Artificial Neural Networks that Classify Musical Chords

Vanessa Yaremchuk, University of Alberta, Canada

Michael R.W. Dawson, University of Alberta, Canada

ABSTRACT

An artificial neural network was trained to classify musical chords into four categories—major, dominant seventh, minor, or diminished seventh—independent of musical key. After training, the internal structure of the network was analyzed in order to determine the representations that the network was using to classify chords. It was found that the first layer of connection weights in the network converted the local representations of input notes into distributed representations that could be described in musical terms as circles of major thirds and on circles of major seconds. Hidden units then were able to use this representation to organize stimuli geometrically into a simple space that was easily partitioned by output units to classify the stimuli. This illustrates one potential contribution of artificial neural networks to cognitive informatics: the discovery of novel forms of representation in systems that can accomplish intelligent tasks.

Keywords: artificial neural networks; chord classification; representation

INTRODUCTION

Cognitive informatics is a field of research that is primarily concerned with the information processing of intelligent agents (Wang, 2003). One way to characterize cognitive informatics is in terms of an evolving notion of information (Wang, 2007). When it originated six decades ago, conventional accounts of information were concerned about using probability theory and statistics to measure the amount of information carried by an external signal. This in turn developed into the notion of modern informatics which studied information as "properties or attributes of the natural world that can be generally abstracted, quantitatively represented, and mentally processed" (Wang, 2007, p. iii). The current incarnation of cognitive informatics recognized that both information theory and modern informatics defined information in terms of factors that were external to brains. Cognitive informatics has replaced this with an emphasis on exploring information as an internal property.

This emphasis on the internal processing of information raises fundamental questions about how such information can be represented. One approach to answering such questions—and for proposing new representational accounts—

would be to train a brain-like system to perform an intelligent task, and then to analyze its internal structure to determine the types of representations that the system had developed to perform this intelligent behavior. The logic behind this approach is that when artificial neural networks are trained to solve problems, there are few constraints placed upon the kinds of internal representations that they can develop. As a result, it is possible for a network to discover new forms of representation that were surprising to the researcher (Dawson & Boechler, 2007; Dawson & Zimmerman, 2003).

Cognitive informatics has been applied to a wide variety of domains, ranging from organization of work in groups of individuals (Wang, 2007) to determining the capacity of human memory (Wang, Liu & Wang, 2003) to modeling neural function (Wang, Wang, Patel & Patel, 2006). The research below provides an example of this approach in a new domain, musical cognition. There is a growing interest in the cognitive science of musical cognition, ranging from neural accounts of musical processing (Jourdain, 1997; Peretz & Zatorre, 2003) through empirical accounts of the perceptual regularities of music (Deutsch, 1999; Krumhansl, 1990) to computational accounts of the formal properties of music (Assayag, Feichtinger, & Rodrigues, 2002; Lerdahl & Jackendoff, 1983). Because music is characterized by both many formal and many informal properties, there has been an explosion of interest in using artificial neural networks to study it (Griffith & Todd, 1999; Todd & Loy, 1991). The simulation below illustrates one intriguing possibility for such research: the discovery of previously unknown representations of formal musical structures. As such, it illustrates that artificial neural networks can be used as a medium to explore a "synthetic approach" to psychology and make important representational contributions to cognitive informatics and cognitive science (Dawson, 1998, 2004).

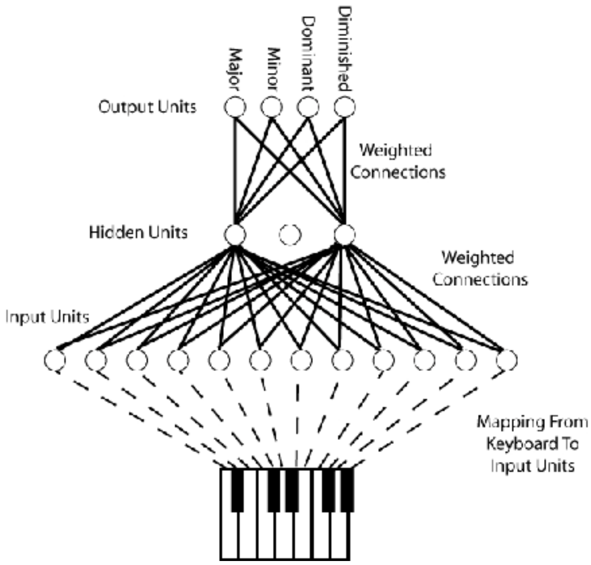
Chord Classificiation By Neural Networks

In a pioneering study, artificial neural networks (ANNs) were trained to classify musical stimuli as being major chords, minor chords, or diminished chords (Laden & Keefe, 1989). Laden and Keefe's networks used 12 input units, where each unit represented a particular note or "piano key" in an octave range—a so-called pitch class representation (see Figure 1). They created a training set consisting of 36 different chords: the major triad for each of the 12 different major key signatures, as well as the minor triad and a diminished seventh triad for each of the 12 different minor key signatures. They examined a number of different networks by manipulating the number of hidden units (three, six, eight, or nine), and by manipulating the pattern of network connectivity.

In general, Laden and Keefe (1989) found that the performance of their simple networks was disappointing. Their most successful simple network used three hidden units, and had direct connections between input and output units, but was still able to correctly classify only 72 percent of the presented chords. Other small networks had accuracy rates as low as 25 percent. Laden and Keefe improved network accuracy to a maximum level of 94 percent by using a more complex network that had 25 hidden units and which used output units to represent distances between musical notes (i.e., musical intervals) rather than chord types.

The current study is an extension of Laden and Keefe's (1989) research. It examines a simple network that uses a pitch class representation and three hidden units. Laden and Keefe used a traditional sigmoid-shaped activation function (the logistic equation) in the processing units of their networks. We instead used a Gaussian activation function because previous research has demonstrated that networks that employ this activation function are adept at solving complex problems, and also lend themselves to detailed interpretation (Berkeley, Dawson, Medler, Schopflocher, & Hornsby, 1995; Dawson, 2004, 2005). Our working hypothesis was that this change in network architecture would

Figure 1. An example of an artificial neural network. Each input unit corresponds to a note on a 12-note keyboard. Signals from these input units are passed along for processing by three different hidden units through sets of weighted connections. For simplicity's sake, only the connections involving two of the hidden units are depicted. Activations from the hidden units are then passed through a second layer of connections to produce activity in a set of four output units. After training, an input stimulus will only turn on one of the output units, which will classify the stimulus as being a particular type of chord. This particular network was the one used in our simulation study.



permit simple networks to classify chord types, and would also permit the internal structure of such networks to be interpreted in the search for new musical representations.

METHOD

Training Set

Networks were trained to identify four different types of musical chords: major, minor, dominant seventh, and diminished seventh. The training set was constructed as follows: First, a root note was selected (e.g., C). Second, the major chord based on this root note was created by activating the three input units that defined the component notes of this chord (e.g., C, E, and G). Third, the minor chord based on this root note was created by activating the three input units that defined its component notes (e.g., C, E♭, and G). Fourth, the dominant seventh chord based on this root was defined by activating the four input units that represented its component notes (e.g., C, E, G, and B♭). Fifth, the diminished seventh chord based on this root was defined by activating the four input units that represented its components (e.g., C, D#, F#, A). This process was repeated until four chords had been constructed for each of the 12 possible root notes in a dodecaphonic note system, resulting in a training set of 48 different chords.

Network Architecture

The network had four output units, three hidden units, and 12 input units. Each of the output units represented one of the four types of musical chords, and each of the input units represented a particular musical note in a pitch class representation, as was illustrated in Figure 1. Three hidden units were used because pilot simulations had indicated that this was the smallest number of hidden units that would permit a network to correctly classify the training stimuli. All of the output units and all of the hidden units were value units that used the Gaussian activation function described by Dawson and Schopflocher (1992): $G(net_i) = \exp(-\pi(net_i - \mu_i)^2)$. In this equation, $G(net_i)$ is the activation being calculated for unit i, net_i is the net input for that unit, and μ_i is the Gaussian mean. When the net input to the equation is equal to the mean (i.e., equal to μ_i), the activity that is generated is equal to 1.0. As net input moves away from the mean in either direction, unit activity quickly drops off to near-zero levels.

Network Training

The network was trained to classify chords by turning the appropriate output unit "on," and the other three output units "off," for each stimulus in the training set. Training was conducted using a variation of the generalized delta rule for value units (Dawson, 2004, 2005; Dawson & Schopflocher, 1992). The software used to perform this training is available as freeware from http://www.bcp.psych.ualberta.ca/~mike/Software/Rumelhart/index.html.

Prior to training, all of the connection weights were randomly assigned values ranging from –0.10 to +0.10. The biases of processing units (i.e., the μs of the Gaussian activation functions) were all initially assigned a value of 0.00. The network was trained with a learning rate of 0.005 and zero momentum. During a single epoch of training each of the 48 chords was presented to the network in random order. Connection weights were updated after each stimulus presentation.

Training proceeded until the network generated a "hit" for every output unit on every pattern. A hit was operationalized as an activation of 0.90 or higher when the desired activation was 1.00, and as an activation of 0.10 or lower when the desired activation was 0.00. The network converged on a solution to the problem—generating a correct response for each of the 48 chords—after 3,964 epochs of training.

RESULTS

One of the potential contributions of artificial neural networks to the study of music is the ability of ANNs to reveal novel or surprising regularities in musical stimuli (Bharucha, 1999). Indeed, we believe that revealing novel representations of information is one of the primary contributions that ANNs can make to cognitive informatics (Dawson, 2004; Dawson & Boechler, 2007; Dawson, Medler, & Berkeley, 1997; Dawson, Medler, McCaughan, Willson, & Carbonaro, 2000; Dawson & Piercey, 2001). In order for such a contribution to be realized, the internal structure of a trained network must be interpreted. The sections below present our interpretation of the current network, and show that it developed a simple, elegant—and surprising—representation of the relationships between musical notes that could in turn be used to classify the chord types.

Interpretation of Weights from Input Units

The first step in interpreting the network was to examine the connection weights from the input units to the hidden units. Because each input unit was associated with a particular note, each of these connection weights could be viewed as a numerical "note name." An inspection of the first layer of connection weights (see Table 1) revealed that the network had converted the pitch class representation of the input units into a smaller set of equivalence classes that assigned the same connection weight or "name" to more than one note. For example, the notes A, C#, and F are all represented by different input units in the pitch class representation. However, the network assigned the same numerical "note name" (i.e., a weight of –1.27) to each of the connections between these input units and Hidden Unit 1. As a result, each of these three notes was treated as being the same by this hidden unit.

An examination of Table 1 reveals that both Hidden Units 1 and 3 convert the pitch

Table 1. Correspondence between input notes and connection weight values for three hidden units in the chord classification network. As noted in the text, the first four rows of the table assign notes to four different circles of major thirds, while the last two rows of the table assign notes to two different circles of major seconds.

Note Name	Hidden Unit 1	Hidden Unit 3	Hidden Unit 2
A, C#, F	-1.27	-0.28	
A#, D, F#	-0.61	-0.10	
B, D#, G	1.28	0.28	
C, E, G#	0.63	0.10	
A, C#, F, B, D#, G			-0.76
A#, D, F#, C, E, G#			-0.68

class representation of 12 different notes into 4 equivalence classes that each contains 3 notes. Each of these equivalence classes can be described in formal musical terms as a "circle of major thirds." That is, each of the notes in one of these classes differs from the other two by a musical interval of a major third, or four semitones. For instance, if one moves a major third up from A, then one reaches the note C#. If one then moves another major third up from C#, then one reaches the note F. If one finally moves yet another major third up from F, then one completes the circle and returns to the note A. The first four rows of Table 1 identify four different groups of three notes that can each be described as a circle of major thirds.

Hidden Unit 2 also employs connection weights from input units that group notes into equivalence classes, but not according to circles of major thirds. Instead, notes are classified as belonging to one of two groups that correspond to circles of major seconds. In this representation, each note in the group is exactly two semitones apart from the next on a scale. The connection weights that define these two equivalence classes are also presented in Table 1; the final two rows of this table identify two groups of six notes that can each be described as a circle of major seconds.

Using Hidden Unit Responses to Classify Chords

How does the network use circles of thirds and seconds representations to classify chords? When we examined hidden unit responses, we found that each hidden unit responds to some stimuli, but not to others. For instance, Hidden Unit 1 generates a strong response (i.e., an activity level of 1.00) to all of the diminished seventh tetrachords, as well as to half of the minor triads (i.e., those associated with the minor keys of a, b, c#, e♭, f, and g). It generates a very weak response to any other type of stimulus. The selective responding of the hidden units is directly related to the circles of thirds and seconds.

For example, any diminished seventh tetrachord is defined by four notes. With respect to the four circles of major thirds that were presented in Table 1, each of these notes comes from a different circle of major thirds. That is, one note from each of the first four rows in Table 1 is required to define any diminished seventh chord. The connection weights that encode these classes for Hidden Unit 1 are such that when each class is represented in a stimulus, the resulting signal to the hidden unit causes it to generate a maximum response.

As well, every minor triad can be defined as having two notes belonging to the same circle of major thirds, and the third note belonging to one of the other circles of major thirds. Furthermore, the two circles of major thirds that are required to define the triads for the minor keys of a, b, c#, e♭, f, and g minor are associated with connection weights that also result in a signal being sent to Hidden Unit 1 that produces near maximum activity in it. However, the two circles of major thirds that are required to define the remaining minor triads (for the minor keys of b♭, c, d, e, f#, and a♭) are associated with weights that send a signal to Hidden Unit 1 that does not cause it to generate a strong response.

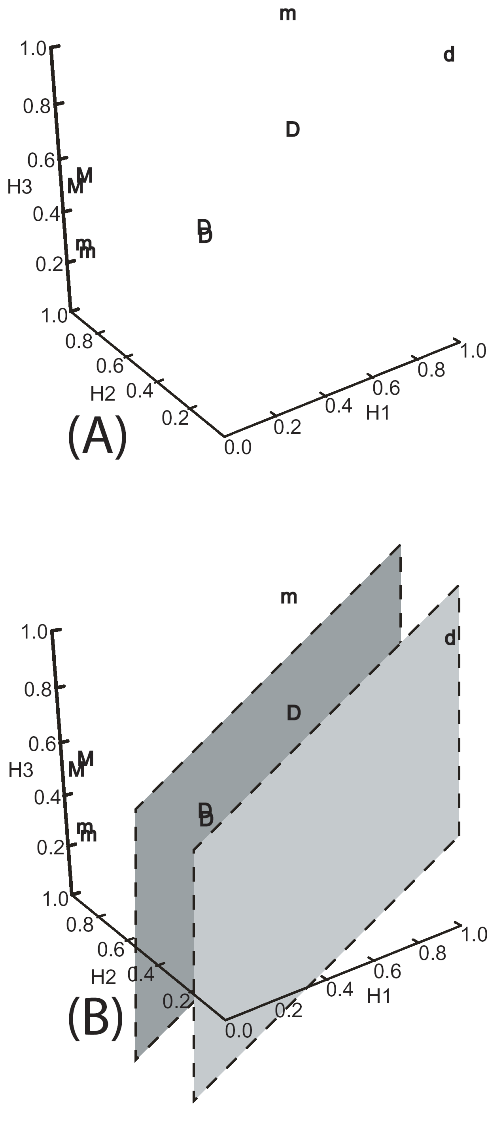
The fact that hidden units a) selectively respond to stimuli, but b) respond to more than one type of stimulus (e.g., diminished seventh tetrachords and minor triads) indicates that the hidden units are representing stimuli using a coarse code. In general, coarse coding means that an individual processor is sensitive to a variety of features or feature values, and is not tuned to detect a single feature type (e.g., Churchland & Sejnowski, 1992). As a result, individual processors are not particularly accurate feature detectors. However, if different processors are sensitive to different feature varieties, then their outputs can be pooled, which often produces an accurate representation of a specific feature.

If hidden units are coarse coding musical chords, then the network's response to each stimulus requires considering the activity that it produces in all three hidden units at the same time. To explore this possibility, one can graph a *hidden unit space* of the patterns in the training set. In such a graph, each pattern is represented as a point in space, where the coordinates of the point are provided by the activity produced by the pattern in each hidden unit. Figure 2A presents the three-dimensional hidden unit space of the music chord patterns for the current network.

Figure 2A indicates that the hidden unit space provides a very simple representation of the 48 patterns in the training set. In particular, many different chords are mapped into the identical location in this graph. For example, all 12 diminished seventh tetrachords are located at the same position in Figure 2A, the location indicated by the symbol d.

How is such a hidden unit space used to classify patterns? An ANN can be described as a tool that "carves" such a space into distinct decision regions (Lippmann, 1989). All of the stimuli (i.e., all of the points in a space) that fall into one decision region are classified as belonging to the same type, and lead the network to generate one kind of classification response. In order to solve a problem, the network must learn how to partition the hidden unit space in

Figure 2. (A) The hidden unit space for the network. Major chords are represented by M, minor chords are represented by m, dominant seventh chords are represented by D, and diminished seventh chords are represented by d. (B) An example of how an output value unit could partition the space of Figure 2A in such a way to separate the dominant seventh chords from all others. The other output units classify chords by adopting a similar partitioning of the hidden unit space.



such a way that it makes the correct response for every stimulus.

Recall that the output units in the current network were all value units that used the Gaussian activation function. Value units partition spaces by making two parallel cuts to separate some patterns from others (e.g., Dawson, 2004; see also Figure 2B). This type of partitioning could easily be used to separate the hidden unit space of Figure 2A into four different decision regions that could be used to correctly classify all 48 of the stimuli. For example, Figure 2B illustrates the orientation of two straight cuts that could be used by the output unit to separate dominant seventh chords from the other three chord types. The other three output units could make similar cuts to separate their chord type from all of the others.

Discussion

In their original study of chord classification, Laden and Keefe (1989) were unable to train a network to correctly classify 100 percent of their training set of 36 major, minor, and diminished seventh triads. One important result of our simulation was that our network was able to classify all of the chords in our training set, including 12 dominant seventh chords that were not studied by Laden and Keefe. This result was achieved even though our network was simpler than the best performing network that Laden and Keefe reported.

There are several likely explanations for this result. As was noted earlier, while our simulation was conducted in the spirit of Laden and Keefe's (1989) research, there were a number of differences between our simulation and theirs. First, we used value units in our networks, instead of processors that use sigmoid-shaped activation functions. Because value units carve up pattern spaces in a different fashion than do these latter units, they are likely more suited to the chord classification problem. Second, we defined diminished seventh and dominant seventh stimuli as tetrachords instead of triads. As was shown in the analysis of the behavior of both Hidden Units 1 and 3, the network was able to take advantage of the "balancing" of four incoming signals to respond to stimuli (e.g., to generate high responses to diminished seventh tetrachords). In short, moderate changes to the network architecture and to stimulus representation produced conditions that made the chord classification problem easier for our network than was the case in Laden and Keefe's earlier simulations.

A more important result concerns the interpretation of network structure. One of connectionism's potential contributions to the psychology of music is its ability to reveal novel regularities in stimulus structure, or to suggest new approaches to represent musical patterns. In order for this potential to be realized, it must be possible to analyze the internal structure of a network after it has been trained. Our study demonstrated that the internal structure of the chord classification network could be interpreted. It revealed that the network classified chord structure first by representing individual notes in terms of circles of major thirds and major seconds, and then by combining these representations to position chords in a three-dimensional hidden unit space. To our knowledge, the only previous occurrence of this kind of representation was the distributed coding scheme that was used by Franklin (2004) to represent inputs to a network. While there is a growing body of evidence concerning specialized neural processing of tones and chords (e.g., Peretz & Zatorre, 2005), this evidence is not yet sufficiently precise to indicate whether distributed representations based on tone circles are used by the brain. We know of no example in the literature of a study that has shown an ANN reorganizing an input encoding scheme into this type of representation. This raises the question of whether circles of thirds and seconds are pertinent to human subjects' representation of musical stimuli, an issue that we are currently exploring.

ACKNOWLEGMENT

This research was supported by NSERC and SSHRC grants awarded to MRWD. The authors would like to thank Margaret Tompkins for her consultation on this project. Correspondence

should be addressed to Dr. Michael Dawson, Department of Psychology, University of Alberta, Edmonton, Alberta, CANADA T6G 2E9

REFERENCES

Assayag, G., Feichtinger, H. G., & Rodrigues, J.-F. (2002). *Mathematics and music: A Diderot mathematical forum*. Berlin/New York: Springer.

Berkeley, I. S. N., Dawson, M. R. W., Medler, D. A., Schopflocher, D. P., & Hornsby, L. (1995). Density plots of hidden value unit activations reveal interpretable bands. *Connection Science, 7*, 167-186.

Bharucha, J. J. (1999). Neural nets, temporal composites, and tonality. In D. Deutsch (Ed.), *The Psychology Of music,* 2nd ed. San Diego, CA: Academic Press, pp. 413-440.

Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, MA: MIT Press.

Dawson, M. R. W. (1998). *Understanding cognitive science*. Malden, MA: Blackwell.

Dawson, M. R. W. (2004). *Minds and machines: Connectionism and Psychological modeling*. Malden, MA: Blackwell.

Dawson, M. R. W. (2005). *Connectionism: A hands-on approach* (1st ed.). Oxford, UK/Malden, MA: Blackwell.

Dawson, M. R. W., & Boechler, P. M. (2007). Representing an intrinsically nonmetric space of compass directions in an artificial neural network. *International Journal of Cognitive Informatics and Natural Intelligence, 1*, 53-65.

Dawson, M. R. W., Medler, D. A., & Berkeley, I. S. N. (1997). PDP networks can provide models that are not mere implementations of classical theories. *Philosophical Psychology, 10*, 25-40.

Dawson, M. R. W., Medler, D. A., McCaughan, D. B., Willson, L., & Carbonaro, M. (2000). Using extra output learning to insert a symbolic theory into a connectionist network. *Minds and Machines, 10*, 171-201.

Dawson, M. R. W., & Piercey, C. D. (2001). On the subsymbolic nature of a PDP architecture that uses a nonmonotonic activation function. *Minds and Machines, 11*, 197-218.

Dawson, M. R. W., & Schopflocher, D. P. (1992). Modifying the generalized delta rule to train networks of nonmonotonic processors for pattern classification. *Connection Science, 4*, 19-31.

Dawson, M. R. W., & Zimmerman, C. (2003). Interpreting the internal structure of a connectionist model of the balance scale task. *Brain & Mind, 4*, 129-149.

Deutsch, D. (1999). *The Psychology of music* (2nd ed.). San Diego: Academic Press.

Franklin, J. A. (2004). *Recurrent neural networks and pitch representations for music tasks*. Paper presented at the Seventeenth International Florida Artificial Intelligence Research Symposium Conference, Miami Beach, FA.

Griffith, N., & Todd, P. M. (1999). *Musical networks: Parallel distributed perception and performace*. Cambridge, MA: MIT Press.

Jourdain, R. (1997). *Music, the brain, and ecstasy*. New York: William Morrow & Co.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Laden, B., & Keefe, B. H. (1989). The representation of pitch in a neural net model of pitch classification. *Computer Music Journal, 13*, 12-26.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Lippmann, R. P. (1989). Pattern classification using neural networks. *IEEE Communications magazine, November*, 47-64.

Peretz, I., & Zatorre, R. J. (2003). *The cognitive neuroscience of music*. Oxford/New York: Oxford University Press.

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology, 56, 1-26*.

Todd, P. M., & Loy, D. G. (1991). *Music and connectionism*. Cambridge, MA: MIT Press.

Wang, Y. (2003). Cognitive informatics: A new transdisciplinary research field. *Brain & Mind, 4*, 115-127.

Wang, Y. (2007a). Cognitive informatics: Exploring the theoretical foundations for natural intelligence, neural informatics, autonomic computing, and agent systems. *International Journal of Cognitive Informatics and Natural Intelligence, 1*, i-x.

Wang, Y. (2007b). On laws of work organization in human cooperation. *International Journal of Cognitive Informatics and Natural Intelligence, 1*(2), 1-15.

Wang, Y., D. Liu, & Y. Wang (2003). Discovering the capacity of human memory. *Brain and Mind*, *4*(2), 189-198.

Wang, Y., Wang, Y., Patel, S., & Patel, D. (2006). A layered reference model of the brain. *IEEE Transactions on Systems, Man, and Cybernetics (C), 36*(2), 124-133.

Vanessa Yaremchuk is a PhD student in psychology at the University of Alberta.

Michael R.W. Dawson received his PhD in psychology from the University of Western Ontario in 1986, and is currently a full professor in the psychology department at the University of Alberta. His research interests include pure and applied research on artificial neural networks and the relationship of this research to empirical and theoretical issues in cognitive science.