SOCIAL MODELING THEORY: A POSSIBLE FRAMEWORK FOR UNDERSTANDING AVIAN VOCAL LEARNING

IRENE M. PEPPERBERG

Department of Anthropology, Northwestern University, Evanston, Illinois 60201 USA

ABSTRACT.—Currently accepted paradigms that demonstrate numerous parallels between human language learning and species-specific song acquisition fail to account adequately for recent findings on avian vocal learning. Human social modeling theory, with its emphasis on live, interacting tutors to effect otherwise inhibited learning, may provide an appropriate framework for understanding a wide range of avian vocal behaviors, particularly the flexible strategies of song acquisition observed in species generally thought to have restrictions on their learning capabilities. Social modeling theory also is proposed as a consistent schema for relating additional human and avian learning behaviors, and as a means of providing clues as to which aspects of song might serve as isolating mechanisms. Received 10 April 1984, accepted 20 May 1985.

UNDERSTANDING the acquisition of communicative competence has been a persistent problem in behavioral research, particularly with respect to the vocal mode. Similarities between human and avian vocal behaviors and the possible parallels between the underlying mechanisms governing their development (Marler 1970, Petrinovich 1972, Shiovitz and Lemon 1980, Nottebohm 1981, Marler and Peters 1981a, Payne et al. 1981) have led to the suggestion that the processes involved in avian vocal learning may provide models for the study of human language acquisition. Rarely, however, have human psychological studies been applied to understanding developmental processes in birds. I do not purport to examine all the human psychological processes that may provide functional parallels with song acquisition; however, I propose that human social modeling theory, with its emphasis on the efficacy of live, interacting social tutors, is well suited for analyzing the factors involved in the development of a wide range of avian vocal behaviors, particularly in light of recent findings that question the ability of currently accepted theories to characterize adequately the mechanisms of song acquisition (e.g. Baptista and Petrinovich 1984; Kroodsma and Pickert 1984a, b).

I will focus on three issues. First, I consider how human social modeling theory may provide a basis for understanding “exceptional” song acquisition. This term refers both to the incorporation of allospecific songs, syllables, or phrases into the repertoire of birds thought to acquire only conspecific vocalizations, and to the time-independent acquisition of vocalizations in species generally recognized as having a limited “sensitive” period for learning. Second, through a reexamination of previous findings, I demonstrate that social modeling theory is a consistent schema for relating additional human and avian vocal behaviors. Finally, I suggest how the types and degrees of social interaction needed to effect exceptional learning may provide clues as to which aspects of song could serve as isolating mechanisms.

“EXCEPTIONAL” SONG ACQUISITION AND SOCIAL MODELING THEORY

Wild birds can be divided into roughly two groups with respect to vocal behaviors: (1) species with learning generally restricted to species-specific vocalizations or confined to limited sensitive periods (for some species, both factors are significant; see Marler 1970, Kroodsma 1981), and (2) species (e.g. mimids, sturnids, psittacids) that can continuously incorporate into their repertoires numerous diverse acoustic patterns (Borror and Reese 1956, Güttinger 1974, Howard 1974, Tenaza 1976, Dowsett-LeMAIRE 1979, Feekes 1982; for reviews see Dobkin 1979, Kroodsma and Baylis 1982). However, reports exist of particular birds of the first group that have demonstrated exceptional learning in field and laboratory settings (Baptista 1972, 1983; Kroodsma 1972, 1973; Borror 1977; Baptista et al. 1981; Kroodsma et al. 1983; Slater 1983a; Baptista and Petrinovich 1984;
Kroodsma and Pickert 1984a; Payne et al. 1984; van Buskirk 1984; see also Kroodsma and Baylis 1982). In each case, exceptional learning seems facilitated by the presence of a live, interacting tutor, and the acquired vocalizations apparently are employed as effective forms of communication (Baptista 1974, Dobkin 1979, Kroodsma 1981, Cooper and Murphy 1985). However, these reports fail to posit underlying mechanisms or a theoretical framework to suggest why social interactions are particularly effective in enabling a bird to learn allospecific song or in extending the learning period. Presently accepted paradigms (see Marler 1970) suggest only that exceptional song acquisition should be an inhibited behavior, regulated by limited sensitive phases and restrictive sensory templates. These mechanisms also are presumed relevant for the development of human communicative competence. The paradigm to be examined here does not dispute the validity of these mechanisms, but seeks to extend them to account for observed exceptional behaviors by drawing parallels to principles derived from human psychological studies. Specifically, might the procedures for circumventing these mechanisms in humans function equally well in birds?

There is substantive evidence that live, social interactions with other humans are necessary to overcome existing strong inhibitions toward specific learning tasks. Moreover, the more inhibited the targeted behavior, the more intense and referential must be the social interaction to effect change. Researchers in social modeling theory (Mowrer 1966; Bandura 1969, 1971a, 1977) have demonstrated that tutor videotapes, and even verbal instruction without concrete referents (i.e. without contextual applicability), may be as effective as live, interacting human tutors for teaching naturally occurring behaviors or those for which only slight inhibitions exist. But training behaviors for which the subject has strong inhibitions (e.g. phobias: Bandura and Menlove 1968, Bandura et al. 1969, Litvak 1969, Rimm and Mahoney 1969), or those for which the subject may not be developmentally ready (Whitehurst et al. 1974, Brown 1976), requires the use of live, interacting models whose actions are strongly referential. Thus, for birds that appear to have a predisposition for species-specific learning, choice of conspecific song from a selection either of conspecific and allospecific tutor tapes or of live tutors (see Kroodsma 1978, 1981 for reviews) is not remarkable: such learning does not involve overcoming any inhibitory mechanism. Overcoming strong inhibitions toward allospecific song learning, however, is likely to require a live, interacting tutor in the presence of behavioral constraints—e.g. such situations may occur in allospecific song matching for possible territorial defense (Kroodsma 1972, Dobkin 1979, Krebs and Kroodsma 1980, Morton 1982) or during the sensitive phases of young birds in the wild that are exposed only to allospecific song (Baptista 1974, Becker 1977, Kroodsma 1981). Similarly, for species where acquisition appears strongly limited by sensitive phases, subsequent learning also should require a live referential tutor [e.g. the White-crowned Sparrow (Zonotrichia leucophrys); Baptista and Petrinovich 1984; see below].

Because capacities for, and constraints on, song acquisition are best viewed along a continuum, the particular features of the social interaction crucial for exceptional learning are likely to differ among species. Immelmann (1969) found that a nurturant interaction (feeding) was particularly effective for learning in Zebra Finches (Poephila guttata); Mundinger (1970) suggested a similar mechanism for call learning in cardueline finches. Bandura and Huston (1961) found that children who interacted with a highly nurturant model were more likely to imitate her behavior than did children with whom the same model had interacted in a “distant” manner. In contrast, Baptista and Petrinovich (1984) observed significant learning in White-crowned Sparrows during aggressive interactions with the allospecific social tutor (see below). Bandura et al. (1963) showed that children readily imitate aggressive acts, but additional human studies suggest that the influence exerted by a social tutor may be a function of the degree and not the type of emotional arousal of the participants (Bandura 1977). However, experiments can be designed to determine which features may be relevant for a particular avian species. In White-crowned Sparrows, for example, the presence of a live tutor had no major effect on song development during or after the generally accepted sensitive phase for learning if the pupil/tutor ratio was high and the aggressive interactions between pupils and tutor were limited (Cunningham and Baker 1983). These data contrast strongly with the findings of Baptista and Petrinovich (1984), in whose study pupil/tutor interactions were
far less constrained and on a 1:1 basis. Subsequent experiments continue to provide data on the aspects of the social interaction that affect learning for this species (see Petrinovich 1985).

For species with sizeable repertoires, a live tutor may not be necessary for exceptional learning. This behavior is also consistent with social modeling theory: Kroodsma and Pickert (1984b) suggested that the extensive learning capacity required for a species to acquire a large conspecific repertoire may be correlated with an overall reduced selectivity toward what is learned, i.e. a weaker innate template. Such a species, like the Marsh Wren (Cistothorus palustris), would therefore have fewer inhibitions, and consequently could learn allospecific vocalizations even from tutor tapes.

Evidence from studies with both communicative and noncommunicative humans (Brown 1976, Schuler 1980) and with psittacids (see below) also suggests a correlation between the efficacy of modeling and the referentiality, or contextual applicability, of the modeled act. Under certain conditions, strong referentiality in the components of the modeled act may compensate for a model that is not strongly interactive (Bandura et al. 1963; Bandura 1971a, b). This seems consistent with the ability of some avian species to acquire allospecific features of territorial songs but not of courtship songs from tapes. For birds such as the North American warblers (Parulinae, Muscicapidae), which appear to use different songs in different contexts (Ficken and Ficken 1967, Morse 1970, Kroodsma 1981, Kroodsma et al. 1984; cf. Lein 1978), referentiality in conjunction with the presence or absence of a live tutor may differentially affect learning of the different songs. When hand-raised Chestnut-sided Warblers (Dendroica pensylvanica) were tutored with tapes of allospecific vocalizations, the songs they developed were all classified as “territorial” (“unaccented,” “type II”; Kroodsma et al. 1983). Alternatively, a wild Chestnut-sided Warbler that appeared to learn a song from live Indigo Buntings (Passerina cyanea) seemed to use this vocalization as a courtship song (“accented,” “type I”; Payne et al. 1984). If threat or territorial defense is indeed enhanced by song matching (see Kroodsma 1979), then there may be a selective advantage for birds to learn allospecific songs with strongly referential components even in the absence of a live tutor: an isolation chamber can be considered a territory worth defending; a competitor need not be observed to be perceived as a threat. Production of neighbors’ territorial songs is most likely to cause interruption of their foraging or mating (Morton 1982), and such songs are also likely to be best adapted for sound propagation in a given territory (Hansen 1979). Furthermore, a bird may obtain reliable information as to the whereabouts of a sympatric allospecific competitor by matching the upgraded song he has learned to sing to the degree of degradation in the song heard (Morton 1982). In contrast, there would be little selective advantage in learning allospecific courtship song except under considerable social interaction or constraint.

Nonexceptional Song Acquisition and Social Modeling Theory

**Mimetic birds.**—Social interaction may affect what is learned even by habitual mimics. Such possibilities have been of particular interest to me in teaching an anomalous form of vocal behavior—referential communication in English—to an African Gray Parrot (Psittacus erithacus; Pepperberg 1981, 1983, 1985). This species is well known for its ability to reproduce non-species-specific sounds. I was the first to demonstrate referential, communicative use of such sounds (Pepperberg 1979, 1981, 1983) and, interestingly, to make extensive use of live, interactive (human) tutors. The study subject now employs English vocalizations to request, refuse, identify, categorize, or quantify more than 60 items and has acquired functional use of “no” and phrases such as “come here,” “I want X,” and “wanna go Y.” In previous studies, mynahs and parrots failed to exhibit significant vocal learning when training of non-species-specific vocalizations was performed either with tutor tapes in social isolation (Grosslight et al. 1964, Grosslight and Zaynor 1967, Gossette 1969) or through live vocal repetitions by humans with no active, referential aspects (Mowrer 1958). Todt (1975), however, successfully trained African Gray Parrots to produce, and possibly comprehend, varied human vocalizations by using a procedure that involved two live, interacting human models to demonstrate the vocal behaviors to be acquired. Recently, auditory exposure has been
shown to produce mimicry in starlings only when presented in conjunction with social interaction (West et al. 1983).

Outside of the laboratory, the effects of social interaction on learning in mimetic birds is less clear. Certain allospecific vocalizations of mockingbirds seem to be learned during aggressive interactions with other species (see Hatch 1967; Glase, as discussed by Brenowitz 1982; and Baylis 1982), and much of the observed vocal mimicry of lyrebirds and bowerbirds may be of aggressive or predatory allospecifics (Robinson 1975; see review by Klump and Shalter 1984). Might these birds, like mynahs and starlings, be incapable of significant vocal learning in social isolation (e.g. see Hatch 1967)? Note that anecdotal reports of extensive learning of human vocalizations by psittacids through ostensibly nonreferential procedures inevitably describe strong social bonds between the avian subjects and their human tutors (Amsler 1947, Hensley 1980). Previous failures to inculcate non-species-specific, referential vocalizations in mimetic birds might have been closely connected to the choice of training procedures and not due to inherent limitations in the avian subjects. As is the case with certain learning in humans (see Bandura 1977), it seems necessary to use social, referential models that actively demonstrate the interactive responses to be acquired (Pepperberg 1981). Nonmimetic birds.—Social learning theory also suggests parallels between other human and avian vocal behaviors. In these instances, social interaction appears to facilitate or occasionally to modify—rather than significantly alter—the course of development. For example, some positive correlation between high model status and efficacy of acquisition (rate and amount) of modeled behavior was found in humans (Mischel and Liebert 1967; Bandura 1973, 1977). Because avian vocalizations may provide an indication (to males and to females) of the relative "vigor" or "fitness" of the various songsters in an area (Kroodsma 1976, 1979; Catchpole 1980; McGregor et al. 1981; Yasukawa 1981), it is of interest to examine which birds serve as vocal models in the wild. Payne (1978, 1981) and Payne and Payne (1977) suggested that a "fit," "strong," or "respected" male (as, for example, demonstrated by age or the quality of his territory) is more likely to be a model than a "subordinate" male (see also Krebs and Kroodsma 1980; Baptista and Morton 1982; Payne 1982, 1983; possibly also Snow and Snow 1983). Mundinger (1979) presented tentative evidence for similar effects of social status on call learning in cardueline species during flock integration. While the adaptive significance of such behavior is not yet resolved (Payne 1983, McGregor and Krebs 1984), children who emulate peer models that exemplify successful social behavior are able to achieve and maintain a significant rise in sociometric status and peer acceptance (Hansen et al. 1969).

But what happens when no single model is presented as "superior," and the modeled patterns are diverse? Evidence from human studies predicts behavioral innovation (Bandura 1971a, b, 1977). Birds that normally have large repertoires, when exposed equally to many songs during their learning period, might therefore be expected to exhibit considerable vocal innovation [what Marler and Peters (1982c) term "invention"]. Marler and Peters (1981b, 1982a) observed such innovation in Swamp Sparrows (Melospiza georgiana) that were presented with a large number of equally emphasized songs on tape. Kroodsma (1978) and Kroodsma and Verner (1978) found that although hand-reared Sedge Wrens (Cistothorus platensis) did not copy specific songs from tutor tapes, they did develop repertoires similar to those found in wild populations. Such vocalizations could represent an innate component of the repertoire, but might they instead be innovations based on those heard on tape? Similar mechanisms also may exist in Yellow-eyed Juncos (Junco phaeonotus; Marler 1967), Zebra Finches (Innemman 1969, Böhner 1983), Northern Cardinals (Cardinalis cardinalis; Lemon 1975), Common Nightingales (Luscinia megarhynchos; Todt et al. 1979), and Marsh Wrens (Kroodsma and Pickert 1984a, b) when modeled patterns are diverse.

In humans, a familiar live model will often facilitate learning when other attempts fail (Litvak 1969, Rimm and Mahoney 1969, Bandura 1971a, Snow 1979); data on avian vocal learning suggest that certain species are incapable of significant learning without live tutors (Thielcke 1970, 1972; Price 1979; see also Slater 1983b). Sometimes a live tutor may simply be a more effective stimulus; certain birds are likely to learn more from live tutors than from tapes, and some birds choose songs of live tu-

Studies in human social modeling also have demonstrated that behaviors acquired through modeling procedures may not be produced until particular circumstances (e.g. developmental maturity, appropriate social context) release such learning (see Bandura 1971b). Similarly, birds are able to store vocalizations learned from models for long periods prior to production (Baldwin 1914, Grosslight and Zaynor 1967, Kroodsma and Pickert 1980, Marler and Peters 1982b). Even some females that normally do not sing have this storage capacity, because appropriate song may be produced under hormone treatment (Baker et al. 1982, Baptista and Morton 1982). As there appear to be neural areas of specialization in the avian brain for this storage ability (Nottebohm 1981, Nottebohm et al. 1981, Kroodsma and Canady 1982), the behavioral similarities may justify using these avian species as animal models for the human processes. [Note that stimulus-evoked production from long-term storage capacity may be consistent with findings on White-crowned Sparrows that (1) switch dialects in successive years and (2) may be “bilingual” with respect to dialect early in the breeding season (see Baptista 1974; Petrinovich et al. 1981; Marler and Peters 1982a, c; Baptista and Petrinovich 1984).]

Superficial similarities appear to exist between human infant babbling and vocal precursors to avian song (Marler 1970, Marler and Peters 1977). While the precise relationship between the human babbling or “practice” period during the onset of vocal learning and the emergence of understandable speech is yet to be ascertained (de Villiers and de Villiers 1978), in humans the overt rehearsal operations increase retention of behaviors acquired from a model (Gerst 1971, Bandura and Jeffery 1973, Bandura 1977). Whatever the neural or neuromuscular processes involved, “babbling” periods thus appear consistent with modeling theory predictions, and may serve as aids to memory and recall in both avian and human subjects.

“SECOND LANGUAGE” ACQUISITION

Noninteractive modeling (e.g. use of tapes, pictured models, etc.) has been questioned as an effective training procedure in normal (first) human language acquisition (see de Villiers and de Villiers 1978), but there is significant evidence for the efficacy of a live, interacting tutor who uses concrete examples [see Rosenthal et al. 1970, Whitehurst et al. 1974, White 1978, Snow 1979; Walker (1981) even suggests that human language is “wired into” a receptive area of the brain through differential activation of certain neurons by “social-linguistic” input]. An even stronger case can be made for the efficacy of live models for the instruction of second languages in adults. Whether parallels actually exist between sensitive phases for vocal learning in birds and humans (Marler 1970, Lenneberg 1973), a general belief is that second language acquisition, much like exceptional song acquisition, does not occur as readily or as easily in adulthood as in youth. But several studies suggest that acquisition of communicative competence (if not phonological accuracy) in a second language is greatly facilitated in adult humans by live, interacting social tutors (Asher and Price 1967, Snow and Hoefnagel-Höhle 1978, Snow 1979, Scarcella and Higa 1982). Baptista and Petrinovich (1984) proposed a similar mechanism for “bilingual” birds or those that switch dialects in later years. The assumed advantage of early exposure to a second language may be, in part, a consequence of the type, rather than the time, of exposure: children are likely to be exposed to language in an interactive, modeled form, as opposed to the static (nonreferential, repetitive) classroom situations generally employed with adults (Asher and Price 1967, Ervin-Tripp 1981). Interestingly, children exposed to a second language only through television fail to achieve communicative competence (Snow et al. 1976), but adults who learn in interactive settings appear to achieve competence comparable to, and occasionally surpassing, that of children.

DOES SOCIAL MODELING THEORY HAVE PREDICTIVE VALUE?

Research on avian vocal learning has demonstrated the existence not only of restrictive innate templates and sensitive periods, but also of the social forces that quite often act to nullify their effects. Much of the value of social modeling theory lies in its ability to integrate these apparently incompatible features into a consistent mechanism for song acquisition. Social modeling theory also may provide predic-
tions for the role of restrictive song development in various assortative mating hypotheses: by integrating features of social interaction and the song template–sensitive phase hypothesis, modeling theory may provide clues to which aspects of a species' vocal repertoire serve as an isolating mechanism.

Most attempts to correlate certain features of song with species recognition have relied on measures of song invariance. Although those songs or song features used in species recognition are usually among the least variable in a repertoire, many such invariant features have been shown not to be used for species recognition by the birds themselves (see Becker 1982). Furthermore, researchers often have failed to separate two subtle aspects of invariance, one based on strong restrictions as to what can be learned and when learning can occur, and one derived from a limited availability of suitably referential, interactive alternative models.

These considerations lead to the hypothesis that invariance of song features per se is not the relevant factor in determining species isolation, but rather that information on species recognition is related to the degree to which inhibitions exist either toward incorporation of allospecific elements or toward changes in vocal behavior after initial acquisition. In this hypothesis, the isolating mechanism may consist of specific song parameters that can be learned or altered only in the presence of a live tutor under intense referential social interaction. Other song features that are relatively invariant under normal conditions, but that could be affected by less interactive or referential means, probably would not be used in such a manner.

This hypothesis may be relevant for the present debate over the use of song dialect as an isolating mechanism in the White-crowned Sparrow (see Baker et al. 1984, Baptista and Petrinovich 1984, Petrinovich and Baptista 1984). The particular aspects of song used for male-female communication in this species have not been determined; some researchers suggest that dialects may serve in this capacity (e.g. Baker 1982, Baker et al. 1982). The prediction from social modeling theory is that dialects should not act directly as an isolating mechanism, at least with respect to male-female communication. Although dialects are specific to a given location and relatively invariant within that location, any dialect can be learned from tape by hand-raised juveniles from any other location (Marler 1970, Baptista and Petrinovich 1984), and dialects may not be as resistant to change as previously thought (Petrinovich et al. 1981, Petrinovich and Baptista 1984). Aspects of song other than dialect are likely to be important: allospecific song in the White-crowned Sparrow, which appears to occur only in the presence of live, strongly interacting tutors, could put a male at an initial disadvantage with respect to female recognition. Note that this hypothesis does not exclude dialect from acting as an isolating mechanism in other ways, such as with respect to territoriality (see above); it does, however, suggest that parameters employed in male-male (intra- or interspecific) interactions are likely to be distinct from aspects used for male-female communication (Becker 1982).

Tentative support for these ideas also may be drawn from studies of two often sympatric species, Swamp Sparrows and Song Sparrows (Melospiza melodia). Marler and Peters (1977, 1981a, b) exposed hand-raised Song and Swamp sparrows to tapes of conspecific, allospecific, and synthetic song; the synthetic songs were various combinations of allospecific and conspecific syllables, syllable arrangements, song structures, and timing patterns. Neither species learned songs from tapes of normal allospecifics. However, certain aspects of song, often different for each species, were acquired from the synthetic tapes, while others were not. Interestingly, allospecific aspects that were not learned or were only partially learned from tape (such as song structure) appeared to be at least weakly correlated with aspects used by adults, especially females, for species recognition (Searcy et al. 1981). Clearly, further study is needed to determine just how the predictive power of social modeling theory corresponds to specific elements of song acquisition.

**SUMMATION**

In both obvious and subtle ways, many factors influence learning abilities in birds and humans. Differences exist in the degree of plasticity of learning in different avian species, in learning abilities of individual humans and birds, and certainly between most human and avian behaviors. Nevertheless, I propose a synergistic relationship between human social
modeling and avian vocal learning. Through detailed investigation of both, interesting parallels may be found that might aid in the understanding of each. Future investigations must examine the adaptive significance of the effects of social interaction on various forms of learning, and the degree to which social modeling theory can be applied to other animal behaviors.

ACKNOWLEDGMENTS

This paper was written under the auspices of a grant from the Harry Frank Guggenheim Foundation and NSF grant BNS 8414483. I thank Luis Baptista and Donald Kroodsma for stimulating discussions about their research, Peter Bjarkman for references to second language learning, and Donald Kroodsma, Robert Lemon, Nancy Mangini, Eugene Morton, Robert Payne, and several reviewers for helpful comments on earlier versions of this manuscript.

LITERATURE CITED


———, & K. Payne. 1977. Social organization and mating success in local song populations of vil-


WEST, M. J., A. N. STRAUD, & A. P. KING. 1983. Mimicry of the human voice by European Starlings:

