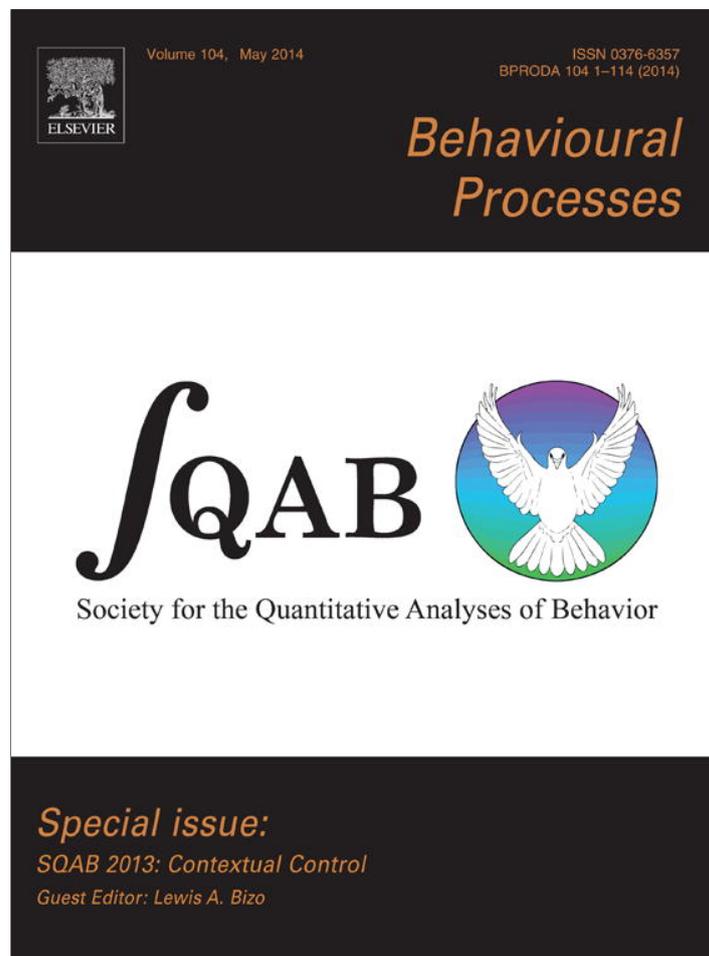


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>



Contents lists available at ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

Individual differences during acquisition predict shifts in generalization



Matthew G. Wisniewski*, Barbara A. Church, Eduardo Mercado III

University at Buffalo, The State University of New York, Buffalo, NY, United States

ARTICLE INFO

Article history:

Received 1 September 2013
 Received in revised form
 23 December 2013
 Accepted 5 January 2014
 Available online 18 January 2014

Keywords:

Generalization gradient
 Peak shift
 Signal detection theory
 Individual differences
 Perceptual learning
 Artificial neural network

ABSTRACT

Learning to distinguish subtle differences in objects or events can impact how one generalizes. In some cases, training can cause novel events to appear more familiar or attractive than those actually experienced during training: the *peak shift* effect. This study examined whether individual differences in learning led to systematic patterns of generalization. Participants were trained to identify simulated bird-songs, and then tested on their ability to identify a target song presented among several similar songs that differed in pitch. Initial analysis showed that those attaining moderate proficiency at discriminating songs during training were more likely to shift than those performing poorly or proficiently. However, a neural network trained to output individuals' gradient dynamics using only performance during training as input found an additional set of training variables that predicted shift. Specifically, one subset of shifters had highly conservative response biases accompanied by very little change to perceptual sensitivity in training. These findings suggest that discrimination learning may only lead to generalization shifts in some individuals, and that all individuals who shift may not do so for the same reason.

© 2014 Elsevier B.V. All rights reserved.

Learning to distinguish perceptual experiences varying along continuous dimensions can distort generalization to favor novel stimuli that accentuate training-relevant features. For example, people familiar with rock musicians may identify a cartoon depiction of Mick Jagger with an extremely large mouth faster and more accurately than a drawing that recreates his actual facial proportions, the so called *caricature effect* (Benson and Perrett, 1991, 1994). Similarly, learning to identify a particular note in a song-bird's call from higher-pitched notes leads birds to preferentially respond to notes lower in pitch, an example of the *peak shift effect* (Guillette et al., 2010; for review see Purtle, 1973; Rilling, 1977). Both effects involve shifts along a continuum from some 'Target' stimulus in a direction opposite competing stimuli (e.g., a prototypical face or 'Non-Target' birdsong note), and can be observed across a wide range of stimulus dimensions (Ghirlanda and Enquist, 2003; Gibson, 1969). This ubiquity has led to their popular use as explanations for naturally occurring shifts in stimulus preference in which it can be reasoned that some intra-dimensional discrimination was previously learned (Barrett, 2010; Costa and Corazza, 2006; De Block and Du Laing, 2010; Derenne et al., 2008; Doyle, 2009; Dunsmoor et al., 2009; Enquist and Arak, 1993; Martindale,

2006; Pinker, 1997; Ramachandran and Hirstein, 1999; Rowe and Skelhorn, 2004).

It is often assumed that learning-related shifts emerge consistently and similarly across individuals. For instance, Ramachandran and Hirstein (1999) propose that well learned discriminations lead to peak shifts that drive commonly held esthetic preferences (e.g., liking of feature exaggeration in artwork). Modeling efforts have largely ignored individual differences in generalization, focusing instead on validating a given theoretical framework's capacity to predict these effects in group-level measures of performance (e.g., Blough, 1975; Ghirlanda and Enquist, 1998; Lynn, 2010; Saksida, 1999; Thomas, 1993). However, learning-related shifts can vary considerably across individuals. In Hanson's (1959) classic demonstration of peak shift, only 3 out of 6 pigeons rewarded for pecks to 550 nm (S+), but not to 590 nm (S-), showed a shift in the mode of their generalization gradients toward wavelengths lower than 550 nm. Later studies revealed substantial variability in the presence and degree of peak shift in animal studies (for review see Rilling, 1977), and in human peak shift and caricature effects (Derenne, 2010; Jones and McLaren, 1999; Livesey and McLaren, 2009; Nicholson and Gray, 1972; Kauffman et al., 2013).

Attempts to explain such individual variations in generalization have been few in number. Some studies have tried and failed (e.g., Derenne, 2010), while others focused on individual differences in human relational/rule-based learning (e.g., respond to 'yellowish' colors; Livesey and McLaren, 2009) have had some success explaining differences in shifts along simple dimensions (e.g., Jones and McLaren, 1999; Livesey and McLaren, 2009; also see Newlin et al.,

* Corresponding author at: Bldg. 441, 2610 7th Street, Wright-Patterson Air Force Base, OH 45433, United States. Tel.: +1 937 255 3169.

E-mail addresses: Matthew.Wisniewski.ctr@us.af.mil, mgw@buffalo.edu (M.G. Wisniewski).

1979). However, these findings may not be generalizable to species that arguably have less sophisticated rule-based/relational learning capabilities (e.g., Smith et al., 2011). Furthermore, variations exist between humans trained and tested on multi-dimensional continua (e.g., faces; Derenne, 2010) that are less susceptible to relational/rule-based learning effects (Spetch et al., 2004). One other early study with human children (Nicholson and Gray, 1972) found a positive correlation between individuals' "neuroticism" and generalization shift. This finding was related to differences in non-reward frustration associated with the S- 'Non-Target' stimulus (cf. Spence, 1937; Terrace, 1966). Yet, learning-related shifts are often seen using forced choice training tasks wherein feedback can be both positive and negative for 'Target' and 'Non-Target' stimuli (e.g., McLaren et al., 1995; Spetch et al., 2004; Wisniewski et al., 2009, 2010). Thus, individual differences in generalization shifts remain largely unexplained.

Past learning theories suggest that between group differences reflect differences in 'what' and how much is acquired during learning; similar explanations may also apply to individuals. For instance, standard single- and multilayer perceptron connectionist networks show stronger peak shift as they become increasingly sensitive to differences between 'Target' and 'Non-Target' training stimuli (Wisniewski et al., 2012). This trend is caused by a set of learned connection weights that disproportionately emphasizes input units more strongly activated by shifted stimuli than the 'Target,' a finding repeatedly reported in associative model simulations of peak shift and caricature effects (Ghirlanda and Enquist, 1998, 2007; Guillette et al., 2010; Saksida, 1999; Tanaka and Simon, 1996). Simulations using modified perceptrons employing activation functions that mimic the tuning properties of sensory neurons important for perceptual learning (e.g., Weinberger, 2007) show shifts only at intermediate levels of learning (Wisniewski et al., 2012). Other perceptual learning models allowing input representations to become less overlapping over the course of learning (e.g., Ahissar et al., 2009; McLaren and Mackintosh, 2000, 2002; Saksida, 1999; Wisniewski et al., 2013) may predict similar non-monotonic trends. This can happen for instance, when the weights learned by associative mechanisms decrease emphasis on features activated more strongly by shifted than trained stimuli as input representations become more distinctive (McLaren and Mackintosh, 2000, 2002; Saksida, 1999; Wisniewski et al., 2013). Although the dynamics of predicted generalization shifts differ between such models, they generally predict that how well an individual learns to become sensitive to differences between stimuli should be related to the degree of shift observed.¹

It is also possible that individual differences in shift correlate with "decision" biases (e.g., Blough, 1969; Boneau and Cole, 1967; Lynn, 2010; Lynn et al., 2005, 2012; Thomas and Vogt, 1983). Assuming a certain degree of uncertainty as to whether or not a given stimulus is the 'Target', subjects/participants may set conservative response criteria so as to avoid false alarms during training. In turn, this could lead to shift if novel stimuli more often exceed criteria than the 'Target'. It has been argued that this explanation of shift in most consistent with many natural situations in which a response to a 'Non-Target' stimulus is costly (e.g., effort extracting nutrients from a poor food source; Lynn et al., 2005), making a conservative bias optimal. In regards to individual differences, those

learning to be conservative in training may shift more than those with unbiased or liberal criteria. Individual differences in sensitivity between training stimuli only play a role insofar as they govern the optimal location of response criteria (Lynn, 2010; Lynn et al., 2012) and/or correlate with overall sensitivity along continua (i.e., how distinctive the 'Target' is from other stimuli).

Here, we examine how individual differences in discrimination learning relate to post-training generalization. Human participants were trained to distinguish a 'Target' birdsong from a 'Non-Target' birdsong that was slightly higher in pitch. Afterward, generalization gradients were obtained by presenting several birdsongs varying in pitch, with instructions for participants to only respond 'Target' to the 'Target' song heard in training. We expected shifts to appear as asymmetries in the generalization gradient favoring 'Target' responses to low-pitched songs. As others have shown large individual variations in performance with complex acoustic stimuli (e.g., Liu et al., 2008; Kidd et al., 2007), we also expected to see variation in learning dynamics and the extent/presence of peak shifts in generalization.

In an initial analysis of individual differences, a post hoc blocking approach was used wherein individuals were split into three equally sized groups based on final training block d' and c . These signal detection measures were used because they reflect an individual's perceptual sensitivity and decision bias respectively (see Macmillan and Creelman, 1991). According to many associative and perceptual learning models, groups based on d' should show significantly different generalization gradients with either intermediate or high values of d' associated with the largest gradient shifts. In contrast, if learned decision biases better predict shift, significant differences should be seen in groups split by c , with the most conservatively biased group associated with the largest shift. It could also be that some combination of training measures best predicts shift. To examine this possibility, we trained an artificial neural network to reproduce gradient measures from input patterns defined by individuals' training data. The internal structure of the network was then analyzed to determine which combination of training measures best predicted generalization.

1. Methods

1.1. Participants

Seventy-six introductory psychology students at the University at Buffalo, The State University of New York, participated for course credit. Two participants were dropped for excess target responding during the generalization test (>90%). Fourteen were dropped for large decreases in discrimination performance during training (d' decreased at least 0.75 from the first to last training block) or low d' in the last training block ($d' < 0.1$).

1.2. Stimuli and apparatus

An excerpt of zebra finch song was used to create stimuli. First, a pitch contour of the song was extracted in Praat (Boersma and Weenink, 2005) to create a stimulus that mimicked frequency and amplitude modulations of song. Adobe® Audition® 3.0 was used to make eleven songs varied in frequency from the original. Songs altered to have spectral peaks at 2368 and 3359 Hz were used for pre-training trials. Other songs with spectral peaks at 2540, 2583, 2670, 2756, 2842, 2928, 3057, 3143, and 3229 Hz were used for training and/or generalization testing. With the exception of the lowest and highest pitched songs, we hereafter refer to songs by their rank from low to high in pitch (songs 1–9). A segment of white noise of equal duration to songs was generated with Adobe® Audition® 3.0 for use in a song vs. noise control training condition.

¹ Connectionist models, like those cited, are able to simulate "decision" biases in addition to changes in perceptual sensitivity (e.g., Ahn et al., 2008). In the Wisniewski et al. (2012) simulations, calculations of network "decision" biases using the signal detection measure c showed that there was also a change over the course of training favoring liberal biases (i.e., a bias to respond 'Target'). However, learned liberal biases in training presumably decrease the probability of observing shift (for review see Lynn, 2010).

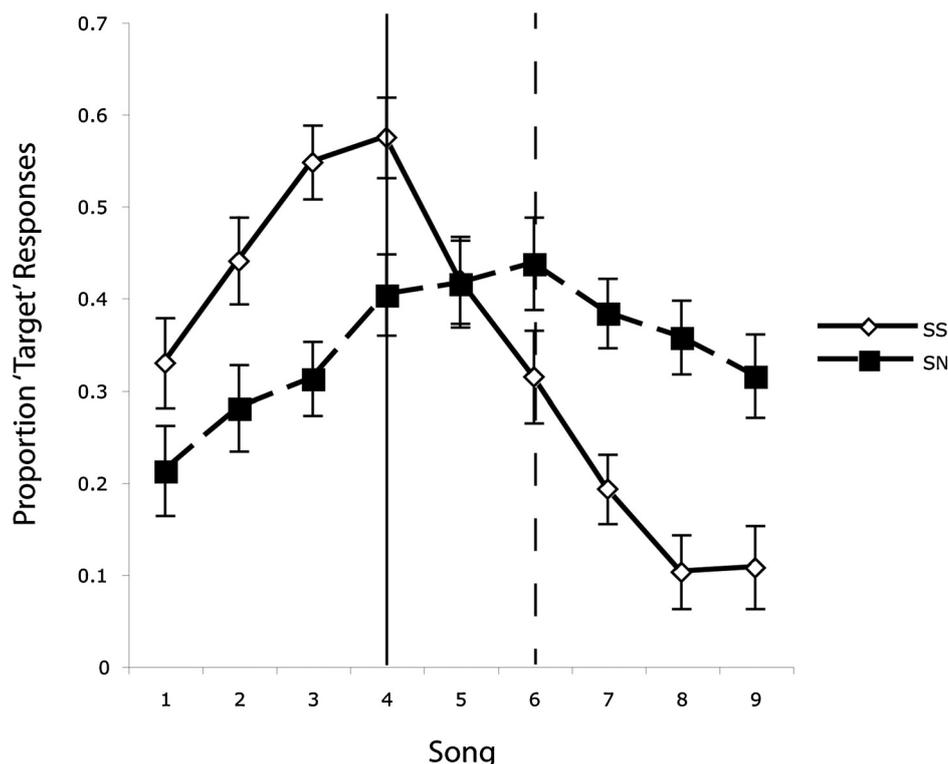


Fig. 1. The proportion of 'Target' responding to each song during generalization testing. Gradients for the SS and SN conditions are shown. The solid vertical line displays the location of the target. The dashed vertical line shows the location of the SS group's non-target. Error bars show the standard error of means.

Procedures were executed with DMDX experimental software (Forster and Forster, 2003). Participants heard songs over Audio-Technica ATH-M40fs headphones and responded using a keyboard.

1.3. Design and procedure

A mixed 2 (training condition) \times 9 (song) design was used. In the song vs. song (SS) training condition ($n=30$), participants learned to distinguish the 'Target' song (song 4) from a higher-pitched song (song 6). In the song vs. noise (SN) training condition ($n=30$), participants were trained to distinguish the same 'Target' from white noise. The SN condition served as a control for possible non-learning related tendencies to respond 'Target' more to one side of the pitch continuum. The levels of the song factor correspond to songs tested in generalization (songs 1–9).

Participants were told to hit a key marked 'Target' if they heard the 'Target' song, and a key marked 'Not Target' for anything else. To make instructions clear, a pre-training phase was used in which the 'Target', the 2368 Hz peak frequency song (lower than the target), and the 3359 Hz peak frequency song (higher than the target) were presented while the appropriate response to each was displayed on the computer screen.

Participants were then trained to distinguish the 'Target' (song 4) from their condition's respective 'Non-Target' stimulus for 4 blocks of 40 trials (160 trials total). Trials within training blocks were pseudo-randomized so that no more than 10 repetitions of the same stimulus occurred consecutively. Participants were given 4 s after the onset of a trial to respond. The word "wrong" was presented on screen after incorrect responses and "correct" appeared after correct responses. After training, songs 1–9 were presented (10 presentations each) in pseudorandom order such that no more

than two of the same song occurred consecutively.² No feedback was given during this testing.

2. Results

2.1. Grand average data

Means for proportion of responding to each song in the SS condition compared to the SN condition showed greater responding on the low-pitch side of the continuum and less responding on the high side (Fig. 1), suggesting a learned generalization shift. A mixed 2 (condition) \times 9 (song) ANOVA using a Huynh–Feldt correction to degrees of freedom (Mauchly's test of sphericity was significant, $p < .05$) revealed a significant main effect of song, $F(3.3, 189.8) = 13.69$, $p < .001$, $\eta_p^2 = .19$, confirming a generalization gradient. There was also a significant condition \times song interaction, $F(3.3, 189.8) = 13.10$, $p < .001$, $\eta_p^2 = .18$, showing that gradients differed between conditions. The main effect of condition was not significant, $F < 2$.

To examine the possibility of shift in the SS group's grand-average gradient, individuals' gradient peak (mode of the gradient) and gradient mean were calculated. Gradient means were determined by multiplying the number of responses to each song by its rank, summing the values, and then dividing by the total number of responses. Only songs 1–7 were used in this calculation so as not

² In one popular model of peak shift (Thomas, 1993) participants learn to base responding on the relationship between a stimulus and the mean rank (adaptation level). Changes to the mean rank from training to test cause shifts according to that model. Note that the mean rank for both training and generalization testing for the SS group was 5. This was purposeful and serves to reduce the capability of the model to predict shifts in our dataset. For further discussion of relational and associative learning issues in studies of learning-related shifts see Lazareva (2012), Spetch et al. (2004), and Thomas (1993).

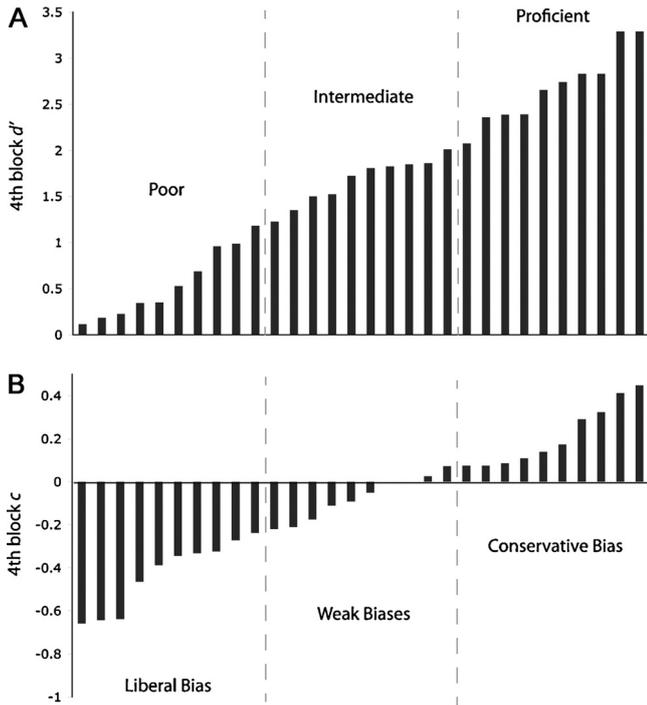


Fig. 2. Individuals sorted by d' (A) and c (B) in the last block of training. Dashed lines indicate tertile splits of the data. Each bar represents an individual.

to skew the mean away from the rank of the 'Target'. Final values for peak and mean were determined by subtracting these gradient peak and gradient mean values from the 'Target' song rank (4). Thus, positive values reflect shift away from the 'Non-Target' and negative values reflect shift toward it. Planned comparison one sample t -tests against 0 were performed on the SS groups' gradient peaks ($M = .29$, $SE = .20$) and means ($M = .31$, $SE = .13$). Only the gradient mean was significantly shifted, $t(29) = 2.48$, $p = .019$, Hedge's $g = .47$. The gradient peak was not, $t(29) = 1.44$, $p = .16$, Hedge's $g = .26$.³ Although the grand average data show only a mean shift and no statistical evidence for a traditional peak shift, this does not preclude the possibility that significant shifts in the peak of generalization gradients exist for individuals with certain learning profiles, a possibility we explore in the next section.

2.2. Individual differences

Fig. 2 shows individual values for d' (sensitivity) and c (decision bias) calculated by using the 4th block of training's proportion of 'Target' responses to the target song as the hit rate and 'Target' responses to the 'Non-Target' song as false alarm rate.⁴ Individuals (bars in the figure) are sorted from low-to-high in value. Participants varied largely on these parameters with sensitivity ranging from poor ($d' = .11$) to proficient ($d' = 3.29$) and decision biases ranging from liberal biases to respond 'Target' ($-.66$) to conservative biases to respond 'Non-Target' (.45). Individuals were partitioned

³ The t -tests were planned because values for gradient peak and mean should be significantly larger than 0 if training led to shift in the generalization gradient, as reviewed in the introduction.

⁴ The c signal detection parameter was used for a measure of decision bias rather than the β signal detection measure for the reason that c is independent of d' (Macmillan and Creelman, 1991). An analysis using β revealed qualitatively similar results. Equations for calculating signal detection measures were: $d' = [z(H) - z(F)]/\sqrt{2}$; and $c = .5[z(H) + z(F)]$. Sensitivity (d') was adjusted downward by $\sqrt{2}$ to compensate for the multidimensionality of the simulated birdsongs (Macmillan and Creelman, 1991).

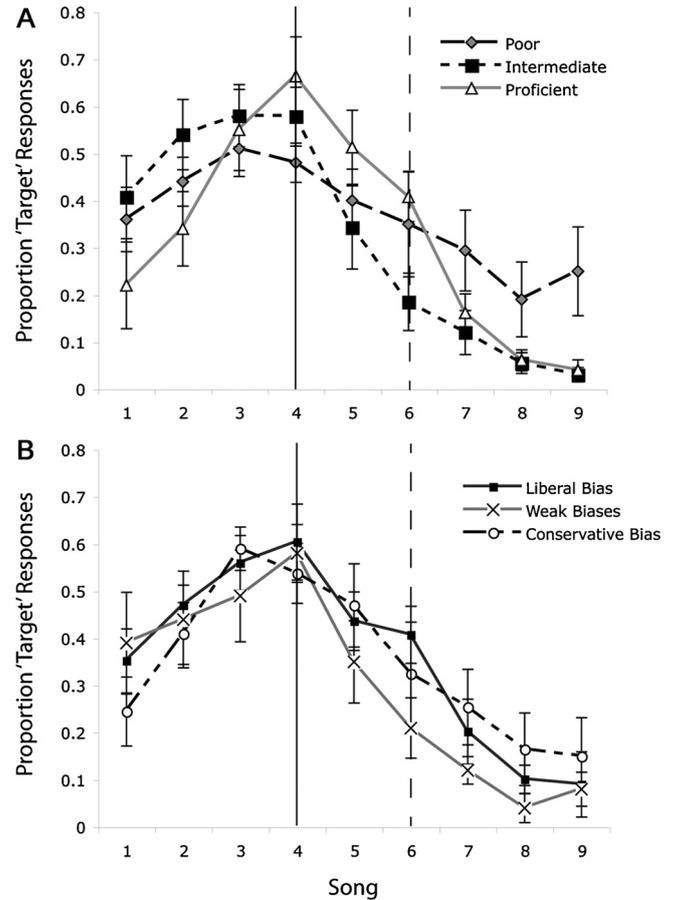


Fig. 3. Generalization gradients of groups based on tertile splits of 4th block training d' (A) and c (B) signal detection measures. Error bars show standard error of means.

into tertiles based on either d' or c . For clarity we refer to tertiles in terms of their relative performance as poor, intermediate, and proficient (for d') and liberal bias, weak biases, and conservative bias (for c).

The generalization gradients of tertiles based on 4th block d' are shown in Fig. 3A. A 3 (tertile) \times 9 (song) ANOVA interpreted with Huynh–Feldt corrections to degrees of freedom revealed a significant main effect of song, $F(8.14, 111.64) = 28.37$, $p < .001$, $\eta_p^2 = .51$, and tertile \times song interaction, $F(8.27, 111.64) = 2.60$, $p = .011$, $\eta_p^2 = .16$, indicating generalization differences between tertiles. There was no main effect of tertile, $F < 2$. Planned comparison one-sample t -tests showed that only the intermediate learners showed significant shifts in gradient peak ($M = .63$, $SE = .25$), $t(9) = 2.49$, $p = .034$, Hedge's $g = .79$, and mean ($M = .68$, $SE = .19$), $t(9) = 3.55$, $p = .006$, Hedge's $g = 1.12$. It is not surprising that poor learners show no statistical evidence of shift. Most learning theories would predict this. A more compelling result is that the only significant shift was found for intermediate learners, as several learning theories predict that the proficient tertile should have shifted the most (Blough, 1975; Chirlanda and Enquist, 1998, 2007; Tanaka and Simon, 1996). This analysis is therefore more consistent with learning models that may allow the dissipation of shift with extended perceptual learning (McLaren and Mackintosh, 2000, 2002; Saksida, 1999; Wisniewski et al., 2012, 2013).

The generalization gradients of tertiles based on 4th block c are shown in Fig. 3B. Qualitatively, it seems as though the conservative bias tertile shows a larger gradient shift than the others, but no significant effects were revealed by ANOVA. Additionally, planned

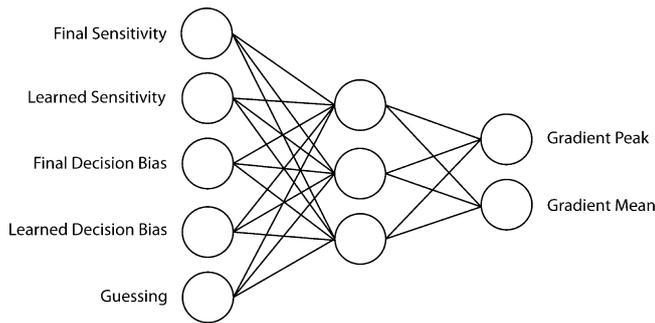


Fig. 4. Neural network architecture trained to produce gradient peaks and means from individual training profiles.

comparison single sided one-sample *t*-tests failed to find significant shifts in the peak or means for any group, $p > .05$.

2.3. Neural network analysis

The above analyses suggest that d' in training may be a better predictor of generalization than decision bias (c). However, other studies have shown that bias can lead to shift in at least some cases. Bumblebees trained to discriminate colors of artificial flowers show peak shift consistent with the relative payoffs for hits and missed detections, a result supporting optimal adjustment of response criteria (Lynn et al., 2005; also see Thomas and Vogt, 1983). Also, even though d' relates to generalization in the current experiment, previous studies have shown that individuals trained to equal levels of discrimination performance still show varying degrees of shift. For instance, in Hanson's (1959) initial demonstration of peak shift all subjects were trained to discriminate their respective S+ and S- to the point where no responses were made to S- at the end of training (given that responding to S+ continued), yet there were still differences in the extent of peak shift.

Here, we use an artificial neural network model as an analysis tool to further assess the capacity of training data to predict an individual's generalization (for similar approaches to examining individual differences see Kidd et al., 2007; Mercado, 2011). A multilayer perceptron was trained to reproduce individuals' peaks and means using measures obtained from training data. By analyzing the internal structure of the trained network, the predictive capability of training variables leading to shift could be further examined.

The multilayer perceptron architecture used is shown in Fig. 4.⁵ Hidden and output units of the network had sigmoid activation functions, squashing the sum of weighted input in between -1 and +1. Individuals' training dynamics corresponding to final sensitivity (block 4 d'), learned sensitivity (block 4–block 1 d'), final bias (block 4 c), learned bias (block 4 c –block 1 c), and guessing (overall false alarm rate) were z-scored and normalized to fall between -1 and 1. These values corresponded to activations in input units.⁶ Gradient peaks and means were similarly z-scored and normalized. These values served as target activations for the two output units of the network. Each individual therefore had a pattern of input activation corresponding to their training performance and a target pattern of output activation corresponding to their generalization. The

⁵ The network was constructed and trained using the Neural Network Toolbox for Matlab R2010a (The MathWorks Inc., Natick, MA).

⁶ These input variables were settled on because pilot simulations with a larger amount of input variables (e.g., d' and c in each block, response time, slope of learning) found that these were the variables most heavily weighted by networks. Several pilot simulations also showed that the addition of more input parameters did not result in large improvements to model accuracy.

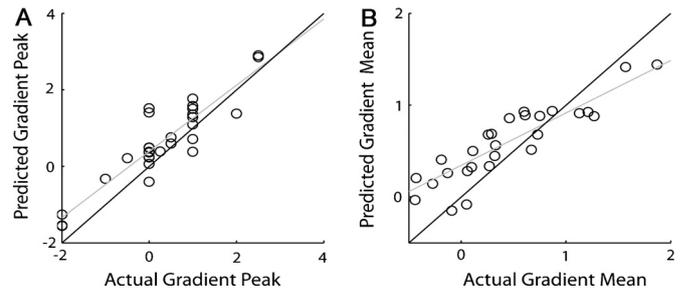


Fig. 5. Network predicted gradient peaks (A) and gradient means (B) as a function of individuals' actual gradient peaks and means. Each circle represents an individual. Dark lines represent unity. The gray lines represent the best linear fit of the data.

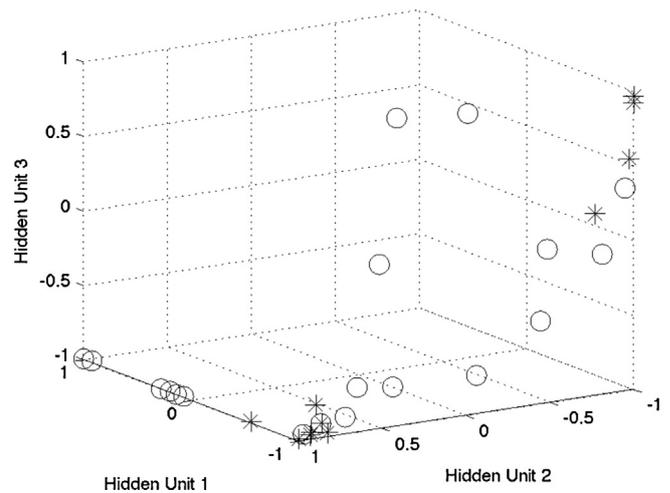


Fig. 6. Hidden unit activations for each individual. Asterisks mark the 10 individuals displaying the largest shift in their generalization gradient.

network was trained to produce the target output activations from input using backpropagation (Rumelhart et al., 1986) for 10,000 iterations with a learning rate of .05.

After training, the output activations were converted from their normalized z-scores back into their original scale to reveal predicted gradient peaks and means. Fig. 5 plots the predicted gradient peaks and means as a function of actual peaks and means for individuals. The dark diagonal line represents unity (i.e., a perfect prediction would fall on this line). The light diagonal line represents the best linear fit of the data. Pearson correlations between predicted and actual gradient peaks ($r = .91$) and means ($r = .91$) were relatively high, showing that the network learned a set of weights that was able to reproduce peaks and means using only data from training.

In order to determine how the network was able to reproduce generalization dynamics from training data the connection weights and hidden unit activities were analyzed. Below, we first describe the hidden unit activations before discussing the weights on training variables that caused them. Fig. 6 plots hidden unit activations for each individual (each point is one individual). For illustrative purposes the points marked by asterisks represent the 10 individuals who had the largest shifts in gradient peak and mean, showing where the network placed the largest shifters in hidden unit space.⁷ These individuals clustered separately into two areas. One cluster ($n = 6$) was in the lower corner of the figure. From here on we will refer to this cluster as CL1. The second cluster ($n = 4$) can be seen in

⁷ The top 10 shifters were identified by averaging the z-scores for gradient peak and mean. The 10 individuals with the highest value were marked with asterisks.

the upper right hand corner. This cluster will be referred to as CL2. The fact that there are two clusters suggests that the network discovered two separate learning profiles that were associated with shift.

For brevity and clarity, we describe qualitatively how the weights of hidden units on training variables affected hidden unit activations. Actual weights are provided in Appendix A. Hidden unit 1 was primarily activated by high learned-sensitivity, a liberal final bias, and a learned change in bias trending toward the conservative end over the course of training. Hidden unit 2 was activated most strongly by low final-sensitivity, high learned-sensitivity, and a final block liberal bias. Hidden unit 3 balanced training variables more equally, but was primarily activated by high final-sensitivity, lower than average learned sensitivity, a conservative final-bias, and a trend for biases to become more conservative over training blocks.

Individuals in CL1 generally correspond to intermediate learners of the 'Target'/'Non-Target' discrimination. As can be seen in Fig. 6, CL1 showed high activation in hidden unit 2, but negative activation in hidden units 1 and 3. Given weights from input to hidden units, one reason individuals clustered together was because they had lower than average levels of final sensitivity. These individuals also had high levels of learned sensitivity, mainly because the negatively activated hidden unit 3 weighted learned sensitivity negatively. The fact that the cluster is associated with high learned-sensitivity, but lower than average final-sensitivity shows that they learned, but did not reach a very high level of d' (not enough to be proficient). Final bias was not an important factor for CL1. However, the activations of hidden units 1 and 3 suggested that individuals in the cluster had a learned bias that moved toward the liberal end over the course of training. Note that this is in the opposite direction predicted to lead to shift from the decisional bias based views of shift (Lynn, 2010; Lynn et al., 2005, 2012; Thomas and Vogt, 1983).

The individuals associated with CL2 seem to be those whose performance is primarily driven by decision bias. Given weights from input to hidden units, individuals in this cluster had high final sensitivity, low learned sensitivity, and a conservative final bias. The generalization performance of these individuals contrasts with previous work suggesting that greater sensitivity can lead to less shift (Terrace, 1966; Cheng and Spetch, 2002; Kahnt et al., 2012; Wisniewski et al., 2010; although see Dukhayil and Lyons, 1973) because individuals in CL2 finished training with high sensitivity, but still shifted. The reason becomes clear when considering that all hidden unit activations were consistent with low learned sensitivity, and a conservative final bias. Thus, participants in CL2 seemed to base their training performance on bias (in a direction that would lead to stronger shift; cf. Lynn et al., 2005) rather than on changes to sensitivity.

3. Discussion

In the current study, participants were trained to discriminate birdsongs varying in pitch and then tested for generalization to other songs on the pitch continuum. Generalization gradients obtained from averaging across individuals showed a significant shift in gradient mean, but no significant peak shift. However, sorting of participants based on their performance during training revealed that a subset of individuals did show peak shift. Specifically, individuals showing intermediate levels of final training block sensitivity shifted away from the 'Non-Target' song experienced during training, whereas the most proficient learners showed no evidence of shift. These findings confirmed the predictions of prior neural network simulations that only individuals with moderate discrimination capacities will show learning-related shifts (Wisniewski et al., 2012). The results also appear to be consistent

with past models used to simulate the effects of perceptual learning on generalization (e.g., McLaren and Mackintosh, 2000, 2002; Saksida, 1999; Wisniewski et al., 2013).

Generalization depends on how discriminable stimuli are, and the sensitivity to differences between stimuli can change over the course of learning (for review see Gibson, 1969; Goldstone, 1998). The absence of shift in some individuals with better discrimination abilities probably reflects the impact of this perceptual learning on generalization gradients. In recent studies of auditory perceptual learning it has been found that training regimens establishing proficient performance also lead to better generalization of that learning to novel sounds. Poorer learning groups generalize less (Church et al., 2013; Liu et al., 2008). Similarly, Amitay et al. (2005) found that individuals given identical frequency discrimination training for several sessions varied vastly in their ability to learn frequency differences. Those that were poor learners did not generalize as well to untrained frequencies as proficient learners. In the current study, some proficient learners may not have shown a shift in generalization because they increased their perceptual acuity during learning in a manner that generalized across the pitch continuum. As a result, these individuals may have been less easily fooled into mistakenly identifying lower-pitched birdsongs as 'Targets'.

Although initial analysis suggested that sensitivity to song differences developed through training was a better predictor of generalization than decision bias, a neural network analysis revealed that for at least one subset of individuals, learned conservative biases may have contributed to generalization shifts. Furthermore, these individuals showed activations in hidden units of the network consistent with high final training sensitivity. This shows that, in some cases, when sensitivity is high, shift can still be seen (cf., Hanson, 1959). Recent work suggests that people can have trouble calibrating bias to sensitivity, favoring non-optimal response criteria (Lynn et al., 2012). However, that study employed a task where reward contingencies skewed the optimal bias to the conservative side. In contrast, in the current study 'Non-Target' and 'Target' stimuli were equally reinforced and occurred with equal probability, nullifying any benefit to having a conservative or liberal bias. Why a subset of our participants showed a conservative bias thus remains unclear. This issue should be addressed in future research (Kantner and Lindsay, 2012).

3.1. Multiple mechanisms of shift

Data from the current study do not clearly support any one model of learning-related shifts. Instead, there are two different types of models that seem to account for portions of the data. Some learning models that have been shown to accurately simulate the effects of perceptual learning on generalization (e.g., McLaren and Mackintosh, 2000, 2002; Saksida, 1999; Wisniewski et al., 2012, 2013) may account for intermediate learners showing large shift, and proficient learners showing little. However, they fail to predict the cluster of individuals who show large shift with high sensitivity and conservative decision biases. This does not mean that the correct set of parameters for those models could not reproduce the data (Ahn et al., 2008). Also, the means by which decision biases are adjusted during learning might be driven by associative mechanisms similar to those built into such models (Ahn et al., 2008; Blough, 1975). However, current decision models of peak shift based on bias seem to do a better job of predicting the latter subgroup a priori, even though they are nonspecific in regards to the particular mechanisms that drive bias adjustment.

Although both types of learning profiles revealed by the neural network showed relatively strong shifts in gradient peak and mean, the dynamics of generalization over the course of time may differ between the two. Those who develop conservative decision biases

during training may show long-lasting and stable shifts in generalization. In fact, recent experiments using recognition memory paradigms have shown that decisional criteria can remain consistent over multiple tests, even when tests are spaced a week apart (Kantner and Lindsay, 2012). In contrast, sensitivity can change greatly over the course of learning and those that rely on perceptual learning during discrimination training may sometimes show transient shifts in generalization because of this. Given that most theories of the effects of learning-related shifts in the real world assume stability over time, the adjustment of decisional criteria may be a more plausible account for these proposed effects. However, our data also show that there are large individual differences in biases (Fig. 2B). In order for decision bias to be a good explanation of widespread real-world phenomena, most individuals would have to show similar conservative biases. We see no evidence for this in the current study, consistent with other studies showing that biases can vary largely between individuals given similar tasks (e.g., Kantner and Lindsay, 2012). It is possible that a consistent and stable conservative bias could develop if the reward contingencies in training were different for “Target” and “Non-Target” stimuli. For instance, shift could remain after extended training if there are more presentations of ‘Non-Target’ than ‘Target’ stimuli in training (making it useful to adopt conservative response criteria; Thomas and Vogt, 1983), but not if ‘Target’ and ‘Non-Target’ stimuli presentations are equalized (cf., Wisniewski et al., 2010). Future studies that mimic real-world reward contingencies to skew optimal response biases (e.g., Lynn et al., 2005; Thomas and Vogt, 1983) may shed further light on the feasibility of learning-related shifts as explanation for real-world phenomena.

3.2. Analyzing individual differences

Researchers have given little attention to differences in shift across individuals, especially in human studies. Several factors may have played a role in this trend. First, participants in past human studies have been trained for very few trials. Some experiments have used as few as 10 training trials (Derenne et al., 2008; Galizio, 1980) and the majority have used less than 100 (Baron, 1973; Bizo and McMahon, 2007; Derenne, 2006, 2010; Derenne and Breitstein, 2006; Doll and Thomas, 1967; Dunsmoor et al., 2009; Galizio, 1980, 1985; Galizio and Baron, 1979; Hedges et al., 1983; Lewis and Johnston, 1999; Newlin et al., 1979; Thomas et al., 1991, 1992; Thomas and Vogt, 1983; White and Thomas, 1979; Wisniewski et al., 2009). Even when participants learn to adequately discriminate objects or events, their experiences with those inputs may not be extensive enough to generate the perceptual expertise necessary to generalize appropriately or consistently to novel items (Wright et al., 2010). Noise in data due to this early variability may have hindered past attempts to predict individual differences in generalization with training data (e.g., Derenne, 2010). Furthermore, individual differences at such early stages of learning have been shown to correlate more strongly with general cognitive abilities than with perceptual skill (Ackerman, 1990). This may make it difficult to relate individual differences in perceptual learning and learned decision biases to shift if they require more time to develop (cf., Boneau and Cole, 1967; Wright et al., 2010).

Additionally, in studies like the current one, strong conclusions that certain training profiles cause a particular gradient cannot be made. It is likely that most studies have focused on between-group analyses and ignored individual variations for this reason. We show here, however, that grand-averaged data may give a distorted view of generalization and that variability between individuals can be utilized to test theoretically motivated hypotheses. We take the stance, along with others (e.g., Kidd et al., 2007; Livesey and McLaren, 2009; McDermott et al., 2010; Mercado, 2008, 2011), that adequate theories should have some way of accounting for

differences between individuals given comparable treatment. Analyses of these individual differences in learning-related shifts are both needed and useful for theory development.

Acknowledgements

This work was supported in part by NSF Grants SBE 0542013 to the Temporal Dynamics of Learning Center and SBE 0835976 to CELEST, two NSF Science of Learning Centers. MGW was supported by a fellowship from the Oak Ridge Institute for Science and Education and a student grant from the Temporal Dynamics of Learning Center. We thank Joanna Solis, Rachel Tan, Amanda Haskell, and Lucas Lee for assistance with data collection, Estella Liu for help with design, Lauren Guillette and Ronald Weisman for comments on the project, and Alexandria Zakrzewski and Joseph Boomer for comments on previous versions of this manuscript.

Appendix A.

See Fig. A.1.

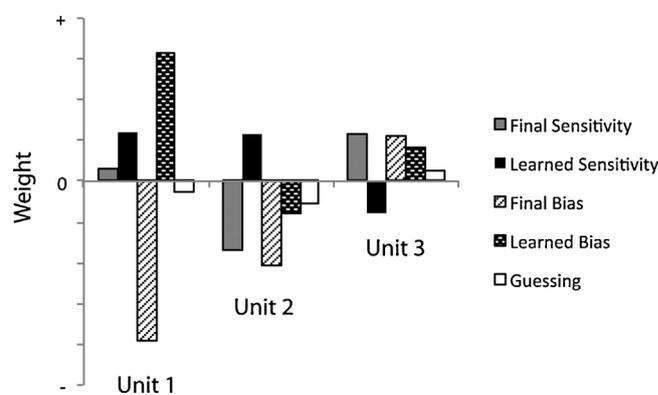


Fig. A.1. Weights of each hidden unit on training variables (input units of the network). Note that activation of a hidden unit is dependent upon the sum of weighted input such that: $net_j = \sum a_i w_{ij}$. In this equation net_j is equal to the sum of each input unit's activation (a_i), multiplied by its weighted connection to unit j (w_{ij}). The determined net_j is added to a unit's bias parameter and then squashed between -1 and $+1$ using a sigmoid activation function (for review see Dawson, 2004). The result is that higher net input to a hidden unit equals greater activation. For example, for the shown set of Unit 1 weights, high final sensitivity (higher than average values of d'), high learned sensitivity, liberal final bias (i.e., lower than average c), a trend in bias becoming more conservative with training, and a lower than average false alarm rate would lead to high activation of Unit 1. In contrast the opposite of these characteristics would lead to large negative activations (e.g., a positive z -score for final bias multiplied the negative weight decreases net_j). Also note that features are differentially weighted. Thus, for Unit 1 final bias has a larger effect on activation than final sensitivity.

References

- Ackerman, P.L., 1990. A correlational analysis of skill specificity: learning, abilities, and individual differences. *Journal of Experimental Psychology: Learning, Memory and Cognition* 16, 883–901.
- Ahissar, M., Nahum, M., Nelken, I., Hochstein, S., 2009. Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B* 364, 285–299.
- Ahn, W.Y., Bussemeyer, J.R., Wagenmakers, E.J., Stout, J.C., 2008. Comparison of decision learning models using the generalization criterion method. *Cognitive Science* 32, 1376–1402.
- Amitay, S., Hawkey, D.J., Moore, D.R., 2005. Auditory frequency discrimination learning is affected by stimulus variability. *Perception & Psychophysics* 67, 691–698.
- Baron, A., 1973. Postdiscrimination gradients of human subjects on a tone continuum. *Journal of Experimental Psychology* 101, 337–342.
- Barrett, D., 2010. *Supernormal Stimuli: How Primal Urges Overran Their Evolutionary Purpose*. Norton, New York.
- Benson, P.J., Perrett, D.I., 1991. Perception and recognition of photographic quality facial caricatures: implications for the recognition of natural images. *European Journal of Cognitive Psychology* 3, 105–135.

- Benson, P.J., Perrett, D.I., 1994. Visual processing of facial distinctiveness. *Perception* 23, 75–93.
- Bizo, L.A., McMahon, C.V., 2007. Temporal generalization and peak shift in humans. *Learning & Behavior* 35, 123–130.
- Blough, D.S., 1969. Generalization gradient shape and summation in steady-state tests. *Journal of the Experimental Analysis of Behavior* 12, 91–104.
- 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.
- Boersma, P., Weenink, D., 2005. Praat: Doing Phonetics by Computer (Version 5.2.21) [computer program]. Retrieved from <http://www.praat.org/>
- Boneau, C.A., Cole, J.L., 1967. Decision theory, the pigeon, and the psychophysical function. *Psychological Review* 74, 123–135.
- Cheng, K., Spetch, M.L., 2002. Spatial generalization and peak shift in humans. *Learning and Motivation* 33, 358–389.
- Church, B.A., Mercado III, E., Wisniewski, M.G., Liu, E.H., 2013. Temporal dynamics in auditory perceptual learning: impact of sequencing and incidental learning. *Journal of Experimental Psychology: Learning, Memory and Cognition* 39, 270–276.
- Costa, M., Corazza, L., 2006. Aesthetic phenomena as supernormal stimuli: the case of eye, lip, and lower-face size and roundness in artistic portraits. *Perception* 35, 229–246.
- Dawson, M.R.W., 2004. *Minds and Machines: Connectionism and Psychological Modeling*. Blackwell, Oxford.
- De Block, A., Du Laing, B., 2010. Amusing ourselves to death? Superstimuli and the evolutionary social sciences. *Philosophical Psychology* 23, 821–843.
- Derenne, A., 2006. Effects of S+ and S− separation on gradient shifts in humans. *Journal of General Psychology* 133, 163–173.
- Derenne, A., 2010. Shifts in postdiscrimination gradients within a stimulus dimension based on bilateral facial symmetry. *Journal of the Experimental Analysis of Behavior* 93, 485–494.
- Derenne, A., Breitstein, R.M., 2006. Gradient shifts with naturally occurring human face stimuli. *The Psychological Record* 56, 365–370.
- Derenne, A., Breitstein, R.M., Chicha, R.J., 2008. Shifts in postdiscrimination gradients within a stimulus dimension based on female waist-to-hip ratios. *The Psychological Record* 58, 51–60.
- Doll, T.J., Thomas, D.R., 1967. Effects of discrimination training on stimulus generalization for human subjects. *Journal of Experimental Psychology* 75, 508–512.
- Doyle, J.F., 2009. A women's walk: attractiveness in motion. *Journal of Social, Evolutionary and Cultural Psychology* 3, 81–92.
- Dukhayil, A., Lyons, J.E., 1973. The effect of overtraining on behavioral contrast and the peak-shift. *Journal of the Experimental Analysis of Behavior* 20, 253–263.
- Dunsmoor, J.E., Mitforr, S.R., LaBar, K.S., 2009. Generalization of conditioned fear along a dimension of increasing fear intensity. *Learning & Memory* 16, 460–469.
- Enquist, M., Arak, A., 1993. Selection of exaggerated male traits by female aesthetic senses. *Nature* 361, 446–448.
- Forster, K.I., Forster, J.C., 2003. DMDX: a windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers* 35, 116–124.
- Galizio, M., 1980. Conditional gradient displacements: the effects of conditional discrimination training on human auditory frequency generalization. *Journal of Experimental Psychology: Human Learning and Memory* 6, 766–777.
- Galizio, M., 1985. Human peak shift: analysis of the effects of three-stimulus discrimination training. *Learning and Motivation* 16, 478–494.
- Galizio, M., Baron, A., 1979. Human postdiscrimination gradients: the effect of three-stimulus discrimination training. *Animal Learning & Behavior* 7, 53–56.
- 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., 2007. How training and testing histories affect generalization: a test of simple neural networks. *Philosophical Transactions of the Royal Society B* 362, 449–454.
- Gibson, E.J., 1969. *Principles of Perceptual Learning and Development*. Appleton-Century-Crofts, New York.
- 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 (*Parus atricapillus*): peak shift in a note-type continuum. *Journal of Comparative Psychology* 124, 109–115.
- Hanson, H.M., 1959. Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology* 58, 321–334.
- Hedges, D.G., Dickinson, J., Modigliani, V., 1983. Stimulus generalization and the peak shift with movement stimuli. *Journal of Motor Behavior* 15, 280–296.
- Jones, F., McLaren, I.P.L., 1999. Rules and associations. In: *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Kahnt, T., Park, S.Q., Burke, C.J., Tobler, P.N., 2012. How glitter relates to gold: similarity-dependent reward prediction errors in the human striatum. *Journal of Neuroscience* 32, 16521–16529.
- Kantner, J., Lindsay, D.S., 2012. Response bias in recognition memory as a cognitive trait. *Memory & Cognition* 40, 1163–1177.
- Kauffman, J.M., Schulz, C., Schweinberger, S.R., 2013. High and low performers differ in the use of shape information for face recognition. *Neuropsychologia* 51, 1310–1319.
- Kidd, G.R., Watson, C.S., Gygi, B., 2007. Individual differences in auditory abilities. *Journal of the Acoustical Society of America* 122, 418–435.
- Lazareva, O.F., 2012. Relational learning in a context of transposition: a review. *Journal of the Experimental Analysis of Behavior* 97, 231–248.
- 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.
- Liu, E.H., Mercado III, E., Church, B.A., Orduña, I., 2008. The easy-to-hard effect in human (*Homo sapiens*) and rat (*Rattus norvegicus*) auditory identification. *Journal of Comparative Psychology* 122, 132–145.
- Livesey, E.J., McLaren, I.P.L., 2009. Discrimination and generalization along a simple dimension: peak shift and rule-governed responding. *Journal of Experimental Psychology: Animal Behavior Processes* 35, 554–565.
- Lynn, S.K., 2010. Decision-making and learning: the peak shift behavioral response. In: *Breed, M., Moore, J. (Eds.), Encyclopedia of Animal Behavior*, vol. 1. Academic Press, Oxford, pp. 470–475.
- Lynn, S.K., Cnaani, J., Papaj, D.R., 2005. Peak shift discrimination learning as a mechanism of signal evolution. *Evolution* 59, 1300–1305.
- Lynn, S.K., Zhang, X., Barrett, L.F., 2012. Affective state influences perception by affecting decision parameters underlying bias and sensitivity. *Emotion* 12, 726–736.
- Macmillan, N.A., Creelman, C.D., 1991. *Detection Theory: A User's Guide*. Cambridge University Press, New York.
- Martindale, C., 2006. Peak shift, behavioral contrast, and phonological evolution with explanations of the great English vowel shift and Grimm's law. *Journal of Cultural and Evolutionary Psychology* 4, 5–26.
- McDermott, J.H., Lehr, A.J., Oxenham, A.J., 2010. Individual differences reveal the basis of consonance. *Current Biology* 20, 1035–1041.
- McLaren, I.P.L., Mackintosh, N.J., 2000. An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior* 28, 211–246.
- McLaren, I.P.L., Mackintosh, N.J., 2002. Associative learning and elemental representation: II. Generalization and discrimination. *Animal Learning & Behavior* 30, 177–200.
- McLaren, I.P.L., Bennett, C.H., Guttman-Nahir, T., Kim, K., Mackintosh, N.J., 1995. Prototype effects and peak shift in categorization. *Journal of Experimental Psychology: Learning, Memory and Cognition* 21, 662–673.
- Mercado III, E., 2008. Neural and cognitive plasticity: from maps to minds. *Psychological Bulletin* 134, 109–137.
- Mercado III, E., 2011. Mapping individual variations in learning capacity. *International Journal of Comparative Psychology* 24, 4–35.
- 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.
- Pinker, S., 1997. *How the Mind Works*. Norton, New York.
- Purtle, R.B., 1973. Peak shift: a review. *Psychological Bulletin* 80, 408–421.
- Ramachandran, V.S., Hirstein, W., 1999. The science of art: a neurological theory of aesthetic experience. *Journal of Consciousness Studies* 6, 15–51.
- Rilling, M., 1977. Stimulus control and inhibitory processes. In: *Honig, W.K., Staddon, J.E.R. (Eds.), Handbook of Operant Behavior*. Prentice Hall, Englewood Cliffs, NJ, pp. 432–480.
- 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: *Rumelhart, D., McClelland, J., The PDP Group (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 1: Foundations. MIT Press, Cambridge, MA, pp. 318–362.
- 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.
- Smith, J.D., Ashby, F.G., Berg, M.E., Murphy, M.S., Spiering, B., Cook, R.G., Grace, R.C., 2011. Pigeons' categorization may be exclusively nonanalytic. *Psychonomic Bulletin & Review* 18, 414–421.
- 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.
- 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 the 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., Vogt, M.B., 1983. Stimulus generalization in humans viewed from a signal detection perspective. *New Zealand Journal of Psychology* 12, 1–6.
- 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.
- Thomas, D.R., Lusky, M., Morrison, S., 1992. A comparison of generalization functions and frame of reference effects in different training paradigms. *Perception & Psychophysics* 51, 529–540.
- Weinberger, N.M., 2007. Auditory associative memory and representational plasticity in the primary auditory cortex. *Hearing Research* 229, 54–68.

- White, G.K., Thomas, D.R., 1979. [Postdiscrimination stimulus generalization in humans: an extension of Galizio and Baron](#). *Animal Learning & Behavior* 7, 564–565.
- Wisniewski, M.G., Church, B.A., Mercado III, E., 2009. [Learning-related shifts in generalization gradients for complex sounds](#). *Learning & Behavior* 37, 325–335.
- Wisniewski, M.G., Church, B.A., Mercado III, E., 2010. [Temporal dynamics of generalization and representational distortion](#). *Psychonomic Bulletin & Review* 17, 809–814.
- Wisniewski, M.G., Radell, M.L., Guillette, L.M., Sturdy, C.B., Mercado III, E., 2012. [Predicting shifts in generalization gradients with perceptrons](#). *Learning & Behavior* 40, 128–144.
- Wisniewski, M.G., Mercado III, E., Church, B.A., 2013. Predicting individual differences in learning-related generalization shifts. Paper presented at the 36th annual meeting of the Society for the Quantitative Analysis of Behavior, May.
- Wright, B.A., Wilson, R.M., Sabin, A.T., 2010. [Generalization lags behind learning on an auditory perceptual task](#). *Journal of Neuroscience* 30, 11635–11639.