learn from an observer more often when they are hungry than when they are satiated (Dorrance & Zentall, 2001), which supports data on the social contingencies of imitative learning in related species (Lefebvre, 2000). Even species known for mimetic abilities, such as Grey parrots, imitate only under certain conditions. Thus preexperimental generalizations about imitative abilities may be premature. Might supposedly nonimitative species imitate under appropriate conditions?

Of interesting is that the conditions under which Grey parrots acquire referential English labels (goal-directed imitation) are those that produce (1) exceptional learning—learning unlikely during normal development but possible with certain input (Pepperberg, 1985, 1988b, 1993, 1997) and (2) acquisition of language by children with specific impairments and

who lack concomitant social skills (e.g., M. Rice, 1991; see Pepperberg & Sherman, 2000, 2002). Normal children, for example, but not impaired children or Grey parrots, learn from video in the absence of interactive caretakers (Pepperberg et al., 1999). Normal children, but not Grey parrots, learn labels from a single caretaker (D. Baldwin, 1991; J. de Villiers & de Villiers, 1978; Pepperberg et al., 2000). For echolalic autistic children (who are characterized by their nonreferential use of mimicked sounds; Fey, 1986), acquiring normal communication may involve exceptional learning and be responsive to M/R training (Pepperberg & Sherman, 2000, 2002).

When comparing M/R with other forms of imitative training, another issue arises. Possibly for parrots (and children with disabilities, and maybe other animals), the typical form of instruction, that is, "do as I do," might prevent the subject from separating the targeted behavior pattern or target of the command from the command and thus inhibit building a representation of the required response (Pepperberg & Sherman, 2000). Maybe such subjects must observe a model responding to the command "do as I do" (i.e., or see a response to "do x") to identify behavior patterns to be learned or imitated (and to correlate the action with the specific label "x").

#### 10.6.1 Neurological Correlates and Levels of Imitation

My findings also suggest a search for both correlates and homologies across species for recently discovered neuroanatomical areas that are apparently involved in replicating action (mirror neurons; Iacoboni et al., 1999), and for relations between different areas and levels of imitation (Rizzolatti et al., 2001). Byrne and Russon (1998), among others, suggest that imitation encompasses a range of behavior patterns of varying complexity; they deconstruct imitation into an "action level"—specification of sequential acts—and a "program level"—description of subroutine structure and the hierarchical layout of a behavioral program. I propose three levels: simple mimicry, low-level imitation, and high-level imitation, and their connections to mirror neurons.

Simple mimicry describes replication of actions (e.g., manual or vocal) without direction—with unclear referential and functional aspects—but that involve some social interaction. Examples might be newborns' imitative acts (Meltzoff & Moore, 1977) and adults' unconscious replication of others' motions in social settings (e.g., touching their hair in specific ways; Chartrand & Bargh, 1999), also behavior like Alex's in M/R variant-1. These patterns most likely do not recruit the mirror neuron system activated when deliberate actions are viewed (what I'll call MN1), but corroborate neurobiological data on involuntary levels of imitation (Arbib et al., 2000;

what I'll call MNO)—levels that involve little or no cognitive processing, intentional reproduction, or integration of observed acts in which “an executive sector of the motor system ‘resonates’ during observation of an action. As a consequence the observed action is usually emitted. In this type of resonance phenomenon understanding of the meaning is neither essential nor required” (Rizzolatti & Arbib, 1999, p. 152). Thus I propose connecting the absence of reference, functionality, and intentionality, the lack of recruitment of MN1, and the lack of overt learning of a novel behavior (or, at best, replication without meaning), and the absence of “theory of mind” (TOM; see following discussion).

Low-level imitation most likely involves MN1 and some reference, functionality, and social interaction, but not improbable behavior or TOM. Examples are replication of motor acts such as movement toward objects (Rizzolatti et al., 2001) or Alex's learning the word “carrot” after trainers ate and labeled the object only a few times, because he had phonemes from “key” and “parrot” (Pepperberg, 1999). When a monkey sees a human arm move, for example, its cortical neurons homologous to those of the human become active, even if the monkey's limb remains inactive (e.g., Fogassi, 2000); the observed action, however, must be goal directed (Buccino et al., 2001; Rizzolatti et al., 2001). In humans, some activation occurs even when intransitive movements are observed (Fadiga et al., 1995). Monkey mirror neurons are not likely involved in Thorpe's improbable acts because their activation also seemingly requires experience of, or capacity for, making the observed movement (Chaminade et al., 2001b; Rizzolatti et al., 2001). How can an existing action be improbable? Of interest is that in humans some different brain areas activate during means versus goal actions (Chaminade et al., 2002), suggesting that behavior involving these neurons is not simple mimicry; instead, these neurons could be involved in the recruitment of sets of acts from one context for use in another; that is, for emulation (e.g., Tomasello, 1999), not imitation. A consistent interpretation of monkey data would be that their mirror neuron system, an MN1, recruits *innate* grasping patterns when the monkeys observe human actions, and that novel patterns—and true imitation—are not part of the system.

Higher-order imitation most likely involves considerable reference, functionality, social interaction, and creation of improbable acts. Alex's referential acquisition should qualify. Such behavior relates to, but I believe differs from, the motor actions described by Rizzolatti and co-workers—patterns requiring “refinement of the motor act or of the motor sequence” (Rizzolatti et al., 2001, p. 668), Rizzolatti et al., like Byrne (2002a), argue

that new patterns arise as subjects dissect observed behavior into a string of already existing simpler sequential components (motor acts), then reconstitute them appropriately. In contrast, I argue that precisely because these are Thorpe's improbable acts, the observer is *creating* these motor patterns, a representation. The issue may be of degree, in that I am discussing creation from the most elementary motor units. In any case, these novel patterns take longer to construct than those involved in low-level imitation; they require strong levels of reference, functionality, and interaction for their inception (e.g., Pepperberg, 1985), and, I suggest, involve the construction of neural pathways that only then are recruited into a mirror system, MN2. These actions might also involve beginnings of TOM.

Might data about levels of imitation thus provide insights into TOM, assuming various TOM levels also exist? Possibly building MN2 activity allows observers to both see and somehow feel how actions look—a means of representing and categorizing actions for future replication (e.g., Hurford, 2002); the beginning of TOM. MN1 activity alone cannot be enough or monkeys would exhibit TOM. Understanding TOM requires teasing apart confounded issues. We must start by determining what neural substrates are involved in integrating information; what is needed for inferential understanding of how replicating a specific observed behavior is a specific means to obtaining a specific goal, even under delayed reward (e.g., Bandura, 1969); and what brain areas control an observer's ability to replicate behavior patterns physically, versus how much processing and what areas are involved in observers' *choosing* to replicate the behavior (e.g., so the imitative act is deliberately executed in pursuit of some goal). Of interesting is that when an observer chooses not to act, spinal (as opposed to cortical) neuronal excitation occurs in the order opposite to cortical mirror neuron excitation, possibly to prevent physical replication (Baldissera et al., 2001).

Different human mirror neurons seem to be involved with means than with goals (Chaminade et al., 2002). Are different mirror neurons involved in the awareness of achieving a different goal with a similar arm movement, or the same goal with the same arm movement, the latter being needed for intentional imitation (e.g., MN1 versus MN2)? Is awareness or recognition intentional or is learning necessary? People learning American Sign Language signs improve their acquisition and retention by delaying rather than concurrently imitating the observed motions (Weeks et al., 1996); thus functional learning most likely requires more than replication. Likewise, autistic children's echolalia but lack of controlled imitation suggests that an executive process that is missing in their mirror

neuron system (see J. Williams et al., 2001) exists for true imitation. Does this relate to TOM? When is TOM required for imitation? Is TOM required for higher forms of imitation, or do creatures simply need a nonreflexive consciousness tied to some relatively low intentionality and biologically based hedonic response (Dickinson & Balleine, 2000)? A chimpanzee that attributes goals to moving blocks (Uller & Nichols, 2001) most likely attributes goals—and its mirror neurons respond—to moving human arms, and it can replicate the human action. But does the chimpanzee attribute these goals or intentions to the human connected to the arm (Povinelli & Eddy, 1996)? Data showing that particular neurons fire in monkeys only when the agent being observed attends to the object of its reaching (Jellema et al., 2000) support this inference, but monkeys rarely imitate even intentionally useful acts demonstrated by humans (Visalberghi & Fragaszy, 2002). And attributing intention to moving objects does not require fully developed TOM; children attribute intention early but fail TOM tasks until they are about 3 years of age (Baron-Cohen et al., 1985). Possibly imitation, TOM, and neural pathways involve corresponding levels of complexity.

Now, what primate mirror neuron analogues (MNO, MN1, MN2) could exist in parrots, who cannot exactly replicate human motor patterns? Do parrots have neurons that react similarly when hearing *and* speaking human labels? For oscine birds' own song, some parallels exist with primates. In songbirds, the high vocal center (HVC) sends efferents to both input and output branches of the song system; the HVC is necessary for song production and has neurons showing song-specific auditory responses (H. Williams, 1989). Furthermore, playback of the birds' own song during sleep causes neural activity comparable to actual singing (Dave & Margoliash, 2000). How these findings relate to parrot brains, which are organized differently than those of songbirds (e.g., Striedter, 1994; Jarvis & Mello, 2000), is unclear. Although studies of *ZENK* gene<sup>2</sup> expression show separation of budgerigar (*Melopsittacus undulatus*) response regions for hearing and vocalizing warble song (Jarvis & Mello, 2000), electrophysiological studies in the frontal neostriatum of awake budgerigars show activity during both production of and response to calls (Plumer & Striedter, 1997, 2000). Evidence also exists for additional budgerigar auditory-vocal pathways (e.g., Brauth et al., 2001). Because the *ZENK* response appears to be

2. Expression of the *ZENK* gene, a songbird analogue to a human transcription factor known as *egr-1*, is driven by actions of singing and hearing. Thus it is used to form a functional map of the avian brain for behavior related to both auditory processing and vocal production (see Jarvis & Mello, 2000).

tuned to specific song features (Ribeiro et al., 1998), the relevance of these data for mirror neurons in talking parrots is not known.

### 10.6.2 Connections to Communication and Possibly Language

Imitation and mirror neuron issues should, however, relate to communicative competence. Mirror neuron data suggest homologies between the monkey F5 and human Broca's area (Fogassi, 2000; Rizzolatti et al., 2001; see Grèzes & Decety, 2001) that might explain the codevelopment of gestural and verbal syntactic combinations occurring in humans and apes (i.e., simultaneous emergence of rule-governed ordering of labels and serial combination of different-sized cups; Johnson-Pynn et al., 1999). Both areas are activated by both action and observation of hand or mouth gestures, suggesting a system that not only matches observation of an action with its execution, but that could also subserve imitation and interindividual communication (Fogassi, 2000; Rizzolatti & Arbib, 1999). A similar simultaneous development of vocal and physical combinatorial actions in parrots (Pepperberg & Shive, 2001) must, however, involve nonprimate neural structures. Are these systems analogous? Or homologous (Medina & Reiner, 2000)? I have previously described parallel avian-primate developmental systems for communication (Pepperberg, 2001), but unrelated to mirror neurons or imitation. The following is a short review in the context of this discussion.

I suspect that for most creatures that learn their repertoire, early babbling mostly involves the MNO system (e.g., Rizzolatti & Arbib, 1999), corresponding to mimicry. These utterances can be triggered by hearing, are meaningless, and involve no cognitive processing, but do involve some social interaction. They may be accompanied by simple physical actions that also "mirror" caretakers' actions, and, for both humans and parrots, may increase with adult attentiveness (Locke, 2001; Pepperberg, 1999). Data demonstrating that listening to a word produces phoneme-specific activation of speech motor centers (Craighero et al., 2002) support a mimetic interpretation. Subsequent maturation of the nervous system allows voluntary control and precise replication; possibly maturation relates to the spinal (and other) inhibitory parts of the mirror neuron system (MN1?).

The next stage, I believe, involves transition to high-, not low-level imitation. For parrots and children, first labels qualitatively differ from later labels. First labels are learned slowly, may be acoustically biased by and based on prenatal or prehatching exposure to sounds, and are predisposed, probably evolutionarily, to refer to whole objects (review in Pepperberg, 2001). First labels are often indexical (they refer to specific items, not a



class), may lack communicative intent and true meaning (P. de Villiers & de Villiers, 1979), but can be goal directed. They are not based on simple associations and can be transferred across items (Pepperberg, 2001). What seems to be missing is the use of representation. Subjects might be unable to store images long enough to form representations or might not sort early labels into the categories used to form representations (i.e., form categorical images). But subjects are beginning to build such representations and the concomitant motor and vocal patterns; that is, they are beginning to construct relevant neural pathways (MN2?).

As learning of labels continues, children (Hollich et al., 2000)—and most likely parrots—begin processing information within a different context that arises through an understanding of social systems; quite likely neural development underlies such understanding and includes the functional recruitment of MN systems. Learning still appears self-directed—driven by the need to influence others and to have basic needs met—but advances because subjects now attend to others' intentions and recognize others as information sources separate from themselves. I suggest that recruitment of the MN2 system, correlated with MN1, precipitates the explosion in learning labels that occurs at about 18 months in children. MN2 provides the learning pathways and MN1 allows emulation (e.g., reorganization of sounds to create new labels) as the developing cognitive processing forms categories and representations.

Of interest is that most studies on labeling and concept formation involve older children (18–24 months, e.g., Bloom, 2000; Tomasello, 2001), ignoring transitional stages in humans and animals (cf. Hollich et al., 2000; Cheney & Seyfarth, 1990; Pepperberg, 1999). But by 18 months, a child can already take others' perspectives and see them as information sources. The child can play with one toy, notice an experimenter playing with and labeling another, and change its focus to the experimenter's toy when it hears the experimenter's label. At 12–14 months, the child is more self-centered; in the same situation, it prefers looking at its toy when it hears the experimenter's label (Hollich et al., 2000), and at about 10 months it usually associates the experimenter's label with its toy (Hirsh-Pasek et al., 2000). Thus in a brief transitional period, the child loses its self-centered bias, possibly laying the groundwork for high-level imitation. Note that autistic behavior and its communicative deficits often appear at this transition (Tager-Flusberg, 2000)—the point at which self-awareness, the need to understand the self as separate from others, and others as information sources become critical for learning. Some autistic children never move beyond the 10-month-old stage in terms of label and object associations (D.

Baldwin & Tomasello, 1998); that is, they cannot convert the actions of another into their own perspective, which is crucial for imitation and may involve MN2.

Is awareness necessary for this transition? Might its development enable the transition? Possibly, creatures learn to generalize and initially separate the self from others by categorizing and generalizing emotions first with respect to environmental events (Humphrey, 2000), then in intersubject interactions (Damasio, 1999). Damasio argues that “core” consciousness (the basic form involving awareness of the present, not the future or past) emerges when we interact with an object (including others), and is a feeling that accompanies the making of a mental image, even one retained only momentarily. His theory might explain why social interaction assists learning. The mental image allows the categorization of events with respect to their emotional content, which in intact individuals leads to categorization of involved objects and actions. A child, for example, initially does not label emotions, but talks about objects, expressing emotion by displaying positive or negative affect (Bloom, 2000). How categorizations lead to a representation of objects and actions that can be manipulated to allow advanced learning is unclear, but categorization may relate to the development of MN2 from MN1 or MN0. In this view, selfhood (not necessarily full consciousness) begins in the emotional domain, before the emergence of language, and seemingly leads to categorization, which then leads to understanding and use of representation. Our parrots’ behavior—like that of children—suggests clues about the transition to advanced learning as they begin recognizing others as information sources.

Evidence for our parrots’ transitions from self-centered learning comes from vocal actions like those of children, particularly sound play (described earlier), when trainers referentially map novel utterances. Apparently, Grey parrots—like children (Brown, 1973)—begin testing humans as information sources for the reference of sound patterns. They see humans in this context during training, then adapt the situation. They play with label phonetics; they also utter a label used in specific contexts, such as “wool” for a woolen pompon, while pulling at a trainer’s sweater. It is rare for such action to happen by chance; by then the bird has at least three or four other labels. Our responses—of high affect, excitement, and repetition of a label, which stimulate the birds further and possibly activate mirror neurons—affirm the power of their utterances and reinforce their attempts at categorization. Even if the birds err in initial categorizations, they still receive positive reinforcement; we provide a correct, new label (e.g., state that an almond isn’t a “cork,” but suggest “cork nut”; Pepperberg, 1999).

Of importance is that the birds' behavior is not trial-and-error learning, but guided invention (A. Lock, 1980), from initial mapping of a label to generalization to imaginal syntax. Parrots, like children, have a repertoire of desires and purposes driving them to form and test ideas in dealing with the world; these ideas can amount to early stages of representation (categorization) in cognitive processing. And manipulation of representations is a syntax of imagery, which Damasio insists requires some self-awareness. We have creation of somewhat novel acts, maybe mirror neuron involvement.

Can training foster such emergent behavior? Preliminary data on dysfunctional children, including autistics, suggest that they respond positively to M/R input (Pepperberg & Sherman, 2000, 2002). The children in these studies have not achieved full age-appropriate behavior, but their communication abilities improve after M/R sessions. Work in progress (D. Sherman, personal communication) suggests that children with severe disabilities need to be prepared before they can accept M/R input. Data from such studies may provide additional understanding of transitional processes and their relevance to imitation.

Of course, sensitivity to input, separation of the self from others, and transition to advanced learning may not be the cause but the outcome of maturation of the brain and neural development. Many neural connections existing at birth die off early (Changeux & Danchin, 1976). Do systems used in early simple learning of labels die off, or do they get pruned and relegated for use only in mimicry? Many new connections are formed in the early years of life. Given that neural categorization occurs when a neural ensemble provides the same output from different inputs, is this connectivity—which is almost absent in year-old babies—what grows with age? Is the failure to form new connections as old ones die related to emergent autistic behavior? Possibly (see Chugani, 1999); but if behavioral shifts and the transition to different learning forms do indeed result from changes in neural connections, such reorganization is unlikely to be specific to humans because it also occurs in parrots, with their significantly different brain architecture (Medina & Reiner, 2000).

### 10.7 Summation

Input often affects whether imitation, like other forms of learning (e.g., Pepperberg, 1999), occurs at all, and what achievement levels exist. Careful experimentation is thus necessary to determine the extent and range of any learning ability in any species, including humans (Pepperberg & Sherman,

2000). Moreover, data on apparent true imitation in species as phylogenetically diverse as parrots and humans suggest that we must reevaluate our understanding of the phenomenon. Either such behavior arose independently multiple times, or the evolutionary precursor is further back than researchers have presumed. In any case, for parrots as well as primates, various forms and levels of imitation seemingly exist. We can label them as different phenomena, or as different aspects of one behavior, but the behavior that specifically develops is contingent upon the input received.<sup>3</sup>

#### **Acknowledgment**

The writing of this chapter was supported by the American Foundation.

3. See comments on this chapter by Byrne, vol. 1, ch. 12.3, p. 290, and by Pickering, ch. 12.4, p. 293. ED.