

Predicting shifts in generalization gradients with perceptrons

Matthew G. Wisniewski · Milen L. Radell ·
Lauren M. Guillette · Christopher B. Sturdy ·
Eduardo Mercado III

Published online: 8 October 2011
© Psychonomic Society, Inc. 2011

Abstract Perceptron models have been used extensively to model perceptual learning and the effects of discrimination training on generalization, as well as to explore natural classification mechanisms. Here, we assess the ability of existing models to account for the time course of generalization shifts that occur when individuals learn to distinguish sounds. A set of simulations demonstrates that commonly used single-layer and multilayer perceptron networks do not predict transitory shifts in generalization over the course of training but that such dynamics can be accounted for when the output functions of these networks are modified to mimic the properties of cortical tuning curves. The simulations further suggest that prudent selection of stimuli and training criteria can allow for more precise predictions of learning-related shifts in generalization gradients in behavioral experiments. In particular, the simulations predict that individuals will show maximal peak shift after different numbers of trials, that easier contrasts will lead to slower development of shifted peaks, and that whether generalization shifts persist or dissipate will depend on which stimulus dimensions individuals use to distinguish stimuli and how those dimensions are neurally encoded.

M. G. Wisniewski (✉) · M. L. Radell · E. Mercado III
Department of Psychology, Park Hall, University at Buffalo,
The State University of New York,
Buffalo, NY 14260, USA
e-mail: mgw@buffalo.edu

L. M. Guillette
Department of Psychology, University of Alberta,
Edmonton, Alberta, Canada

C. B. Sturdy
Department of Psychology and Centre for Neuroscience,
University of Alberta,
Edmonton, Alberta, Canada

Keywords Peak shift · Discrimination learning ·
Representation · Perceptual learning · Neural network ·
Similarity

When an organism learns that a stimulus results in some consequence, it will often generalize that learning to similar novel stimuli (Shepard, 1987). For instance, Watson and Raynor (1920) famously demonstrated, in experiments with Little Albert, that fear associated with a white rat can generalize to other stimuli, such as a rabbit, a fur coat, or even a piece of cotton. Numerous theoretical efforts have focused on explaining and predicting generalization patterns, with varying degrees of success (for a review, see Ghirlanda & Enquist, 2003; Thomas, 1993). Computational models of discrimination learning, in particular, have proven to be adept at simulating many of the empirically observed generalization patterns (Ghirlanda & Enquist, 1998, 2006; Guillette et al., 2010; Livesey, Pearson, & McLaren, 2005; Saksida, 1999; Shepard, 1990; Staddon & Reid, 1990; Suret & McLaren, 2002; Thomas, 1993). Such models are becoming increasingly useful tools for generating hypotheses about the mechanisms and cues that participants use during learning and generalization. In the present study, we assessed the ability of a neural network model of discrimination learning and generalization to account for recent observations of learning-related shifts in generalization observed during an auditory learning task.

A phenomenon commonly observed after discrimination learning is that the highest levels of responding may occur for stimuli other than those experienced during training. Generally, a *peak shift* results when an individual is trained to respond to one stimulus (S+) and not to some other stimulus (S-) that varies along a common dimension. When generalization is measured after training, responding is strongest not to the trained S+, but to a stimulus that is

shifted along the dimension even further from S⁻. In a classic example of this, Hanson (1959) trained pigeons to peck a key when presented with a 560-nm light (S⁺), but not when presented with a 570-nm light (S⁻). During generalization tests, pigeons responded most strongly to wavelengths other than 560 nm (such as 540 nm) that were farther along the continuum from the trained S⁻, 570 nm light. A similar phenomenon (*area shift*) is seen when a generalization gradient shows a peak at the S⁺ but an asymmetry in responding to stimuli on either side of the S⁺, such that stimuli on the side away from S⁻ are responded to more frequently. Such learning-related shifts have been well documented experimentally (Baron, 1973; Bizo & McMahan, 2007; Cheng, Spetch, & Johnston, 1997; Derenne, 2010; Galizio & Baron, 1979; Guillette et al., 2010; Hanson, 1959; Landau, 1968; Lewis & Johnston, 1999; Moye & Thomas, 1982; Newlin, Rodgers, & Thomas, 1979; Nicholson & Gray, 1972; Purtle, 1973; Rowe & Skelhorn, 2004; Spetch, Cheng, & Clifford, 2004; Terrace, 1966; Thomas, Mood, Morrison, & Wiertelak, 1991; Verzijden, Etman, van Heijningen, van der Linden, & ten Cate, 2007; Wills & Mackintosh, 1998; Wisniewski, Church, & Mercado, 2009, 2010) and have been replicated using artificial neural networks (Ghirlanda & Enquist, 1998, 2006; Guillette et al., 2010; Livesey et al., 2005; Saksida, 1999; Suret & McLaren, 2002).

Most of the experimental studies of learning-related shifts have focused on the presence, size, or generality of the effects. For instance, several studies have shown that how far a peak in generalization shifts after training depends on the similarity of the stimuli that the subject initially learned to discriminate (Baron, 1973; Ghirlanda & Enquist, 1998; Hanson, 1959; Purtle, 1973; Thomas et al., 1991). Typically, the more similar the stimuli being discriminated are, the further the learner will shift (Baron, 1973; Ghirlanda & Enquist, 1998; Hanson, 1959; Spence, 1937). Shifts occur not only along simple, single dimensions like wavelength (Hanson, 1959) and sine wave frequency (Baron, 1973), but also along more complex acoustic (Guillette et al., 2010; Verzijden et al., 2007; Wisniewski et al., 2009, 2010), and visual (Derenne, 2010; Livesey et al., 2005; Spetch et al., 2004; Suret & McLaren, 2002; Wills & Mackintosh, 1998) dimensions. The ubiquitous nature of learning-related shifts have made it possible to use the effects to identify natural dimensions along which stimuli vary (Guillette et al., 2010; Rowe & Skelhorn, 2004). For instance, Guillette et al. used the peak shift effect as a way to identify a continuum along which chickadees discriminated notes that make up their species' call and were also able to use similar generalization shifts in an artificial neural network to identify the acoustic features that defined that continuum.

Most theories of discrimination learning's impact on generalization have focused on explaining the basic experi-

mental effects. For example, several different associative learning theories (Blough, 1975; McLaren & Mackintosh, 2002; Saksida, 1999; Spence, 1937) adequately explain the direction of shift and the changes to the size of the effect that result from variations in stimulus similarity. Computational models have been used to test how well conflicting associative theories explain peak shift and related generalization phenomena under different conditions (Ghirlanda & Enquist, 1998; Livesey et al., 2005; Saksida, 1999), to generate new hypotheses about how stimuli may be represented (Ghirlanda & Enquist, 1998; Livesey et al., 2005; Suret & McLaren, 2002; Tijsseling & Gluck, 2002), and to determine what aspects of stimuli may be given the most associative weight (Guillette et al., 2010).

Although there has been much work exploring the basic predictions of current theories regarding peak shift, relatively few studies, experimental or theoretical, have looked at the dynamics of generalization over time. A few studies have shown that extinction can occur during testing, such that shifts in generalization dissipate as more nondifferentially reinforced testing trials are experienced (Cheng et al., 1997; Purtle, 1973). For instance, pigeons that exhibit a strong peak shift in the first block of testing show a reduction in the strength of that shift in the following test blocks (Cheng et al., 1997). Other studies have shown that both in humans (Wisniewski et al., 2009) and in nonhumans (Moye & Thomas, 1982), peak shift is stable over time and lasts at least 24 h post-discrimination-training. Recently, in an effort to more fully characterize how learning-related shifts change over the course of experience, Wisniewski et al. (2010) trained humans for 60, 100, 140, 180, 220, or 260 trials on a task requiring the discrimination of two complex sounds that varied in their rate of periodic frequency modulation. Participants who were trained with the fewest trials (60 or 100) or the most trials (220 or 260) did not show a peak shift effect. However, participants who were trained with an intermediate number of trials (140 or 180) did exhibit shifts. These results suggest that in at least some training conditions, peak shift may occur only at intermediate levels of learning.

The dynamics of generalization over the course of learning have been explored in past computer simulations of discrimination learning. For example, the shape of generalization gradients produced by connectionist networks change from being Gaussian to exponential as more training iterations are experienced (Shepard, 1990; Staddon & Reid, 1990). However, simulations investigating how learning-induced generalization shifts vary over different amounts of training are lacking. One study found that a multilayer connectionist model trained to recognize faces along a continuum showed a gradually emerging shift in generalization such that exaggerations of trained faces were recognized more easily than the original faces (Tanaka &

Simon, 1996). With extensive training, however, this advantage disappeared. Nevertheless, the network continued to respond to novel faces falling along the continuum of trained facial features just as strongly as it did to the originally-trained faces. The basic trajectory of generalization changes in this network included the emergence of a peak shift effect at intermediate levels of training that later transformed into an area shift after extensive training.

The purpose of the present study was to assess whether a simple connectionist model could replicate the gradual emergence and dissipation of a peak shift effect that occurs over the course of auditory discrimination training in humans (Wisniewski et al., 2010). Toward this goal, variants of a previously developed connectionist model (Dawson, 2004, 2005, 2008; Dawson & Schopflocher, 1992) used to study auditory perception in chickadees and known to exhibit peak shift effects (Guillette et al., 2010) were used to simulate the learning of complex sounds by humans. The hypothesis was that the model would show transitory peak-shifts in generalization. Additionally, because there can be large individual differences in stimulus-evoked neural representations (Miglioretti & Boatman, 2003; Orduña, Mercado, Gluck, & Merzenich, 2005), and because different stimulus representations can yield drastically different generalization patterns (Enquist & Ghirlanda, 2005; Ghirlanda, 2002; Ghirlanda & Enquist, 1998), we tested how altering the representations used to train the model would affect the dynamics of generalization. Because many neural network studies have focused on replicating the effect of stimulus similarity on the strength of shift without examining how that strength varies with the amount of learning, we also tested how stimulus similarity would affect the dynamics of shifts in networks. Finally, because past studies of generalization shifts have often involved different amounts of training (Baron, 1973; Bizo & McMahan, 2007; Derenne, 2010; Galizio & Baron, 1979; Lewis & Johnston, 1999; Newlin et al., 1979; Thomas et al., 1991; Wisniewski et al., 2009, 2010), and because there can be large individual differences in performance gains during training and generalization (Nicholson & Grey, 1972; Withagen & van Wermeskerken, 2009), we assessed how different criteria for ending training affected the variability of generalization gradients.

Simulation 1

In simulation 1, we attempted to produce a transient peak shift during the learning of a discrimination task, using variants of a neural network known as a *perceptron*. The perceptron is a simple connectionist model developed in the 1950s that transforms a set of inputs to generate a desired output (Rosenblatt, 1957). In *single-layer perceptrons*, several input

units are connected to a single output unit (Fig. 1a). Each input unit connection is associated with a numerical value (called a *weight*) that determines the contribution of that input to the final output. Multilayer perceptrons are essentially a series of sequentially applied single-layer perceptrons (Fig. 1b), in which each layer typically contains a different number of processing units. The computations performed by output units, which are the same for both single- and multilayer perceptrons, are described in the [Appendix](#).

Associative learning theories often characterize discrimination learning as a process in which links between specific inputs (or elements of inputs) and particular outputs or associated stimuli become stronger or weaker during learning (e.g., Blough, 1975; McLaren & Mackintosh, 2002; Saksida, 1999; Spence, 1937). When properties of an S+ overlap with those of the S–, these theories predict that features of the S+ that are not shared with the S– gain associative weight as training progresses. When such theories have been instantiated as connectionist models (Ghirlanda & Enquist, 1998, 2006; Livesey et al., 2005; Suret & McLaren, 2002), they have proven to be able to explain the direction of shifts in generalization gradients, as well as the finding that learning to discriminate more similar stimuli can lead to larger shifts in generalization (Ghirlanda & Enquist, 1998). Because few experiments have attempted to track changes in generalization shifts over time, past modeling studies have seldom reported simulations of such changes.

As was noted earlier, a recent study examining the development of generalization shifts over time revealed that such learning-induced shifts in generalization can be transient (Wisniewski et al., 2010). In Wisniewski et al. (2010), participants learned to identify one sound by pressing a particular key on a keyboard to indicate that sound and to press another key to indicate that a different sound had occurred. Unlike most past studies with animals, this task reinforces two behavioral responses equally, rather than reinforcing responses to only one of the stimuli (i.e., the “S–” was not a nonreinforced stimulus). Nevertheless, a peak shift effect was observed after training, such that sounds further from the nontarget provoked stronger responses in generalization tests (Fig. 2). The finding that shifts in generalization were observed only after intermediate amounts of training had not been previously described in the literature, and did not appear to be predicted by existing models of discrimination learning. The current simulation tested whether an existing computational model, which was known to show a peak shift effect, would also show a dissipation of this effect with extensive training.

We attempted to simulate this effect with a basic perceptron because this neural network has previously been shown to produce a peak shift effect in simulations of auditory discrimination by chickadees (Guillette et al., 2010) and because perceptron learning algorithms are

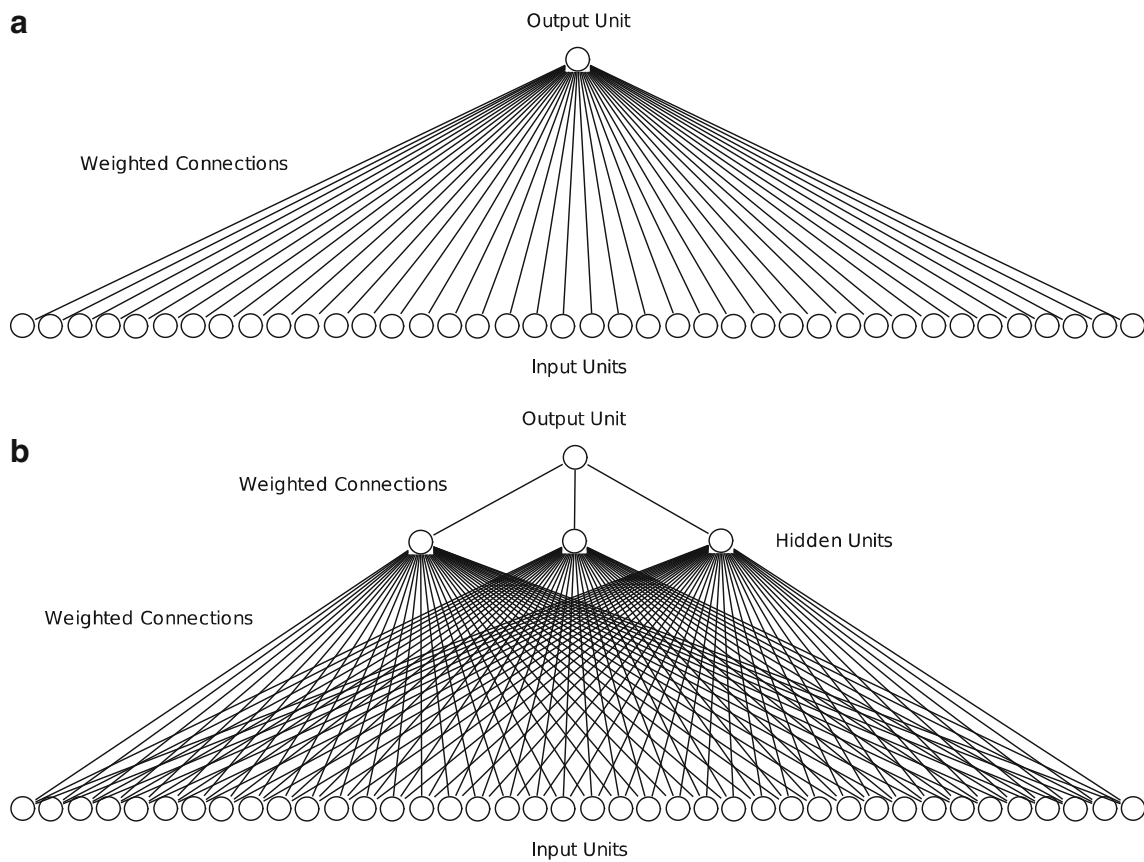


Fig. 1 Illustrations of the architectures of **(a)** single-layer and **(b)** multilayer perceptrons used in the simulations

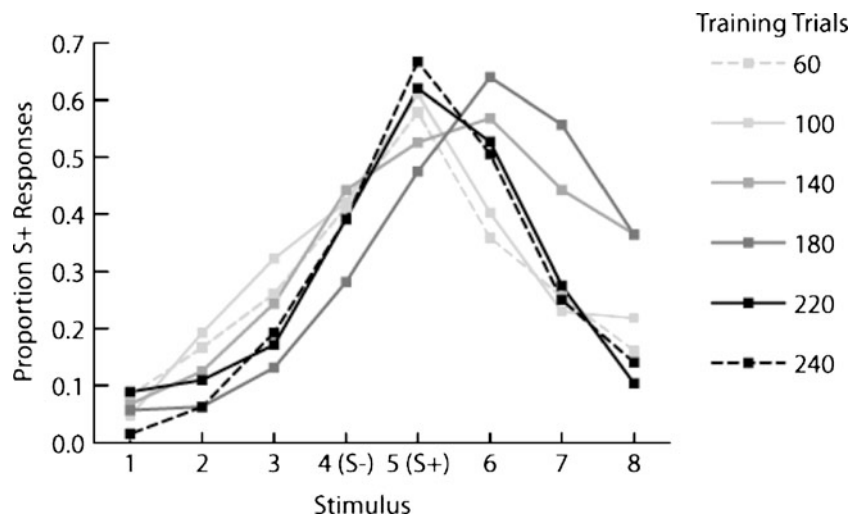
known to capture many of the main features of associative learning theory (Dawson, 2008; Gluck, 1991).

Method

Network architectures Networks consisted of either a single-layer or a multilayer perceptron. In both architec-

tures, the input layer consisted of 40 units. For the single-layer perceptrons (Fig. 1a), each unit in the input layer was connected to a single output unit. For the multilayer perceptrons (Fig. 1b), each unit in the input layer was connected to three units in a hidden layer (the layer between the input layer and the output unit), and every hidden layer unit was connected to a single output unit.

Fig. 2 The generalization gradients reported by Wisniewski, Church, and Mercado (2010) for groups of participants trained for different amounts of trials. Participants trained for either 140 or 180 trials showed both a peak shift and an area shift. Shifts were also reflected in the empirical data by the gradient means and modes



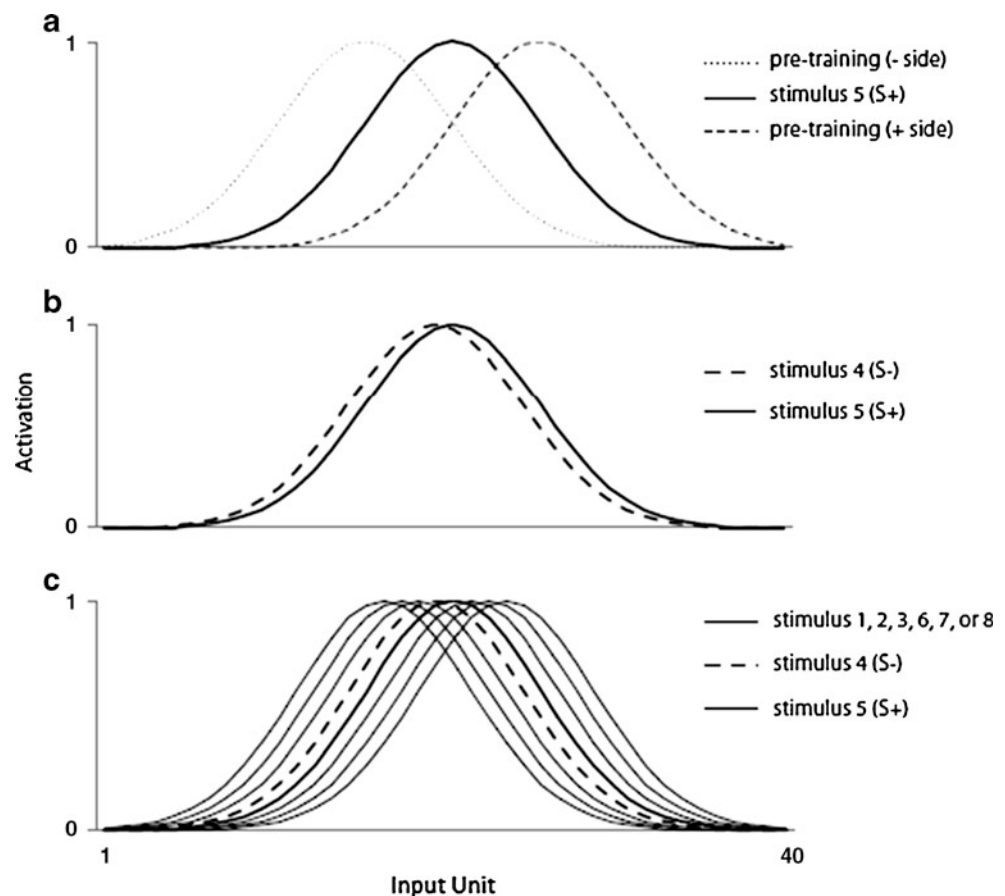
There were no direct connections between input units and the output unit in the multilayer networks. Networks used sigmoid activation functions (see the [Appendix](#) for details). This type of function is often used in perceptron models and has been used extensively to model generalization and peak shift (Dawson, 2004, 2005, 2008; Ghirlanda & Enquist, 1998, 2006; Guillette et al., 2010; Livesey et al., 2005; Suret & McLaren, 2002; Tanaka & Simon, 1996).

Stimulus representations The stimulus set used by Wisniewski et al. (2010) consisted of eight stimuli, rank ordered 1–8, with stimulus 5 used as the “S+” (target stimulus) and stimulus 4 used as the “S–” (nontarget stimulus). Two additional stimuli, which were not part of the generalization distribution, were used during pretraining. Here, we use overlapping patterns of Gaussian-shaped inputs to represent the stimuli used by Wisniewski et al. (2010). These inputs had a variance of 5 and a maximum value of 1. Similar representations have been used previously in connectionist models of peak shift (Ghirlanda & Enquist, 1998, 2006; Livesey et al., 2005; Suret & McLaren, 2002). The input stimulus sets used for pretraining, S+/S– discrimination, and generalization testing are shown in Fig. 3.

Training and testing Single-layer networks were trained with the standard perceptron learning algorithm for sigmoid units, and multilayer networks were trained with error backpropagation (see the [Appendix](#) for details). Connection weights were adjusted to minimize activation error between the actual activation of the output unit and the desired activation (Dawson, 2004, 2005, 2008; Rumelhart, Hinton, & Williams, 1986). Initial network weights were set at random between -0.1 and 0.1 .

Networks were initially pretrained with a desired value of 1 in the output unit for the S+ and 0 for two other stimuli that were displaced along the dimension to either side of the S+. Network pretraining continued until the sum of the squared error (*SSE*) in the output unit response was at least .05. Error was a measure of the difference between the model's output and the desired output. For instance, if the output unit had a desired response of 1 and produced an output of .75, then the squared error would be .05. This is because the difference between the desired and actual response of the output unit ($1 - .75$) is equal to .25. The square of .25 is then .05. *SSE* is the sum of this calculation for each input pattern. The stimuli used to pretrain the networks, with the exception of the S+, were not part of the S+/S– discrimination training or generalization stimulus

Fig. 3 Representations of stimuli used in (a) pretraining (b) S+/S– discrimination, and (c) generalization testing. For the generalization stimulus set, stimuli are ranked from low to high (1–8) on the basis of their position from left (low) to right (high)



sets. The pretraining procedure is analogous to the pretraining used previously in experimental studies (Spetch et al., 2004; Wisniewski et al., 2009, 2010). It also allows the networks to start S+/S- discrimination with some discrimination ability between stimuli on the dimension, rather than being completely naive.

After pretraining, all networks were given S+/S- discrimination training. The desired output for the networks was set at 1 for the S+ and 0 for the S-. In order to compare each model's generalization after different levels of training experience, we trained groups of networks to six different criteria that were defined by (*SSE*) in the output unit. The six criteria were *SSEs* of .5, .3, .2, .1, .05, and .02. Network training was stopped after the respective *SSE* level was reached. After training, generalization was assessed by presenting networks with the S+, the S-, and six novel stimuli. In the generalization phase, networks were tested only for their responding to stimuli and did not experience more training trials with the new stimuli. Five networks were trained per network type (multi- or single-layer) and *SSE* criterion.

Results and discussion

The mean output unit activation of each group of five single layer perceptrons for all test patterns is shown in Fig. 4a. The mean output unit activation was also used to compute

the gradient mean and mode for each criterion level, shown in Fig. 4b. In this case, gradient means and modes reflect the strength of shift such that values higher than 5 are shifted from the S+, in a direction away from the S-. The higher the value is, the further the shift. Even though several networks were trained per condition, the standard errors for output activations, gradient means, and gradient modes were extremely low (0–.02). For this reason, we report the values without standard errors for simulation 1 and the rest of the simulations in this article.

Single-layer perceptrons with sigmoid units showed an increase in shift as *SSE* decreased on the S+/S- discrimination. For instance, the networks trained to an *SSE* of .02 responded most strongly to stimuli 7 and 8. The gradient means and modes also shifted from the S+, even in the most extensively trained group of networks. The gradient mean increased with each level of decrease in *SSE*. The gradient mode peaked at *SSEs* of .2 and .1 and then decreased at *SSEs* of .05 and .02, but these modes are still displaced from the S+ in a direction opposite S-.

Generalization gradients from multilayer networks are shown in Fig. 4c, d. The multilayer perceptrons showed an increasing amount of shift with decreased levels of *SSE* that was highly similar to the pattern observed in single-layer perceptrons.

These simulation results are not specific to the parameter settings, input representations, or number of network units used. Numerous additional simulations were conducted

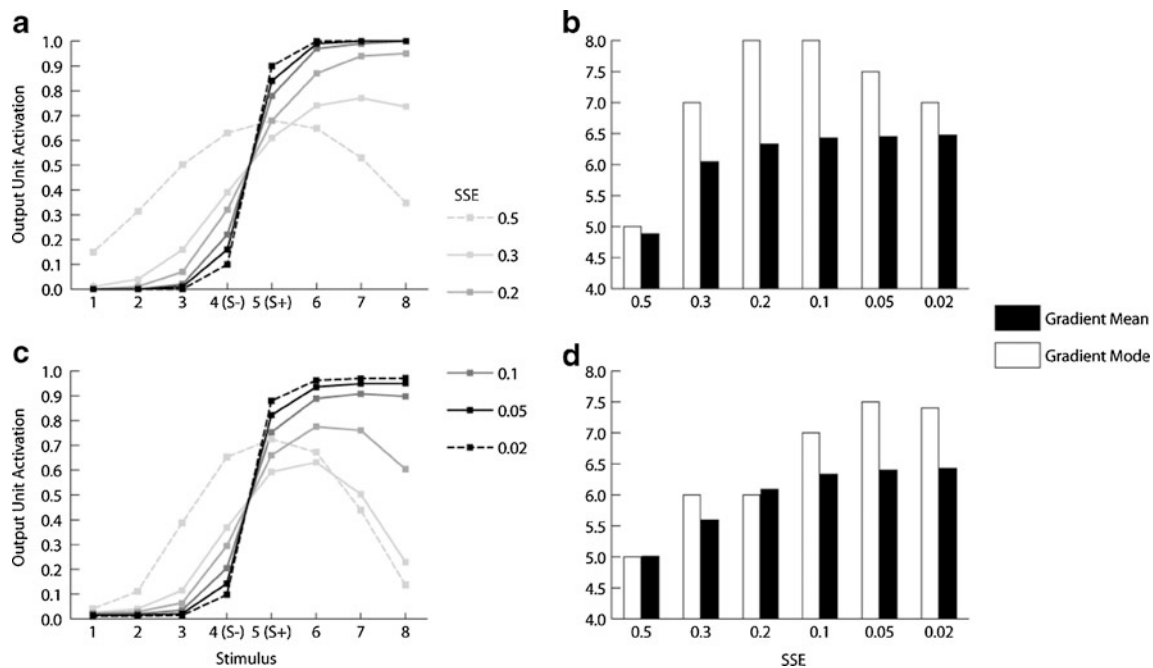


Fig. 4 Generalization gradients (a) and gradient modes and means (b) for single-layer perceptrons with sigmoid units, as well as for the multilayer perceptrons (c and d)

using different learning rates, numbers of input and hidden units, and input representations, all of which produced similar patterns of generalization (unpublished data). Furthermore, published simulations of face recognition learning (Tanaka & Simon, 1996) produce comparable patterns of generalization after different amounts of training. Collectively, these results suggest that standard perceptron models of discrimination learning (and related error-correction-based associative models of learning) predict that learning-related shifts in generalization are a stable asymptotic end state of learning. These models predict that increased training should lead to either greater shifts in generalization or no changes in generalization once a maximally shifted state has been reached. Consequently, these models appear to be unable to account for transient shifts in generalization that first appear and then dissipate as learning progresses, as reported by Wisniewski et al. (2010). Further attempts to modify perceptrons to enable them to replicate the observed pattern of shift revealed that the use of an activation function other than the standard sigmoid function (described below in simulation 2), could account for this pattern of learning-induced shifts in generalization.

Simulation 2

Past connectionist models of discrimination learning provide an elegant and efficient way of instantiating associative learning theories and of describing a wide range of empirical data. It would thus be advantageous if these models could also account for the temporal dynamics of generalization shifts without requiring major modifications or increased computational complexity. One feature of perceptrons that is known to impact how they learn is their activation functions. These functions determine how weighted input values are transformed into output values (see the Appendix for details). In fact, the extensive use of sigmoid activation functions in current perceptron simulations reflects the fact that this function makes efficient training of multilayer perceptrons feasible using error backpropagation. Because activation functions are known to affect how artificial neural networks learn (e.g., Dawson, 2004, 2005, 2008; Dawson & Schopflocher, 1992), we explored whether using an activation function other than the sigmoid function might change the trajectory of generalization shifts observed during learning.

One activation function that has proven effective in past pattern recognition systems is the Gaussian function (similar to a bell curve). This activation function is used extensively in a type of neural network known as a *radial basis function network* (e.g., Park & Sandberg, 1993) and

has also been used in simple perceptron simulations (Dawson, 2004, 2005, 2008; Dawson & Schopflocher, 1992); in this context, units with Gaussian activation functions have been described as *value units*. Value units are like sigmoid units in that they transform the sum of the weighted values from each input unit into an output value that ranges between 0 and 1. Unlike sigmoid units (and more like a dose response curve), value units transform both very low and very high sums into outputs smaller than those for intermediate sums. This allows units to become selective to a range of input values, as is seen in the receptive fields of many cortical neurons (e.g., Elhilali, Fritz, Chi, & Shamma, 2007; Linden, Liu, Sahani, Schreiner, & Merzenich, 2003). Earlier work comparing the effects of stimulus coding on learning-induced shifts in generalization found that simulating different neural codes for stimuli at the input level could dramatically impact how generalization shifts developed with training (Enquist & Ghirlanda, 2005; Ghirlanda, 2002; Ghirlanda & Enquist, 1998). We similarly expected that switching from sigmoid units to value units might significantly impact the patterns of generalization exhibited by perceptrons, even if the use of value units did not affect their ability to learn the trained discrimination.

Method

Network architectures Networks consisted of either a single-layer or a multilayer perceptron, the architectures of which were identical to those described for simulation 1. Networks used Gaussian activation functions for all layers (see the Appendix for details).

Stimulus representations The stimulus set used for these simulations was the same as that used in simulation 1.

Training and testing Single-layer networks were trained with the perceptron learning algorithm for value units developed by Dawson and Schopflocher (1992), and backpropagation was used to adjust weights between hidden and input layer units in multilayer perceptrons (see the Appendix for details). All other training and testing methods were identical to those used in simulation 1.

Results and discussion

The generalization gradients and means and modes of those gradients for value unit networks are shown in Fig. 5. Single layer perceptrons ($n = 5$) made up of value units that were trained to an *SSE* of .02 responded strongest to the S+. Value unit networks that were trained to the criteria of .3, .2,

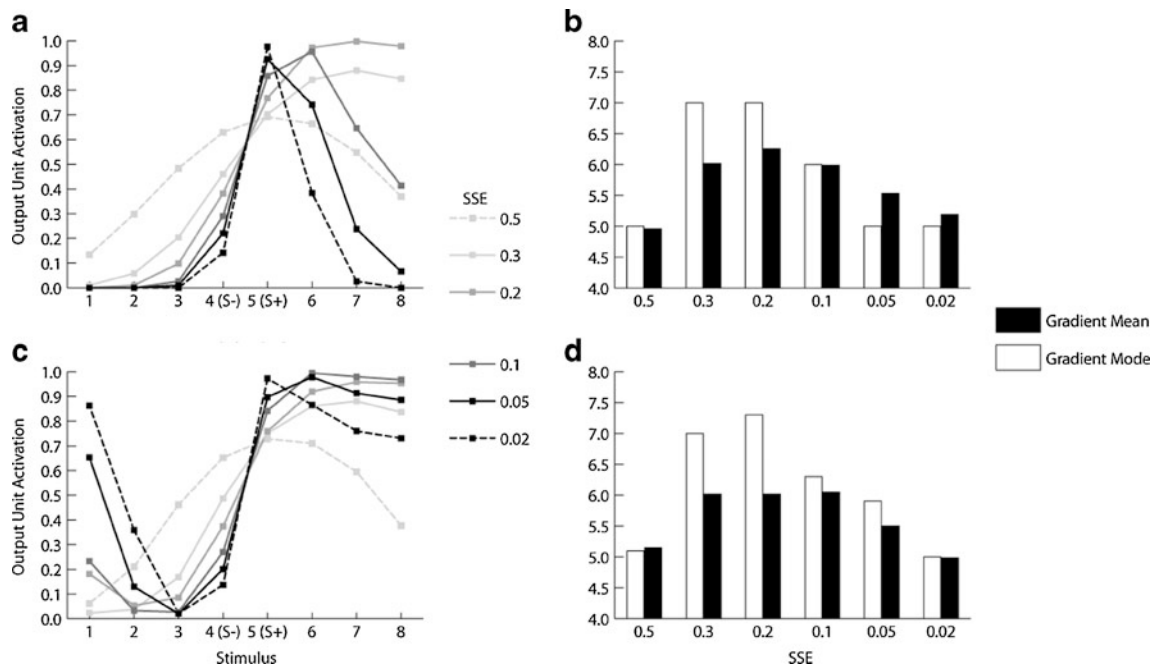


Fig. 5 Generalization gradients (a) and gradient modes and means (b) for single-layer value unit networks. Results for multilayer networks trained with value units (c and d)

and .1 SSE showed gradient shifts, with the most shift seen in networks trained to .2 SSE. The modes and means of value unit trained networks also showed a trajectory of generalization shift such that the largest shifts in the modes and means occurred at intermediate levels of SSE.

Multilayer perceptrons composed of value units also showed a quadratic trend in generalization shift such that the strongest shifts were seen at intermediate levels of SSE. It should be noted here that even though the multilayer networks of value units showed a quadratic trend in mode and mean shift, the responses to stimuli 1 and 2 are qualitatively higher than in the empirical data or the single-layer perceptron simulations. Thus, the single-layer perceptron provided a better qualitative fit to the transient shifts in generalization reported by Wisniewski et al. (2010). This finding suggests that the better fit of the value-unit-based perceptron is not due to greater computational power (because the multilayer perceptron is more powerful) and that the use of Gaussian functions within a perceptron does not guarantee that the network will generalize symmetrically both early and late in training. In general, the differences in generalization behavior of single- and multilayer perceptrons constructed from value units were much greater than observed in simulation 1.

The results of these simulations demonstrate that switching from a sigmoid activation function to a Gaussian activation function enables the single-layer perceptron model to account for the appearance and later disappear-

ance of a peak shift effect during discrimination learning. To our knowledge, this is the first report of a computational model showing temporary shifts in generalization during discrimination learning. It remains possible that other learning theories or computational models (e.g., radial basis function networks) might predict similar trajectories in generalization shifts, but if so, these properties of existing models have yet to be explicitly noted. Why perceptrons with Gaussian activation functions are better able to account for transient shifts in generalization gradients is considered in more detail in the [General Discussion](#) section.

Simulation 3

Manipulations to the type of stimulus representations given to computational models can strongly impact the way they generalize. For instance, within the same model, using different shapes of representations can determine whether a generalization gradient shows peak shift or a tendency to respond to stimuli with the most input layer activation after learning (Ghirlanda, 2002; Ghirlanda & Enquist, 1998). The degree to which stimulus representations overlap can impact generalization as well (Ghirlanda & Enquist, 1998). We performed an additional set of simulations to test how representing stimuli differently would impact the temporal dynamics of generalization shifts that were seen in simulations 1 and 2.

Method

Network architectures Networks were single-layer perceptrons with the same architectures as single-layer perceptrons in simulations 1 and 2. Networks had a single output unit of the sigmoid or value unit type.

Stimulus representations Different stimulus representations were tested in different sets of networks. Triangle- and center-surround-shaped input layer representations were used. We also tested modifications to the Gaussian used in simulations 1 and 2 by reducing the resolution of the original Gaussian or by using Gaussians with other variances. The resolution of the stimulus representation was reduced by taking every other value of the Gaussian, but maintaining the peak at 1, as well as its symmetry. Different variances of the Gaussian were tested by keeping the peak activation for a stimulus representation at its original input unit, but replacing the variance of the Gaussian with the values 1, 3, 7, and 10. Thus, representations with low variances were narrow and overlapped little, whereas representations with high variances were wide and overlapped greatly. Figure 6 display shapes of the center-surround, triangle, and reduced Gaussian representations.

Training and testing Training and testing was identical to single-layer perceptron training and testing in simulations 1

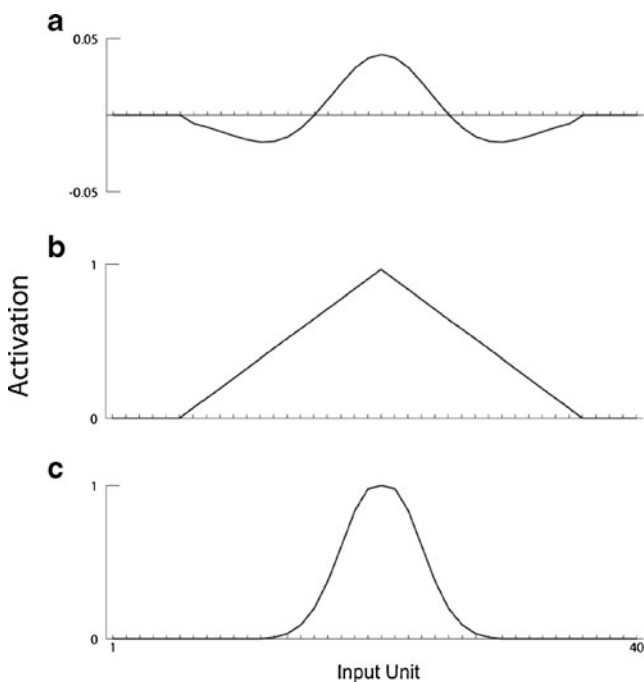


Fig. 6 (a) Center-surround, (b) triangle, and (c) reduced resolution Gaussian representations

and 2. Two single-layer networks were trained for each type of representation, unit type (sigmoid or value), and criteria level.

Results and discussion

For sigmoid unit networks trained with different types of representational shapes and with different resolutions of the original Gaussian, there was an increase in shift with decreasing *SSE* on the S+/S- discrimination similar to the results from simulation 1. Networks trained extensively (i.e., to a criterion of .02 *SSE*) still showed shift from the S+. Value unit networks trained with these representations showed generalization patterns similar to those of simulation 2. That is, these networks showed maximum shift at intermediate *SSE* levels. There were few differences between networks (with the same unit type) trained with different representations in how they generalized. The means and modes for these networks are shown in Fig. 7.

The means and modes of generalization gradients for sigmoid and value unit networks trained with Gaussian representations that had different variances are shown in Fig. 8. For sigmoid unit networks, the trend in shift with decreasing *SSE* on the S+/S- discrimination was such that the lowest *SSE* conditions still showed shift. This was true for all variances. Sigmoid unit networks trained with Gaussian representations that had a variance of 1 or 10 shifted later on in learning than networks trained with variances of 3 and 7. For value unit networks, the quadratic trend we observed using a variance of 5 in simulation 2 did not hold when variances were much lower (i.e., 1), or much higher (i.e., 10). The gradient means increased with decreasing *SSE*, and the modes still show shift at the lowest level of *SSE*. It thus seems that when representations are narrow with little overlap, or wide and highly overlapping, the trend is an increase in shift. The networks did show a quadratic trend, however, when variances of 3 and 7 were used.

In summary, the generalization patterns of networks in simulation 3 demonstrate that the trends in gradient shifts over the course of learning that we found in simulations 1 and 2 are repeatable using different shapes of stimulus representation. In the case of sigmoid unit networks this was also true for input representations that overlapped to different degrees (Gaussians with different variances). However, the time course of shifts was slightly delayed in networks trained with narrow or wide Gaussian representations. For value unit networks, the degree of overlap did matter for whether or not the quadratic trend in generalization shifts was observed. Part of the reason for this effect could be that training stimuli with different degrees of

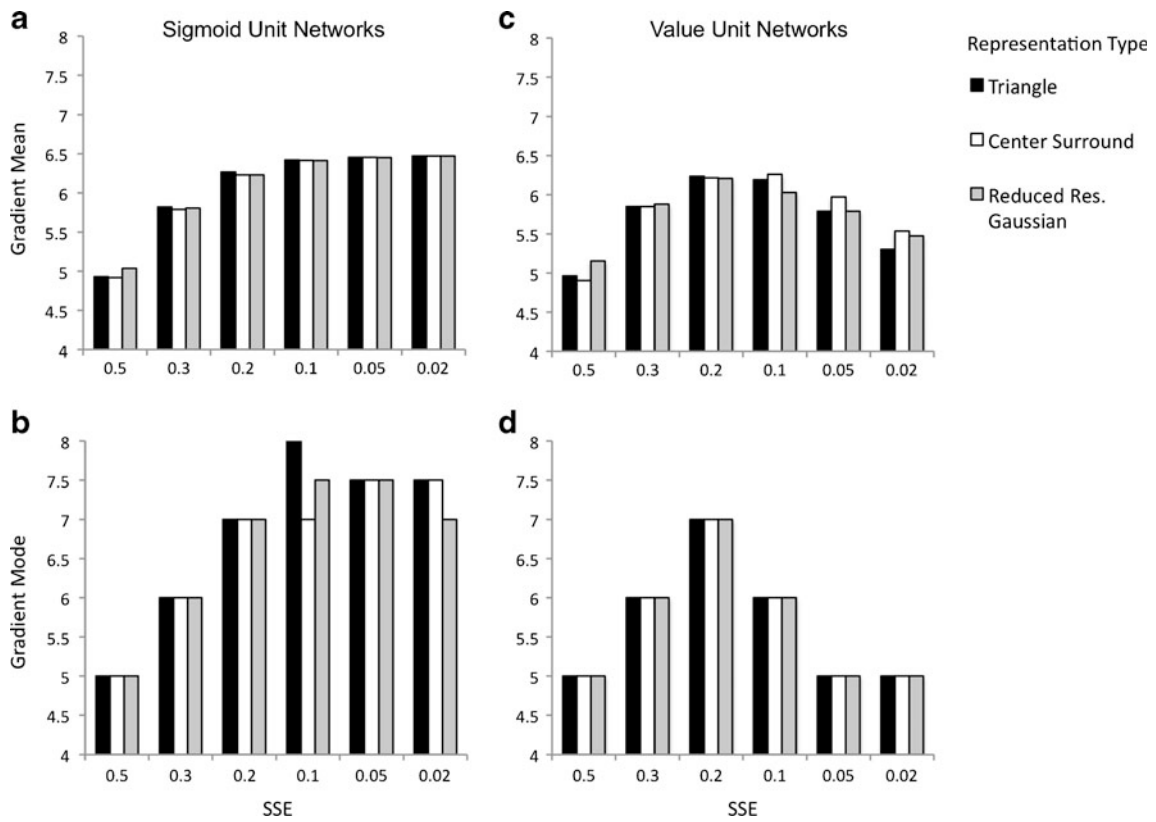


Fig. 7 Gradient means (a, c) and modes (b, d) for sigmoid and value unit networks trained and tested with triangle, center-surround, or reduced resolution Gaussian representations

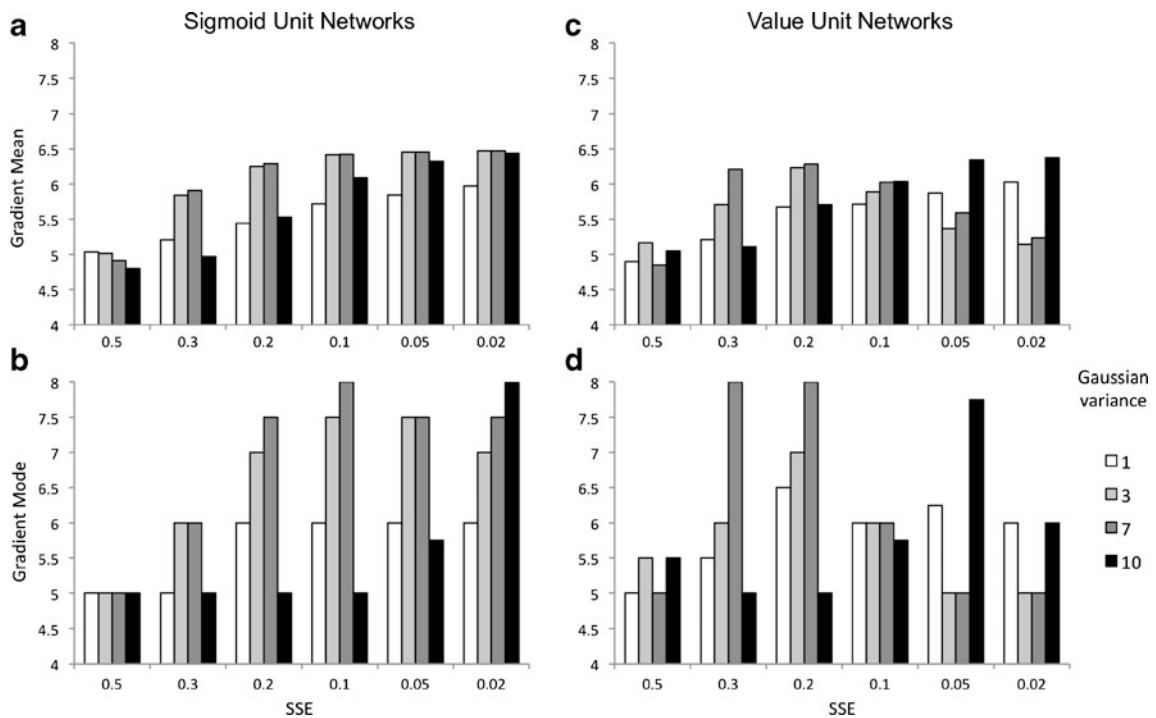


Fig. 8 Gradient means (a, c) and modes (b, d) of sigmoid and value unit networks obtained for different levels of SSE with different variances of the Gaussian used to represent stimuli in the input layer

overlap yield different generalization patterns (Baron, 1973; Ghirlanda & Enquist, 1998; Hanson, 1959; Spence, 1937). We further explored the effect of S+/S- overlap and similarity in simulation 4.

Simulation 4

As was mentioned previously, the similarity between S+ and S- during discrimination learning can affect how much the peak of generalization shifts. In Hanson's (1959) classic peak shift demonstrations with pigeons, more similar S+ and S- discrimination wavelengths led to a larger shift in generalization away from the S-. The effect of S+/S- similarity on the peak shift effect is a consistent empirical result (Baron, 1973; Hanson, 1959; Purtle, 1973) that adequate discrimination learning models should replicate. Although perceptrons with sigmoid units do replicate the effect of S+/S- similarity on generalization (e.g., Ghirlanda & Enquist, 1998), to our knowledge, similar replications with value unit networks have not been reported. For this reason, in simulation 4, we tested whether a single-layer perceptron with a value output unit would produce this effect. Also, there are no studies that have yet examined how fixed stimulus similarity during discrimination training affects the trajectory of learning-related shifts either empirically or theoretically. Thus, in simulation 4, we also tested the generalization of single-layer perceptrons with sigmoid and value units trained to different *SSE* criteria.

Method

Network architectures Network architectures were identical to the single-layer perceptrons used in simulations 1, 2, and 3. Networks had either a sigmoid or a value output unit.

Stimulus representations The stimulus representations for pretraining and generalization testing were the same as those used in simulations 1 and 2. The stimulus representation for the S+ in S+/S- discrimination training was the same as it was in simulations 1 and 2, but the S- stimulus was stimulus 2, 3, or 4 from the set of generalization stimuli shown in Fig. 3c.

Training and testing Training criteria of .5, .3, .2, .1, .05, .02, and .01 *SSE* were used. Pilot testing revealed that some value unit networks did not show the dissipation of shift after being trained to an *SSE* of .02 but did show this effect after being trained to an *SSE* of .01. For this reason, we included the additional *SSE* criterion of .01. Two networks were trained per S- condition, criterion, and unit type. All other training and testing methods were identical to those used in simulations 1, 2, and 3.

Results and discussion

The means and modes of the gradients for each group of networks are shown in Fig. 9. For both sigmoid and value unit trained networks, S+/S- discrimination training with an S- more similar to the S+ led to larger gradient shifts. For instance, in sigmoid unit networks, the values for the largest gradient modes in each S- condition were 7.0 (S- = stimulus 2), 7.5 (S- = 3), and 8 (S- = 4). The largest gradient means for each group of sigmoid unit networks were 5.81 (S- = 2), 6.17 (S- = stimulus 3), and 6.48 (S- = 4). For value unit networks, the largest gradient modes were: 6 (S- = 2), 6 (S- = 3), and 7 (S- = 4); the largest gradient means were 5.65 (S- = 2), 6.00 (S- = 3), and 6.26 (S- = 4). The replication of the effect of S+/S- similarity on peak shift in the value unit neural networks indicates that value unit perceptron networks capture other effects of discrimination learning on generalization, in addition to the temporal dynamics of shift over the course of training.

Although the general trend over the course of learning in networks was such that perceptrons composed of a sigmoid output unit increased shift with learning and value unit networks showed a quadratic trend, stimulus similarity did have an effect on how shifts progressed. Greater similarity of stimuli resulted in generalization shifts appearing earlier in training. For example, for value unit networks that were trained with stimulus 4 as the S-, the peak mean shift was at an *SSE* of .2. However, for networks trained with stimulus 2 as the S-, the peak of mean shift was seen at an *SSE* of .02. To our knowledge, these predicted differences in the temporal dynamics of generalization have yet to be empirically tested.

Simulation 5

The previous simulations show that generalization shifts can depend greatly on the degree to which neural networks learn an S+/S- discrimination. Many past studies of generalization and peak shift involved training participants for a specific number of trials (Baron, 1973; Bizo & McMahan, 2007; Derenne, 2010; Lewis & Johnston, 1999; Newlin et al., 1979; Thomas et al., 1991; Wisniewski et al., 2009, 2010), rather than to specific performance criteria. Recent experimental data also show that generalization differs strongly and, sometimes, nonmonotonically with changes in discrimination performance (Wisniewski et al., 2010). Therefore, training to a specific trial criterion may lead to more variability between participants in generalization than training to a performance criterion. To examine this possibility, in simulation 5, single-layer perceptrons were trained with different learning rates, to criteria defined by *SSE* or by the number of trials.

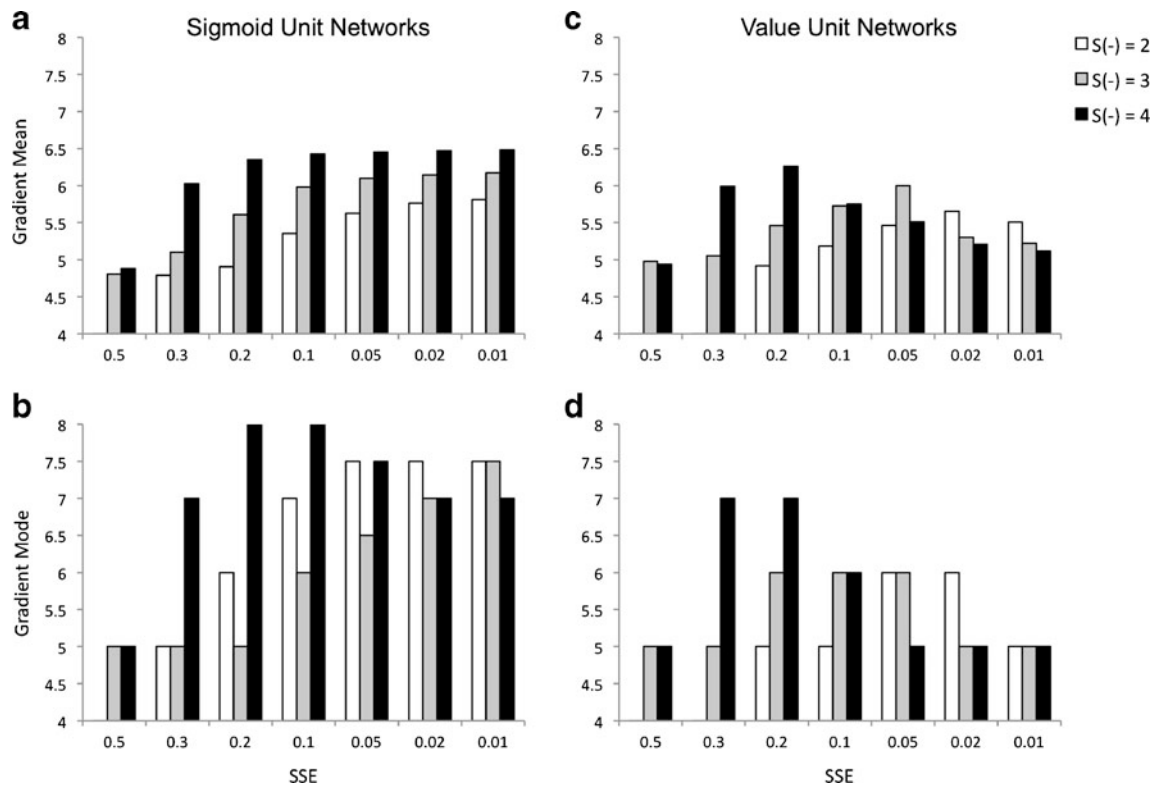


Fig. 9 Displayed are the generalization gradient means (a, c) and modes (b, d) for sigmoid and value unit networks trained to different levels of SSE with differing similarity between S+/S- stimuli in training

Method

Network architectures Network architectures were identical to the single-layer perceptrons used in simulations 1–4.

Stimulus representations Stimulus representations were identical to those used in simulations 1 and 2.

Training and testing Training and testing procedures were the same as those used for single-layer perceptrons in simulations 1–3, with the following exceptions. Networks were trained to the six criterion levels, rank-ordered from lowest to highest, and defined by the SSEs previously used in simulations 1–3 or for a prespecified number of training trials (10, 160, 320, 640, 1,280, or 2,560 trials). Different learning rates of .01, .02, .04, and .08 were used to simulate individual differences in learning capacity. One network was trained per unit type, learning rate, and criterion level defined by SSE or number of training trials.

Results and discussion

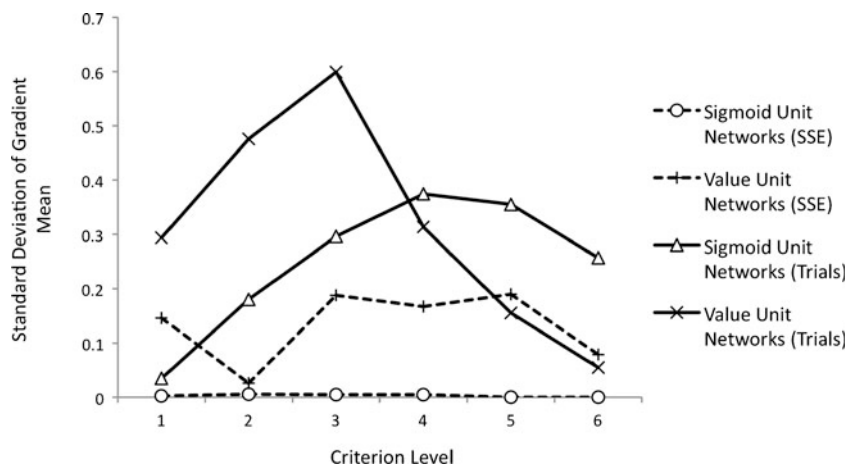
Figure 10 shows the standard deviation of the gradient means calculated by averaging over learning rate at each criterion level. Generally, the standard deviation of generalization

gradient means was higher for sigmoid and value unit networks that were trained for a specified number of trials than for networks that were trained to criterion levels 1–4. Criterion levels 5 and 6 for value unit networks show similar standard deviations for networks trained to a criterion defined by SSE and number of trials. Overall, however, the results suggest that averaging over individuals who have learned to different degrees can yield more variable generalization gradients and that the degree of performance on S+/S- discrimination is a more accurate predictor of generalization.

General discussion

The present simulations demonstrate that perceptron models of discrimination learning can replicate the quadratic trend for shifts in generalization gradients reported by Wisniewski et al. (2010), in which a peak shift effect emerges and then dissipates during the course of learning. However, perceptrons constructed with sigmoid units were not successful in capturing this trend. Sigmoid units have been popular in previous connectionist models of peak shift, which is understandable given that they predict peak shift after learning (Ghirlanda & Enquist, 1998, 2006; Livesey et al., 2005; Tanaka & Simon, 1996), but here they appear to result in stronger shifts with more training on the S+/S-

Fig. 10 Standard deviations of gradient means for sigmoid and value unit networks that were trained to different levels of *SSE* or different amounts of trials



discrimination. This was true for both single- and multilayer architectures. In contrast, single- and multilayer networks trained with value units qualitatively replicated the quadratic trend in gradient shifts. In perceptrons with Gaussian activation functions, shifts in gradients occurred only after intermediate levels of performance were reached on the S+/S− discrimination. Discrimination performance on the S+/S− discrimination that was too poor, or too good, led to little or no shift in the generalization gradients of networks. Previous simulations have shown that networks constructed with Gaussian activation functions have several computational advantages (Dawson, 2004, 2005, 2008; Dawson & Schopflocher, 1992). Our results demonstrate that, in at least some cases, they may also better capture the temporal dynamics of learning and generalization after discrimination training.

Comparison with earlier models of generalization shifts in humans

In the past, special attention has been given to how well findings obtained in experiments with nonhumans generalize to human behavior. For instance, Thomas and colleagues (Newlin et al., 1979; Thomas, 1993; Thomas et al., 1991) suggested that humans are much more sensitive to manipulations to the range of stimuli presented during generalization testing than are other animals. Specifically, they noted that humans tend to respond to the center of a range of generalization stimuli. If the range of stimuli presented during generalization tests is shifted, the peak of generalization gradients for humans can shift as well. These *range effects* have led some researchers to question whether learning-related shifts in humans reflect the same mechanisms as shifts observed in nonhumans (Bizo & McMahon, 2007; Thomas, 1993; Thomas et al., 1991), whereas others have challenged this idea (Galizio & Baron, 1979; Ghirlanda & Enquist, 2006; Spetch et al., 2004; Wisniewski et al., 2009).

Models developed to account for range effects in humans have utilized relational representations to account for generalization shifts.¹ For instance, Thomas (1993) extended Helson's (1964) adaptation-level theory to computationally specify a relational process that replicated the range effects seen in human generalization. According to Thomas, humans develop a prototype of all stimuli along a dimension, which he called the *adaptation level (AL)*; he defined adaptation level as the mean of all stimuli experienced. In this framework, participants learn to identify the S+ during discrimination learning by responding to the adaptation level plus some amount ($AL + X$). During generalization testing, a change in stimulus range can alter the mean, and as a result, the rule $AL + X$ leads to a peak of responding that is at a stimulus other than the S+. The adaptation level model has successfully replicated the central tendency effect and predicted learning-related shifts in many conditions for which they occur.

In the present simulations, we were able to account for the temporal dynamics of human generalization shifts with a simple perceptron that has been previously used to model songbird perceptual discrimination of naturally occurring stimuli (Guillette et al., 2010). Importantly, Thomas's (1993) adaptation-level model cannot account for the shifts in generalization observed by Wisniewski et al. (2010), because generalization testing in that experiment was designed to maintain a fixed adaptation level during both training and testing.² Successful simulation of transient

¹ For an alternative modeling approach to range effects in humans, see Ghirlanda and Enquist (2006).

² Thomas (1993) also suggested that adaptation level theory could explain peak shifts when the computed adaptation level did not shift between training and testing periods. This could be done if the value of X increases from training to testing due to the greater abundance of stimuli during generalization. However, Thomas specified that values of AL and X would change to best represent the generalization stimuli, making changes to values unidirectional. Thus, this explanation cannot explain the quadratic trend in shift seen empirically or in value unit networks.

peak shift effects with a perceptron shows that relational mechanisms are not necessary to explain human peak shift effects in cases where range effects are controlled for (Galizio & Baron, 1979; Spetch et al., 2004; Wisniewski et al., 2009, 2010) and suggests that perceptrons can adequately model discrimination learning and generalization by both humans and nonhumans (Ghirlanda & Enquist, 2006; Mercado, 2008; Spetch et al., 2004; Wisniewski et al., 2009).

Impacts of stimulus encoding on generalization shifts

Value units may have been better for simulating the empirical results of Wisniewski et al. (2010) for at least two reasons. First, the receptive fields of many neurons in the cortex that code for the features of sound selectively respond to specific features of an input, with response decreasing to properties that are dissimilar to those features (e.g., Elhilali et al., 2007; Linden et al., 2003). Sigmoid activation functions cannot capture this type of receptive field. It could be that Gaussian activation functions simulate these types of receptive fields and that these receptive fields are important for discrimination training. Second, Wisniewski et al. (2010), and many others studying generalization (Baron, 1973; Bizo & McMahon, 2007; Derenne, 2010; Lewis & Johnston, 1999; Spetch et al., 2004; Thomas et al., 1991; Wisniewski et al., 2009), used a task for which participants were told to make a response only to the trained stimulus and nothing else. Gaussian activation functions may be best for modeling tasks in which responses should be withheld in the presence of stimuli that are different from the trained stimulus in both directions along the relevant dimension (Dawson, 2004, 2005, 2008; Dawson & Schopflocher, 1992). In contrast, sigmoid functions may be more appropriate for modeling tasks in which participants are instructed to generalize responses to novel stimuli (as in studies of the caricature effect; Tanaka & Simon, 1996).

The simulations also suggest that differences between stimuli in training and testing can impact the temporal dynamics of generalization shifts. When the representations of stimuli used to train networks were changed such that the representations were very wide and overlapping, or very narrow and far apart, a monotonic increase in shift with learning was seen. It was additionally found that the overlap of stimulus representations during discrimination learning impacted the strength and the time course of gradient shifts such that greater overlap led to larger shifts with a faster time course (i.e., shifts developed earlier in learning). Neural representations evoked by stimuli vary from individual to individual (Miglioretti & Boatman, 2003; Orduña et al., 2005), and these representations can change over the course of learning (Elhilali et al., 2007; Goldstone, 1998; Saksida, 1999). It could be the case that

transient shifts may not be seen for all participants and that the state of initial stimulus representations in an individual restricts how he or she can associate those representations to responses, as well as how they generalize (Gibson, 1969; Mercado, 2008).

We also found that generalization was more variable when networks were trained to a criterion based on the number of training trials than to a criterion based on discrimination performance. This finding suggests that when investigating trajectories of changes in generalization, the method of training for a certain number of trials (Baron, 1973; Bizo & McMahon, 2007; Derenne, 2010; Galizio & Baron, 1979; Lewis & Johnston, 1999; Newlin et al., 1979; Thomas et al., 1991; Wisniewski et al., 2009, 2010) may be less effective than training to a performance criterion (Spetch et al., 2004; Wills & Mackintosh, 1998). In addition, there are large individual differences in how participants generalize (Guttman & Kalish, 1956; Landau, 1968; Nicholson & Gray, 1972; Orduña et al., 2005). Some of the previously reported differences in generalization may stem from participants' not reaching similar levels of performance.

The present simulations lead to several novel predictions about the temporal dynamics of generalization shifts: (1) Shifts in generalization over the course of learning should gradually increase and then reach a stable asymptotic state when individuals learn to discriminate stimuli along a dimension that is encoded by absolute neural firing rate but may be transient when the relevant dimension is encoded by selectively tuned neurons; (2) in the latter case, transient shifts will occur only when stimulus representations are moderately overlapping, but not when there is no overlap in encoding or when there is extensive overlap; (3) more similar stimuli should lead to quicker transitions in generalization shifts; (4) training individuals to a performance criterion should lead to less variable generalization than training for a fixed number of trials; and (5) individuals with different learning trajectories (e.g., different learning rates) should show different shifts in generalization with training. These predictions not only provide a way to assess the adequacy of current models of discrimination learning but also suggest that many of the previously reported differences in generalization between humans and other animals (and across different stimulus dimensions) may reflect differences in stimulus encoding, training regimens, and learning capacities, rather than differences in generalization mechanisms.

Acknowledgments This work was supported in part by NSF Grant SBE 0542013 to the Temporal Dynamics of Learning Center, an NSF Science of Learning Center. This research was also supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant and Discovery Accelerator Supplement, an Alberta Ingenuity Fund (AIF) New Faculty Grant, a CFI New Opportunities Grant, along with start-up funding and CFI partner

funding from the University of Alberta, Edmonton, Alberta, Canada. L.M.G. is supported by an Izaak Walton Killam Memorial Scholarship at the University of Alberta. We thank Michael Dawson and Barbara A. Church for comments on the project, Alexander Dovgopoly for assistance with simulations, and Joseph Boomer for reading a previous version of the manuscript.

Appendix

Architectures

In all perceptrons, units had a net input value (net_j) that reflected the sum of their input activations (α_i) after being multiplied by their weighted connection (w_{ij}) with the output unit. Equation 1 shows this computation:

$$net_j = \sum_{i=1}^M \alpha_i w_{ij} \tag{1}$$

Thus, the output unit in the single-layer networks received a weighted sum of all input unit activations. This was also the case for the hidden units in the multilayer networks. However, the output unit of multilayer networks received a weighted sum of the hidden unit activations and did not have any direct connections to the input units.

Units transformed net_j into a value that ranged between zero and one, using activation functions. The activation function for sigmoid unit networks is shown in Eq. 2:

$$\alpha_j = f(net_j) = \frac{1}{1 + e^{-(net_j + \theta_j)}} \tag{2}$$

In this equation, low net_j values are converted to low activations (α_j) (near 0) and high net_j values are converted to high activations (near 1).

For value unit networks the activation function was Gaussian (Eq. 3):

$$\alpha_j = G(net_j) = e^{(-\pi(net_j - \mu_j)^2)} \tag{3}$$

α_j represents the activation for a value unit, net_j is the net input for that unit, and μ is the Gaussian mean. When net_j equals μ , α_j is equal to 1. As net_j becomes dissimilar to the mean in either direction, α_j becomes closer to 0.

Learning algorithms

Sigmoid and value unit single-layer networks were trained using gradient descent learning algorithms. For sigmoid units, the change in a weighted connection between input unit i and output unit j after a sweep of training is defined in Eq. 4:

$$\Delta w_{ij} = \eta \cdot \alpha_i \cdot \alpha_j \cdot (1 - \alpha_j) (t_j - \alpha_j) \tag{4}$$

In this equation, η is the learning rate, α_i is the activation of the input unit, α_j is the activation of the output unit, and t_j is the desired activation of the output unit. It should also be noted that the term $(1 - \alpha_j)$ is the first derivative of the sigmoid activation function. Thus, the change in the weight of the connection between an input and an output unit is a function of the learning rate parameter, the activation function, and the difference between the desired and actual outputs of the output unit. This algorithm was modified in the gradient descent rule defined by Dawson and Schopflocher (1992) that was used to train value units with Gaussian activation functions. Part of their modification involved an elaborated error term that included two parameters for error. These can be seen in Eqs. 5 and 6:

$$\delta_{pj} = (t_{pj} - \alpha_{pj}) \cdot -2\pi \cdot (net_{pj}) \cdot e^{(-\pi(net_{pj} - \mu)^2)} \tag{5}$$

$$\varepsilon_{pj} = t_{pj} \cdot (net_{pj} - \mu_{pj}) \tag{6}$$

Note that δ_{pj} is the same as the error term used to adjust the weight of sigmoid units (Eq. 4), except that it uses the derivative of the Gaussian rather than of the sigmoid activation function. Namely, it is the difference between the output unit's intended activation (t_{pj}) and the actual activation (α_{pj}) for a given input pattern (p) multiplied by the derivative of the Gaussian. The second error term, ε_{pj} , is a measure of error between the net input of an output unit and the mean of the Gaussian activation function multiplied by the desired activation of the unit. Dawson and Schopflocher (1992) found that the elaborated error term was needed to reliably train perceptrons with Gaussian activation functions in order to avoid weights that approached negative or positive infinity. The equation with the additional error term that was used to adjust weights between units is shown in Eq. 7:

$$\Delta w_{ij} = \eta \cdot (\delta_{pj} - \varepsilon_{pj}) \cdot \alpha_{pj} \tag{7}$$

For training of multilayer networks, error backpropagation was used to adjust the weights between hidden layer units and input units. The standard generalized delta rule defined by Rumelhart et al. (1986) was used. For value unit networks, the only difference was that the error used in backpropagation was the elaborated error term defined by Dawson and Schopflocher (1992). More detailed descriptions of network architectures and learning algorithms can be found in Dawson (2004, 2005, 2008; Dawson & Schopflocher, 1992). All simulations in this paper were conducted using software provided by Dawson (2004) for which it was possible to use the learning algorithms and activation functions listed above.

Biological plausibility

It should be noted that perceptrons with different architectures vary in how biologically plausible they are. For instance, the activation functions used in processing units within networks can mimic the type of encoding that happens neurally (Dawson, 2004, 2005, 2008; Enquist & Ghirlanda, 2005). Gaussian activation functions mimic the receptive fields of neurons coding for rearrangement dimensions like frequency, whereas sigmoid activation functions mimic the receptive fields of neurons coding for intensity dimensions like the loudness of sound. Other activation functions, such as those proposed originally by Rosenblatt (1957), can mimic the all-or-none firing of a neuron. The learning algorithms used in perceptrons also vary in biological plausibility. For instance, there are unsupervised learning algorithms (e.g., Kohonen, 1990) that do not require feedback of what the desired response is. These algorithms may be more biologically plausible than supervised learning algorithms, because much learning occurs when there is not a supervisor indicating to the learner the intended response. There are also alternatives to standard backpropagation for training multilayer networks, such as recirculation methods that generate error terms based on internal network activation (e.g., O'Reilly, 1996).

References

- Baron, A. (1973). Postdiscrimination gradients of human subjects on a tone continuum. *Journal of Experimental Psychology*, *101*, 337–342.
- Bizo, L. A., & McMahon, C. V. (2007). Temporal generalization and peak shift in humans. *Learning & Behavior*, *35*, 123–130.
- Blough, D. S. (1975). Steady state data and a quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, *1*, 3–21.
- Cheng, K., Spetch, M. L., & Johnston, M. (1997). Spatial peak shift and generalization in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 469–481.
- Dawson, M. R. W. (2004). *Minds and machines: Connectionism and psychological modeling*. Oxford: Blackwell.
- Dawson, M. R. W. (2005). *Connectionism: A hands-on approach*. Oxford: Blackwell.
- Dawson, M. R. W. (2008). Connectionism and classical conditioning. *Comparative Cognition and Behavior Reviews*, *3*(Monograph), 1–15.
- 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.
- Derenne, A. (2010). Shifts in postdiscrimination gradients within a stimulus dimension based on bilateral facial symmetry. *Journal of Experimental Analysis of Behavior*, *93*, 485–494.
- Elhilali, M., Fritz, J. B., Chi, T., & Shamma, S. A. (2007). Auditory cortical receptive fields: Stable entities with plastic abilities. *Journal of Neuroscience*, *27*, 10372–10382.
- Enquist, S., & Ghirlanda, M. (2005). *Neural networks and animal behavior*. Princeton, NJ: Princeton University Press.
- Galizio, M., & Baron, A. (1979). Human postdiscrimination gradients: The effects of three-stimulus discrimination training. *Animal Learning & Behavior*, *7*(1), 53–56.
- Ghirlanda, S. (2002). Intensity generalization: Physiology and modeling of a neglected topic. *Journal of Theoretical Biology*, *214*, 389–404.
- Ghirlanda, S., & Enquist, M. (1998). Artificial neural networks as models of stimulus control. *Animal Behaviour*, *56*, 1383–1389.
- Ghirlanda, S., & Enquist, M. (2003). A century of generalization. *Animal Behaviour*, *66*, 15–36.
- Ghirlanda, S., & Enquist, M. (2006). How training and testing histories affect generalization: A test of simple neural networks. *Philosophical Transactions of the Royal Society B*, *B362*, 449–454.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Meredith.
- Gluck, M. A. (1991). Stimulus generalization and representation in adaptive network models of category learning. *Psychological Science*, *2*, 50–55.
- Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*, 585–612.
- Guillette, L. M., Farrell, T. M., Hoeschele, M., Nickerson, C. M., Dawson, M. R. W., & Sturdy, C. B. (2010). Mechanisms of call note-type perception in black-capped chickadees (*Poecile atricapillus*): Peak shift in a note-type continuum. *Journal of Comparative Psychology*, *124*, 109–115.
- Guttman, N., & Kalish, H. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology*, *51*, 79–88.
- Hanson, H. M. (1959). Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology*, *58*, 321–334.
- Helson, H. (1964). *Adaptation level theory*. New York: Harper & Row.
- Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*, 1464–1480.
- Landau, J. S. (1968). Postdiscrimination generalization in human subjects of two different ages. *Journal of Experimental Psychology*, *76*, 656–663.
- Lewis, M. B., & Johnston, R. A. (1999). Are caricatures special? Evidence of peak shift in face recognition. *European Journal of Cognitive Psychology*, *11*, 105–117.
- Linden, J. F., Liu, R. C., Sahani, M., Schreiner, C. E., & Merzenich, M. M. (2003). Spectrotemporal structure of receptive fields in areas A1 and AAF of mouse auditory cortex. *Journal of Neurophysiology*, *90*, 2660–2675.
- Livesey, E. J., Pearson, L. S., & McLaren, I. P. L. (2005). Spatial variability and peak shift: A challenge for elemental associative learning? In *Proceedings of the XXVIIth Annual Convention of the Cognitive Science Society* (pp. 1302–1307). Hillsdale, NJ: Erlbaum.
- McLaren, I. P. L., & Mackintosh, N. J. (2002). Associative learning and elemental representation: II. Generalization and discrimination. *Animal Learning & Behavior*, *30*, 177–200.
- Mercado, E., III. (2008). Neural and cognitive plasticity: From maps to minds. *Psychological Bulletin*, *134*, 109–137.
- Miglioretti, D. L., & Boatman, D. (2003). Modeling variability in cortical representations of human complex sound perception. *Experimental Brain Research*, *153*, 382–387.
- Moye, T. B., & Thomas, D. R. (1982). Effects of memory reactivation treatments on postdiscrimination generalization performance in pigeons. *Animal Learning & Behavior*, *10*, 159–166.
- Newlin, R. J., Rodgers, J. P., & Thomas, D. R. (1979). Two determinants of the peak shift in human voluntary stimulus generalization. *Perception & Psychophysics*, *25*, 478–486.

- Nicholson, J. N., & Gray, J. A. (1972). Peak shift, behavioral contrast, and stimulus generalization as related to personality and development in children. *British Journal of Psychology*, *63*, 47–62.
- O'Reilly, R. C. (1996). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation*, *8*, 895–938.
- Orduña, I., Mercado, E., III, Gluck, M. A., & Merzenich, M. M. (2005). Cortical responses in rats predict perceptual sensitivities to complex sounds. *Behavioral Neuroscience*, *119*(1), 256–264.
- Park, J., & Sandberg, I. W. (1993). Approximation and radial basis function networks. *Neural Computation*, *5*, 305–316.
- Purcell, R. B. (1973). Peak shift: A review. *Psychological Bulletin*, *80*, 408–421.
- Rosenblatt, F. (1957). *The perceptron: A perceiving and recognizing automaton (project PARA) (Tech. Rep. 85-460-1)*. Ithaca, NY: Cornell Aeronautical Laboratory.
- Rowe, C., & Skelhorn, J. (2004). Avian psychology and communication. *Proceedings of the Royal Society B*, *271*, 1435–1442.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal errors by error propagation. In D. Rumelhart, J. McClelland, & the PDP Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations* (pp. 318–362). Cambridge, MA: MIT Press.
- Saksida, L. M. (1999). Effects of similarity and experience on discrimination learning: A nonassociative connectionist model of perceptual learning. *Journal of Experimental Psychology. Animal Behavior Processes*, *25*, 308–323.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.
- Shepard, R. N. (1990). Neural nets for generalization and classification: Comment on Staddon and Reid (1990). *Psychological Review*, *97*, 579–580.
- Spence, K. W. (1937). The differential response in animals to stimuli varying within a single dimension. *Psychological Review*, *44*, 430–444.
- Spetch, M. L., Cheng, K., & Clifford, C. W. G. (2004). Peak shift but not range effects in recognition of faces. *Learning and Motivation*, *35*, 221–241.
- Staddon, J. E. R., & Reid, A. K. (1990). On the dynamics of generalization. *Psychological Review*, *97*, 576–578.
- Suret, M., & McLaren, I. P. L. (2002). An associative model of human learning on an artificial dimension. In *Proceedings of the 2002 International Joint Conference on Neural Networks* (pp. 806–811). Piscataway, NJ: IEEE.
- Tanaka, J. W., & Simon, V. B. (1996). Caricature recognition in a neural network. *Visual Cognition*, *3*, 305–324.
- Terrace, H. S. (1966). Behavioral contrast and peak shift: Effects of extended discrimination training. *Journal of Experimental Analysis of Behavior*, *9*, 613–617.
- Thomas, D. R. (1993). A model for adaptation-level effects on stimulus generalization. *Psychological Review*, *100*, 658–673.
- Thomas, D. R., Mood, K., Morrison, S., & Wiertelak, E. (1991). Peak shift revisited: A test of alternative interpretations. *Journal of Experimental Psychology. Animal Behavior Processes*, *17*, 130–140.
- Tijsseling, A. G., & Gluck, M. A. (2002). A connectionist approach to processing dimensional interaction. *Connection Science*, *14*, 1–48.
- Verzijden, M. N., Etman, E., van Heijningen, C., van der Linden, M., & ten Cate, C. (2007). Song discrimination learning in zebra finches induces highly divergent responses to novel songs. *Proceedings of the Royal Society B*, *274*, 295–301.
- Watson, J. B., & Raynor, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, *3*, 1–14.
- Wills, S., & Mackintosh, N. J. (1998). Peak shift on an artificial dimension. *Quarterly Journal of Experimental Psychology*, *51B*, 1–31.
- Wisniewski, M. G., Church, B. A., & Mercado, E., III. (2009). Learning-related shifts in generalization gradients for complex sounds. *Learning & Behavior*, *37*, 325–335.
- Wisniewski, M. G., Church, B. A., & Mercado, E., III. (2010). Temporal dynamics of generalization and representational distortion. *Psychonomic Bulletin & Review*, *17*, 809–814.
- Withagen, R., & van Wermeskerken, M. (2009). Individual differences in learning to perceive length by dynamic touch: Evidence for variation in perceptual learning capacities. *Attention, Perception, & Psychophysics*, *71*, 64–75.