Auditory Scene Analysis by Songbirds: Stream Segregation of Birdsong by European Starlings (*Sturnus vulgaris*)

Stewart H. Hulse, Scott A. MacDougall-Shackleton, and Amy B. Wisniewski
Johns Hopkins University

Three experiments examined the capacity of European starlings to segregate perceptually 2 superimposed, intermixed auditory stimuli. The stimuli were 10-s song samples from 2 of 4 songbird species: European starling, brown thrasher, mockingbird, and nightingale. The birds first learned a discrimination between the intermixed song pairs. Then, they maintained the discrimination with novel song exemplars in the mixtures and when song stimuli for each species were presented alone. Performance fell, but remained above chance, when song pairs were mixed with the dawn chorus of bird song. The results show that starlings were identifying the songs of individual species within the baseline superimposed song pairs, a process of auditory stream segregation and scene analysis (A. S. Bregman, 1990).

In this article, we demonstrate that a species of songbird, European starlings (*Sturnus vulgaris*), has the perceptual capacity for auditory scene analysis. By *auditory scene analysis*, we mean that the species can discriminate and selectively attend to one ongoing sound that occurs simultaneously with one or more other sounds. The *cocktail party effect*, by which one identifies one speech source amidst a mixture of other speech sources (Cherry, 1953, 1954; Moray, 1959; Wood & Cowan, 1995), is a familiar example of auditory scene analysis in human perception.

Bregman (1990) summarized a wealth of such evidence from the literature on human auditory perception and on speech and music perception showing that people organize their ongoing perceptual world into an auditory scene in which certain sources of sound are heard separately from others. *Auditory scenes* consist of multiple auditory *objects* that are each unique, organized, collections of sounds that share, for example, a common pitch, spectral composition, temporal organization, and location in space. The auditory scene may consist of many of these objects that often overlap in time. Common auditory scenes include a walk through the springtime forest, a crowded room with the intermixed voices of many people, a symphonic rendition with the overlapping sounds of many different musical instruments, and so on. The total auditory scene becomes parsed into different auditory objects (different animal and forest sounds, different voices, different instruments) on the basis of many principles, of which Gestalt organizing principles are often paramount. Thus, in Cherry's work, for example, female and male voices could be segregated in perception because the voices differed in overall pitch height and spectral composition and could be grouped perceptually on the basis of those acoustic cues as the speech samples ran their course.

Although the capacity for auditory scene analysis has apparently not been demonstrated experimentally in animals other than man, nonhuman animals face the same problem that humans do in this regard. It goes without saying that birds, song birds especially, use acoustic information to make their way in the world. Catchpole and Slater (1995, chapters 6, 7) summarize a number of functions for which birdsong seems to be especially adaptive, ranging from territorial defense, through species and individual recognition, to sexual selection. Most important, sounds uttered and perceived by birds are subject to all the distortions characteristic of any meaningful signal embedded in environmental noise. Consider, for example, the dawn chorus, the conglomeration of noise from the contributions of many species singing at the break of day (Catchpole, 1973; Kacelnik & Krebs, 1983). If birdsong is indeed involved in breeding and territory defense, songbirds must be able to select and identify important signals regarding species, gender, and so on from that cacophonous background if they are to survive and reproduce. Although, a priori, that possibility would seem to be a virtual certainty (Lewis, 1991), it should not necessarily be taken for granted. We do not yet have computer algorithms that will disentangle consistently and accurately superimposed samples of human speech, for example, and also the capacity for auditory stream segregation may be uniquely a high-order, human perceptual process. If, as seems likely, auditory scene analysis does occur in nonhuman animals, however, the question remains whether the process functions as it does in humans.

To start answering that question, biologically relevant stimuli seem especially appropriate. Therefore, we designed experiments with starlings in which samples of two species' songs, A and B, were superimposed and intermixed to form a single sound sample; samples of two other species songs,

C and D, were similarly combined to form a second sound sample. Starling song itself was included as the A song; whereas songs B, C, and D were the songs of other species. For example, starling song was intermixed with brown thrasher (*Toxostoma rufum*) song for one stimulus, and mockingbird (*Mimus polyglottos*) song was intermixed with nightingale song (*Luscinia megarynchos*) for the other stimulus. Starlings were then trained to discriminate 15 exemplars of A and B superimposed pairs of sounds from 15 exemplars of C and D superimposed pairs of sounds. Following initial discrimination training, transfer experiments then analyzed the basis on which the initial discriminations were made.

If auditory stream analysis occurred during initial training, then perceptual segregation of the song of one or more individual species' songs from the mixtures should occur. That analysis would presumably take place on the basis of whatever acoustic properties distinguished the songs as auditory objects. It seems likely that the distinctive flow of acoustic information through time, grouped by pitch and spectral structure, for example, would be especially important, but that is an experimental question. Of course, given stream segregation, further research would have to identify the acoustic basis on which the process was, in fact, taking place.

One prediction to test in the transfer experiments would be that the birds ought to maintain the discrimination on the basis of either component song of both song pairs used in initial training because they, presumably, identified and used each song as the effective discriminative stimulus on some subset of the training trials. At the same time, the ability to parse the songs might be biased by the songs' immediate biological significance; conspecific song might be a more salient part of the auditory scene than heterospecific song; therefore, conspecific song might be attended to and discriminated, whereas heterospecific song would not. In terms of scene analysis, the former might take place because of the simple, low-level acoustic properties of the sounds moving in time: *primitive scene analysis* in Bregman's (1990) terms. To the extent that learning modulated perception, however, the discrimination might be controlled by properties of the acoustic stimulus as a functional signal: *schema-driven scene analysis* in Bregman's terms.

Experiment 1: General Method and Baseline Discrimination

Starlings were first shaped to peck for food in an operant task, then trained in a baseline discrimination to distinguish between superimposed combinations of two species' song from superimposed combinations of two other species' song. Given the initial discrimination, a series of additional experiments then used transfer procedures to test for the starlings' ability to perceive and to process otherwise song components of the baseline stimuli.

Method

Birds

The European starlings were six male, wild caught, adult birds captured approximately 40 miles north of Baltimore, MD. They were housed in individual cages in an aviary of about 50 starlings and 20 zebra finches (*Taeniopygia guttata*) throughout the experiment. Fluorescent lights with a daylight color spectrum in the aviary were turned on and off automatically at times that mimicked the annual changes in photoperiod at Baltimore. Prior to the experiment, the starlings were fed a commercial poultry starter (Purina, St. Louis, MO) supplemented with raw spinach twice a week. All animals had ad lib access to water and an appropriate grit.

When not under test in the experiment, the birds were housed in the aviary. They were maintained at approximately 85% of their ad libitum body weight. The birds obtained much of their daily ration of food during testing in the experiments. Supplemental food was given in the aviary following the daily session as necessary to maintain the appropriate weight.

Apparatus

The starlings were trained and tested with a standard operant conditioning apparatus housed inside an IAC (Industrial Acoustics, New York, NY) Model AC-3 sound-attenuating chamber (80 cm wide × 60 cm high × 60 cm deep). A response panel was suspended from the ceiling of the chamber, forming one end of the test cage. The panel consisted of three horizontally aligned, translucent response keys. The keys were 2 cm in diameter and were spaced 6 cm apart, center to center. Gerbrands (Cambridge, MA) Model G5610 food hoppers delivered food to a 6.0 cm × 4.5 cm opening located 4.5 cm below the right and left response keys. The bird could reach food through the openings. The birds were transported to the testing room in weld-wire cages, measuring 28 cm × 20 cm × 30 cm, that were attached to the response panel. The birds gained access to the response panel once a sliding door on the transport cage was removed. Pecks to the response keys were detected by microswitches. Indirect illumination of the test chamber was provided by a houselight located behind a translucent screen on the back wall of the chamber. A Bose (Framingham, MA) Model 101 speaker was located immediately above and behind the response panel. Stimulus, response and reinforcement contingencies were controlled on line by a PC-type computer. The computer also recorded data.

Stimuli

Stimulus components in the experiments consisted of segments of birdsong, approximately 10 s long. Most of the experiments used combinations of stimulus components that consisted of superimposed segments of birdsong from two or more species. The superimposed stimuli were also approximately 10 s long. Tape recordings of European starling, northern mockingbird, nightingale, and brown thrasher song were used. Ten-second samples from a tape of the "dawn chorus" were also used in addition to the songs of individual species. The tapes of individual species' songs and samples from them were selected to minimize background noise as much as possible. Tape recordings of starling, northern mockingbird, and brown thrasher songs, and of the dawn chorus were obtained from the Borror Laboratory of Bioacoustics (Ohio State University, Columbus, OH). Tapes of starling song were

kindly supplied by MarthaLeah Chaiken (Institute of Animal Behavior, Rutgers University). Tapes of nightingale song were kindly supplied by Henrike Hultsch (Institute für Verhaltensbiologie, Freie Universität, Berlin). The species used were chosen because these birds are all continuous singers with complex songs and large repertoires. Furthermore, mockingbird and brown thrasher songs are quite similar and would provide a stringent test of song discriminability. Nightingale song was chosen because the species is not native to North America and its song was certainly unknown to starlings from Maryland. Song segments for each species were recorded from at least four different birds, but the starling exemplars did not include song from the birds participating in the experiment. The song combinations that were superimposed and mixed to form the stimuli, together with the *n*s exposed to each combination, are shown in Table 1.

All song segments were digitized at a sampling rate of 22050 Hz using a Macintosh IIx computer with SoundDesigner II (version 2.1) software and a SoundTools 16-bit analog-to-digital converter (Digidesign, Menlo Park, CA). Digitized sounds were edited to remove long silent gaps between songs. If a gap was greater than 3 s, it was reduced to a maximum of 2 s. Each sound segment ended with the termination of a song note, so there was some small variation in the 10-s length of the song stimuli. However, all segments were 9.0 to 10.0 s long.

To create stimuli consisting of dual superimposed birdsongs, we used the mixing board function in the SoundDesigner II software. All sound files were then resampled at 20 kHz and converted to PC sound files using Soundview software (Peabody Computer Music Department, Johns Hopkins University, Baltimore, MD). Because of the nature of the stimuli, there were large fluctuations in sound levels. We normalized the peak amplitude of each song segment digitally and again normalized the peak amplitude of mixed songs before converting the stimuli to files stored digitally on the PC computer. Some variation in loudness levels still existed among the stimuli when played in the experimental chamber, but all stimuli in the baseline training had peak loudness of 65 to 68 dB as measured with a Rion (Tokyo, Japan) sound level meter placed at the approximate level of the bird's head directly in front of the center key in the apparatus.

Representative sonograms of stimuli appear in Figure 1 for (a) European starling, (b) nightingale, (c) European starling mixed and superimposed on nightingale, and (d) European starling and nightingale superimposed on the dawn chorus. The figure provides a visual analogue of the increasing complexity of the acoustic stimulus as additional songs were mixed together.

Table 1

Song Pairings and the Correct Response (A or B) for the Superimposed Stimulus Mixtures During Baseline Training of Experiment 1

No. of starlings	Song pairs with starling song Peck A	Song pairs without starling song Peck B
2	Starling + brown thrasher	Nightingale + mockingbird
2	Starling + nightingale	Mockingbird + brown thrasher
2	Starling + mockingbird	Nightingale + brown thrasher

Note. The correct key, A or B, assigned to each song pair category was counterbalanced across subjects.

Procedure

Shaping. All starlings had prior experience in operant conditioning experiments involving arbitrary, synthetic stimuli, but no birdsongs. Shaping for starlings in the experiment lasted two sessions. Trials during these sessions consisted of two types that were selected randomly ($p = .50$). In one type, a peck to the center key followed by a peck to the right key was rewarded with food by illuminating the right hopper opening and raising the right hopper for 2 s. Pecks to the left key on this trial type were neither reinforced nor punished. The second trial type rewarded pecks to the center key followed by pecks to the left key by providing food access at the left food opening. The shaping procedure assured that the starlings were trained to start a trial with a peck to the center key and then to move to one or the other of the side keys to obtain reinforcement.

Food rewards for starlings in shaping and throughout the experiments consisted of the standard diet used in the aviary. Throughout the experiments, each bird was tested in the experimental apparatus for 2 hr per day, 6 days per week.

Baseline training. During training, the birds learned to discriminate between two sets of stimuli. One set (e.g., starling–thrasher; see Table 1) was constructed from five starling song segments mixed with five brown thrasher song segments in unique combinations. Fifteen superimposed stimuli were then selected from the total set of 25 possible combinations. The other set of 15 stimuli (e.g., mockingbird–nightingale) were drawn from the 25 possible unique combinations of five mockingbird song segments mixed with five nightingale song segments. The three possible combinations of song type pairs were counterbalanced across starlings, with 2 birds tested under each combination. Half of the birds were trained to peck the right key in response to the set containing the starling song and to peck the left key in response to the nonstarling set. The other birds were trained with the opposite key assignment.

The birds initiated a trial by pecking the center key. In response to this, one of the stimuli was randomly chosen and played once through the speaker. Pecks to any key during the first 5 s of the stimulus presentation (observation period) had no consequences. Following the observation period, a peck to the key appropriate for the stimulus stopped the stimulus and produced food reward. If the bird pecked the inappropriate key, the stimulus stopped and the house lights were turned off for a 10-s "timeout." The same stimulus was then used for the next trial (a correction trial). Correction trials were repeated until a correct response was made.

Measures of performance. Discrimination was measured by calculating the proportion of correct trials divided by the total number of trials (excluding the correction trials) for each daily session. Birds were trained with this procedure until they performed at a level above 85% correct for five consecutive sessions. At that time, the birds produced from 36 to well over 200 total trials per session.

Results and Discussion

Five starlings learned the baseline discrimination, reaching the learning criterion in 21, 35, 24, 20, and 22 sessions, respectively. One starling (P58) failed to meet the criterion and was transferred to other conditions after 36 sessions. At that time, the starling was performing above chance, with a mean over its last five sessions of 71% correct overall.

None of the starlings showed any greater skill in learning to discriminate the stimulus containing the starling song
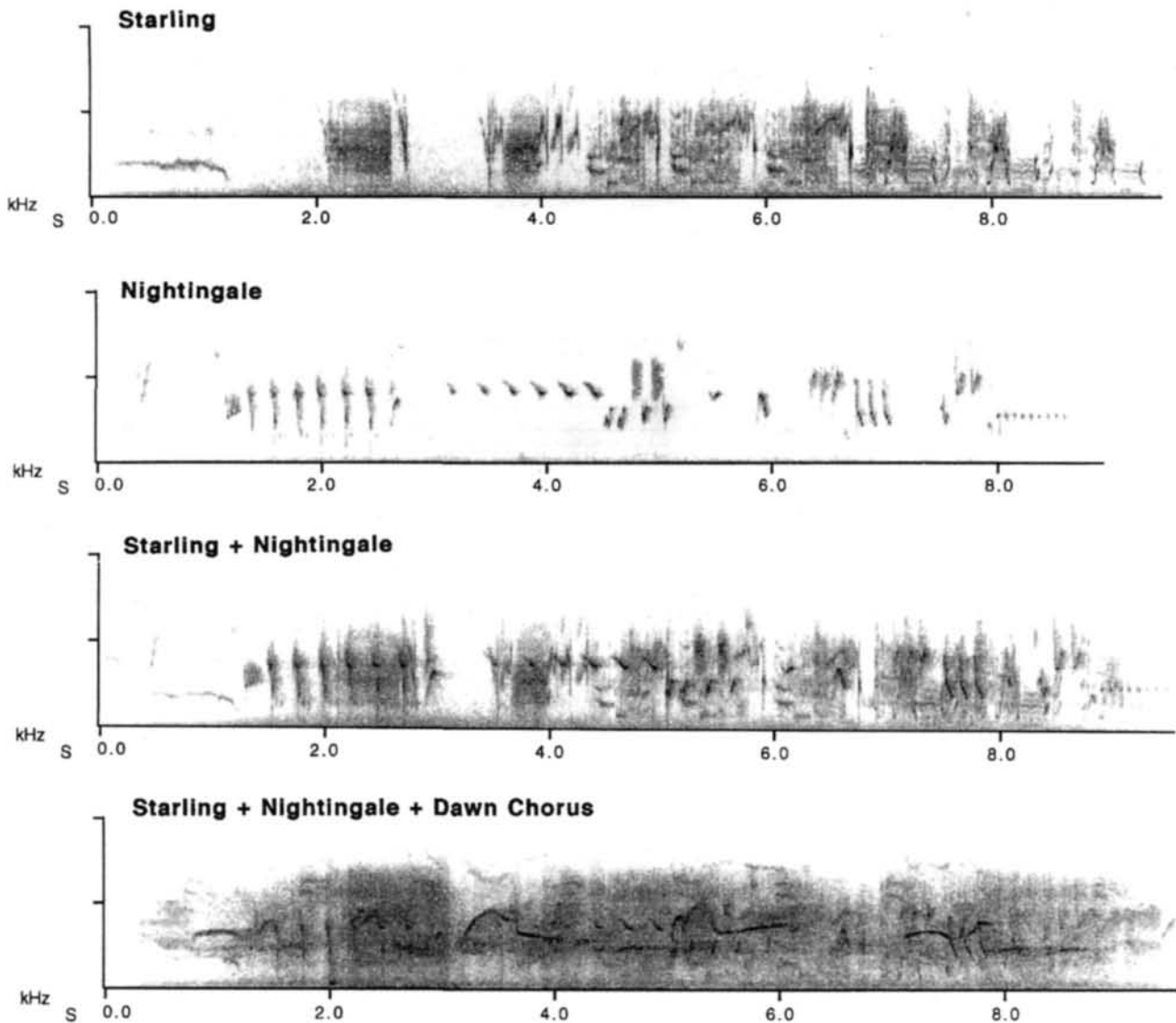
Figure 1. Representative sonograms of the song of the European starling, nightingale, European starling superimposed on nightingale, and European starling and nightingale superimposed on the dawn chorus. Frequency (in Hz) appears on the ordinate, and time (in seconds) appears on the abscissa. The high-amplitude tracing in the dawn chorus sonogram, which is especially salient after about 3 s into the sonogram, is the song of a wood pewee (*Contopus virens*) that was close to the microphone when the recording of the dawn chorus was made.

than they demonstrated for any other song combination. Except for P58, which responded 60% correct on the starling–thrasher combination and 82% correct for the nightingale–mockingbird combination, all the starlings were about equally adept at correctly identifying stimuli containing the starling song (range 84% to 96% correct) as compared with stimuli that did not (range 85% to 93% correct). Similarly, inspection of improvement in performance over successive sessions revealed no special advantage of including conspecific song in the stimuli for the starlings. For example, there were no consistent differences

in number of sessions to the first session in which the birds achieved 75% correct overall (range 6 to 31 sessions). That result was true for any song combination, whether the combination contained starling song or not.

Although the initial baseline discrimination was learned by the starlings, indicating the stimuli within the two 15-exemplar sets of superimposed bird songs could be discriminated and classified, the results do not tell us on what basis the birds were making the discrimination and classification. That issue, and some related questions, were addressed in Experiments 2, 3, and 4.

Experiment 2: Novel Song Combinations

Perhaps the birds had learned the baseline discrimination by singling out a significant acoustic feature of the super-imposed songs unique to each stimulus and simply memorizing the feature. That could happen if the features were distinctive merely on an acoustic basis (e.g., a unique "blip" of some particular frequency or spectral structure). Although the birds would have to memorize at least 30 such distinctive features (at least one for each of the 15 stimuli in the two sets to be discriminated) to learn the discrimination task to criterion, it is possible that they did so. On that basis, one could argue that the birds were not perceiving the stimuli as bird song with the structural characteristics typical of that behavioral trait. Experiment 2 addressed these issues by transferring the birds to an entirely new set of superimposed song stimuli. If the birds had memorized features unique to each of the stimuli in the original baseline set, their performance on the novel stimuli ought to be impaired because the features, in their original form at least, would be missing. However, if the birds had learned the discrimination on the basis of some more overarching structural property of song characteristic of one (or more) species, they ought to generalize to the new set of stimuli rapidly. That process, in turn, might depend at least to some extent on the ability to hear out the song of one (or more) species from the mixture of two songs.

The transfer experiment was also an incidental test for the superimposed songs (or some perceptual property associated with them) acting as exemplars from an open-ended stimulus category (Astley & Wasserman, 1992; Bhatt, Wasserman, Reynolds, & Knauss, 1988; Gray, 1995; Herrnstein, 1990; Wasserman, Kiedinger, & Bhatt, 1988). Thus, the individual song stimuli were all different acoustically, but the birds appeared to treat them as exemplars from two stimulus "classes" because they could assign them appropriately to the right or left response keys during the initial discrimination. If the stimulus sets had become true open-ended stimulus categories, presumably on the basis of properties common to all the stimuli in each category, the birds should show immediate transfer of discrimination to novel exemplars of the same types.

Method

Animals and Apparatus

The same birds served, and the same apparatus was used as in Experiment 1.

Procedure

The procedure was essentially the same as that used in the baseline discrimination of Experiment 1. The only difference was that new superimposed combinations of birdsongs were constructed. Five novel song segments from the same singers for each of the four species were uniquely combined to produce 15 new stimuli in both the starling and nonstarling sets.

The novel stimuli were introduced all at once at the beginning of

a daily session. The new stimuli replaced the original baseline stimuli. The trials in each daily session proceeded identically to the baseline training trials. The birds pecked right or left for song combinations. They were rewarded with food for correct choices and punished with 10-s time outs for incorrect responses. Correction trials were used as before. The birds were maintained on this procedure until they demonstrated 85% correct responses for five consecutive sessions. They were then returned to the baseline discrimination for at least five sessions. At that time, the ratio of reinforcement for correct responses was reduced from 100% to 90% and the birds' performance was restabilized at or above the 85% correct criterion in preparation for the experiments to follow.

In all other regards, the procedure and the preparation of results for analysis were the same as in baseline training. Like Experiment 1, the slowest starling produced at least 35 total responses during any session entering into the analysis, and total responses ranged across birds to over 200 responses per session.

Results

All six starlings showed virtually immediate transfer to novel superimposed song combinations that they had never heard before. Figure 2 shows their performance on the baseline discrimination trials averaged over the five sessions just preceding the transfer and the first session of the transfer itself.

An analysis of variance (ANOVA) compared mean percentage correct for the last five pretransfer sessions and for the first posttransfer session for each of the three combina-
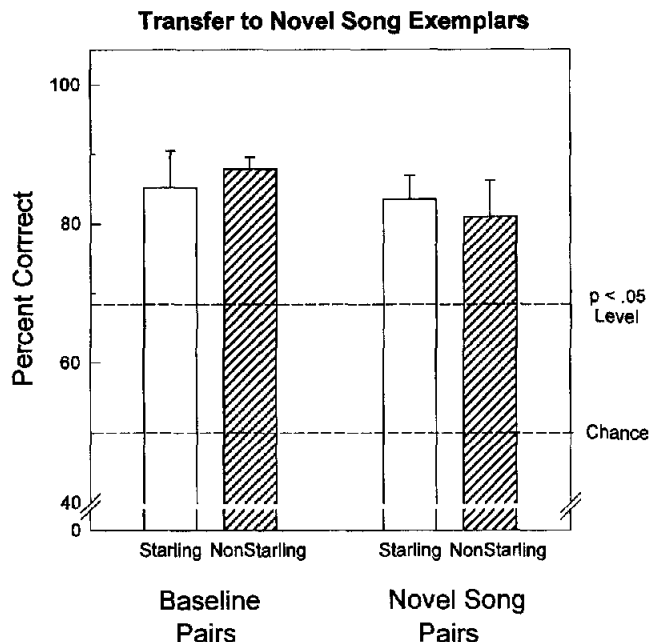
Transfer to Novel Song Exemplars

Figure 2. Mean percentage correct (±*SEM*) for the discriminative stimuli over the last five baseline sessions with the original baseline pairs of superimposed song stimuli and on the first transfer session with novel exemplars of the baseline song stimuli. The data are divided into those pairs of stimuli that contained starling song and those that did not. The chance (50%) and the confidence interval above chance (*p* < .05) of 69% are displayed in the figure.

tions of starling song with other species' song. The analysis showed that neither main effect nor their interaction was statistically reliable (all $Fs < 1.0$).

The binomial was used to ascertain whether performance on the first session following the transfer was reliably above chance. Starling 283 completed the fewest number of trials (28) during this session (the other birds ranged from 36 to 216 trials), and 28 observations were used as a conservative n to calculate an upper confidence interval ($p < .05$) of 69% for Transfer Session 1 performance relative to chance of 50% correct. Figure 2 shows the upper confidence interval for posttransfer performance on Session 1. The birds were performing reliably above chance.

Discussion

There are several important conclusions to be drawn from these results. First, the starlings readily transferred the discrimination to completely new exemplars of song combinations. This result indicates that the birds had not merely memorized some particular sound feature or features unique to each of the baseline songs and solved the original discrimination that way. That is so because such features presumably changed substantially with the introduction of the new exemplars.

Discrimination performance was the same regardless of the combination of starling song with any of the other species' song. This fact means that all species' song (conspecific song in particular) was equally well discriminated in this transfer.

Finally, although Experiment 2 does not tell us whether or not the birds were segregating and discriminating one or more species' songs as such, the results are certainly in line with that proposition. The question therefore remains as to whether or not the birds were, in fact, discriminating the songs of individual species and making the categorization on that basis, and if so, which songs they were using. That issue was addressed in the next experiment, as we shall see.

The data of Experiment 1 are consistent with the proposition that the starlings formed an open-ended concept for the classes of songs combined to form the baseline discriminative stimuli. That is so because they quickly transferred to novel exemplars from the baseline stimulus classes (Herrnstein, 1990). Of course, further research will be necessary to establish open-ended categorization firmly because Experiment 1 did not contain necessary control groups, that is, groups trained on sets of the same stimuli assigned arbitrarily to pseudocategories (Astley & Wasserman, 1992; Bhatt et al., 1988). Some data obtained by Gray (1995) with designs appropriate for the study of concept formation demonstrate unequivocally, however, that starlings can form open-ended concepts, at least, for starling song stimuli, although Gray's stimuli were much shorter than the ones used here. Most of the work on concept formation in both humans and nonhumans has been based on visual stimuli. The available data, sparse though they are at present, suggest that studies of concept formation with acoustic stimuli, especially naturally occurring acoustic stimuli, will be fruitful.

Experiment 3: Discrimination of Unmixed Single-Species Songs

If starlings were performing an auditory scene analysis and hearing out the song of individual species in the baseline discrimination, they ought to be able to maintain the discrimination if they heard just one of the species' songs in isolation. That is, for example, if starling and brown thrasher songs were superimposed in the baseline discrimination, the species' songs ought to be assigned to the same response choice if the individual songs were played by themselves—on the assumption the starlings were segregating and parsing the two songs in the baseline discrimination and, functionally, hearing them as separate auditory objects. However, to the extent that the starlings were responding to the superimposed songs as a single fused structure of acoustic information, the use of exemplars based on single-species song should change that perceptual structure dramatically. As a consequence, the discrimination ought to be substantially impaired if not lost. Experiment 3 tested these predictions.

Method

Animals and Apparatus

The same birds served, and the apparatus was the same as that of the earlier experiments.

Procedure

In most regards, the procedures were also the same as those of the earlier experiments, with the following exceptions. Song stimuli for the probe trials were the same five 9- to 10-s song segments used for each of the four species songs used during baseline training. Here, however, the song segments were not superimposed and mixed together. Instead, the song segments from the four species were used by themselves, providing 20 probe stimuli, five for the song of each species. It is important to note that, although never heard in isolation, the songs were potentially familiar to the starlings from baseline and later training, on the assumption that the birds had isolated the songs perceptually and heard them as separate auditory objects.

Each day during Experiment 3, the test birds were maintained on baseline discrimination with the original baseline stimuli of superimposed songs for 90% of the trials. The other 10% of the trials (randomly selected) were probe trials in which the birds heard single-species songs. On probe trials, stimuli were selected randomly from the set of 20. Responses, regardless of choice, were treated nondifferentially. That is, a response to either key was rewarded with a 2-s access to food on 90% of the probe trials. Birds were tested with this procedure over a number of sessions sufficient to assure they had experienced at least 20 probe trials for each of the four single-species songs. The test birds were then returned to the initial baseline discrimination for at least five sessions before the beginning of the next experiment.

In all other regards, the procedure and the preparation of the results for analysis were the same as those in Experiment 1. The smallest number of trials, both correct and incorrect, produced by any bird for baseline stimuli was 44, and that total increased to over 200 trials for other birds.

Results

Starlings classified all the isolated song types with above-chance performance. There was also evidence that they were more accurate in identifying isolated conspecific starling song than songs of the other three species.

Figure 3 displays mean performance on the baseline stimuli and mean performance on the probes over the sessions (range 5 to 9) required to reach 100 probe trials for all the probe stimuli taken together. At this time, each starling had responded to at least 20 probes for each isolated song type. The figure shows that performance was above chance (50% correct) for both the baseline stimuli and each of the probe stimuli. However, starlings responded more accurately to conspecific song probes than to the probes of any other species' song. In fact, the starlings classified their own isolated song just as accurately as they classified the superimposed stimulus combination that contained starling song in the baseline stimulus set.

An ANOVA on mean percentage correct collapsed across the baseline stimuli and mean performance collapsed across the probe stimuli showed that baseline performance was reliably more accurate in general than probe performance, $F(1, 5) = 12.158, p < .05$. An ANOVA on mean percentage correct for probe stimuli taken by themselves was also reliable, $F(3, 9) = 5.62, p < .05$. Subsequent correlated t tests showed that mean performance for the starling probe was reliably better ($ps < .05$) than that for the nightingale probe, $t(5) = 4.01$, and the brown thrasher probe, $t(5) = 4.47$, but (because of chance performance by one bird) not the mockingbird probe, $t(5) = 2.55, p > .05$.
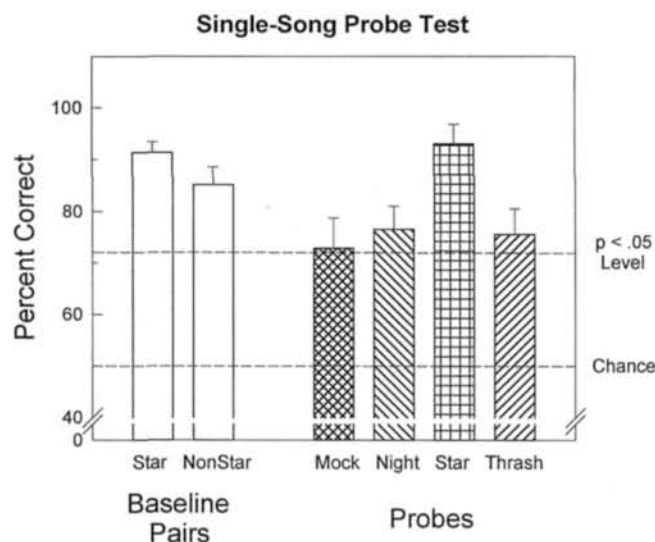

Figure 3. Mean percentage correct ($\pm SEM$) for the discriminative stimuli on the last five baseline sessions for the baseline pairs of superimposed song stimuli and for the probe stimuli, composed of song exemplars for individual species, that were introduced on 10% of the trials during probe sessions. The chance (50%) and the confidence interval above chance ($p < .05$) of 72% are displayed in the figure. Star = starlings; nonStar = nonstarlings; Mock = mockingbirds; Thrash = brown thrashers.

The performance on all probes was reliably ($p < .05$) above chance of 50% as determined by a binomial test based on an n of 20 observations (the minimum number obtained for any bird for any of the four single-song probes). The upper confidence boundary of 72% correct for $p < .05$ appears in Figure 3. Performance on all probes was reliably above chance.

Discussion

The results of Experiment 3 demonstrate that the starlings were capable of maintaining the discrimination that they had learned in baseline between the sets of superimposed songs when exemplars from only one species' song in the pair was played in the probe tests. This was true for all probe song types. Also, the starlings were more accurate in classifying their own song in the probes than the songs of the other three species, reliably so for all but the mockingbird probes (because of poor performance by 1 bird on the mockingbird exemplars). Finally, performance for all probes was reliably ($p < .05$) above chance of 50% correct.

The starlings could maintain their accurate song classification on the basis of acoustic information contained in just one of the component song stimuli. These results are perfectly consistent with the proposition that the starlings were identifying the songs of the individual species from trial to trial in the baseline stimuli and solving the classification task on that basis. That is, the starlings demonstrated auditory stream segregation of the four individual songs comprising the two sets of baseline stimuli.

Given good evidence for stream segregation for the starlings, the results do not tell us, in any definitive manner, precisely what the cues were in the isolated songs that permitted stream segregation to take place. For humans, stream segregation may take place on the basis of differences in spatial location, pitch, spectral structure, and loudness among auditory objects, and any or all of these parameters, except for spatial location of the auditory objects (because all stimuli came from one speaker, hence one location in space), may have been at work in this experiment. It is also possible that the starlings were listening to some memorized acoustic feature of the baseline stimuli in the probes because the same exemplars for each species' song were used in the baseline and probe stimuli. If so, however, that feature or features would still have to be recognized when removed from the "background" features of the other species' song in the baseline song mixture. This in itself would be a case of stream segregation.[1]

Experiment 4: Discrimination With Added Auditory Objects

In the real world, songbirds must be able to distinguish songs and other ecologically relevant auditory objects

---

[1] In identical work preliminary to the experiments reported here, however, we did test starlings with probe trials using stimuli that they had never heard before. The results were virtually the same as those reported in Experiment 3.

amidst a much more complicated acoustic melange than the song-pair combinations of Experiments 1 and 2. Therefore, it is of interest to see how a discrimination between baseline stimuli of the types used in the first two experiments might fare when mixed with a much "noisier" acoustic signal. One way to do this would be to bury the song pairs in some form of artificial noise, such as white noise, that would potentially mask all frequencies in the pairs. That would be merely an acoustic masking experiment, on the basis of a psychophysical model, of relatively little interest from the point of view of auditory streaming (Bregman, 1990). Another, more interesting way to add "noise" to the baseline song stimuli would be to mix them with the complicated signal afforded by the dawn chorus: the potpourri of birdsong and other natural sounds characteristic of an early spring morning. This procedure is not only more ecologically relevant, it also introduces noise by adding many potential auditory objects from which the target objects must be segregated. As a collection of many auditory objects, the dawn chorus represents the acoustic world that the birds must actually deal with when they have to single out functional sound sources of immediate survival or adaptive value. To this end, Experiment 4 was a simple transfer experiment in which the starlings were confronted with the same baseline stimuli used in the earlier experiments, this time mixed with recorded samples of the dawn chorus.

If the starlings could still hear out the baseline stimuli in the mixture and classify them accurately, that would be evidence for a segregation process in their auditory perception. This outcome would be the most interesting. If they lost the discrimination, however, that would be evidence that the added "noise" interfered effectively with those cues in the overall signal that permit stream segregation to take place. Clearly, there must be some form of an added noise, like that of the dawn chorus, that would be sufficiently complex or of a sufficiently high level to render signals of the types under study here effectively indiscriminable.

Method

Animals and Apparatus

The same birds served, and the apparatus was the same as that of the earlier experiments.

Procedure

The procedure in this experiment was like that used in Experiment 2 with the exception of the stimuli. Here, the same superimposed song pairs used as baseline stimuli in Experiment 1 were mixed with samples of the dawn chorus. The method of mixing the stimulus components was described earlier under Experiment 1.

As in Experiment 2, the starlings were transferred at the beginning of a daily session from the baseline stimuli to the superimposed paired stimuli now mixed with the dawn chorus. Responses were reinforced or not on the basis of the baseline stimulus key assignments. Incorrect responses were followed by correction trials. The starlings were maintained on the transfer for a minimum of 5 days.

In all other regards, the procedure and the preparation of the data for analysis were the same as in Experiment 2.

Results

The starlings maintained their discrimination performance when the superimposed song stimuli were mixed with the dawn chorus, although there was a decrease in overall discrimination accuracy. The relevant data appear in Figure 4, which reports mean percentage correct for the last 5 days of baseline prior to the transfer and the mean percentage correct on the first day of transfer.

In all cases, both before and after the transfer, discrimination accuracy was above chance. An ANOVA including baseline versus transfer performance, and the three combinations of starling song with the other species' song showed a reliable effect only for baseline versus transfer, $F(1, 5) = 10.70$, $p < .05$. There was no evidence that exemplars containing starling song were especially privileged in making the transfer discrimination.

Transfer performance was reliably above chance for exemplars that contained starling song and for those that did not. A binomial test (again using the smallest number of trials—32—completed by any bird on the first day of transfer as a conservative value) established the upper confidence interval ($p < .05$) at 67% correct.
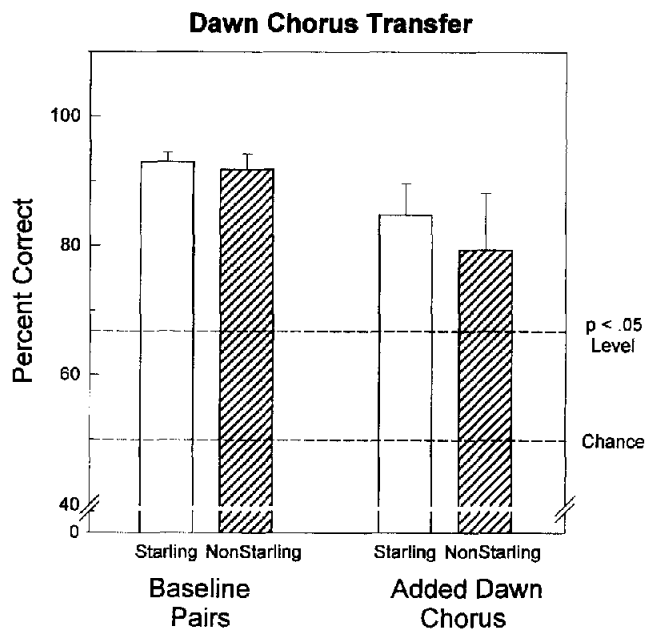

Dawn Chorus Transfer

Figure 4. Mean percentage correct ($\pm SEM$) for the discriminative stimuli when the baseline superimposed song stimuli were mixed with recordings of the dawn chorus. Results appear for means over the last five baseline sessions and the first transfer session. The starlings showed a decline in accuracy with transfer to the new stimuli, but nevertheless respond above chance. The chance (50%) and the confidence interval above chance ($p < .05$) of 67% are displayed in the figure.

Discussion

These results add to those of Experiment 3 indicating that the starlings were able to segregate the various species' songs in spite of the addition of a substantial number of new ecologically relevant auditory objects. Again, the data do not tell us on what acoustic basis the starlings were making the discrimination, but the mechanism must be some analytic process that is very robust in the face of an added functional stimulus that ought to provide substantial interference with the baseline stimuli.

At the same time, it is important to remember that we do not claim that the starlings could maintain the discrimination in the face of any true acoustic masker, whether functionally relevant like the dawn chorus or not. No doubt, as we have already noted, suitable increases in the loudness or complexity of any stimulus used as a true masker would eventually render the discrimination impossible.

General Discussion

Taken together, the experiments provide evidence for stream segregation and auditory scene analysis by European starlings. Thus, the data extend to songbirds the phenomenon of auditory scene analysis observed in humans (Bregman, 1990).

Although the baseline discrimination between two 15-exemplar stimulus sets could have been learned by memorizing one or more acoustic features characteristic of each baseline stimulus, the immediate generalization of the discrimination to novel exemplars suggests that a more global perceptual process took place spanning each exemplar and overarching all the stimuli in the two baseline sets. Most important, the data are entirely consistent with the proposition that the starlings were hearing out the songs as functional auditory objects and responding to them individually on that basis. Experiment 3 showed directly that the starlings could make the discrimination on the basis of *any* of the component species' songs that had been melded to construct the baseline stimuli; in the process, there was an interesting bias to do better with conspecific starling song. Finally, Experiment 4 showed that the baseline discrimination could be maintained when the baseline exemplars were further melded with the dawn chorus of birdsong, which is a complex multitude of individual species' songs with potential functional relevance.

Implications

Given the evidence for auditory scene analysis in a songbird, there are several perceptual principles and other facts associated with that process worth noting in some detail. The information may be helpful, potentially, in seeking other parallels between human and nonhuman auditory scene analysis and in additional analysis of the auditory scene by birds and other nonhumans animals. However, this is not the place to describe auditory scene analysis in detail. Besides Bregman's (1990) seminal book, thoughtful recent discussions, again based on human perception, also appeared in Darwin and Carlyon (1995) and Handel (1995).

Gestalt Organizational Principles

First of all, auditory scene analysis stresses the importance of Gestalt principles of perceptual organization and grouping in the formation of auditory objects and their participation in auditory streams. Thus, sounds of any kind (such as the sounds forming a birdsong syllable or the units forming an entire song) group together perceptually on the basis of temporal proximity. Also, sound patterns containing alternating sequences of different pitches will, under the right temporal conditions, group together to form separate auditory streams on the basis of pitch, one stream of high pitches and another of low pitches. In nature, for example, if songs from two different species were sung at the same time and if they were at different overall pitch heights, they might be segregated into two distinct auditory objects (species' songs) at least partly on the basis of pitch. Or the songs' temporal patterns, once again on the basis of grouping principles, might serve similarly to disambiguate them. Thus, Gestalt principles of organization abound, potentially, in birdsong. Thorpe and Hall-Craggs (1976) identified the possible significance of Gestalt organizational principles in the structure of birdsong some 20 years ago, yet little seems to have been made of their suggestions. Perhaps, auditory scene analysis is, in fact, an expression in birdsong of the functional utility of such an organizational system.

Selective Attention

Given that auditory scene analysis was taking place in the experiments, it follows that selective attention may have been at work as well. The early work of Cherry (1953, 1954) and Moray (1959) on the "cocktail party" phenomenon (see also Wood & Cowan, 1995) demonstrated that selective attention was an active process in those experiments. People were biased to identify one voice stream or their own names in a mixture of other names and acoustic events. More generally, however, the principles of auditory scene analysis, like those for other sensory systems such as vision, hold that organisms can attend to only one primary auditory object at a time and that attention shifts from one object to another but cannot be applied simultaneously to two or more objects.

Experiment 3 provides evidence for selective attention in starlings. In that experiment, birds could maintain accurate discrimination performance on probe trials in spite of the fact that they were hearing—for the first time in their experience—only one of the two auditory objects that had been intermixed to form a baseline stimulus. Presumably, they had in fact heard the isolated stimuli as such by selectively attending to one or the other from trial to trial in baseline training. It is intriguing, in this regard, that they did better with samples of starling song, song with which they were presumably more familiar and to which they should

presumably be especially attentive functionally, than with the other songs used in the experiments.

Common Organizational Principles for Birdsong?

It is surprising that the starlings were able to discriminate accurately the isolated nightingale song exemplars at all. It is a virtual certainty that they had never heard this species' song prior to the experiment because the nightingale is a species alien to North America, and the starlings were captured near Baltimore. This fact leads, in turn, to the larger question of how the starlings did perceive nightingale song as an auditory object subject to stream segregation. Perhaps they did not hear nightingale song as birdsong, but merely as an auditory pattern characterized by a jumble of distinguishable acoustic events moving in time to form an auditory object. Perhaps, however, there is some organizational principle that is common to the structure of all birdsong. Although songbirds may not recognize that a particular collection of sounds is the song of, say species X, they may recognize that the collection does belong to a class of acoustic structures characteristic in some sense of birdsong in general. That recognition might be based on purely bottom-up acoustic features: harmonic structure and an appropriate range of pitches, perhaps combined with a characteristic dynamic range and so on. More likely, however, the organization might also be based on the general top-down principles of organization and pattern recognition discussed earlier (see also, Sinnott, 1989, for other ideas about top-down organization in birdsong). To the extent that such organizational principles do exist, they ought to encourage further the binding together of acoustic information to form auditory streams in song perception; and they might, in the process, create acoustic patterns that are in some sense universal for all birdsong. Such a hypothesis could account for the present results.

Field Studies

Anyone who has tried to record birdsong in the field is well aware of the problem of background noise that obscures the song of the target species. To the extent that extraneous noise exists that cannot be avoided in nature, acoustic filters are used later when analyzing recordings in the laboratory to purify further the signal of interest. But from the recorded bird's point of view, recording song under conditions that are as quiet as possible may remove the acoustic background from which the bird must ordinarily search for relevant acoustic signals. Ignoring that background and the effect that it may have on auditory scene analysis and acoustic communication may miss some interesting things that take place. A songbird contributing to the dawn chorus (and trying to hear out significant signals in turn), for example, faces a complicated auditory analysis and could conceivable alter the manner in which it transmits (or attends to) acoustic information accordingly. We know that some species of songbird do vary their song on the basis of responses from a listener as, for example, in the phenom-

enon of song matching (e.g., Falls, 1985; Lemon, 1968; Payne, 1982). Whether song is further altered (or perhaps evolved historically) to take advantage of scene analysis in communication is an interesting question. There are instances in which song varies with habitat (Handford, 1988; Nottebohm, 1975; Wiley, 1991) and incorporates acoustic features designed to facilitate sound transmission in a habitat (Wiley & Richards, 1982). Furthermore, the type and organization of noise in which a significant communicative signal is rendered may be just as important for acoustic communication as the signal itself—from the perceptual side at least and perhaps from the production side as well. That problem has received little, if any, experimental attention so far as we are aware. In any case, Experiment 4 showed that starlings can solve the problem of identifying birdsong signals in a background of functionally relevant noise, and that raises some interesting questions regarding both the proximate and ultimate development of communication under field conditions.

Questions Remaining

Scene Analysis With Synthetic Stimuli

If songbirds can analyze the auditory scene, there are some fundamental issues that remain for future study. First of all, the experiments reported here make use of birdsong, a highly complex communicative signal about which we have, from a functional standpoint, much yet to learn regarding both its acoustic and syntactic structure. It would be useful to study auditory scene analysis in birds and other animals using a synthetic, but a more readily managed, stimulus arrangement. That would lead to direct examination of the stimulus parameters that might affect scene analysis in these species. Among humans, for example, the rate at which stimuli are presented is an important stimulus parameter that has major effects on whether or not a sequence of sounds will break apart and stream as independent auditory objects. The same holds for the magnitude of the pitch difference between neighboring elements in a stream of sounds. Finally, of course, the demonstration of auditory scene analysis with synthetic stimuli would provide converging evidence for the very existence of the phenomenon in nonhuman animals.

Monophonic Versus Multiphonic Sound Sources

As one reviewer of this article pointed out, our stimuli were delivered to the birds through one speaker, and evidence exists that monophonic (as compared with stereophonic) sound sources reduce the "cocktail party" effect (Speith, Curtis, & Webster, 1954). In fact, separate spatial locations for sounds facilitate their perceptual analysis into separate auditory objects (Bregman, 1990). We were well aware of these facts, but equipment limitations prevented our use of separate speakers for delivering to-be-segregated stimuli to our birds. In spite of this, the starlings did very well in parsing the intermixed stimuli into separate auditory

objects. The effect of spatial separation of sound sources on scene analysis remains an interesting issue to study with nonhuman animals in the future.

What Is Noise?

From the perspective of auditory scene analysis in natural settings, how should we define noise? White noise, for example, is appropriate for the bottom-up analysis of low-level psychoacoustic functions, but that is not the noise in which the presumably high-level processes of auditory scene analysis typically take place. Instead, songbirds—indeed, any species that uses acoustic information—must perform the analysis in the real world of environmental noise and the natural signals produced by other species. In the present experiments, we used birdsong signals embedded in the noise of birdsong from other avian species; but at present, we lack general principles for defining such functional noise. As studies of auditory scene analysis progress, especially in natural settings, this issue will become increasingly interesting and important.

References

Astley, S. L., & Wasserman, E. A. (1992). Categorical discrimination and generalization in pigeons: All negative stimuli are not created equal. *Journal of Experimental Psychology: Animal Behavior Processes, 18,* 193–207.

Bhatt, R. S., Wasserman, E. A., Reynolds, W. F., Jr., & Knauss, K. S. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 219–234.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Catchpole, C. K. (1973). The functions of advertising song in the sedge warbler (*Acrocephalus schoenobaenus*) and reed warbler (*A. Scirpaceus*). *Behavior, 46,* 300–320.

Catchpole, C. K., & Slater, P. J. B. (1995). *Bird song: Biological themes and variations.* Cambridge, England: Cambridge University Press.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 25,* 975–979.

Cherry, E. C. (1954). Some further experiments upon the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 26,* 554–559.

Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. J. Moore (Ed.), *Hearing* (pp. 387–424). San Diego, CA: Academic Press.

Falls, J. B. (1985). Song matching in Western meadowlarks. *Canadian Journal of Zoology, 63,* 2520–2524.

Gray, C. B. (1995). *Categorization of artificial and natural auditory stimuli by the European starling* (Sturnus vulgaris). Unpublished doctoral dissertation, Johns Hopkins University, Baltimore, MD.

Handel, S. (1995). Timbre perception and auditory object identification. In B. C. J. Moore (Ed.), *Hearing* (pp. 425–461). San Diego, CA: Academic Press.

Handford, P. (1988). Trill rate dialects in the rufous-collared sparrow, *Zonotrichia capensis,* in northwestern Argentina. *Canadian Journal of Zoology, 66,* 2658–2670.

Herrnstein, R. J. (1990). Levels of stimulus control: A functional approach. *Cognition, 37,* 133–166.

Kacelnik, A., & Krebs, J. R. (1983). The dawn chorus in the great tit (*Parus major*): Proximate and ultimate causes. *Behavior, 83,* 287–309.

Lemon, R. E. (1968). The relation between organization and function of song in cardinals. *Behaviour, 32,* 158–178.

Lewis, E. R. (1991). Convergence of design in vertebrate acoustic sensors. In D. B. Webster, R. R. Fay, & A. N. Popper (Eds.), *The evolutionary biology of hearing* (pp. 163–184). New York: Springer Verlag.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instruction. *The Quarterly Journal of Experimental Psychology, 11,* 56–60.

Nottebohm, F. (1975). Continental patterns of song variability in *Zonotrichia capensis:* Some possible ecological correlates. *American Naturalist, 109,* 605–624.

Payne, R. B. (1982). Ecological consequences of song matching: Breeding success and intraspecific song mimicry in indigo buntings. *Ecology, 63,* 401–411.

Sinnott, J. (1989). Internal cognitive structures guide birdsong perception. In R. J. Dooling & S. H. Hulse (Eds.), *The comparative psychology of audition* (pp. 447–463). Hillsdale, NJ: Erlbaum.

Speith, W., Curtis, J. F., & Webster, J. C. (1954). Responding to one of two simultaneous messages. *The Journal of the Acoustical Society of America, 26,* 391–396.

Thorpe, W. H., & Hall-Craggs, J. (1976). Sound production and perception in birds as related to the general principles of pattern recognition. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology* (pp. 171–189). Cambridge, England: Cambridge University Press.

Wasserman, E. A., Kiedinger, R. E., & Bhatt, R. S. (1988). Conceptual behavior in pigeons: Categories, subcategories, and pseudocategories. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 235–246.

Wiley, R. H. (1991). Associations of song properties with habitats for territorial oscine birds of eastern North America. *American Naturalist, 138,* 973–993.

Wiley, H., & Richards, D. G. (1982). Adaptions for acoustic communication in birds: Sound transmission and signal detection. In D. E. Kroodsma & E. H. Miller (Eds.), *Acoustic communication in birds* (Vol. 2, pp. 131–181). New York: Academic Press.

Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 255–260.