Signal Decoding and Receiver Evolution

An Analysis Using an Artificial Neural Network

Michael J. Ryan   Wayne Getz

Section of Integrative Biology C0930, University of Texas, Austin, Tex., and Department of Environmental Science and Policy Management, University of California, Berkeley, Calif., USA

Key Words

Artificial neural networks · Animal communication · Receiver evolution · Signal decoding

Abstract

We use a connectionist model, a recurrent artificial neural network, to investigate the evolution of species recognition in sympatric taxa. We addressed three questions: (1) Does the accuracy of artificial neural networks in discriminating between conspecifics and other sympatric heterospecifics depend on whether the networks were trained only to recognize conspecifics, as opposed to being trained to discriminate between conspecifics and sympatric heterospecifics? (2) Do artificial neural networks weight most heavily those signal features that differ most between conspecifics and sympatric heterospecifics, or those features that vary less within conspecifics? (3) Does selection for species recognition generate sexual selection? We find that: (1) Neural networks trained only on self recognition do not classify species as accurately as networks trained to discriminate between conspecifics and heterospecifics. (2) Neural networks weight signal features in a manner suggesting that the total sound environment as opposed to the relative variation of signals within the species is more important in the evolution of recognition mechanisms. (3) Selection for species recognition generates substantial variation in the relative attractiveness of signals within the species and thus can result in sexual selection.

Copyright © 2000 S. Karger AG, Basel

Introduction

Many animal communication systems are involved in discriminating between self and others. This is especially true in species recognition, in which individuals discriminate between conspecifics and heterospecifics. The evolution of species recognition mechanisms has long been of interest to animal behaviorists and evolutionary biologists alike due to their importance in processes of speciation and sexual selection [Dobzhansky, 1940; Blair, 1958, 1964; Mayr, 1963; Alexander, 1975; Andersson, 1994]. Several issues exist regarding the behavioral processes involved in the evolution of species recognition. These issues have proven difficult or intractable to investigate empirically, and include: the degree to which the evolution of the recognition mechanism is influenced by response to heterospecifics; the salience of the various signal features in recognition; and the degree to which the evolution of species recognition has pleiotropic effects or unintended consequences for recognition of individuals within the species, thus potentially generating sexual selection.

We address these issues in the context of auditory communication systems, which are critically important to mate recognition in a number of species, especially song birds, frogs, and insects [Andersson, 1994; in the context of kin selection see Getz, 1981, 1982; Lacy and Sherman, 1983; Getz and Page, 1991; Hepper, 1991]. We use an Elman neural network model (see appendix) to conduct our analyses [Elman, 1990; Demuth and Beale, 1997].

The first issue we address is how animals form categories of self and others. In order to discriminate between self and

others an individual must have a set of sensory rules or concepts to which they refer when forming these two categories. Different referential rules or 'self-concepts' have been implied in the speciation literature with little understanding of how these mechanisms might influence the process of species recognition. At the two extremes are suggestions by Dobzhansky [1937, 1940] and Paterson [1978, 1982, 1985]. In Dobzhansky's [1937,1940] hypothesis of reproductive character displacement or reinforcement [Butlin, 1987], mate recognition mechanisms begin to diverge when the incipient species are geographically isolated, but there is subsequent selection to discriminate between conspecifics and heterospecifics if and when the species come back into contact. Selection acts against those females that mate with heterospecifics due to the reduced vigor of offspring that are able later to mate and reproduce themselves. (Note that selection can act on both signal-production and perceptual mechanisms, we treat the latter here.) In the character displacement (reinforcement) scenario, therefore, the evolution of the recognition mechanism is influenced by sampling the difference in signals between conspecifics and heterospecifics; this is true whether the selection against mismatings is generated from hybrid disadvantage or other factors such as ineffective syngamy or increased search time [cf. Butlin, 1987]. In contradistinction, Paterson [1985] suggests there is strong selection for self recognition, which then results incidentally in individuals distinguishing between self and other; he also suggests there is little empirical support for character displacement. Thus Paterson posits that there is no need for selection against heterospecific matings to result in conspecific versus heterospecific recognition.

One strength of Paterson's argument comes from the lack of much empirical support for character displacement [but see Coyne and Orr, 1989; Gerhardt, 1994; Ryan et al., 1996; Saetre et al., 1997]. Paterson and others [e.g. Passmore, 1981] seem to assume that species recognition logically could have evolved equally effectively with or without selection generated by interaction with heterospecifics. Their argument addresses how species recognition actually evolved.

We investigate the influence of heterospecific signals on the evolution of recognition mechanisms by using four different training regimes; the training sessions mimic the evolutionary processes of selection and mutation. In the self referential assessment, training is based on reference to a 'typical' or mean conspecific signal without any reference to heterospecifics, as suggested by Paterson. In the mean referential assessment, training involves a comparison between the mean conspecific signal and the mean (or typical) heterospecific signal in the same environment. In the variance referential assessment, training involves comparisons between a population sample of the conspecific and heterospecific signals in the sound community. In the noisy variance referential assessment, training is similar to that in the variance referential approach but ambient noise is added to the signal to assess the degree to which it might increase the difficulty of achieving recognition [Ryan and Brenowitz, 1985; Klump, 1996].

The second issue we address is feature weighting. Most signals are parsed by organisms into multivariate arrays representing different components or features. It is known, however, that the receiver does not equally attend to all the potential information encoded by each component, and it has been of interest to determine those features salient in discriminating signals [e.g. Emlen, 1972; Brenowitz, 1983; Nelson, 1988; Nelson and Marler, 1990; Wilczynski et al., 1995; Miller, 1996]. Feature weighting is a question that involves both the mechanisms of communication and the process of evolution: how does the receiver decode information, and how did it come to rely on certain signal parameters for decoding?

The statistical distributions of signal components within the sound environment are likely candidates for influencing how a receiver decodes signals; how it weights various features of the signal. Nelson and Marler [1990] explored this issue by contrasting two hypotheses that predict feature weighting in conspecific (acoustic) recognition. The feature invariance hypothesis suggests that those signal features with relatively less variation within the population will be most heavily weighted in discrimination tasks. The sound environment hypothesis predicts that those features that best statistically discriminate between conspecific versus others in a sound community will be most heavily weighted. For any given data set (i.e. the multivariate distribution of signals in an acoustic community), however, these hypotheses might not be mutually exclusive. Nelson and Marler [1990] tested these hypotheses in a study of a song bird community. Unfortunately, signal dominant frequency was both the feature that tended to have less variation within a species (feature invariance hypotheses) and best predicted species identity in a discriminant function analysis (sound environment hypothesis). Relative to the importance of this sound feature, these hypotheses did not make mutually exclusive predictions; the examination of receiver discrimination of other variables, however, tended to support the sound environment hypothesis.

The third issue we address is the pleiotropic effects of species recognition on sexual selection. The referential system that is the basis of the recognition mechanism, the

statistical distributions of signal features, and the influence of environmental noise might all influence how well the receiver is able to evolve species-discriminating abilities. There are, however, likely to be incidental but important consequences of the receivers on how conspecific signals are perceived. For example, selection to avoid a heterospecific signal could cause females to be less responsive to conspecific signals that more closely resemble heterospecific ones. Such receiver biases can result in reproductive character displacement and sexual selection [Fisher, 1930; Dobzhansky, 1937; Ryan and Rand, 1993; Gerhardt, 1994; Pfennig, 1998; Ryan, 1998]. Thus it is interesting to understand how the above variables influence not only the accuracy with which the receiver can discriminate between conspecific and heterospecifics, but also how the receiver responds to conspecific signal variation.

An Artificial Neural Network Model

Artificial neural networks or connectionist models consist of simple units or 'neurons' that are connected to one another in varying degrees. Each unit is computationally simple, but the network derives its computational power from the inter-connectedness of the units. It is these properties of the connectionist models that result in the analogy to a nervous system [e.g. Churchland and Sejnowski, 1992]. Although artificial networks have been used for more than a decade in fields such as computational neurobiology, computer science, and linguistics, their uses have been more restricted in animal behavior and evolutionary biology [but see Enquist and Arak, 1993, 1994; Bateson and Horn, 1994; Johnstone, 1994; Hurd et al., 1995; Phelps and Ryan, 1998, 2000; Holmgren and Getz, 2000].

When applied to behavior and evolution, these models can provide important insights into both the mechanisms underlying sensory decoding as well as the processes by which sensory mechanisms evolve, even though in many studies the networks themselves do not reflect the architecture of the brains of the animals of interest [e.g. Elman, 1990; Churchland and Sejnowski, 1992]. The Elman net is designed to investigate perceptual questions involving temporally patterned input signals in which the recent past structure of the signal provides an essential context for the current structure of the signal. In particular, an Elman net performs much better than many other types of connectionist nets when it comes to parsing auditory information [Elman, 1990]. Of course, the issues we address could also be analyzed in the context of visual, olfactory, or other sensory systems, using other types of neural nets more appropriate to the particular sensory system under consideration.

The purpose of this study is to use an artificial neural network model to explore how this particular recognition mechanism and the statistical distribution of signal properties we employ might influence a receiver's ability to discriminate between conspecifics and heterospecifics, how these receivers weight signal features to make such discriminations, and to uncover consequential biases in response to conspecific signal variation.
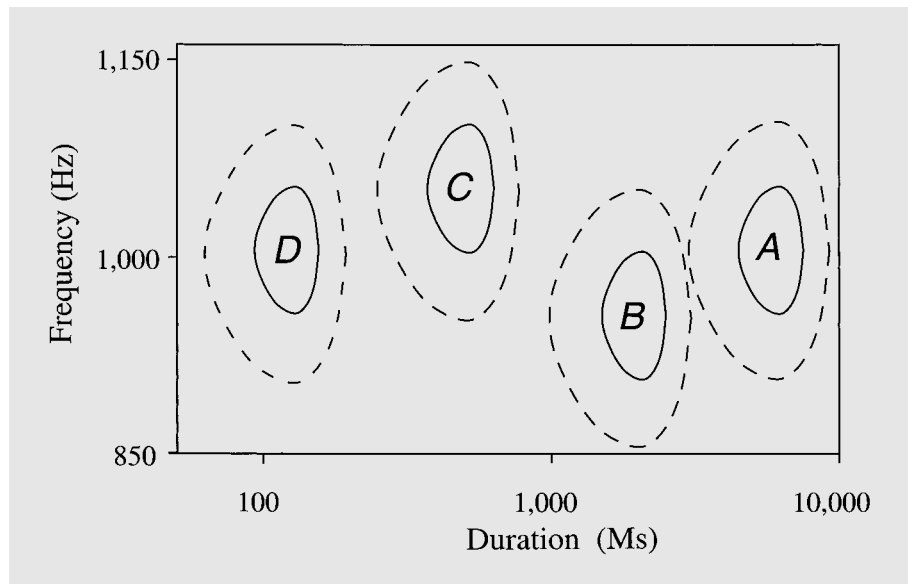
Materials and Methods

Statistical Distribution of Signal Features

We constructed populations of stimuli in which two signal parameters or features were varied in four populations. Frequency, the less variant feature, overlapped the most between populations, whereas temporal duration, the more variant feature, overlapped the least between populations (fig. 1). This resulted in the feature invariance and sound environment hypotheses making mutually exclusive predictions as to which features should be weighted more heavily by the receiver because the less variant feature was not the best predictor of species identity.

We synthesized signals for four populations. For each population the coefficient of variation for frequency was 5% and the coefficient of variation for duration was 25%. These values are within the range that Nelson [1988] found for species used in the analysis by Nelson and Marler [1990]. The mean frequency and time values for each population were: population A, 1,000 Hz, 6,000 ms; population B, 950 Hz, 2,000 ms; population C, 1,050 Hz, 500 ms; population D, 1,000 Hz and 125 ms (fig. 1). We synthesized twenty five signals for each population, whose relative values are shown in figure 2.

Signals were synthesized in the software applications package Signal [Beeman, 1996]. All signals were centered in a time bin of 11,500 ms, large enough to contain the longest signal (6,000 ms $\times$ 3 SD, where SD = $0.25 \times 6,000$; fig. 1). The sampling rate was 4 kHz, giving a Nyquist frequency which was above the highest frequency value ($1,050$ Hz + 3 SD = $1,050 + 3 \times 0.05 \times 1,050 = 1,207.5$; fig. 1). (The Nyquist frequency is half the digital sampling rate, i.e. 2,000 Hz, and is the highest frequency that can be accurately represented in digital reconstruction of a signal as one cycle of a sine wave requires a minimum of two points to define it.) All signals had the same peak amplitude and relative amplitude envelope, a simple cosine function. A sonogram was calculated for each signal representing a pure tone (i.e. produced by only one frequency). A sonogram is a frequency-by-time matrix in which the relative amplitude of each point is represented. We selected a frame length of 256 for the fast Fourier transform. The frequency resolution we used was 15.6 Hz and the temporal resolution was 64 ms because the sampling frequency was 4 kHz. Temporal and frequency resolution are inversely related in digital spectral analysis. The sonogram data were saved as a matrix of 104 rows (in 15.6 Hz increments) by 125 columns (in 64 ms increments), with frequency varying among rows and time varying among columns. The cell values, which represent amplitude, were scaled to a maximum of 1.0. This matrix was then trimmed to 28 by 125, eliminating frequency bands (rows) not used by any of the signals. This was the matrix used for the stimulus input.
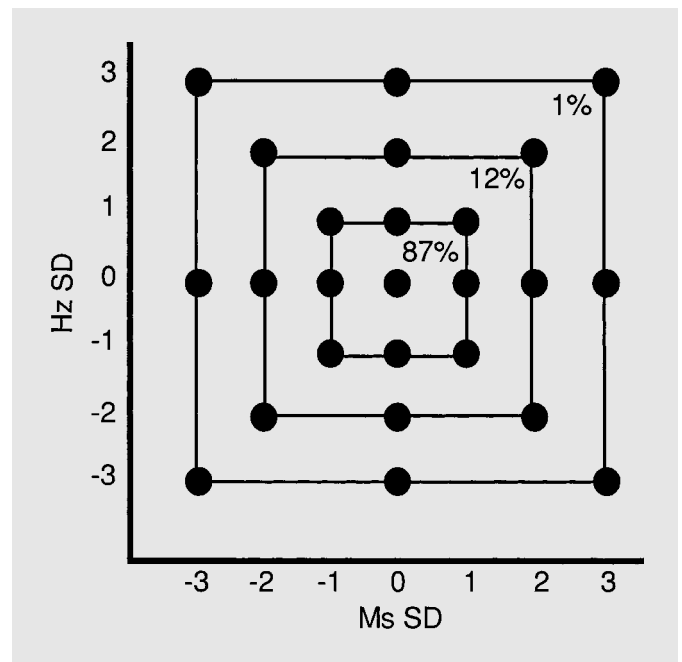
Fig. 1. A representation of the distribution of signal parameters for four populations, A–D, used in the artificial neural network studies. The letters indicate population and mean values, the inner solid circles represent one standard deviation around the mean, and the dashed outer circles two standard deviations around the mean. Note that the axes are on a logarithmic scale.
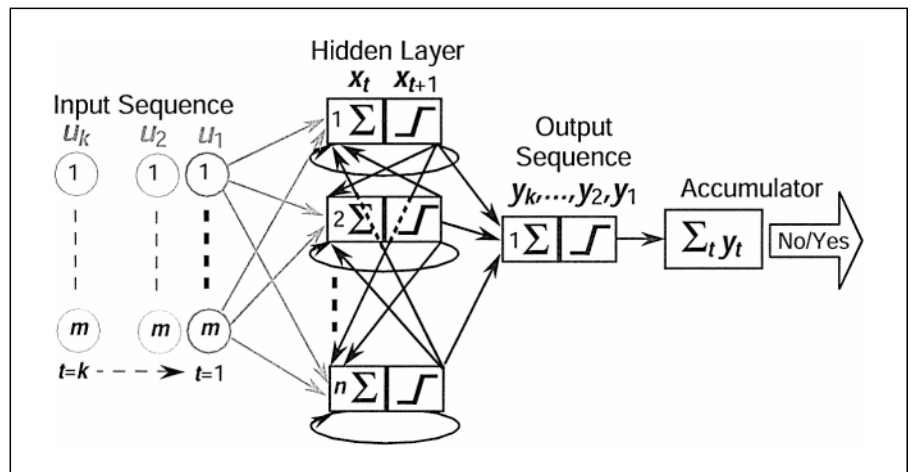
The Neural Network

All simulations were run in MatLab version 5.1 in a PC environment. The network architecture is an Elman net [Elman, 1990; Demuth and Beale, 1997; see also Appendix]. This type of net has an input layer that receives the stimulus input and feeds information forward to a hidden layer and then an output layer (fig. 3). Units of the hidden layer also feed information back onto itself and all other units of the hidden layer. It is this recurrence (feedback) with time that provides the Elman net with the time-dependent memory that is critical for decoding temporally structured signals such as sound. As with other neural networks, the activity of a unit is a product of the stimulus input or activity of other units connecting to that unit and any weights on the connections and biases of the units themselves. It is the difference in the weights and biases that result in networks of the same architecture giving different responses to the same stimulus.

The architecture of the net we used consisted of 28 input neurons that corresponded to the 28 frequency bands (rows of the input matrix) of the signal (i.e. $m = 28$, see Appendix). Each of these input units fed forward to each of 20 units in the hidden layer (i.e. $n = 20$, see Appendix). This ratio of input neurons to neurons in the hidden layer is in a range of values that has proven to be effective in other studies [Elman, 1990; Demuth and Beale, 1997; Phelps and Ryan, 1998, 2000]. In our model, the units in the hidden layer feed forward onto a single output unit. Each of the units in the hidden and output layer were influenced by bias elements of the vector c (equation 1, Appendix), and each of the connections between units were influenced by a weight (elements of matrix A, equation 1, Appendix). There were 960 connections (number of elements in the matrix A is $(20 + 28) \times 20 = 960$, see Appendix) and thus the same number of weights between the input and hidden layer (forward and backward); there were 20 connections (number of elements in matrix B is 1×20, see Appendix), each with a weight, between the hidden and output layer. There was no weight between the stimulus and the input units, and the input units were not influenced by a bias. The activity of the input units was determined merely from the stimulus input. The activity of the units in the hidden layer was deter-

Fig. 2. Points show the stimulus values synthesized for each population as standard deviations from the mean (A–D; 0 = mean population value). The percentage values show the number of signals selected from each area in a simulation of 1,000 runs in which values were selected between –3 and +3 from a normal random distribution with a mean of 0.

Fig. 3. A discrete-time Elman Net consists of an input layer of dimension m, a hidden layer of dimension n, and an output layer of dimension r. A series of m-dimensional input vectors u_t, are feed into the net at each time step t = 1,2,3…, k. In effect, the input is an mxk dimensional matrix. At each time step, each input node in the hidden layer receives a weighted sum of inputs (different weights for each node), as well as the output from the hidden layer in the previous time step (see Appendix for details). The summed information is then transformed using an activation function (represented here by a ramp) to scale the variable in question to lie between 0 and 1. At each time step, besides information in the hidden layer being fed back into this layer, it is also fed forward to the output layer. Here we only show an output layer of dimension r = 1, because that is what we used in our study. In addition to this layer, we have a fourth layer which we used in our study. In addition to this layer, we have a fourth layer which simply sums the output y_t over time where the final value of this sum at time t = k determines whether the receiver identifies the caller (represented by the input matrix) as a conspecific or as a heterospecific.

mined from the interaction of the activity and weights of connections to it as well as the bias, and was calculated using a hyperbolic tangent function as a squashing function [Demuth and Beale, 1997]. The activity of the output unit was also determined from the activity and bias of its connections but was calculated using a pure linear function [Demuth and Beale, 1997].

Evolution (Training) of the Network

We created one hundred networks of the architecture described with randomly chosen weights and biases that were constrained to be within values of –1 to 1. Each network was trained with a conspecific signal and several heterospecific signals (except for the self referential training regime, see below for details). Prior to training, we added noise to each signal, except in the training regime in which noise was specifically eliminated. The noise was a matrix the same size as the signal matrix. We chose each cell value from a random uniform distribution between 0 and 0.10. Thus for any cell the noise level was a maximum of 10% the peak signal value. In some training regimes (variance referential and self referential) noise was not added to the stimuli. The addition of noise is biologically realistic because animals communicate in a noisy world [Klump, 1996].

We assigned fitnesses to each net based on the differences in their response to conspecific versus heterospecific signals. In particular, the heterospecific response was subtracted from the conspecific response for all six pairs of conspecific-heterospecific comparisons, and the average of these six differences was designated to be the fitness of the network in question.

After each generation we selected nets for use in the next generation. The probability of selecting a particular net was weighted in proportion to its relative fitness and 100 nets were otherwise selected at random with replacement. Additionally, we assured that the net with maximum fitness was always one of those selected. These nets were then subject to mutation. Weights and biases were selected for mutation with a probability of 0.01. If selected, the weight or bias was perturbed by a number randomly selected from a uniform distribution between –0.5 and 0.5 subject to the constraint that the final values were always between –1 and 1. One net with the maximum fitness was not mutated.

We ran each training session for 500 generations. Preliminary results showed that fitness tended to plateau within this time frame. At the end of the training session we saved the network with the maximum fitness (i.e. its weights and biases were retained); we refer to this as the trained network. We repeated these training sessions 20 times to obtain a sample of 20 trained networks for each of the four types of analyses we undertook to address the four different issues raised above (i.e. self referential, mean referential, noisy variance referential, variance referential).

Population Comparisons: Reference and Noise

We ran four sets of 20 training sessions for all four populations (A–D) whose signal parameters are detailed above (fig. 2). In each of these 320 training sessions one of the populations is designated the conspecific (e.g. population A), whereas all others are designated the heterospecifics (e.g. populations B, C, D). We conducted training sessions in which each population was treated as the conspecific.

The four different sets of 20 training sessions differed in their training regimes. In the self referential (sr) training regime, we trained networks to discriminate between the mean conspecific signal and noise. The noise in this case was constructed by assigning values from 0 to 1 from a random uniform distribution to a matrix that was the same size as the conspecific signal. (This white-noise stimulus is not to be confused with the noise that was added to the signals in the mean referential and variance referential training sessions.) These values were then normalized so that the column sums of the noise were equal to the column sums of the signal. Thus the noise had the same amplitude envelope (column sums), the same total amplitude, and the same root-mean-square amplitude as the signal but the distribution of frequencies (row columns) did not match that of the signal. Six comparisons were made for each net. Populations in these training sessions are denoted with subscripts indicating self-referential (e.g. A_{sr}). Noise was added to each stimulus prior to training the net.

In the mean-referential (mn) training regime, we used only the mean signals in each population. Noise was added to each signal. Each net was trained with (i.e. its output was determined in response to) the mean conspecific signal, and with each of the mean signals from the three heterospecific populations. Thus for each net there were three conspecific-heterospecific contrasts that we used to determine the net's fitness. Noise was added to the signals as described above. Populations in these simulations are denoted with subscripts mn (e.g. A_{mn}).

In the noisy variance referential (vn) training regime, the full population variation of signals was used for each population and noise was added to these signals. For each population, signals had frequency and time values that were various combinations of the mean (denoted as 0) and –3 to +3 SDs of the mean value (fig. 2). The conspecific signal to be used in training was selected by randomly choosing a number between –3 to 3 from a standard normal distribution. A simulation of 1,000 selections showed that 87% of the signals chosen had values for both frequency and time that were within ± 1 SDs of the mean, 12% had values in which at least one value was ± 2 SDs of the mean, and only 1% of the signals had at least one value that was +3 or –3 SDs from the mean (fig. 3).

In the noisy variance referential (vn) training sessions we randomly chose a signal to be tested from the pool of the full range of population variation represented by the 25 signals in each of the three heterospecific populations. Each net was trained with one conspecific signal, chosen at random from the full range of conspecific variation, and six heterospecific signals. Populations in these simulations are denoted with subscripts vn (e.g. A_{vn}).

The fourth training regime was the same as the noisy variance referential, but without the overlay of noise on the signals. Thus for this variance referential (vs) we used the symbols s to indicate silence (i.e. no background noise) and the appropriate populations are denoted with subscripts indicating this (e.g. A_{vs}).

Testing the Nets

After completing each set of 20 training sessions for each training regime with each population, we had 20 'best' nets from each regime/population that were retained for further testing; these we refer to as the trained nets. We tested each trained net with (i.e. allowed them to respond to) all of the 25 conspecific signals (fig. 3). We calculated the response to each signal relative to the mean signal of the same population by taking the sum of the difference between the output vectors, as described above. We then calculated z scores for the mean responses. Thus the relative response to the mean call is always 0; a positive value

for a call indicates that the response is greater than the response to the mean and a negative value shows a lesser response to that call.

Statistical Analysis
Self Recognition versus Species Discrimination. We calculated the mean and the standard error of the maximum fitnesses for each population in each set of training sessions. Maximum fitness of all populations reached a plateau before 500 generations were reached.

We determined to what extent the trained networks could distinguish between the population of conspecific signals versus the other heterospecific signals. This was a simple discrimination task in which conspecific recognition was scored if the net showed a greater response to the conspecific signal versus a heterospecific one. The signals to be compared were drawn at random from the conspecific and heterospecific populations as described above. We determined the responses of each of the 20 trained nets to 50 pairs of conspecific-heterospecific signals for a total sample size of 1,000 for each population.

Feature Weighting. We assessed feature weighting by calculating the differences in the networks' responses as a z score between neighboring signals (i.e. within 1 SD) along four signal axes through the conspecific population (cf. fig. 2). These axes were: Hz, signals varied in frequency from –3 to +3 SDs and were all of mean duration; Ms, signals varied in duration from –3 to +3 SDs but had the mean frequency; Ds, all signals varied in the same magnitude on both axes of SD from –3 to +3 and bear the same sign; Df, all signals varied in the same magnitude on both axes of SD from –3 to +3 but have a different sign. All four axes contained the mean signal of the population (i.e. 0 SD Hz, 0 SD Ms; fig. 2). For the 20 trained nets for each training regime/population we determined the average response difference between neighboring signals on each axis. We then compared the mean response differences for each axis within a population with an analysis of variance. Fisher's least-significant-difference test was used as a post hoc test to compare differences among the individual axes. This analysis allows us to examine variation in response to conspecific signals relative to each signal feature we varied, as well as a direct comparison to the results of Nelson and Marler in their test of the feature invariance versus sound environment hypotheses [Nelson and Marler, 1990].

Incidental Sexual Selection. We also examined how the trained nets responded to variation in the conspecific signals. We standardized the responses to conspecific signals as z scores or standard normal deviates with a mean of 0 and a variance of 1. In these analyses it is irrelevant how strong the responses were to heterospecific signals; we are concerned with how the receivers responded to differences only among conspecific signals.

A Control
An uncontrolled feature of the artificial neural networks just described is that time and frequency are processed differently by the networks. Time is processed in columns with a feedback loop between columns whereas frequency is processed in rows. Although this is biologically realistic, we asked if the processing mode itself made a difference. We rotated the stimulus matrix by 90°, thus the columns of cells in the matrix now became rows, and the rows became columns. The coefficients of variation, however, remained the same. Because the number of columns in the input matrix changed, so did the number of units in the network. The trained networks were then tested with their population of signals to determine the parameter to which they were more sensitive.

The training and testing of the networks with the rotated stimuli were conducted as described above. We used these data to test two

Table 1. Mean and (standard error) of the maximum fitness of the 20 trained networks in the four populations for each simulation

Simulation type	Populations			
	A	B	C	D
Self-referential (sr)	1,985 (65.7)	1,798 (59.5)	1,687 (31.3)	1,743 (31.9)
Mean-referential (mn)	2,586 (41.3)	2,330 (33.6)	2,286 (50.0)	2,369 (32.5)
Noisy-variance (vn)	2,293 (48.8)	1,982 (69.6)	2,031 (101.1)	1,989 (95.5)
Variance-referential (vs)	2,263 (38.5)	1,830 (35.5)	1,819 (40.0)	1,598 (49.6)

alternative predictions. First, if processing mode was critical we would expect the networks to differ in their relative responsiveness to frequency and time variation from the previous study as the parameters are being processed in different modes even though the variation of these signal parameters in the populations remained the same. We note however that if the results differ from the previous study it could also be due to the change in the size of the network. This is an uncontrolled variable. Second, if the mode of processing does not make a difference, and the coefficients of variation do, then we would predict the networks to be similar in their relative responses to frequency and duration regardless of whether these signal parameters are represented in rows or columns.

Results

Evolution of Nets

The mean and the standard error (SE) of the maximum fitnesses for the 20 trained nets in four populations in each of the types of training are shown in table 1. There were significant differences among mean maximum fitnesses. A multiple analysis of variance shows a statistically significant effect for training types across populations ($F_{3,304} = 48.1$, $p < 0.001$) and populations across training types ($F_{3,304} = 23.7$, $p < 0.001$) as well as an interaction effect ($F_{9,304} = 8.9$, $p < 0.001$). We conducted one-way analyses of variance for each of the main effects to allow post-hoc comparisons. Training types again had a significant effect ($F_{3,304} = 33.4$, $p < 0.001$). The Fisher's least-significant-difference test showed that the average maximum fitness of the self-referential training regime was significantly lower than that of the mean-referential training regime, and that the fitness of the mean-referential training regime was significantly greater than the other two training regimes (table 1). There were significant differences among populations across training regimes ($F_{3,304} = 14.1$, $p < 0.001$). Population A differed from C and D, whereas population D differed from populations A, B, and C (table 1).
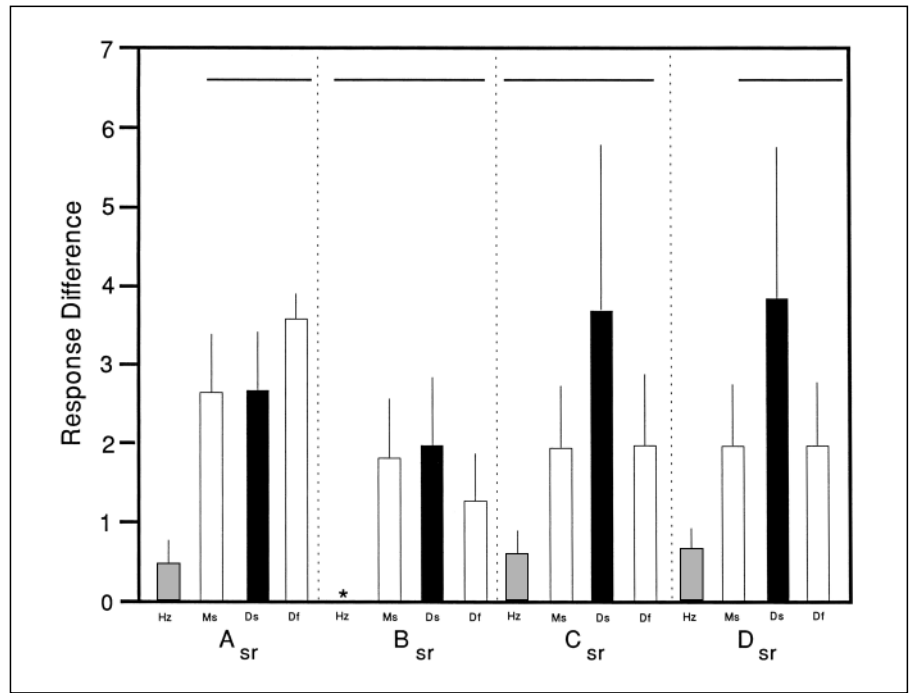
These results show that nets trained in the self referential training regime, that is, those nets trained to distinguish the mean population signal from noise with no reference to heterospecific signals, do not perform this task as well as nets trained to distinguish between conspecific and heterospecific signals as indicated by the average maximum fitnesses (table 1). Of the latter, the nets trained to distinguish the mean conspecific versus heterospecific signals, that is, the mean referential training regime, perform that task better than nets trained to discriminate between samples of conspecific and heterospecific signals (table 1). Another trend is that regardless of the training task, the populations at the edge of the acoustic space (A and D; fig. 1) perform their tasks better than those nearer the center of the acoustic space (table 1).

The maximum fitnesses obtained by each population of nets during its training is specific to the training task: for example, discriminating between self and noise (sr), between the mean conspecific call versus mean heterospecific calls (mn), or between a sample of conspecific call variants versus heterospecific call variants (vn and vs). A more appropriate comparison is to ask how each population of networks fared in their ability to discriminate among the varieties of calls in the acoustic environment once its nets reached maximum fitness. To assay this parameter, each of the trained nets were tested with 50 pairs of calls, a conspecific and heterospecific chosen from the random normal distribution of population variation and scored according to whether the conspecific or heterospecific signal elicited the greatest response. The percent correct and incorrect classifications were compared with a χ^2 likelihood ratio.

In the self-referential training regime the percent of signals correctly classified as conspecific was not much better than the random expectation of 50%: A_{sr}, 53%; B_{sr}, 56%; C_{sr}, 58%; D_{sr}, 62%. In the mean-referential training regime the percent correct classifications for each population were: A_{mn}, 83%; B_{mn}, 71%; C_{mn}, 77 %; D_{mn}, 87%. When the training relied on all of the call variation in the populations and noise was added to all signals prior to discrimination (noisy variance referential), the nets of maximum fitness exhibited the following measures of classification accuracy: A_{vn}, 95%;

Fig. 4. The mean response difference in the responses to neighboring calls along one of four transects through conspecific call variation by the 20 best nets from the self referential training regime (sr). The populations A–D differ in their signal parameters as indicated in figure 2 and the text. The four transects are: Hz = signals vary in sequency but have the mean duration; Ms = signals vary in duration but have the mean frequency; Ds = all signals have the same magnitude and sign SD for Hz and Ms; Df = all signals have the same magnitude but different sign SD for Hz and Ms. The bars represent the mean response difference plus one standard deviation. The horizontal lines indicate response differences that were not significantly different from one another by Fisher's least-significant-difference post hoc test. In these simulations only the mean calls among populations were compared. *Indicates a value of 0.003.

B_{vn}, 78%; C_{vn}, 72%; D_{vn}, 86%. Finally, when the total variation was utilized in training, but noise was not added to signals (variance referential), the percent correct classifications were: A_{vs}, 93%; B_{vs}, 74%; C_{vs}, 78%; D_{vs}, 84%.

Among training regimes there was a clear trend in that the greater the amount of signal variation encountered during training, the greater the average classification accuracy: self-referential, 57.5%; mean-referential, 79.5%; variance referential, 82.2%; noisy variance referential, 84.2%. The classification accuracy was significantly lower in the self-referential training regime than in the other three types of training regimes (χ^2 likelihood ratio, criterion p = 0.05).

There was also a trend in the accuracy of conspecific recognition among populations across training regimes. The two populations bordered by a population on one side in the duration dimension had greater accuracy in classification (A, 81%; D, 79.8%), whereas the populations bordered by populations on both sides had lower accuracy in recognizing the conspecific signal (B, 69.7%; C, 71.2%). There were no significant differences in the accuracy of classification among populations within the self-referential training regimes ($\chi^2 = 2.3$, p = 0.517), but in the mean referential training regimes the differences were almost significant ($\chi^2 = 6.80$, p = 0.079), and were significant in both the noisy variance referential ($\chi^2 = 23.2$, p < 0.001), and the variance referential training regimes ($\chi^2 = 15.4$, training regimes p =

0.001). When comparing the populations within the latter two training regimes (likelihood ratio χ^2, criterion p = 0.05), populations A (95%) and D (86%) showed greater classification accuracy than B (78%) and C (72%) in the variance referential noise training regimes, and population A (93%) showed greater classification accuracy than populations B (74%), C (78%), and D (84%) in the variance referential training regimes.
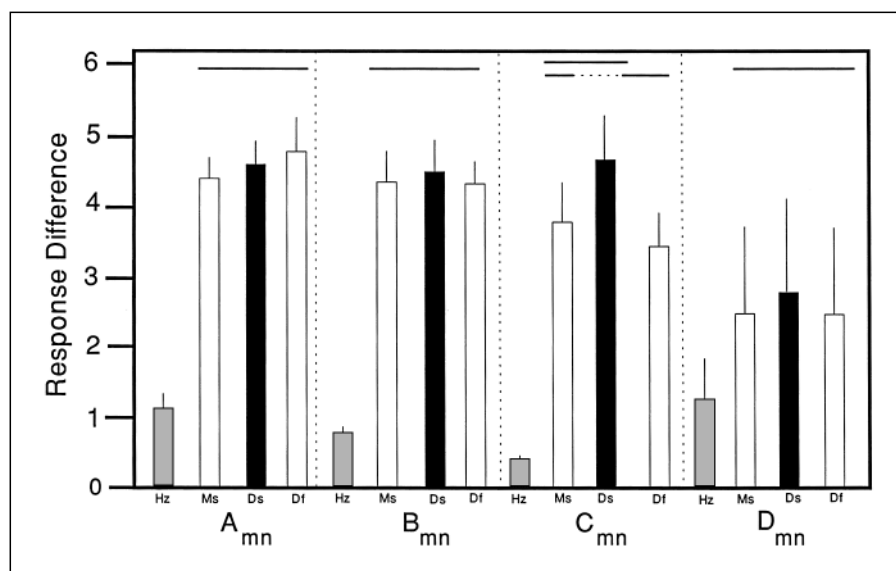
The general results of the classification tests tend to parallel the results of the differences in fitness. Neural networks that were trained with no reference to heterospecific signals had the lowest accuracy, and populations at the edge of acoustic space tended to exhibit the highest accuracy.

Feature Weighting

In the self-referential training regime (sr), nets were trained only to discriminate the mean population signal from noise. In two of these populations, B_{sr} and C_{sr}, there were no significant differences in the mean response difference of trained nets between neighboring signals on the same axis (B_{sr}, $F_{3,76} = 1.81$, p = 0.150; C_{sr}, $F_{3,76} = 1.32$, p = 0.271; fig. 4). In the other two populations, A_{sr} and D_{sr}, the response difference along axes of signal variation showed significant differences (A_{sr}, $F_{3,76} = 3.01$, p = 0.035; D_{sr}, $F_{3,76} = 4.11$; p = 0.009; fig. 4). In both of these populations, the response differences along the Hz axis were significantly

Fig. 5. The mean response difference in the responses to neighboring calls along one of four transects through conspecific call variation by the 20 best nets from the mean referential training regime (mn). All abbreviations and statistical analyses are as in figure 4.

smaller than along the Ms, Ds, and Df axes, and there were no significant differences among the responses along those three axes (Fisher's least-significant-difference test, criterion $p = 0.05$; fig. 4).
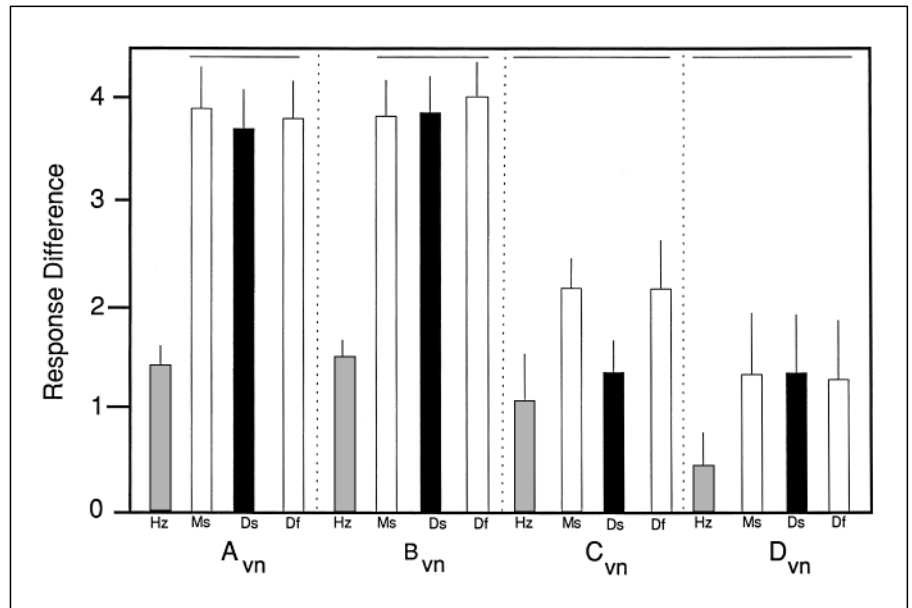
In the mean-referential training regime (mn) only the mean signals among populations were compared during training, as opposed to sampling conspecific and heterospecific calls from the distribution indicated in figure 3. For all four populations in this training regime, there were statistically significant differences in the mean response among the axes for each of the four populations (A_{mn}, $F_{3,76} = 78.4$, $p < 0.001$; B_{mn}, $F_{3,76} = 73.8$, $p < 0.001$; C_{mn}, $F_{3,76} = 21.0$, $p < 0.001$; D_{mn}, $F_{3,76} = 31.8$, $p < 0.001$; fig. 5). Fisher's least-significant-difference post hoc tests showed that in all populations the mean response difference along the Hz axis is significantly smaller than the response along the other three axes. In populations A_{mn}, B_{mn}, and D_{mn} there were no significant differences among the other three axes, whereas in population C_{mn} the mean response difference along axes Ds and Df are significantly different, but the others are not significantly different from one another (fig. 5).

In the noisy variance referential training regime (populations A_{vn}-D_{vn}), the conspecific and heterospecific calls were chosen from a population distribution and noise was added to each stimulus. Qualitatively, the result are very similar to the previous training regime: the response difference along the Hz axis was always less (but not always statistically different, see below) than the response along the other three axes, and the response differences along those three axes were not different from one another. More specifically, in
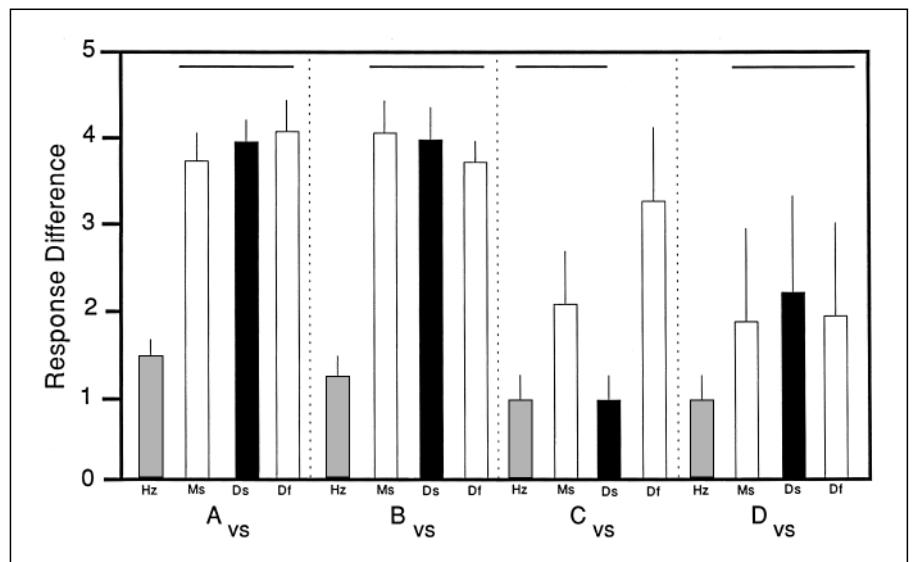
populations A_{vn} ($F_{3,76} = 31.8$, $p < 0.001$) and B_{vn} ($F_{3,76} = 40.2$, $p < 0.001$), in which responses differed among the four axes, the post-hoc test showed that the response differences along the Hz axis were significantly different from the other axes, and the response differences among the Ms, Ds and Df axes were not significantly different from one another (fig. 6). The responses in populations C_{vn} and D_{vn} showed the same trends but they were not statistically significant (population C_{vn}, $F_{3,76} = 1.6$, $p = 0.21$; population D_{vn}, $F_{3,76} = 0.9$, $p = 0.44$; fig. 6).

The same trend in response differences along the axes within the populations is apparent in the training regime in which calls were chosen from population distributions but no noise was added to the stimuli. In populations A_{vs} ($F_{3,76} = 31.7$, $p < 0.001$) and B_{vs} ($F_{3,76} = 30.9$, $p < 0.001$) there were significant response differences among the axes. Population D_{vs} showed the same pattern but it was not statistically significant ($F_{3,76} = 0.4$, $p = 0.73$). The post-hoc test shows that the response difference along the Hz axis is significantly lower than those along the Ms, Ds, and Df axes and those three axes are not significantly different from one another (fig. 7). Population C_{vs} also shows significant differences in the response differences along the four axes (C_{vs}, $F_{3,76} = 5.47$, $p = 0.002$). This is the only population in all sets of training regimes in which when there are differences, the response differences are not the smallest along the Hz axis. The post hoc test shows that in population C_{vs} the response differences along the Df axis is significantly larger than the response differences along the other three axes, and that those three axes do not differ in this measure (fig. 7).

Fig. 6. The mean response difference in the responses to neighboring calls along one of four transects through conspecific call variation by the 20 best nets from the noisy variance referential training regime (vn). All abbreviations and statistical analyses are as in figure 4.

Fig. 7. The mean response difference in the responses to neighboring calls along one of four transects through conspecific call variation by the 20 best nets from the variance referential training regime (vs). All abbreviations and statistical analyses are as in figure 4.
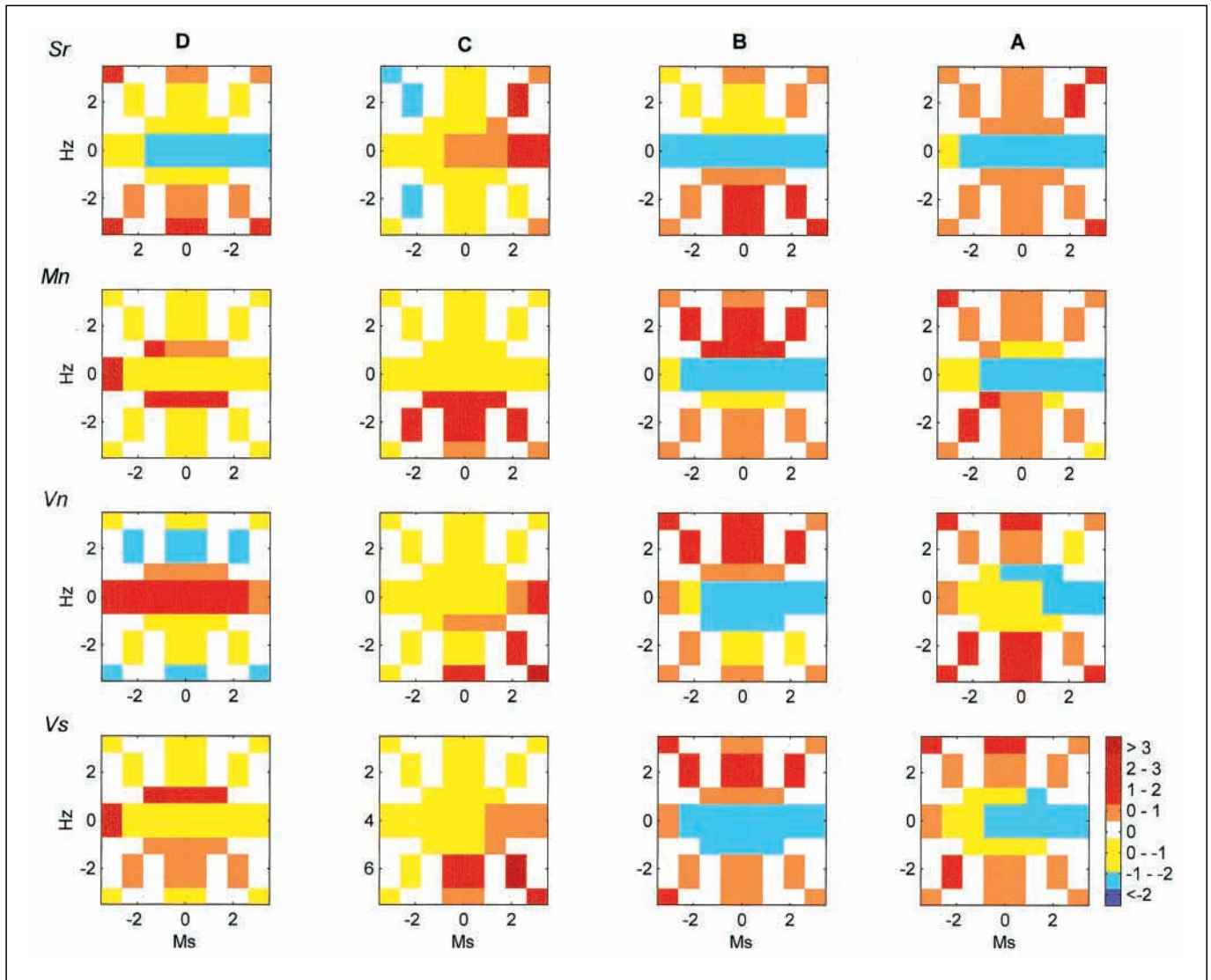
These results show quite clearly that the responses of networks differ along different axes in acoustic space. The general conclusion is that the networks are more sensitive to changes in duration than in frequency. Frequency is the less variant of the two features but is not a good predictor of species identity, whereas duration is the more variant feature but is a good predictor of species identity (fig. 2).

Preference Landscapes and Pleiotropy

We examined the strength of preferences for signals distributed over the variation of conspecific calls (i.e. the

acoustic landscape); we refer to these distributions of response variation superimposed on the acoustic landscape as preference landscapes. The acoustic landscape plots signals in units of SD relative to the mean, and the entire set of signals for a population consisted only of calls that were the population mean or differed from the mean in units of 1–3 SDs (fig. 3). Onto this acoustic landscape we plotted the relative strength of response of the networks; thus we determined the mean strength of response of the nets in the population to each signal. The responses were standardized to z scores using the mean and standard deviation of

Fig. 8. The mean relative responses of the trained networks to call variation within the conspecific population. The X and Y axes of each plot show how the conspecific signals varied in standard deviations of the mean duration (Ms) and frequency (Hz; see fig. 2), which is denoted as 0,0. Twenty five of the possible 49 calls were synthesized and evaluated by the nets. The 24 white cells that do not correspond to dots in figure 2 were not synthesized and evaluated. Populations are represented across columns and training regimes across rows (see text for abbreviations of training regimes). The average responses of each of the 20 trained networks for each population/simulation were measured in response to each conspecific call relative to the mean conspecific call. These response values were then normalized to z scores and coded by the color of each cell in reference to the color bar on the right of each plot. The red end of the color spectrum signifies responses greater than the mean, and the blue end of the spectrum indicates responses weaker than the mean. White patches (0 on the color bar) are empty spaces in the acoustic space for which signals were not synthesized and thus responses are not quantified, as noted above.
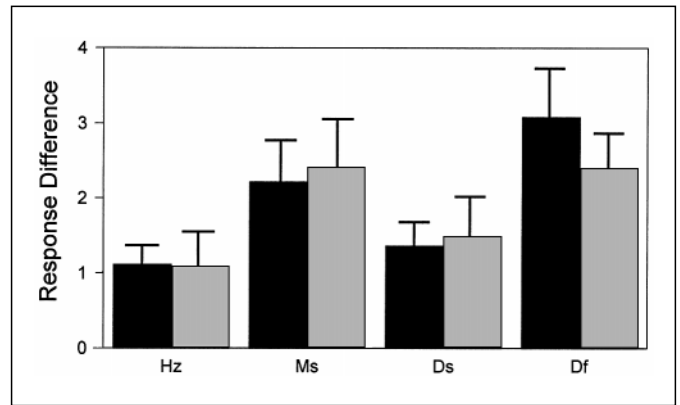
responses for each population. The z scores of the responses are by definition measured in units of standard deviation. This is an appropriate way to analyze the response data when we ask to what degree might the nets generate selection on the distribution of signals within the populations. Thus we can compare the relative response of the networks to mean signals and signals that differ from the mean signals in measures of standard deviation. In figure 8 the z scores are coded in integer units of standard deviations.

In the self-referential populations (sr), as in all the populations we examined, there is variation in the responses of the nets across the population variation. In population A_{sr}, variation in responses were between –2 and +2 SD. In the other populations in this training regime, the strengths of response ranged from –2 to +3 SDs. In none of these four populations were the strongest preferences for the mean signal, and the distribution of responses was usually displaced from the mean along both the duration and frequency axes. Population B_{sr} exhibited a heightened response toward longer durations and lower frequency, whereas population C_{sr} exhibited a heightened response toward longer durations and higher frequencies. In both populations A_{sr} and B_{sr} there was divergence in the most preferred signals, longer times at lower or higher but not intermediate frequencies in A_{sr}, whereas the picture in D_{sr} was rather mixed.

In the mean referential training regime (mn) the amount of variation is marginally less than that seen in the self referential training regime. One population, D_{mn}, varied in the strength of response between –1 and +3 SDs, C_{mn} between –1 and +2 SDs, and A_{mn} and B_{mn} vary between –2 and +2 SDs. Unlike the previous training regime, there was not a clear displacement of stronger responses along the time axis. B_{mn} and C_{mn} showed stronger responses for higher and lower frequencies, respectively; A_{mn} showed more divergent preferences.

In the noisy variance referential training regime (vn), the variation in responses was similar to the mean referential training regimes. A_{vn}, B_{vn}, and C_{vn} varied in strength of response between –2 and +2 SDs, whereas population C_{vn} varied from –1 to +3 SDs. D_{vn} showed a strong preference for mean frequencies across most of the time axis, B_{vn} for higher frequencies across time, C_{vn} for longer signals with lower frequencies, and A_{vn}, once again, showed preferences divergent in frequency.

The responses of networks derived from the variance referential training regime (vs) were quite similar to those from the noisy variance referential training regime. A_{vs}, B_{vs}, and C_{vs} showed the same range of variation in response strength (–2 to +2 SDs), with D_{vs} ranging from –1 to +3 SDs. Also, the directions by which the strengths of response were dis-

Fig. 9. The responses of the neural net in the control tests in which the stimulus matrix was rotated by 90°. Networks were trained in the noisy variance referential simulations (vn) and only trained to calls from population C. The mean response difference in the responses to neighboring calls along one of four transects through conspecific call variation by the 20 best nets from the networks trained under the control conditions and the networks whose results are reported in figure 6. All abbreviations are as in figure 4.

placed from the mean were identical with the exception of more divergent preference in B_{vs} along the frequency axis.

A Control

Figure 9 shows the response of neural networks that were trained in the noisy variance referential regime (nv) in response to calls from population C in the control experiments. In this case the stimulus matrix was rotated 90°. These results are shown together with the neural networks that were trained in the same training regime to the same calls from population C but in which the stimulus matrix was not rotated. The responses were quite similar. This suggests that the different sensitivities to Ms and Hz are not due to processing-mode but to CV. Although the network size was different, the similar responses suggest that network size was not critically important in determining the relative responsiveness of the nets to different signal parameters.

Discussion

The purpose of this study is to use an artificial neural network model to investigate how signal recognition might evolve given the distribution of signal parameters of several species in the same acoustic environment. We are specifically interested in how the referent used for recognition influences discrimination abilities, how the variance of sig-

nal features within and among species influence the weighting of these features in signal decoding, and if evolution of species recognition has pleiotropic effects on discrimination of signals within species.

If the relevance of artificial neural networks models to these issues is in question, the results we obtain certainly provide deeper insights and a better understanding than pure verbal arguments that have no way of assessing complexities that arise when information is processed by networks, whether they are real brains or highly stylized mathematical representations of brains. Furthermore, population genetic models used to explore these same questions tend not to capture the various contingencies and response biases that result when complex nervous systems solve difficult problems. Such response biases are proving to be of some value in understanding the evolution of animal receiver systems [Ryan 1998], are commonly seen in artificial neural networks [e.g. Enquist and Arak, 1993, 1994; Johnstone, 1994], and can predict quite accurately the response biases of real animal receiver systems [Phelps and Ryan, 1998, 2000].

Effect of Category Formation

This study suggests the internal reference that an animal uses in signal recognition influences its ability to discriminate between conspecific and heterospecific signals. Paterson [1978, 1982, 1985] suggested that there is strong selection for self recognition and that discrimination against heterospecifics is an incidental consequence. This is in contrast to Dobzhansky [1937, 1940] and others [e.g. Blair, 1964; Butlin, 1987; Littlejohn, 1988; Coyne and Orr, 1989; Gerhardt, 1994] who suggest that interactions with heterospecifics influence the evolution of the signal recognition strategy.

In our study, recognition strategies that evolved with no reference to heterospecific signals were not as successful in discriminating between conspecific and heterospecific signals as were those in which the recognition strategy evolved with specific reference to those signals that would later have to be discriminated. Amongst the latter, the more signal variation the nets encountered while being trained, the better they later discriminated. Making reference to a sample of conspecific and heterospecific signals during training resulted in better discrimination than making reference only to the mean conspecific and heterospecific signals.

These results rnight seem intuitive but they highlight the crucial difference in emphasis between two major views on the evolution of species recognition – the specific mate recognition system versus the isolation concept [reinforce-ment/reproductive character displacement, e.g. Coyne et al. 1988]. A prediction of Paterson's hypothesis is that the recognition system is under stabilizing selection. Some have pointed out, however, that there can be substantial variation in species mate recognition signals, which would seem to invalidate the specific mate recognition hypothesis [Coyne et al., 1988; Ryan and Wilczynski, 1991]. Paterson [1993] states that such criticisms are unwarranted because the authors utilize a different species concept than he does, and that stabilizing selection does not predict lack of variation [p. 214, Paterson, 1993]. It is not clear how to resolve these issues between species concepts and the precise threshold of variation that the specific mate recognition hypothesis can accommodate.

A more fundamental assumption of Paterson's model, however, is that a recognition system based only on self-recognition will result, although incidentally, in effective discrimination between conspecifics and heterospecifics. This study shows that this need not be true, at least under the admittedly restrictive limitations and assumptions of the artificial neural network models utilized here. The ability to evolve effective conspecific-heterospecific discrimination in this study is reduced when the training regime excludes the potential for false recognition of heterospecific signals to influence fitness during training. Our results cannot be viewed as a firm rejection of Paterson's hypothesis. These artificial networks certainly do not accurately represent the recognition strategies used by all organisms in all situations, and there are other limitations to these models discussed below. But we believe this is the first study to directly assess a critical assumption of this debate – that selection for self recognition will incidentally yield an effective conspecific-heterospecific discrimination system. In these models self-recognition is not effective whereas in nature it might be, but we suggest that perhaps one should not make that assumption.

Feature Weighting

The second issue we addressed was how the variance in signal features influences the degree to which they are weighted in signal decoding strategies. Nelson and Marler [1990] contrasted the feature invariance hypothesis, which states that features least variant within the taxa should be weighted most heavily, and the sound environment hypothesis, which states that those features that best discriminate among taxa in the environment should be weighted most heavily. Their study tended to support the sound environment hypothesis, but given the limitation of their data, these two hypotheses did not always make exclusive predictions.

We constructed signals so that the most invariant feature, frequency, was not a good predictor of species, whereas the most variant feature, duration, did accurately predict species. Our results are consistent with those of Nelson and Marler [1990] in supporting the sound environment hypothesis. Once trained, the artificial neural networks were more sensitive to variation in duration than in the frequency axis. The nets were also sensitive to variation along the two orthogonal axes in which duration and frequency covaried. Thus the nets were sensitive to variation in the feature that best discriminated among taxa even though this feature was quite variable within each taxa and, conversely, the nets were less sensitive to features that did not predict species status even if these features were relatively invariant within the species. This generalization tended to hold across most populations and training regimes.

A number of studies have examined variation of signals within and among individuals and species to gain some insight into signal function and evolution. For example, Barlow [1977] suggested that mate recognition signals should be less variable than those used in male-male interactions, and Ehret [1990] argued that information in recognition signals should be perceived categorically, whereas information in aggressive signals should be perceived continuously. Ryan [1988] emphasized constraints on signal variation in that the mechanism of signal generation should influence the variance, predicting that variance should be higher in features under behavioral-physiological control and lower in features under morphological control. Gerhardt [1991] classified signal variance into static versus dynamic, and argued that females exhibit stabilizing preferences for static features and directional ones for dynamic features. This study, together with that of Nelson and Marler [1990], show that signal variance relative to other species rather than to other conspecifics is a crucial character in signal decoding. All of these approaches concentrate on the same statistical measure of signals and emphasize vastly different issues. All of these approaches might have some merit, but future work should consider these various effects on signal variation in concert.

We want to be quite clear about how constraints of the neural network architecture relate to some vertebrate sound processing systems. The input matrix of frequency (rows) versus time (columns) is not arbitrary but represents our understanding of some aspects of the physics and biology of auditory processing. We are especially influenced by our knowledge of auditory processing in frogs where the neural processing of mate recognition signals is perhaps best understood [see various chapters in Fritzsch et al., 1988]. Our input matrix has time arranged in columns, and thus is

inputted into the network in sequence. This aspect of the network design reflects a constraint of physics: time is an arrow. The arrangement of frequency in rows, however, reflects a constraint of biology. Most vertebrate auditory receptors contain arrays of hair cells that are differentially sensitive to frequency and vary in their sensitivity in a predictable manner across space; that is, they are tonotopically organized. In anurans, for example, the frequency sensitivity of the auditory receptors, which is a property of the sensitivity of all the hair cells, tends to match the distribution of spectral energy of the conspecific advertisement call [Lewis and Lombard, 1988; Zakon and Wilczynski, 1988]. Thus the arrangement of frequency in rows in the input matrix, in which each row interacts with a single input neuron, mimics the sensitivity of hair cells to specific frequency bands. This aspect of the neural network model is not arbitrary but intentionally constrained by biology and physics.

We conducted one control experiment in which the stimulus matrix was rotated by 90°, thus the frequency data that was processed as rows is now processed in columns, and the temporal data that was processed in columns is now processed as rows. These nets showed the same patterns of responsiveness to frequency and time as the networks trained in the same training regimes and to the same population but using the unmanipulated stimulus matrix. Thus the differential responses to frequency and time that we observed does not appear to result from how frequency and time are processed by the neural networks.

Preference Landscapes and Pleiotropy

We do not have a formal set of null and alternative hypotheses against which we tested the preference landscapes. Our discussion, however, can be guided by expectations that might arise from a purely nonquantitative analysis and verbal discussion of the issues considered above.

No Conspecific Signal Preferences. One idea is that networks that evolve to discriminate between conspecific and heterospecific signals will respond to conspecific signals categorically; that is, they will ignore variation within the species and be hypersensitive to a similar magnitude of variation between species. The rationale for this expectation might be that the training regime did not impose a fitness differential for preferring one conspecific signal over another, thereby creating a preference landscape that is flat across the conspecific call variation. Our results show this is clearly not the case as there was substantial variation in responses to conspecific signal variation (fig. 8).

Mean Template. If the response properties of the receivers resulted strictly from the matching of a template with no generalization, then we might expect the nets' pref-

erence landscapes to reflect this experience during training. When the networks were trained to discriminate among the mean signal of conspecific versus noise (sr), or the mean signals of the conspecific and heterospecific populations (mn), we might expect the networks to show a strong response to only mean conspecific signal and little or no response to the other conspecific signals. This is clearly not the case (fig. 8). Other signals besides the mean signals elicit responses and the mean signal elicits the strongest response in only one of the sixteen training regimes (fig. 8, population D, variance referential).

Population Distribution. Continuing the template metaphor, in the training regimes in which the networks were trained with a variety of signals that reflected their distribution in the conspecific and heterospecific populations, we might expect a preference landscape similar to the signal landscape (vn, vs). In this case, we might expect to observe the strongest response to the mean signal with responses differing from the mean in the same magnitude as did the signals – that is, a signal 2 SDs from the mean would result in a signal that elicits a response of 2 SD less than that of the mean (fig. 8). These predictions are also not supported for the same reasons detailed above. The strongest responses are usually not elicited by signals nearer the mean (fig. 8).

Character Displacement. The preference landscapes could also show a skew from the mean in the direction that would reduce interaction with other species. This is analogous to the hypothesis of reproductive character displacement in which signals or preferences of species are displaced from heterospecific signals in a manner that reduced false recognition of heterospecifics (fig. 8). Our results tend to support the interpretation that the networks are somewhat influenced by a pattern analogous to displacement during their training. Displacement from the mean tends to be in directions that move the responses away from heterospecific signals thus reducing false positives in recognition of conspecifics. For example, displacement of preferences along the frequency axis would be predicted for populations D and A but not in C and B. Furthermore, to reduce false positives we would expect displacement toward lower frequencies in population D and higher frequencies in population A (fig. 2, 8). This pattern tends to hold for the eight training regimes in populations D and B, but only for two of the training regimes in population C. For population A, the direction of displacement from the mean is not consistent with a character displacement effect.

This study was not designed specifically to simulate the effects of character displacement. Instead, we asked the more general question of whether selection for conspecific-heterospecific discrimination could generate sexual selec-

tion on signals within populations. These results show this is clearly the case. Nets do not respond equally to all conspecific variants, and the strongest responses are usually not to the mean signal. Thus sexual selection appears to be directional or diversifying rather than stabilizing, and the displacement of the responses from the mean is sometimes consistent with the direction that would decrease false recognition of heterospecific and thus mimics the effects of reproductive character displacement of preferences.

Summary

This study uses artificial neural network models to investigate the evolution of species recognition among sympatric species. The models assume that there is a fitness benefit associated with correct identification of the conspecific signal relative to heterospecific signals and noise. This is a realistic assumption, and is implicit in many discussions of the divergence of communication systems. A further assumption is that the fitness benefits accrued to the receiver do not vary with the conspecific signal chosen. This is a controversial issue in sexual selection, and it is fair to say that in some systems females increase their immediate fecundity through prudent mate choice, although this is not true in other systems [Andersson, 1994; Ryan, 1997]. An important limitation of this study is that signals are fixed and recognition evolves. A more detailed understanding of the dynamics of some of these processes, such as character displacement, should simulate the evolution of both signal and receiver [e.g. Holmgren and Enquist, 1999].

The addition of neural network models to studies of communication system evolution introduces some advantages not apparent in other approaches. In empirical studies the distribution of signal variables might constrain competing hypotheses from making mutually exclusive predictions. Other types of simulations and population genetic models [e.g. Pfennig, 1998] explicitly define or imply the decoding strategy of the receiver, but the less deterministic generalizations that arise from artificial neural networks might be a better indictor of how brains really work [e.g. Churchland and Sejnowski, 1992; Phelps and Ryan, 1998]. These types of simulations, therefore, appear to be useful tools, which in addition to knowledge of the behavior, neurobiology, and population genetics of real systems can advance our understanding of mechanisms and evolution of animal communication.

Appendix: Elman Net Equations

Elman Nets belong to a class of discrete time recurrent networks that are dynamical systems with the following characteristics:

1. The hidden units are equivalent to an internal state vector representation $x(t)$ which is an n-dimensional vector – that is of the form $x = (x_1, \ldots, x_n)'$ (the prime symbol denotes vector transpose), where x_1, $i = 1, \ldots, n$ are real numbers – at time t. Note t takes on consecutive integer values from initial time $t = 1$ to final time $t = k$ (i.e. $t = 2, 3, \ldots, k$).

2. The input is a sequence of m-dimensional vectors $u(t)$ (also with real values, in our case nonnegative).

3. The internal state dynamics are a concatenation of two transformations: (i) a linear transformation A ($n \times (n + m)$ matrix) of the internal state and input vectors, $x(t)$ and $u(t)$ respectively, biased by an n-dimensional vector, c, to yield the intermediate n-dimensional vector, $z(t)$, given by

$$z(t) = A \begin{pmatrix} x(t) \\ u(t) \end{pmatrix} + c, \tag{1}$$

and (ii) a hyperbolic tangent updating function that transforms the i-th element of z at time t into the i-th element of x at time $t + 1$ – that is,

$$x_i(t + 1) = \tanh(\alpha_1 z_i(t) + \alpha_2), i = 1, \ldots, n. \tag{2}$$

where the first and second elements of $\alpha = (\alpha_1, \alpha_2)'$ are scaling and bias constants respectively. (More generally, these constants could depend on i.)

4. The output is an r-dimensional vector $y(t + l)$ that represents a linear transformation B ($r \times n$ matrix) of the state $x(t)$ translated (biased) by the constant r-dimensional vector d – that is,

$$y(t + 1) = Bx(t + 1) + d. \tag{3}$$

For the Elman net to act as a computing device, we need to first specify the $n \times (n + m) + r \times n + (n + m) + r + 2$ elements making up the two matrices, two vectors, and two updating-function constants thereby defining the parameter set $P = \{A, B, c, d, \alpha\}$. Then for an initial state $x(0)$ and an $m \times k$-dimensional input matrix $U = (u(0), u(1), \ldots, u(k - 1))$, we can calculate the trajectory of the output as represented by the $r \times k$ matrix $Y = (y(l), y(2), \ldots, y(k))$, using equations (1)–(3). Thus, in effect, for a given parameter set P, the Elman Net is a mapping $E_p: R^n \times R^{m \times k} \rightarrow R^{r \times k}$ (where R^n, etc., is an n-dimensional real Euclidean vector space).

In the study we reported here, the dimension of our output vector is of dimension $r = 1$; that is, the output set is the k-dimensional array $Y = (y(l), \ldots, y(k))$. Further, we always set the initial state to be $x(0) = 0$ (null vector) so that the input matrix U_j produces the same repeatable sequence $Y_j = (y_j(1), \ldots, y_j(k))$. If, every time the net is 'tested' by the input 'signal' U_j we calculate the sum

$$v_j = \sum_{t=1}^{k} y_j(t) \tag{4}$$

then V_j represents a scalar-valued output 'response' of the network to the input matrix U_j.

All the parameters in the set P (i.e. all elements of the transformation matrices A and B and bias vectors c and d) are constrained to have values on the interval [–1, 1]. The input signals we use in our simulations, as elaborated in the text, have the dimensions $m = 28$ and $k = 125$. We selected the number of hidden units to be 20 – i.e. the state x has dimension $n = 20$ – and the summed output v_j has the dimension $r = 1$. Further, at each point in time t, equation (2) implies that $x_i(t)$ is constrained to lie between –1 and +1 for each $i = 1, \ldots, n$. Thus, because of the constraints on the elements of B and d and the dimension $n = 20$ of the vector $x(t)$, equation (3) implies that the corresponding output $y(t)$ at time t (which is a scalar rather than vector because $r = 1$) lies on the interval [–21, 21]. Hence, for any given U_j, because 125 values of the corresponding output $y_j(t)$ $t = 1, \ldots, 125$, are summed to obtain the response v_j (see equation (4), the response is constrained to lie on the interval [–2,625, 2,625]. Thus the maximum difference between any two response values v_{j1} and v_{j2} is 5,250.

The fitness of an Elman net E_p is evaluated in terms of its ability to discriminate between two inputs. In particular, we are interested in the net being able to respond strongly to one of these inputs (e.g. when the input is a signal from a conspecific or some appropriate mate) but not another (e.g. when the signal is an input from a heterospecific or some inappropriate mate). If, for example, we regard acceptance of the 'right' signal U_1 as the occurrence of a large positive value v_1, and rejection of the 'wrong' signal U_2, as the occurrence of the most negative value v_2, then we define fitness in terms of the difference $d_{12} = v_1 - v_2$. In particular, larger values of d_{ij}, where U_i, represents a signal that should be accepted and U_j, represents a signal that should be rejected, correspond to fitter individuals. The maximum value for this difference is 5,250, while the minimum is

–5,250. Note, negative values imply that the net is favoring signals that detract from fitness over those that add to fitness.

In our simulation we used the average of six different evaluations d_{ij} (from six different pairs of beneficial and detrimental signals U_i and U_j, respectively) to obtain a fitness measure. Clearly, this average value $\bar{d}$ lies on the interval [–5,250, 5,250]. To obtain a fitness measure ϕ that lies on [–1, 1], we applied the transformation

$$\phi = \tanh(\alpha_3 \bar{v} + \alpha_4)$$

Pilot studies indicated that the values $\alpha_3 = 1/1,500$ and $\alpha_4 = 0$ provides an extensive range of values for our measure of fitness (with these values we obtained $\phi > 0.9$ for the fittest individuals).

References

Alexander, R.D. (1975) Natural selection and specialized chorusing behavior in acoustical insects. *In* Insects, Society and Science (ed. by D. Pimental), Academic Press, New York, pp. 35–77.

Andersson, M. (1994) Sexual Selection. Princeton University Press, Princeton.

Barlow, G.W. (1977) Modal action patterns. *In* How Animals Communicate (ed. by T.A. Sebeok), Indiana University Press, Bloomington, Indiana, pp. 98–134.

Bateson P., and G. Horn (1994) Imprinting and recognition memory: a neural net model. Anim. Behav., *48:* 695–715.

Beeman, K. (1996) Signal. Engineering Design, Belmont, Massachusetts.

Blair, W.F. (1958) Mating call in the speciation of anuran amphibians. Am. Natur., *92:* 27–51.

Blair, W.F. (1964) Isolating mechanisms and interspecies interactions in anuran amphibians. Quart. Rev. Biol., *39:* 334–344.

Brenowitz, E.A. (1983) The contribution of temporal cues to species recognition in the redwinged blackbird. Anim. Behav., *31:* 116–127.

Butlin, R. (1987) Speciation by reinforcement. Trends Ecol. Evol., *2:* 8–13.

Churchland, P.S., and T.J. Sejnowski (1992) The Computational Brain. The MIT Press, Cambridge, Massachusetts.

Coyne, J.A., and H.A. Orr (1989) Patterns of speciation in *Drosophila*. Evolution, *43:* 362–381.

Coyne, J.A., H.A. Ro, and D.J. Futuyma (1988) Do we need a new species concept? Syst. Zool., *37:* 190–200.

Demuth, H., and M. Beale (1997) Neural Network Toolbox. The Math Works Inc., Natick, Massachusetts.

Dobzhansky, T. (1937) Genetics and the Origin of Species. Columbia University Press, New York.

Dobzhansky, T. (1940) Speciation as a stage in evolutionary divergence. Am. Natur., *74:* 312–332.

Ehret, G. (1990) Categorical perception of sound signals: facts and hypotheses from animals studies. *In* Categorical Perception (ed. by S. Harnad), Cambridge University Press, Cambridge, pp. 301–331.

Elman, J.L. (1990) Finding structure in time. Cogn. Sci., *14:* 179–211.

Emlen, S.T. (1972) An experimental analysis of the parameters of bird song eliciting species recognition. Behaviour, *41:* 130–171.

Enquist, M., and A. Arak (1993) Selection of exaggerated male traits by female aesthetic senses. Nature, *361:* 446–448.

Enquist, M., and A. Arak (1994) Symmetry, beauty and evolution. Nature, *372:* 169–170.

Fisher, R.A. (1930) The Genetical Theory of Natural Selection. Clarendon Press, London.

Fritzsch, B., M.J. Ryan, W. Wilczynski, T.E. Hetherington, and W. Walkowiak (eds.) (1988) The Evolution of the Amphibian Auditory System. John Wiley and Sons, New York.

Gerhardt, H.C. (1991) Female mate choice in treefrogs: static and dynamic acoustic criteria. Anim. Behav., *42:* 615–635.

Gerhardt, H.C. (1994) Reproductive character displacement of female mate choice in the grey treefrog, *Hyla chrysoscelis*. Anim. Behav., *47:* 959–969.

Getz, W.M. (1981) Genetically based kin recognition systems. J. Theor. Biol., *99:* 204–226.

Getz, W.M. (1982) An analysis of learned kin recognition in Hymenoptera. J. Theor. Biol., *99:* 565–597.

Getz, W.M., and R. Page (1991) Chemosensory kin communication systems and kin recognition in honey bees. Ethology, *87:* 298–315.

Hepper, P.G. (1991) Kinship, kin discrimination and mate choice. *In* Kin Recognition (ed. by P.G. Hepper), Cambridge University Press, Cambridge, pp. 125–147.

Holmgren, N.M.A., and M. Enquist (1999) The dynamics of mimicry evolution. Biol. J. Linn. Soc., *66:* 145–158.

Holmgren, N.M.A., and W.M. Getz (2000) Evolution of host plant selection in insects under ecological and perceptual constraints: a simulation study. Evol. Ecol. Res., *2:* 81–106.

Hurd, P.L., C.A. Wachtmeister, and M. Enquist (1995) Darwin's principle of antithesis revisited: a role for perceptual biases in the evolution of intraspecific signals. Proc. R. Soc. Lond. Ser. B, *259:* 201–205.

Johnstone, R.A. (1994) Female preference for symmetrical males as a by-product of selection for mate recognition. Nature, *372:* 172–175.

Klump, G.M. (1996) Bird communication in the noisy world. *In* Ecology and Evolution of Acoustic Communication in Birds (ed. by D.E. Kroodsma and E.H. Miller), Cornell University Press, Ithaca, New York, pp. 310–338.

Lacy, R.C., and P.W. Sherman (1983) Kin recognition by phenotype matching. Am. Natur., *121:* 489–512.

Lewis, E.R., and R.E. Lombard (1988) The amphibian inner ear. *In* The Evolution of the Amphibian Auditory System (ed. by B. Fritzsch, M.J. Ryan, W. Wilczynski, T.E. Hetherington, and W. Walkowiak), John Wiley and Sons, New York, pp. 93–123.

Littlejohn, M.J. (1988) Frog calls and speciation; the retrograde evolution of homogamic acoustic signaling systems in hybrid zones. *In* The Evolution of the Amphibian Auditory System (ed. by B. Fritzsch, M.J. Ryan, W. Wilczynski, T.E. Hetherington, and W. Walkowiak), John Wiley and Sons, New York, pp. 613–635.

Mayr, E. (1963) Animal Species and Evolution. Harvard University Press, Cambridge, Massachusetts.

Miller, E.H. (1996) Acoustic differentiation and speciation in shorebirds. *In* Ecology and Evolution of Acoustic Communication in Birds (ed. by D.E. Kroodsma and E.H. Miller), Cornell University Press, Ithaca, New York, pp. 241–257.

Nelson, D.A. (1988) Feature weighting in species song recognition by the field sparrow *(Spizella pusilla)*. Behaviour, *106:* 158–182.

Nelson, D.A., and P. Marler (1990) The perception of bird song and an ecological concept of signal space. *In* Comparative Perception (ed. by E.C. Stebbins and M.A. Berkeley), Wiley, New York, pp. 443–478.

Passmore, N.I. (1981) The relevance of the specific mate recognition concept to anuran reproductive biology. Mon. Zool. Ital., *6:* 93–108.

Paterson, H.E.H. (1978) More evidence against speciation by reinforcement. So. African J. Sci., *74:* 369–371.

Paterson, H.E.H. (1982) Perspective in speciation by reinforcement. So. Afr. J. Sci., *78:* 53–57.

Paterson, H.E.H. (1985) The recognition concept of species. *In* Species and Speciation (ed. by E. Vrba), Transvaal Mus. Monogr. No. 4. Pretoria, Transvaal Mus., pp. 21–29.

Paterson, H.E.H. (1993) Variation and the specific-mate recognition system. Persp. Ethol., *10:* 209–237.

Pfennig, K.S. (1998) The evolution of mate choice and the potential conflict between species and mate-quality recognition. Proc. R. Soc. Lond. Ser. B, *265:* 1743–1748.

Phelps, S.M., and M.J. Ryan (1998) Neural networks predict response biases in female túngara frogs. Proc. Roy. Soc. Lond. Ser. B, *265:* 279–285.

Phelps, S.M., and M.J. Ryan (2000) History influences responses of neural networks and túngara frogs. Proc. Roy. Soc. Lond. Ser. B (in press).

Ryan, M.J. (1988) Constraints and patterns in the evolution of anuran acoustic communication. *In* The Evolution of the Amphibian Auditory System (ed. by B. Fritzsch, M.J. Ryan, W. Wilczynski, T.E. Hetherington, and W. Walkowiak), John Wiley and Sons, New York, pp. 637–677.

Ryan, M.J. (1997) Sexual selection and mate choice. *In* Behavioral Ecology, An Evolutionary Approach (ed. by J.R. Krebs and N.B. Davies), Blackwell Sciences Ltd, Oxford, pp. 179–202.

Ryan, M.J. (1998) Sexual selection, receiver biases, and the evolution of sex differences. Science, *281:* 1999–2003.

Ryan, M.J., and E.A. Brenowitz (1985) The role of body size, phylogeny, and ambient noise in the evolution of bird song. Amer. Natur., *126:* 87–100.

Ryan, M.J., and A.S. Rand (1993) Species recognition and sexual selection as a unitary problem in animal communication. Evolution, *47:* 647–657.

Ryan, M.J., and W. Wilczynski (1991) Evolution of intraspecific variation in the call of a cricket frog (*Acris crepitans,* Hylidae). Biol. J. Linn. Soc., *44:* 249–271.

Ryan, M.J., L. Dries, P. Batra, and D.M. Hillis (1996) Male mate preference in a gynogenetic species complex of Amazon mollies. Anim. Behav., *52:* 1225–1236.

Saetre, G.-P., T. Moum, S. Bures, M. Kral, M. Adamjan, and J. Moreno (1997) A sexually selected character displacement in flycatchers reinforces premating isolation. Nature, *387:* 589–592.

Wilczynski, W., A.S. Rand, and M.J. Ryan (1995) The processing of spectral cues by the call analysis system of the túngara frog, *Physalaemus pustulosus.* Anim. Behav., *49:* 911–929.

Zakon, H.H., and W. Wilczynski (1988) The physiology of the anuran eighth nerve. *In* The Evolution of the Amphibian Auditory System (ed. by B. Fritzsch, M.J. Ryan, W. Wilczynski, T.E. Hetherington, and W. Walkowiak), John Wiley and Sons, New York, pp. 125–155.