

Artificial neural network discrimination of black-capped chickadee (*Poecile atricapillus*) call notes

Carly M. Nickerson, Laurie L. Bloomfield, and Michael R. W. Dawson
Department of Psychology, University of Alberta, Edmonton, Alberta, T6G 2P9, Canada

Christopher B. Sturdy
Department of Psychology, Centre for Neuroscience, University of Alberta, Edmonton, Alberta, T6G 2P9, Canada

(Received 9 March 2006; revised 11 April 2006; accepted 12 May 2006)

Artificial neural networks were trained to discriminate between two different notes from the “chick-a-dee” call of the black-capped chickadee (*Poecile atricapillus*). An individual note was represented as a vector of nine summary features taken from note spectrograms. A network was trained to respond to exemplar notes of one type (e.g., A notes) and to fail to respond to exemplar notes of another type (e.g., B notes). After this training, the network was presented novel notes of the two different types, as well as notes of the same two types that had been shifted upwards or downwards in frequency. The strength of the response of the network to each novel and shifted note was recorded. When network responses were plotted as a function of the degree of frequency shift, the results were very similar to those observed in birds that were trained in an analogous task [Charrier *et al.*, *J. Comp. Psychol.* **119**(4), 371–380 (2005)]. The implications of these results to simulating behavioral studies of animal communication are discussed.

© 2006 Acoustical Society of America. [DOI: 10.1121/1.2211509]

PACS number(s): 43.80Ka., 43.80Lb, 43.80Jz [JAS]

Pages: 1111–1117

I. INTRODUCTION

An artificial neural network (ANN) is a computer program that can learn to accomplish tasks such as pattern recognition (Pao, 1989). It is a simulation of neuron-like processors that accept incoming signals from one or more other processors, convert these signals into a level of internal activity, and finally convert this activity into an output signal. Signals are sent through weighted, modifiable connections that are analogous to synapses. A brief tutorial on ANNs is provided by Dawson *et al.* (2006). For a more extensive introduction to such networks, a number of sources are available (Bachetl and Abrahamsen, 1991; Caudill and Butler, 1992; Dawson, 2004; Gluck and Myers, 2001; Quinlan, 1991; Ripley, 1996).

ANNs are increasingly popular statistical tools in the study of animal vocalization and communication. For example, they have been used to classify bat species on the basis of echolocation calls (Parson and Jones, 2000), to categorize different calls produced by false killer whales (Murray *et al.*, 1998), to identify dolphin echolocation clicks (Houser *et al.*, 1999), to recognize stress calls of domestic pigs (Schon *et al.*, 2001), to distinguish particular female sea lions on the basis of their mother-pup contact calls (Campbell *et al.*, 2002), to assign a variety of Manitoban songbirds to their species on the basis of recorded songs (MaIlraith and Card, 1997), and to classify the component notes of the “chick-a-dee” call of the black-capped chickadee (Dawson *et al.*, 2006).

In addition to their statistical use, ANNs can also serve as models of animal behavior. In particular, ANNs provide interesting accounts of classical conditioning and discrimina-

tion learning (Dawson and Spetch, 2005; Delamater *et al.*, 1999; Gluck and Myers, 2001; Pearce, 1987). In such simulations, a single output unit of a network represents an animal’s response. The unit is trained to turn on when the unconditioned stimulus (US) is presented, and to turn off when the US is absent. A set of input units is used to represent the present or absence of a particular conditioned stimulus (CS). For example, consider a simple network with two input units that are used to represent the presence of conditioned stimuli A and B (e.g., Dawson and Spetch, 2005). To simulate a positive patterning paradigm, this network would be trained to turn on when both input units were activated to represent a compound stimulus (i.e., AB+), and to turn off when only one input unit was activated to represent that only one of the stimulus components was present (i.e., A–, B–).

The purpose of this paper is to examine the ability of such a network to deal with what would appear to be a more complicated discrimination learning task from the animal communication literature (Charrier *et al.*, 2005). In the simulations below, rather than representing individual CSs (local representation), each input unit represents the value of a stimulus feature. An individual stimulus is defined by the entire vector of input unit values (distributed representation). The network is trained to respond to some feature vectors, and to not respond to others. There are two questions addressed in the simulations below. First, can a simple network perform discrimination learning when it is defined in this distributed fashion? Second, if such discrimination learning is possible, how well does the behavior of the networks resemble the choice behavior of animals trained to perform an analogous task?

A. Chickadee discriminations of notes and frequency-shifted notes

The chick-a-dee call of the black-capped chickadee (*Parus atricapillus*) contains four note types (A, B, C, and D) produced in a fixed order (A → B → C → D), but note types can be repeated or omitted to produce chickadee calls with seemingly infinite combinations of notes (e.g., ACCCCD, ABDDD). Many researchers are interested in the composition of this call, as it more closely resembles human language syntax than any other animal system (Haliman and Ficken, 1986). As a result, numerous studies have explored the regularities, and the information contained, in chickadee call syntax (Baker *et al.*, 1987; Charrier and Sturdy, 2005; Clucas *et al.*, 2004; Ficken *et al.*, 1994; Freeberg and Lucas, 2002; Haliman *et al.*, 1985).

However, in addition to the study of the syntax of the chickadee call, an analysis of its constituent notes is also important. This is because the individual notes in a call can themselves communicate crucial information, such as the individual identity and gender of the caller (Charrier *et al.*, 2004). Playback studies have shown that chickadees can discriminate the calls of their own flock from those of foreign flocks (Nowicki, 1983), a function that is apparently mediated by D notes (Mammen and Nowicki, 1981). C notes may indicate the location or availability of food sources (Freeberg and Lucas, 2002). Individual A, B, C, and D notes can be used to distinguish black-capped chickadees from another species, the mountain chickadee *P. gambeli* (Dawson *et al.*, 2006).

Discrimination learning provides one paradigm for studying the ability of chickadees to process component notes of the chickadee call. For example, Charrier *et al.* (2005) first trained black-capped chickadees to discriminate two individual note types (e.g., A and B) by responding after hearing one (by visiting a feeder) and by withholding a response after hearing the other (by refraining from visiting a feeder). Then, they studied the responses of the birds to modified notes in a second phase. Birds were presented notes that had been frequency shifted up or down by between 0.5 and 2.5 standard deviations, in 0.5-SD steps. The responses of the birds to these shifted notes (i.e., whether or not the birds flew to a feeder) were used as an index of how the birds classified the transformed stimuli. Charrier *et al.* found, for example, that shifting the frequency of A notes down resulted in them being classified as being “B-like,” while shifting the frequency of B notes upwards produced behavior that indicated that they were classified as being “A-like.”

The purpose of the present study is to explore the ability of simple ANNs to simulate the experiment conducted by Charrier *et al.* (2005). In a training phase, ANNs learned to discriminate between two different types of notes from the chickadee call. In different conditions, this discrimination learning used the same pairings as those studied by Charrier *et al.* (i.e., A+B-, A-B+, B+C-, and B-C+). In a generalization phase, the ANNs were presented novel notes from the call, as well as notes whose frequency had been shifted up or down. Of interest was whether the networks could learn to discriminate between different note types, and

whether the responses of the networks to the novel and shifted notes would simulate the behavioral data collected previously by Charrier *et al.*

II. MATERIALS AND METHODS

A. Training stimuli

Sixty notes that were originally used by Charrier *et al.* (2005) were used to create training set to teach ANNs to discriminate between different note types (A, B, and C). Twenty notes each of A, B, and C-note types were randomly selected from black-capped chickadee calls obtained from Colorado, USA and Ontario, Canada. Equal representations of both sources were used and all notes were of high quality. The notes were bandpassed filtered (1–10 kHz) to remove background noise and call-note amplitude was equalized using SIGNAL v 4.0 (Beeman, 2002; Engineering Design, CA). Only non-D notes were utilized, as D notes cannot be quantified on the same features as the A, B, and C notes; more specifically, D notes are acoustically dissimilar from non-D notes and are rarely confused with adjacent notes. For example, D notes are longer duration and consist of a harmonic series (Charrier *et al.*, 2004).

For each note, nine acoustic features were measured using a standard methodology (Nowicki and Nelson, 1990) that was also employed by Charrier *et al.* (2004) and Dawson *et al.* (2006). These features consisted of SF (start frequency), EF (end frequency), PF (peak frequency), AD (ascending duration), DD (descending duration), TD (total duration), FM_{asc} (slope of the ascending frequency modulation), FM_{desc} (slope of the descending frequency modulation), and F_{max} (highest amplitude frequency).

The frequency measurements (SF, EF, PF) were obtained in a digital spectrogram (window size 1024 points, frequency precision=43 Hz) using a cutoff amplitude of -35 dB relative to peak amplitude in the note. Maximum frequency (F_{max}) was measured in a power spectrum using a window size of 4096 points (frequency precision=11 Hz). We measured the main (highest amplitude) harmonic of A and B notes, and the fundamental harmonic of C notes. The duration measurements (AD, DD, TD) were quantified in a digital spectrogram (window size 256 points, temporal precision=5.8 ms). FM_{asc} , slope of ascending frequency (Hz/ms), and FM_{desc} , slope of descending frequency (Hz/ms), were calculated using the following formulas: $(PF-SF)/(AD)$ and $(EF-PF)/(DD)$, respectively. Therefore, nine summary features formed a vector that was used to describe each note, and was used as a stimulus for the ANN to discriminate.

Prior to being presented to the network, each of the nine features for the training set was converted into a z score. Dawson *et al.* (2006) had done this to reduce the range of the input features (which can cause problems when networks are trained) while maintaining the essential characteristics of the raw data.

B. Probe and shifted stimuli

A set of 330 test stimuli was also created to examine the responses of ANNs to notes to which they had not been

TABLE I. Examples of feature vectors used to represent notes as inputs to the networks. The examples below are based on one A note (0 SD shift), and are produced by shifting the frequency values of this note upwards or downwards. All feature values have been converted into z scores as described in the text.

Frequency shift in SD	Input unit encoding of note features								
	SF	PF	EF	AD	DD	TD	FM _{asc}	FM _{desc}	F _{max}
-2.5	0.38	-1.04	0.59	-1.43	1.40	-0.92	-0.47	2.01	-0.94
-2.0	0.64	-0.53	1.17	-1.43	1.40	-0.92	-0.47	2.01	-0.42
-1.5	0.89	-0.01	1.75	-1.43	1.40	-0.92	-0.47	2.01	0.10
-1.0	1.15	0.50	2.32	-1.43	1.40	-0.92	-0.47	2.01	0.62
-0.5	1.40	1.02	2.90	-1.43	1.40	-0.92	-0.47	2.01	1.13
0	1.65	1.53	3.48	-1.43	1.40	-0.92	-0.47	2.01	1.65
0.5	1.91	2.04	4.06	-1.43	1.40	-0.92	-0.47	2.01	2.17
1.0	2.16	2.54	4.63	-1.43	1.40	-0.92	-0.47	2.01	2.69
1.5	2.42	3.07	5.21	-1.43	1.40	-0.92	-0.47	2.01	3.20
2.0	2.67	3.58	5.79	-1.43	1.40	-0.92	-0.47	2.01	3.72
2.5	2.92	4.10	6.36	-1.43	1.40	-0.92	-0.47	2.01	4.24

exposed to during training. Thirty of these were probe notes (10 A, 10 B, 10 C) used by Charrier *et al.* (2005) which were standard notes that were randomly sampled according to the same procedures that were described for the training stimuli. These 30 probe notes were then linearly shifted upwards or downwards in frequency to create the remaining 300 test stimuli, using the same shifts as those studied by Charrier *et al.*

Shifted notes were created by raising or lowering each probe note's frequency while maintaining natural amplitude and temporal structure. In essence, the entire spectrogram was shifted in frequency, and then the frequency features used to represent the spectrogram were recomputed. The only features that were changed by this procedure were SF, PF, EF, and F_{\max} . These features were manipulated by either adding or subtracting those values used by Charrier *et al.* (2005) from the SF, PF, EF, and F_{\max} values within each vector for each note (see Table I for examples of positive and negative frequency-shifted vectors). Charrier *et al.* (2005) shifted each probe note up or down by the following standard deviations of frequency: ± 0.5 , ± 1.0 , ± 1.5 , ± 2.0 , ± 2.5 SD. Each of the 30 test notes in the current study was treated in exactly the same fashion, creating 10 new shifted notes for each, and ultimately producing 330 stimuli (10 standard and 100 shifted A, B, and C notes, respectively; see Table I for an example from one note). The SDs used to shift the three different note types were 800 Hz for A notes, 500 Hz for B notes, and 150 Hz for C notes (Charrier *et al.*, 2004, 2005). The nine-feature vectors used to represent each test stimulus were also converted to z scores, using the means and standard deviations of the training stimuli.

C. Network architecture

The networks trained in this study were integration devices (Dawson, 2004). An integration device is a perceptron (Rosenblatt, 1962) that uses the sigmoid-shaped logistic equation as an activation function in its output unit. Each network had one output unit, used to represent the strength of its response to a presented stimulus. Each network also had

nine input units, each of which represented the value of a summary feature when a stimulus note was presented to the network.

D. Discrimination learning

To conduct discrimination learning with a network, 40 patterns (i.e., the vectors representing 40 different notes) were selected from the 60 training stimuli that were described earlier. Twenty patterns were all notes of one type (A, B, or C), and the remaining patterns were all notes of another type. The network was trained to respond to every instance of one note type, and not to respond to every instance of the other. Following Charrier *et al.* (2005), four different combinations of note type/response type pairings were used: A+B-, A-B+, B+C-, and B-C+.

Prior to training an individual network, all of its connection weights were set to random values selected from the range between -0.1 and 0.1. The biases of the output units (i.e., the "thresholds" of the logistic activation functions) were initialized to 0. Because a network began with small, random connection weights, 20 different networks were trained in independent simulations. Each network was viewed as a "subject" in an experiment, with the results of the experiment being averaged over each subject.

The networks were trained with a gradient descent version of the delta rule (Dawson, 2004, 2005). This rule modifies network weights in a fashion that is identical to how the output unit weights in a more complex multilayer network (i.e., a network that has a layer of processors between input and output units) are trained by the generalized delta rule (Rumelhart *et al.*, 1986). Such a learning rule presents a stimulus, measures the error in the network's response to the stimulus, and then uses this error to modify the connection weights in the network. These modifications are such that the next time the pattern is presented to the network, its response error will be smaller. During a training sweep, each pattern is presented once, and connection weights are modified after each presentation. The order in which patterns are presented

is randomized every sweep. A dependent measure of interest is the number of sweeps of training required for a network to converge to a solution.

The networks were trained with the ROSENBLATT program (Dawson, 2005) that is available free of charge from <http://www.bcp.psych.ualberta.ca/~7emike/Software/Rosenblatt/index.html>. The learning rate for the simulations was 0.5, with zero momentum. Training ends once the network generates a “hit” to every pattern. A hit occurs when the network produces a response of 0.9 or more to every pattern to which it is supposed to respond, or when it produces a response of 0.1 or less to every pattern to which it is not supposed to respond.

E. Testing responses to probe notes and shifted notes

After a network was trained according to the procedure described above, it was presented 20 probe notes and 200 shifted notes that had not been used during discrimination learning. As in the Charrier *et al.* (2005) study, the probe notes and the shifted notes were of the same two types as those presented during discrimination learning. For instance, in the A+B– condition, a network would be presented the 10 probe A notes, the 10 probe B notes, the 100 shifted A notes (i.e., each probe note shifted up or down by the amount previously described), and the 100 shifted B notes. The dependent measure for this phase of the study was the strength of a network’s response to each of these 220 new stimuli.

III. RESULTS

A. Discrimination learning

The first question of interest was whether simple ANNs could learn this discrimination task when a distributed code was used to represent stimulus notes. The results indicated that this was not a problem for this architecture, because every network in every condition converged to a solution. The average number of sweeps required for this was 639.1 for the ten networks in the A–B+ condition, 638.9 for the A+B– condition, 259.4 for the B–C+ condition, and 259.3 for the B+C– condition. There were no significant differences in speed of learning between the networks in the two AB conditions, or between the networks in the two BC conditions. When all of the AB condition networks and all of the BC condition networks were treated as being in the same learning condition, it was found that the speed of convergence for the BC networks was significantly faster than that of the AB networks ($t=259.308$, $df=18$, $p<0.0001$).

B. Responses to shifted notes

The second question of interest was how the trained networks would respond to the novel notes (probe notes and shifted notes). In particular, to what extent would network responses to these notes simulate the responses of birds?

Figure 1 presents a plot of the average network responses to novel A notes and B notes after both the A–B+ and the A+B– discrimination learning. Qualitatively, these results are very similar to the bird responses that were re-

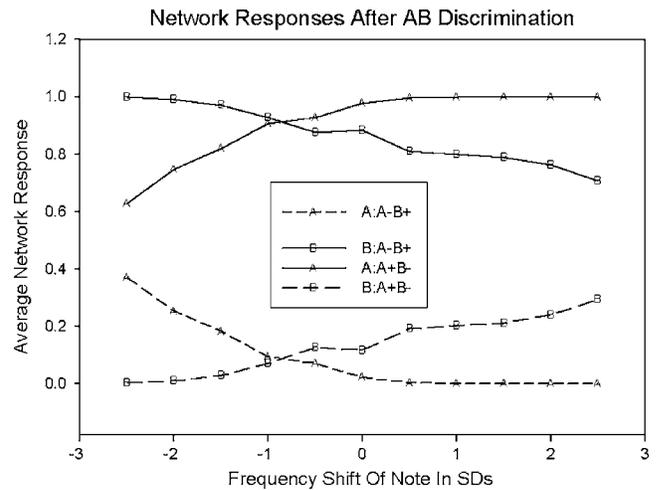


FIG. 1. Average response of networks to shifted notes after AB discrimination learning. The symbol (A, B) indicates the type of note presented to the network. Solid lines represent results from trials in which the network was trained to respond to a note, while dashed lines represent results from trials in which the network was trained not to respond to a note.

ported by Charrier *et al.* (2005). Consider the A notes in the A–B+ condition. When they are shifted up in frequency, there is little change in response: the networks continue not to respond to these notes. However, when the notes are shifted down in frequency, these notes become more “B-like,” because the networks begin to generate weaker responses to them. Now consider the B notes in the same condition. When they are shifted up, these notes become more “A-like,” because the networks reduce their response to them. When they are shifted down, the networks actually increase their response, indicating that the notes seem more “B-like”. A similar account can be made for the other two lines in the graph (from the A+B– condition), which are essentially mirror images of the two lines that have just been discussed.

Figure 2 presents a plot of the average network re-

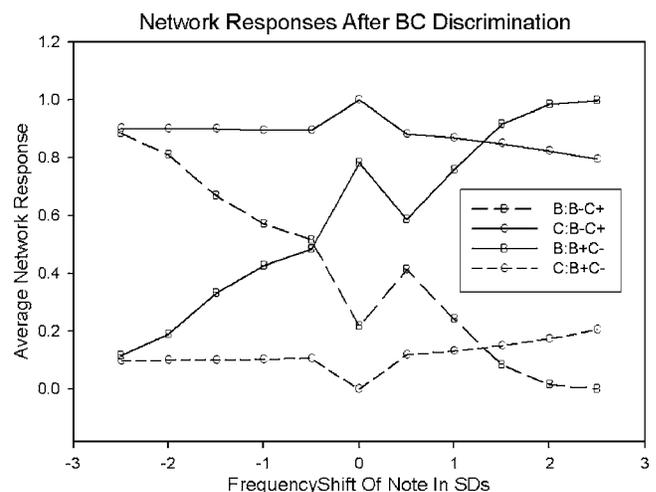


FIG. 2. Average response of networks to shifted notes after BC discrimination learning. The symbol (B, C) indicates the type of note presented to the network. Solid lines represent results from trials in which the network was trained to respond to a note, while dashed lines represent results from trials in which the network was trained not to respond to a note.

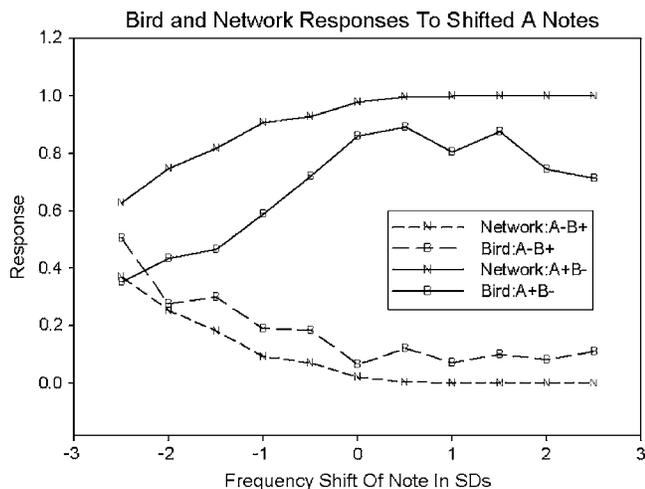


FIG. 3. Comparison of average response of black-capped chickadees and ANN to shifted A notes after AB discrimination learning. The symbol (B, N) indicates whether the function is produced by a bird (B) or an ANN (N). Solid lines represent results from A+B- trials, while dashed lines represent results from A-B+ trials.

sponses to novel B notes and C notes after both the B-C+ and the B+C- discrimination learning. Again, these results are very similar to the bird responses that were reported by Charrier *et al.* (2005). Consider the B notes in the B-C+ condition. When these notes are shifted up in frequency, there is a decrease in network response. However, when the notes are shifted down in frequency, these notes become more “C-like,” because the networks begin to generate stronger responses to them. Now consider the C notes in the same condition. Regardless of whether they were shifted up or down, there was little effect of the shift on network response. This is consistent with the result of Charrier *et al.*, who found no significant effect of frequency shift on bird responses to C notes. Again, a similar account can be made for the other two lines in the graph (from the B+C- condition), which are essentially mirror images of the two lines that have just been discussed (Fig. 3).

To provide an additional quantitative assessment of the relationship between the network and animal responses, correlations were computed between the data used to plot the eight different lines in Figs. 1 and 2, and the analogous data used by Charrier *et al.* (2005) to plot average chickadee responses to shifted notes in their Figs. 4 and 5. Each of these correlations is based upon 11 pairs of responses, and is presented in Table II. As can be seen from this table, for all eight graphs there was a substantially high positive correlation,

indicating that the network responses were excellent simulations of the bird responses. The overall correlation between the two datasets (i.e., between all 88 data pairs taken as a group) was 0.69.

IV. DISCUSSION

To summarize, a number of simple ANNs were trained to discriminate between two different component notes from the chick-a-dee call of the black-capped chickadee. After this training was accomplished, the network was presented novel notes that had been shifted in frequency. When the responses of networks to shifted notes was compared to the responses of birds trained in an analogous task, a high degree of correspondence was observed.

Previous research has shown that the nine summary features used to represent stimuli in the current study can be used to classify important aspects of the chick-a-dee call. For example, Dawson *et al.* (2006) demonstrated that component notes of this call (i.e., A, B, and C notes) could be classified to a high degree of accuracy on the basis of these features. This was true for both ANNs and linear discriminant analysis. Similarly, (Dawson *et al.*, 2006) represented A, B, C, and D notes with these summary features. They found that linear discriminant analysis could use these features, for any of the four note types, to discriminate black-capped chickadees from mountain chickadees.

While this previous research has indicated that the summary features used in the current study are extremely useful with respect to the statistical analysis of these notes, it says little about whether the birds themselves might be sensitive to similar features when processing the different notes in the chick-a-dee call. The simulation results in the current paper are one step towards modeling how the birds themselves might process and respond to the component notes.

The goal of simulation research is to develop models that are strongly equivalent to the agents or the phenomena that are being modeled (Pylyshyn, 1984). A strongly equivalent model of how chickadees discriminate component notes would generate the same performance as the birds, and would do so by exploiting the same procedures, mechanisms, and representations. A fundamental issue in validating simulations is collecting evidence to support the claim that they are strongly equivalent. This is particularly problematic when a researcher does not have direct access to the entity being simulated (e.g., the internal processes used by chickadees to identify notes). What kind of evidence is available to simulation researchers?

TABLE II. Correlations between bird responses and network responses to shifted notes for the eight different lines plotted in Figs. 1 and 2 in the current paper, and in Figs. 4 and 5 from Charrier *et al.* (2005).

	Relations between responses to shifted A and B notes after AB discrimination learning		Relations between responses to shifted B and C notes after BC discrimination learning	
	A-B+ Condition	A+B- Condition	B-C+ Condition	B+C- Condition
A	0.97	0.92	B	0.86
B	0.83	0.79	C	0.85

Pylshyn (1984) has suggested that one type of information that can be used to evaluate a simulation is error evidence: when the simulation makes mistakes, are these mistakes similar to those generated by the modeled subjects? The current results provide error evidence that can be used to evaluate the ANNs. When both the ANNs in the current study, and the birds in the Charrier *et al.* (2005) study, were presented the shifted notes, these were novel and deviant stimuli. The extent to which the responses (e.g., misclassifications) of the birds correspond to the responses of the ANNs is the extent to which error evidence supports these networks as models of the bird behavior. The high degree of correspondence that was observed between the two types of data (Table II) indicates that these ANNs are excellent models of chickadee discriminations between note types.

What, then, do these networks say about this discrimination learning task? First, they indicate that a linear combination of the nine summary features that were taken from the spectrograms provides sufficient information to accomplish this task. Second, the likelihood of responding to a note can be modeled as a nonlinear function of these combined features (e.g., the sigmoid activation function). Third, higher-order featural combinations (of the sort that would be provided by the hidden units in a multilayer network) are not required to accomplish this task.

This last point is relevant to a more general implication of the current simulation results. One important issue in the discrimination learning literature concerns how animals process information contained in compound stimuli. One approach to this issue is elemental (Rescorla, 1973, 1988, 2003; Rescorla and Wagner, 1972; Wagner, 2003). According to this approach, animals independently process the elemental components that define a compound stimulus. A second approach to this issue is configural (Pearce, 1987, 1994, 1997, 2002; Pearce and Bouton, 2001; Wasserman and Miller, 1997). According to this approach, compound stimuli are processed holistically: while a compound stimulus is related to its components, there is additional information that is represented that reflects the notion that a compound stimulus is not merely the sum of each element.

The elemental and configural approaches have strong implications for representing stimuli to be presented to ANNs. A typical elemental representation would have a single input unit representing the presence or absence of the elements of a compound stimulus, and no other input units (e.g., Pearce, 1997, p. 131). For example, a simple network would use two units that could represent the presence of stimulus elements A and B. The presence of the compound stimulus AB would be indicated by turning both of these units on as an elemental representation. In contrast, a typical configural representation includes additional units that are turned on only when a compound stimulus is presented (e.g., Pearce, 1997, p. 132). For example, three units would be required to represent compound stimulus AB: one for element A, another for element B, and a third to represent unique configural properties (AB). This third element is only activated when all of the stimulus compounds are presented. More complex configural representations—in particular,

those that can be changed via learning—are represented in the hidden units of multilayer networks (Delamater *et al.*, 1999).

The current simulations provide a stimulus representation that does not fit neatly into any of the categories described above. The reason for this is that the networks use a representation that appears to be elemental from the description in the preceding paragraph. However, the stimuli that are presented to the network are feature configurations. Indeed, Dawson *et al.* (2006) found that all of the presented features were necessary for classifying chickadee call notes. The fact that the current networks were able to learn to discriminate note types on the basis of vectors of stimulus features suggests that configural learning may not necessarily require the presence of units that explicitly represent configurations of elements. Further studies of this sort of representation are required to explore the implications of networks like ours to the elemental/configural debate.

ACKNOWLEDGMENTS

This research was supported by an NSERC Discovery Grant, an Alberta Ingenuity New Faculty Grant, a Canada Foundation for Innovation (CFI) New Opportunities Grant as well as start-up and CFI partner funding from the University of Alberta to C.B.S. and NSERC and SSHRC research grants to M.R.W.D. L.L.B. was supported by an NSERC PGS-B and an Alberta Ingenuity Studentship. This research was approved by the University of Alberta Biological Sciences Animal Care Committee (Protocol No. 351201, originally approved on 14 November 2001) and the University of Calgary Life and Environmental Sciences Animal Care Committee (Protocol No. BI2001-028 originally approved on 29 November 2001). Birds were captured under an Environment Canada, Canadian Wildlife Service Scientific permit (Permit No. WSA-1-02, originally approved on 25 January 2002) and Alberta Fish and Wildlife Research permits (Nos. 4619 GP, 4621 GP, and 8734 GP) and collection licenses (Nos. 088 CN, 089 CN, and 147 CN) that were originally approved on 21 December 2001. Correspondence can be addressed to C.B.S.

- Baker, M. C., Bjerke, T. K., Lampe, H. U., and Espmark, Y. O. (1987). "Sexual-response of female yellowhammers to differences in regional song dialects and repertoire sizes." *Anim. Behav.* **35**, 395–401.
- Bechtel, W., and Abrahamsen, A. (1991). *Connectionism And The Mind* (Basil Blackwell, Cambridge, MA).
- Beeman, Kim. (2002). *SIGNAL 4.0 Reference Guide* (Engineering Design, Berkeley, CA).
- Campbell, G. S., Gisiner, R. C., Helweg, D. A., and Milette, L. L. (2002). "Acoustic identification of female Steller sea lions (*Eumetopias jubatus*)," *J. Acoust. Soc. Am.* **111**(6), 2920–2928.
- Caudill, M., and Butler, B. (1992). *Understanding Neural Networks* (MIT Press, Cambridge, MA), Vol. 1.
- Charrier, I., and Sturdy, C. B. (2005). "Call-based species recognition in black-capped chickadees," *Behav. Processes* **70**(3), 271–281.
- Charrier, I., Bloomfield, L. L., and Sturdy, C. B. (2004). "Note types and coding in Parid vocalizations. I. The chick-a-dee call of the black-capped chickadee (*Poecile atricapilla*)," *Can. J. Zool.* **82**, 769–779.
- Charrier, I., Lee, T. T. Y., Bloomfield, L. L., and Sturdy, C. B. (2005). "Acoustic mechanisms of note-type perception in black-capped chickadee (*Poecile atricapillus*) calls," *J. Comp. Psychol.* **119**(4), 371–380.
- Clucas, B. A., Freeberg, T. M., and Lucas, J. R. (2004). "Chick-a-dee call syntax, social context, and season affect vocal responses of Carolina

- chickadees (*Poecile carolinensis*)," Behav. Ecol. Sociobiol. **57**(2), 187–196.
- Dawson, M. R. W. (2004). *Minds And Machines: Connectionism And Psychological Modeling* (Blackwell, Malden, MA).
- Dawson, M. R. W. (2005). *Connectionism: A Hands-on Approach*, 1st ed. (Blackwell Pub, Malden, MA).
- Dawson, M. R. W., and Spetch, M. L. (2005). "Traditional perceptrons do not generate the overexpectation effect," *Neur. Info. Proc. – Letters and Reviews* **7**(1), 11–17.
- Dawson, M. R. W., Charrier, I., and Sturdy, C. B. (2006). "Using an artificial neural network to classify black-capped chickadee (*Poecile atricapillus*) call note types," *J. Acoust. Soc. Am.* **199**(5), 3161–3172.
- Dawson, M. R. W., Bloomfield, L. L., Charrier, I., and Sturdy, C. B. (2006). "Using non-D notes of the 'chick-a-dee' call to discriminate the black-capped chickadee (*Poecile atricapillus*) from the mountain chickadee (*Poecile gambeli*)," *J. Comp. Psychol.* **120**(2), 147–153.
- Delamater, A. R., Sosa, W., and Katz, M. (1999). "Elemental and configural processes in patterning discrimination learning," *Q. J. Exp. Psychol. B* **52B**, 97–124.
- Ficken, M. S., Hailman, E. D., and Hailman, J. P. (1994). "The chick-a-dee call system of the Mexican chickadee," *Condor* **96**(1), 70–82.
- Freeberg, T. M., and Lucas, J. R. (2002). "Receivers respond differently to chick-a-dee calls varying in note composition in Carolina chickadees, *Poecile carolinensis*," *Anim. Behav.* **63**, 837–845.
- Gluck, M. A., and Myers, C. (2001). *Gateway to Memory: An Introduction to Neural Network Modeling of the Hippocampus and Learning* (MIT Press, Cambridge, MA).
- Hailman, J. P., and Ficken, M. S. (1986). "Combinatorial animal communication with computable syntax—chick-a-dee calling qualifies as language by structural linguistics," *Anim. Behav.* **34**, 1899–1901.
- Hailman, J. P., Ficken, M. S., and Ficken, R. W. (1985). "The chick-a-dee calls of *Parus atricapillus*—a recombinant system of animal communication compared with written English," *Semiotica* **56**(3–4), 191–224.
- Houser, D. S., Helweg, D. A., and Moore, P. W. (1999). "Classification of dolphin echolocation clicks by energy and frequency distributions," *J. Acoust. Soc. Am.* **106**(3), 1579–1585.
- Mammen, D. L., and Nowicki, S. (1981). "Individual differences and within flock convergence in chickadee calls," *Behav. Ecol. Sociobiol.* **9**(3), 179–186.
- McIlraith, A. L., and Card, H. C. (1997). "Birdsong recognition using back-propagation and multivariate statistics," *IEEE Trans. Signal Process.* **45**(11), 2740–2748.
- Murray, S. O., Mercado, E., and Roitblat, H. L. (1998). "The neural network classification of false killer whale (*Pseudorca crassidens*) vocalizations," *J. Acoust. Soc. Am.* **104**(6), 3626–3633.
- Nowicki, S. (1983). "Flock specific recognition of chickadee calls," *Behav. Ecol. Sociobiol.* **12**(4), 317–320.
- Nowicki, S., and Nelson, D. A. (1990). "Defining natural categories in acoustic signals: Comparison of 3 methods applied to chick-a-dee call notes," *Ethnology* **86**(2), 89–101.
- Pao, Y.-H. (1989). *Adaptive Pattern Recognition And Neural Networks* (Addison-Wesley, Reading, MA).
- Parson, S., and Jones, G. (2000). "Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks," *J. Exp. Biol.* **203**, 2641–2656.
- Pearce, J. M. (1987). "A model for stimulus generalization in Pavlovian conditioning," *Psychol. Rev.* **94**(1), 61–73.
- Pearce, J. M. (1994). "Similarity and discrimination: A selective review and a connectionist model," *Psychol. Rev.* **101**, 587–607.
- Pearce, J. M. (1997). *Animal Learning And Cognition: An Introduction* (Psychology Press, East Sussex).
- Pearce, J. M. (2002). "Evaluation and development of a connectionist theory of configural learning," *Anim. Learn. Behav.* **30**(2), 73–95.
- Pearce, J. M., and Bouton, M. E. (2001). "Theories of associative learning in animals," *Annu. Rev. Psychol.* **52**, 111–139.
- Pylyshyn, Z. W. (1984). *Computation And Cognition* (MIT Press, Cambridge, MA).
- Quinlan, P. (1991). *Connectionism And Psychology* (University of Chicago Press, Chicago, IL).
- Rescorla, R. A. (1973). "Evidence for the unique stimulus account of configural conditioning," *J. Comp. Physiol. Psychol.* **85**, 331–338.
- Rescorla, R. A. (1988). "Pavlovian conditioning: Its not what you think it is," *Am. J. Psychol.* **43**(3), 151–160.
- Rescorla, R. A. (2003). "Contemporary study of Pavlovian conditioning," *Span. J. Psychol.* **6**, 185–195.
- Rescorla, R. A., and Wagner, A. R. (1972). "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement," *Classical Conditioning II: Current Research And Theory*, edited by A. H. Black and W. F. Prokasy (Appleton-Century-Crofts, New York), pp. 64–99.
- Ripley, B. D. (1996). *Pattern Recognition And Neural Networks* (Cambridge University Press, Cambridge, U.K.).
- Rosenblatt, F. (1962). *Principles Of Neurodynamics* (Spartan, Washington).
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). "Learning representations by back-propagating errors," *Nature (London)* **323**, 533–536.
- Schon, P. C., Puppe, B., and Manteuffel, G. (2001). "Linear prediction coding analysis and self-organizing feature map as tools to classify stress calls of domestic pigs (*Sus scrofa*)," *J. Acoust. Soc. Am.* **110**(3), 1425–1431.
- Wagner, A. R. (2003). "Context-sensitive elemental theory," *Q. J. Exp. Psychol. B* **56**(1), 7–29.
- Wasserman, E. A., and Miller, R. R. (1997). "What's elementary about associative learning?," *Annu. Rev. Psychol.* **48**, 573–607.