CrossMark

Applied Psycholinguistics **37** (2016), 1369–1396
doi:10.1017/S0142716415000600

Performance pressure enhances speech learning

W. TODD MADDOX, SETH KOSLOV, HAN-GYOL YI, and BHARATH CHANDRASEKARAN
University of Texas

Received: January 27, 2015 Accepted for publication: November 8, 2015

ADDRESS FOR CORRESPONDENCE
W. Todd Maddox, Department of Psychology, University of Texas, 1 University Station, A8000, Austin, TX 78712. E-mail: maddox@psy.utexas.edu

ABSTRACT
Real-world speech learning often occurs in high-pressure situations such as trying to communicate in a foreign country. However, the impact of pressure on speech learning success is largely unexplored. In this study, adult, native speakers of English learned nonnative speech categories under pressure or no-pressure conditions. In the pressure conditions, participants were informed that they were paired with a (fictitious) partner, and that each had to independently exceed a performance criterion for both to receive a monetary bonus. They were then informed that their partner had exceeded the bonus and the fate of both bonuses depended upon the participant's performance. Our results demonstrate that pressure significantly enhanced speech learning success. In addition, neurobiologically inspired computational modeling revealed that the performance advantage was due to faster and more frequent use of procedural learning strategies. These results integrate two well-studied research domains and suggest a facilitatory role of motivational factors in speech learning performance that may not be captured in traditional training paradigms.

Acquisition of nonnative speech categories is a difficult learning task in adulthood (Iverson et al., 2003). In the real world, speech learning often occurs under pressure-filled conditions. For example, a learner might feel pressure to perform well when attempting to communicate in a foreign country. Although laboratory training paradigms have yielded successful speech learning in adulthood (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Chandrasekaran, Sampath, & Wong, 2010; Lim & Holt, 2011), to our knowledge the influence of performance pressure on speech learning has not been examined directly. An examination of the effects of performance pressure on speech learning is the goal of the current research.

An extensive behavioral, neuropsychological, and neuroscience literature using visually presented stimuli and laboratory defined category structures suggests that the learning of different types of category structures is mediated by

© Cambridge University Press 2015 0142-7164/15

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

different cognitive and neural systems (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Ennis, 2006; Ashby & Maddox, 2005, 2010; Maddox & Ashby, 2004; Poldrack & Foerde, 2008; Seger, 2008; Smith et al., 2012). A dual-learning systems (DLS) model has been proposed that argues for a reflexive and a reflective learning system. The reflexive system is implicit, procedural, relies on processing in the striatum, and is thought to mediate the learning of "information-integration" categories. The reflective system is explicit, hypothesis-driven, relies heavily on frontal processes, and is thought to mediate the learning of "rule-based" categories.

Performance pressure has been shown to have a strong impact on the strategies learners use, with performance pressure enhancing processing in the reflexive system at the expense of processing in the reflective system (DeCaro, Thomas, Albert, & Beilock, 2011; Maddox & Markman, 2010; Markman, Maddox, & Worthy, 2006; McCoy, Hutchinson, Hawthorne, Cosley, & Ell, 2014). Thus, pressure enhances reflexive-optimal (information-integration) category learning and attenuates reflective-optimal (rule-based) category learning. Unfortunately, all of this work uses visually presented stimuli and experimenter constructed category learning problems.

Over the past several years, we have begun to explore the possibility that the DLS approach, originally developed in vision, might be applicable to the auditory domain, in particular the behaviorally relevant task of speech learning. We recently interpreted speech category learning within the DLS framework and across multiple measures provided evidence that speech learning is often dominated by a procedurally based *reflexive* learning system (Chandrasekaran, Koslov, & Maddox, 2014; Chandrasekaran, Yi, & Maddox, 2014; Maddox, Chandrasekaran, Smayda, & Yi, 2013; Maddox et al., 2014) that relies upon dopamine-mediated reward learning in the striatum (Chandrasekaran, Yi, Blanco, McGeary, & Maddox, 2015; Yi, Maddox, Mumford, & Chandrasekaran, 2014).

This article takes the obvious next step by exploring the influence of performance pressure on nonnative speech learning using Mandarin tones as stimuli. Combining the pressure results in the visual domain with the DLS applications in speech, we hypothesize that performance pressure should *enhance* speech category learning by shifting the balance from reflective to reflexive learning processes. We test, and find strong support for, this prediction behaviorally and through the application of neurocomputational models.

In the next section we elaborate on the DLS approach developed in the visual domain and review the behavioral as well as neuroscience evidence in support of these two learning systems. In the third section we review the literature examining pressure effects on learning, with nearly all of this work focused on visually present stimuli and experimenter defined categories. In the fourth section we review the literature that extends the DLS approach (developed in vision) to the auditory domain (in the context of nonnative speech learning) and provide evidence to suggest that speech learning is dominated by the reflexive learning system. In the fifth section we generate predictions regarding the effects of performance pressure on speech learning. We then turn to a presentation of the experimental task and results and some final conclusions.

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

DLS IN VISION

In a typical laboratory-based category learning task, the participant is presented with a single stimulus on each trial. The participant's task is to determine which category is associated with that stimulus by pressing a button on the computer keyboard. Following their response, the participants receive corrective feedback. Generally, the participants receive several hundred trials in the task, and learning curves are constructed. An example of three trials from a visual category learning task that uses Gabor patches (sine-wave gratings that vary across trials in bar width and orientation) is presented in Figure 1a.

Over the past 20 years, a growing body of behavioral research supports the notion that human visual category learning is mediated by multiple category-learning systems (Allen & Brooks, 1991; Ashby et al., 1998; Brooks, 1978; Erickson & Kruschke, 1998; Love, Medin, & Gureckis, 2004; Maddox & Ashby, 2004; Nosofsky, Palmeri, & McKinley, 1994; Regehr & Brooks, 1993). One of the strongest pieces of evidence comes from an examination of the two category structures depicted in Figure 1b and c, and the learning profiles associated with each category structure. Figure 1b denotes a rule-based category learning problem. Each of the Gabor patches in the figure denotes a stimulus to be presented to the participant. The solid black vertical line denotes the decision boundary that separates Gabor patches in Category A from those in Category B. The strategy that maximizes accuracy is to place Gabor patches with narrow bars into Category A and Gabor patches with wide bars into Category B. This is a reflective strategy because it involves a strategy that is verbalizable and available to conscious awareness. Figure 1c denotes an information-integration category learning problem. Each of the Gabor patches in the figure denotes a stimulus to be presented to the participant. The solid black tilted line denotes the decision boundary that separates Gabor patches in Category A from those in Category B. The strategy that maximizes accuracy is not easily verbalizable, so an information-integration strategy implemented via the reflexive system is most optimal for categorizing these stimuli. If one were to attempt to verbalize this rule, it would be akin to saying something like "If the orientation exceeds the bar width say A, otherwise say B." Because orientation and bar width are measured in different units, this is like mixing apples and oranges. Thus, although one could describe the optimal strategy as a "rule," this strategy is not verbalizable, is not available to conscious awareness, and is qualitatively different from the optimal rule for Figure 1b.

It is critical that humans show very different learning profiles and introspection when asked to solve these tasks. When faced with the verbalizable rule-based task depicted in Figure 1b, participants start out near chance and then at some point "get it" and perform nearly optimal. In other words, participants' learning profile is characterized as a step function. In addition, participants are able to describe the strategy that they used explicitly and accurately. When faced with the information-integration task depicted in Figure 1c, participants start out at near chance and then show gradual, incremental learning. Participants are unable to describe the strategy that they used accurately and often say that they went with their "gut" feeling or "gut reflex." In other words, learning is more implicit.

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

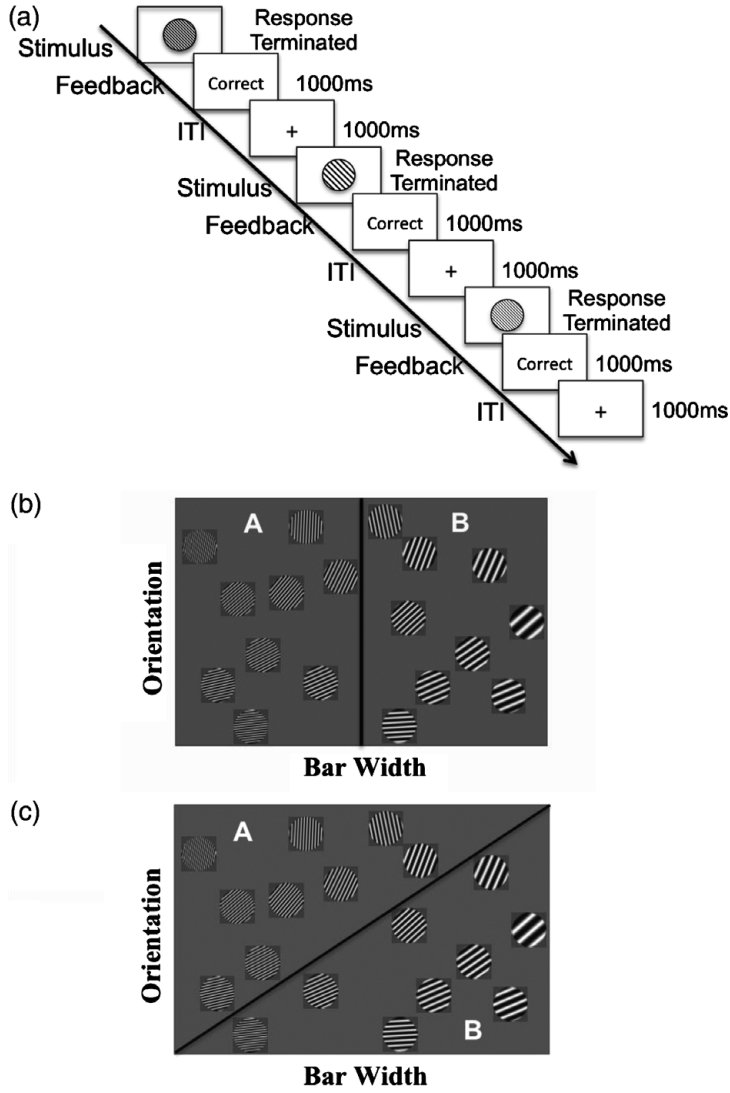


Figure 1. (a) Example of trials in which subjects are shown a stimulus until responding, then given 1000 ms of feedback, followed by a 1000-ms intertrial interval. Three trials are shown. (b) Example of a rule-based category learning task in which narrow bar width Gabor patches are in Category A and wide bar width Gabor patches are in Category B. (c) Example of an information-integration category learning task in which no verbalizable rule can be used to describe the strategy that maximizes accuracy.

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

This qualitative difference in performance led to a number of interesting studies that revealed strong empirical dissociations between the learning of these two category structures, and a plethora of research examining the neural basis of category learning and the development of neurobiologically grounded theories (Ashby et al., 1998; Filoteo, Maddox, Simmons, et al., 2005; Folstein & Van Petten, 2004; Keri, 2003; Knowlton & Squire, 1993; Knowlton, Squire, & Gluck, 1994; Maddox & Filoteo, 2001, 2005; Nomura et al., 2007; Poldrack et al., 2001; Poldrack & Foerde, 2008; Reber, Gitelman, Parrish, & Mesulam, 2003; Seger, 2008; Seger & Cincotta, 2002, 2005, 2006; Seger & Miller, 2010; Shohamy et al., 2004; Shohamy, Myers, Onlaor, & Gluck, 2004; Zeithamova, Maddox, & Schnyer, 2008). One of the theories of visual category learning that specifies the constraints imposed by the underlying neurobiology is the competition between verbal and implicit systems (COVIS; Ashby et al., 1998; Ashby & Maddox, 2005, 2010; Ashby, Paul, & Maddox, 2011) model. COVIS postulates two learning systems, one reflective and one reflexive.[1] The reflective system is an explicit learning system in the sense that it formulates and tests specific categorization rules using cognitive control processes such as executive attention and working memory. The critical neural structures include the prefrontal cortex, anterior cingulate, and anterior caudate nucleus (Ashby, Ell, Valentin, & Casale, 2005; Ashby & Valentin, 2005; Filoteo, Maddox, Simmons, et al., 2005; Lombardi et al., 1999; Monchi, Petrides, Petre, Worsley, & Dagher, 2001; Nomura et al., 2007; Schnyer et al., 2009; Seger & Cincotta, 2006). The reflective system mediates rule-based category learning. In contrast, the reflexive system is implicit and procedural and learns to associate stimuli lying in different regions of perceptual space with specific motor outputs as a result of reinforcement via trial feedback. Learning in this system does not rely on working memory and executive attention, and the critical structures are the posterior caudate, putamen, and the supplementary motor area (Aron et al., 2004; Ashby & Crossley, 2011; Ashby & Waldron, 1999; Filoteo, Maddox, Salmon, & Song, 2005; Maddox & Filoteo, 2001, 2005; Nomura et al., 2007; Seger & Cincotta, 2002, 2005). The reflexive system mediates information-integration category learning.

COVIS assumes that the reflective and reflexive learning systems compete throughout category learning with an initial bias toward reflective dominance. Individuals explicitly test category rules and adjust the weight given to each rule depending on its success or failure. The success or failure of rules is assessed by explicit processing of the feedback. After each trial, the utility of a particular rule is updated. Through this method of hypothesis testing, relevant decision boundaries are learned. The explicit nature of the reflective system requires working memory and executive attention to remember which rules have been used, to process the success or failure of these decision bounds, and to switch between rules. COVIS posits that an accurate reflective system prevents the transfer of control to the striatally mediated reflexive system (Ashby & Maddox, 2010). Learners will therefore continue to use the reflective system until the reflexive system is more accurate.

In comparison, during reflexive learning, a striatal unit implicitly associates an abstract cortical–motor response with sensory cells in the sensory association cortex. Learning occurs at cortical–striatal synapses. Such synaptic plasticity is

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

enhanced by a dopamine-mediated reinforcement signal. The timing of feedback in a categorization experiment is crucial to the effectiveness of the reflexive learning system but not the reflective system, while working memory and executive attention are critical to learning in the reflective system but not the reflexive system. This set of predictions has been tested extensively using experimenter-defined categories with visually presented stimuli. To examine the influence of feedback timing on rule-based and information-integration category learning, researchers have manipulated the time between the participants response and presentation of the feedback. Because corticostriatal synaptic learning is critical to the reflexive system, delaying the feedback should adversely affect information-integration learning because synaptic activity will decay with time. In contrast, because executive attention and working memory can be used to hold a representation of the stimulus and response in memory during a delay period, delaying the feedback should not affect rule-based learning. These predictions are supported by the data from several studies (Maddox, Ashby, & Bohil, 2003; Maddox & Ing, 2005; Worthy, Markman, & Maddox, 2013). To examine the influence of working memory and executive attention on rule-based and information-integration category learning, researchers have manipulated the availability of these resources during task completion or during the time available to process the feedback. Specifically, using dual task procedures, or procedures in which the time available to process the feedback is brief, it has been shown that both of these manipulations adversely affect rule-based learning but not information-integration learning (Filoteo, Lauritzen, & Maddox, 2010; Maddox, Ashby, Ing, & Pickering, 2004; Waldron & Ashby, 2001; Zeithamova & Maddox, 2006, 2007).

PERFORMANCE PRESSURE AND DLS IN VISION

People enter learning environments with trait personalities that affect their learning state (e.g., Grimm, Markman, & Maddox, 2012; Grimm, Markman, Maddox, & Baldwin, 2009; Higgins, 1997), and learning environments often contain situational factors that affect learning (Higgins, 2000). One factor that is often prevalent is performance pressure (Cooper, Worthy, Gorlick, & Maddox, 2013; Markman et al., 2006; Worthy, Markman, & Maddox, 2009a, 2009b). Sometimes we feel pressure because completing a task is a team effort and everyone needs to "carry their weight."

One popular theory of the pressure–cognition relation is the "distraction hypothesis" (Beilock, Bertenthal, McCoy, & Carr, 2004; Beilock & Carr, 2005; Markman et al., 2006). The distraction hypothesis assumes that pressure reduces working-memory capacity by delegating some capacity to worrying about the pressure. This effectively reduces available working-memory capacity for task learning. Within the framework of DLS, and because of the proposed competition between systems, this suggests that performance pressure should attenuate reflective-learning processes and thus should be detrimental to rule-based category learning. Analogously, this suggests that performance pressure should accentuate reflexive-learning processes and thus should accentuate information-integration category learning. Markman et al. (2006) and others (DeCaro et al., 2011; McCoy et al., 2014) applied the distraction hypothesis to rule-based (Figure 1b)

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

and information-integration (Figure 1c) visual category learning and found that pressure did attenuate rule-based category learning while accentuating information-integration category learning. The goal of the current study is to explore the effects of performance pressure on nonnative speech learning. Before we can do that, though, we need to determine the viability of applying DLS in the auditory domain. We address that in the next section.

DLS IN AUDITION: SPEECH LEARNING IN ADULTHOOD IS REFLEXIVE

One exciting possibility that we and others are beginning to explore is the possibility that the DLS approach developed in vision might apply fairly directly to the auditory domain (Chandrasekaran, Koslov, et al., 2014; Chandrasekaran, Yi, et al., 2014; Holt & Lotto, 2008, 2010; Maddox & Chandrasekaran, 2014). In a recent review, we showed that the brain regions associated with auditory processing are interconnected with the brain regions associated with reflective and reflexive category learning, suggesting that the neurobiology associated with COVIS is plausible in the auditory domain (Chandrasekaran, Koslov, et al., 2014). We have also begun to explore some of the cognitive processing correlates of auditory reflective and reflexive category learning. In one study, Maddox, Ing, and Lauritzen (2006) examined reflective-optimal and reflexive-optimal category learning using Gabor patches (i.e., frequency and orientation) as visual stimuli and compared it with reflective-optimal and reflexive-optimal category learning using frequency and duration of a tone as auditory stimuli. The category structures (reflexive-optimal or reflective-optimal) remained the same across the visual and auditory applications; only the specific dimensions changed (from frequency and orientation of a Gabor patch in vision to frequency and duration of a tone in audition). Participants showed similar learning profiles across the visual and auditory versions of the reflective-optimal and reflexive-optimal tasks, suggesting that similar mechanisms were in place (see Smith et al., 2014). In a second test, we examined whether individual differences in working-memory capacity were predictive of individual differences in reflective-optimal and reflexive-optimal nonspeech auditory category learning using simple auditory tones that varied in frequency and duration. In line with the vision work (DeCaro, Thomas, & Beilock, 2008; Tharp & Pickering, 2008), we found that a strong positive correlation between working-memory capacity and reflective-optimal learning with these auditory stimuli, but no correlation between working-memory capacity and reflexive-optimal learning with these auditory stimuli (Chandrasekaran, Koslov, et al., 2014).

This work with simple auditory tones that vary in frequency and duration is important, but one advantage of the auditory domain is that it is rich with natural category learning problems such as the adult learning of nonnative speech categories. Nonnative speech category learning is an excellent starting point for applying theoretical approaches developed in vision to naturalistic problems in audition. Nonnative speech category learning is one of the most difficult human category learning problems and involves taking multidimensional and highly variable acoustic signals and parsing them into discrete phonological representations. Not all nonnative speech categories are difficult to learn in adults, but some categorical distinctions (e.g., /r/ vs. /l/ in Japanese listeners and tone

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

categories in nontone language speakers) are known to be challenging (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Wang, Jongman, & Sereno, 2003; Wang, Spence, Jongman, & Sereno, 1999). Why are these categorical distinctions difficult to learn? Several reasons have been proposed. These include the possibility of interference caused by existing long-term stored representations of native speech categories (Best & Tyler, 2007) as well as interference as a consequence of a "warped" auditory-perceptual space (Francis, Ciocca, Ma, & Fenn, 2008). Laboratory-based training in adults has been shown to be successful in ameliorating difficulties in acquiring new tone categories (Chandrasekaran, Yi, et al., 2014; Maddox et al., 2014; Wang, Jongman, et al., 2003; Wang, Sereno, Jongman, & Hirsch, 2003; Wong & Perrachione, 2007). Training paradigms that have used "high-variability" training approaches have been largely successful in improving tone category learning (Wang, Jongman, et al., 2003; Wang, Sereno, et al., 2003). The thought behind this approach is that high-variability training results in robust category representations that are largely resistant to talker variability. Specifically, they may aid in orienting the focus of attention to relevant dimensions, which shows less variability across talkers. In addition to high-variability training, prior approaches examining tone learning have used trial-by-trial feedback to allow participants to monitor errors and learn from prior trials (Chandrasekaran, et al., 2015; Smayda, Chandrasekaran, & Maddox, 2015). In the proposed training paradigm, we employ elements of both these approaches (high-variability training and trial-by-trial feedback).

The multidimensional and highly variable (due to talker variability) characteristics of speech signals make speech learning a "difficult" categorization problem, especially for individuals learning novel speech categories in adulthood (Bradlow & Bent, 2008). This means that generating and testing hypotheses can be resource intensive. Because the reflective system is dependent on working memory and attention, generating rules/hypotheses for multiple dimensions may not be efficient. Further, the redundancy and variability of cues available during speech perception prevents a simple one-to-one mapping of cues to categories. These suggest that reflexive learning may be most optimal for speech categories. Our hypothesis is therefore that speech learning is reflexive-optimal.

Several lines of evidence support this claim, but we review two that are most relevant to the present discussion. First, in a recent behavioral dissociation study, we (Chandrasekaran, Yi, et al., 2014) showed that delaying the presentation of feedback impairs nonnative speech category learning using the same Mandarin tone learning task used in the present study (described in detail below). As outlined above, delaying feedback interferes with the timing of dopamine release relative to corticostriatal synaptic activation, reducing the effectiveness of the association of stimulus–response with reward (Maddox et al., 2003; Maddox & Ing, 2005). In this same study we also showed that rich, informational, "full" feedback that provides the correctness of the response on each trial as well as information about which category was present attenuates Mandarin tone learning. Full feedback promotes the generation and testing of rules that are critical to reflective learning but disrupts the transfer of control to the reflexive system (Maddox, Love, Glass, & Filoteo, 2008). Second, in a recent imaging study, we showed that greater final block accuracy and greater use of reflexive strategies, as determined from

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

computational modeling (described in detail in the Results section), were associated with increased activation in the putamen (a substructure within the striatum) that is a central brain structure within the reflexive system (Yi et al., 2014). Taken together, these data suggest that nonnative speech category learning, at least with Mandarin tones, is mediated by the reflexive learning system.

PERFORMANCE PRESSURE AND MANDARIN TONE LEARNING: PREDICTIONS

The visual category learning literature suggests that pressure attenuates reflective-optimal visual category learning and enhances reflexive-optimal visual category learning via a mechanism that reduces available executive function resources (DeCaro et al., 2011; Markman et al., 2006; McCoy et al., 2014). Recent work suggests that Mandarin tone learning is reflexive optimal. Thus, we predict that performance pressure will enhance nonnative speech category learning. We tested this hypothesis by having adult, native speakers of American English learn to classify Mandarin tones into one of four linguistically relevant categories under performance pressure or no pressure. In addition to simple measures of accuracy, we utilized a neurobiologically inspired computational modeling approach derived for DLS theory that provides a window onto the nature of learning strategies that is not available through an examination of simple performance measures such as accuracy (Chandrasekaran, Koslov, et al., 2014).

EXPERIMENT

Adult native speakers of American English were trained to learn Mandarin tone categories with feedback and multiple talkers. Mandarin Chinese has four tone categories that differ primarily on the basis of pitch pattern described phonetically as "high-level" in Tone 1 (T1), "low-rising" in Tone 2 (T2), "low-dipping" in Tone 3 (T3), and "high-falling" pitch patterns in Tone 4 (T4). In Mandarin, tone variation within the syllable (for example, /ma/) can change word meaning (ma^1 "mother" [T1], ma^2 "hemp" [T2], ma^3 "horse" [T3], ma^4 "scold" [T4]), making the tonal distinction as phonetically salient as vowels or consonants. Two dimensions (pitch height and pitch direction) serve as primary cues in categorizing tones (Gandour & Harshman, 1978). Although these primary cues explain a large amount of variance per multidimensional scaling studies (Chandrasekaran, Gandour, & Krishnan, 2007; Chandrasekaran et al., 2010; Francis et al., 2008), it is important to note that there are several other cues that may aid in tone categorization. For example, Whalen and Xu (1992) showed that accurate classification is possible on the basis of amplitude contour and duration (in the absence of pitch-related cues).

Participants were randomly assigned to one of three experimental conditions: a no-pressure control condition or one of two performance pressure conditions. The pressure conditions differed only in the target performance level. Performance pressure influences behavior when the target performance level is not too easy and not too hard (Maddox & Markman, 2010; Markman et al., 2006). Average performance in our previous work reached asymptote at 65%–70% accuracy (Chandrasekaran, Koslov, et al., 2014). Thus, we examined the effects of

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

performance pressure across two conditions: a target performance level of 65% accuracy in the final block (P65), and a target performance level of 70% in the final block (P70). Including two target performance levels tests the generality of performance pressure effects.

METHOD

Participants

Seventy-five participants were recruited using flyers and Internet advertisements, were paid $8 for their participation, and were randomly assigned to three groups. This sample size was based on previous work from our lab, and data collection stopped once 75 participants had been tested. Because music training has been shown to enhance speech category learning (Lee & Hung, 2008), individuals with 6 years or more of continuous musical training at any time in life were not allowed to complete the study, as well as those with hearing loss or any significant exposure to tone languages. Participants were divided into the P65 ($N = 22$), P70 ($N = 24$), and the no-pressure control ($N = 29$) conditions. Stimuli were presented at comfortable suprathreshold levels. The participants who did not reach the learning criterion of 30% (chance = 25%) during the final block of trials were excluded from the final analysis: P65 ($N = 2$), P70 ($N = 3$), and no-pressure control ($N = 3$). This research was approved by the University of Texas at Austin Institutional Review Board.

Stimulus characteristics

Stimuli consisted of natural native exemplars of the four Mandarin tones (T1, T2, T3, and T4) in the context of five monosyllabic Mandarin Chinese words (*bu*, *di*, *lu*, *ma*, and *mi*), produced in citation form by native speakers ($N = 4$, 2 females) of Mandarin Chinese originally from Beijing. The five syllables also exist in the American English inventory, circumventing the need to learn phonetic structures additional to the tone distinction. By using different segments and multiple talkers, our aim was to expose learners to variability inherent in natural language. Stimuli were root mean square amplitude and duration normalized (70 dB, 0.4 s; Perrachione, Lee, Ha, & Wong, 2011) using the software Praat (Boersma & Weenink, 2015). The stimuli were rated as highly natural and accurately identified (>95%) by five native speakers of Mandarin. Behavioral studies as well as multidimensional scaling analyses have shown that dimensions related to pitch, especially height and direction, are used primarily to distinguish tone categories (Francis et al., 2008). A scatter plot of the 80 stimuli (4 tones × 4 talkers × 5 words) in the pitch height–pitch direction space is displayed in Figure 2a. Scatter plots of the 40 stimuli spoken by a two male and two female speakers are displayed in Figure 2b and c, respectively.

Procedure

Participants were asked to give informed consent and were seated at a computer in an experimental testing room. All participants were told that they were going to

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

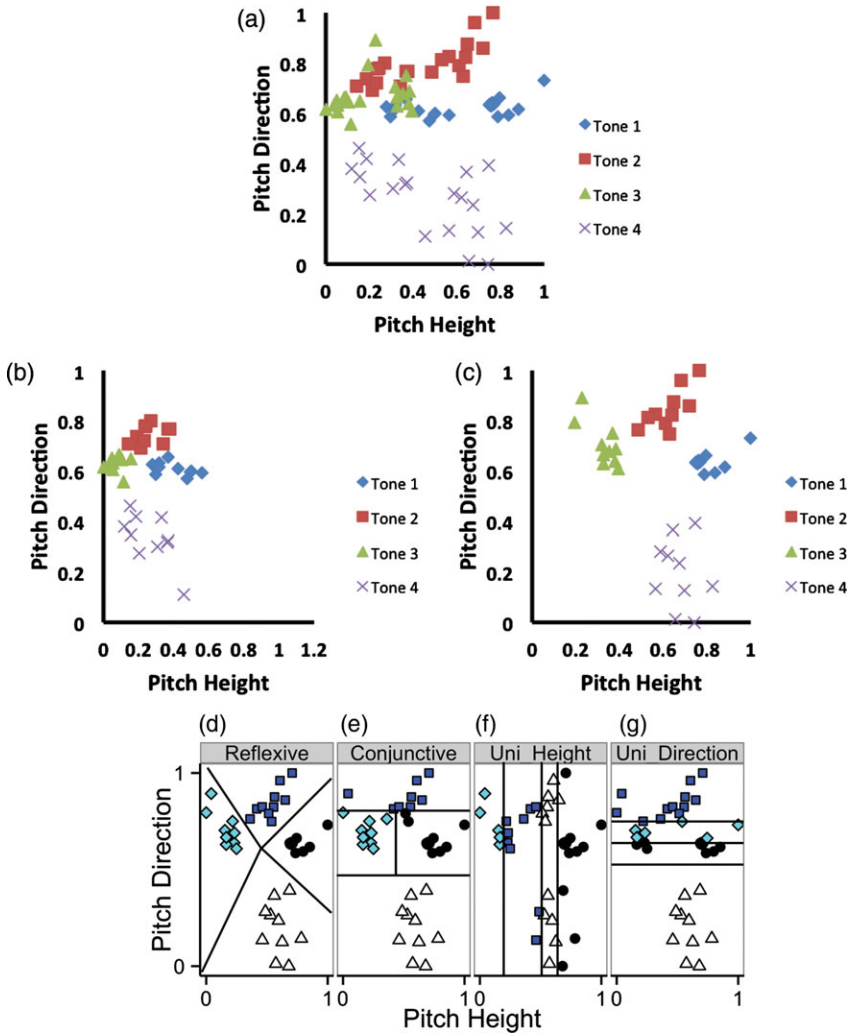


Figure 2. (Color online) Scatter plots of (a) all stimuli from the experiment, (b) male-talker stimuli, and (c) female-talker stimuli. Stimuli dimensions (pitch height and pitch direction) were normalized between 0 and 1. Scatter plots of the responses along with the decision boundaries that separate response regions from a hypothetical participant using a version of the (d) reflexive, striatal pattern classifier; (e) reflective, conjunctive rule-based; (f) reflective, unidimensional_height rule-based; and (g) reflective, unidimensional_direction rule-based models as applied to the female-talker stimuli shown in (c).

be presented with a speech sound on each trial and that their task was to categorize that sound into one of four categories by pressing one of four buttons labeled 1, 2, 3, and 4. Participants were instructed that high levels of accuracy were possible but that the task would be difficult at first. Participants were informed that they

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

would receive corrective feedback on each trial. Feedback was presented for 1000 ms on the computer screen and consisted of the word "Correct" or "No" followed by the label of the tone that was actually presented. For example, on a correct T2 trial the feedback display was as follows: "Correct, that was a category 2." On an incorrect response trial where T3 was the correct response, the feedback display was as follows: "No, that was a category 3." A 1-s intertrial interval followed the feedback. Participants listened to each of the 80 unique stimuli (4 tone categories $\times$ 5 syllables $\times$ 4 talkers) once in each block in a random order, and completed a total of six blocks of training.

The participants in the P65 and P70 conditions were given additional instructions that were modeled directly after those from a number of other published studies (Markman et al., 2006). Participants were informed that they had a chance to earn an extra $8 monetary bonus for good performance. Participants were informed that they would earn points for correct responses and that each correct response would increment a point meter on the right side of the screen. In the P65 and P70 conditions, progress was tracked with a black, 720-pixel tall point meter, which vertically filled with green pixels as subjects categorized tones. With each correct response, the meter was vertically filled by 9 pixels. The point meter remained unchanged following an incorrect response. The bonus criterion was indicated by a line through the point meter that said "Bonus," and was set at 468 pixels, or 65% of the total possible pixels, for the P65 condition and 504 pixels, or 70% of total possible pixels, for the P70 condition. Participants were informed that they must exceed the "Bonus" line on the point meter to be eligible for the bonus, and that they must exceed this in the final block of trials, although they were not informed of the number of blocks to complete. Critically, participants were then informed that they had been paired with a (fictitious) partner, and that each had to complete the task and each had to independently exceed a performance criterion in order for both to receive the monetary bonus. They were informed that should one of them not exceed the performance criterion, then neither would receive the bonus. The participant was then informed that his or her partner had just completed the task and had exceeded the performance criterion. At this point, each participant was asked to rate on a 7-point scale: "How much pressure do you feel to perform well on this task?" In the no-pressure control condition, no instructions pertaining to the monetary bonus or the fictitious partner.[2]

RESULTS

Pressure manipulation check

The average perceived pressure ratings were 3.08, 3.74, and 4.43 in the no-pressure control, P65, and P70 conditions, respectively (95% confidence interval [CI]: 2.353, 3.800 for control; 2.891, 4.583 for P65; 3.624, 5.234 for P70). As expected, the pressure manipulation led to an increase in perceived pressure, F (2, 63) = 3.121, p = .05, partial η^2 = 0.090. Post hoc analyses suggested that the perceived pressure in the P70 condition was significantly higher than that in the control condition (p = .015), but the perceived pressure in the

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

P65 condition was not significantly higher than that in the control condition ($p = .241$). The perceived pressure in the P65 and P70 conditions did not differ ($p = .251$).

Proportion exceeding target performance criterion

Participants in the pressure conditions were required to meet a target performance level in the final block in order to earn a bonus for them and their (fictitious) partner. In the P65 condition, this target was associated with a proportion correct of .65 in the final block. In the P65 condition, 65% of participants reached the target performance level. Although no performance criterion was operative in the no-pressure control condition, we calculated the percentage of participants whose proportion correct in the final block reached .65. Whereas 65% of participants reached this goal in the P65 condition, only 46% of participants reached the same criterion in the no-pressure condition. In the P70 condition, 62% of participants reached the target performance level of .70 proportion correct in the final block. This was compared with only 42% of participants in the no-pressure condition who reached the same criterion. Thus, approximately 20% more participants in each of the pressure conditions reached the relevant performance criterion as compared with participants in the no-pressure condition. Even so, a chi-square test on the control versus P65, χ^2 (1) = 1.275, $p = .25$, and on the control versus P70 frequencies were nonsignificant, χ^2 (1) = 1.785, $p = .18$.

Accuracy results

Figure 3 shows the learning curves across the control, P65, and P70 conditions. Learning performance difference across conditions was assessed using a mixed effects modeling analysis with binomial logit link, using the lme4 package implemented in R (Bates, Maechler, & Bolker, 2012). The dependent variable was trial-by-trial accuracy outcome (correct vs. incorrect: reference level) for each participant. The fixed effects were the between-subjects pressure condition (P70, P65, and no-pressure: reference level), trial number (1 to 480; mean-centered and divided by 100), and their interaction term. The most complex random effects structure as justified the data was by-subject and by-item random intercepts ($p < 2.2 \times 10^{-16}$):

$$\text{outcome} \sim \text{condition} \times \text{trial} + (1|\text{subject}) + (1|\text{item}).$$

We looked at the simple effect of trial on logit odds of producing an accurate response. A positive estimate would indicate that the accuracy increased over successive trials, whereas a negative estimate would indicate the opposite. Because the no-pressure group served as the reference level, the magnitude of the trial effect would pertain to the learning rate in the no-pressure group. The trial effect was significant ($b = 0.003$, $SE = 1.5 \times 10^{-4}$, $z = 20.36$, 95% CI = 0.003, 0.003; $p < 2 \times 10^{-16}$), indicating that accuracy increased over time for the no-pressure group. Next, we looked at the Trial × Condition interaction effects, which would indicate whether the learning rate was different from that of the no-pressure group. The

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

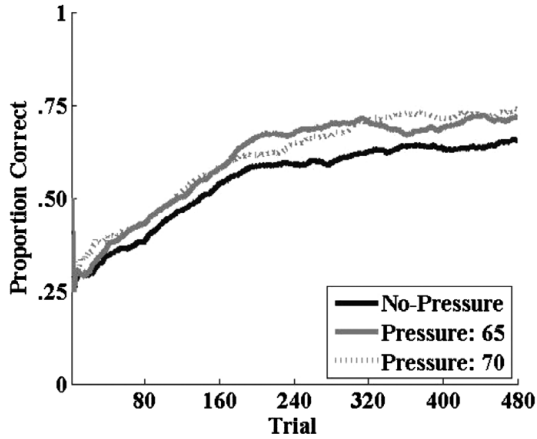


Figure 3. Learning curves across the P70 (dotted line), P65 (solid gray line), and no-pressure control (solid black line) conditions. Each data point on each curve was generated by averaging the accuracy across the previous 80 trials of data (i.e., using a sliding 80-trial uniform kernel). For example, the accuracy rate associated with trial 100 was generated by averaging the accuracy rates on trials 21–100. The only exceptions are the trials prior to trial 80. For those, we used a cumulative average. For example, the accuracy rate associated with trial 60 was generated by averaging the accuracy rates on trials 1–60.

Trial × P70 Condition interaction was positive significant ($b = 7.1 \times 10^{-4}$, $SE = 2.3 \times 10^{-4}$, $z = 3.09$, 95% CI $= 2.64 \times 10^{-4}$, 1.2×10^{-3}; $p = .002$), indicating that the learning rate was higher for the P70 condition than for the no-pressure control condition. The Trial × P65 Condition interaction was not significant ($b = 2.7 \times 10^{-4}$, $SE = 2.3 \times 10^{-4}$, $z = 1.19$, 95% CI $= -1.8 \times 10^{-4}$, 7.3×10^{-4}; $p = 0.233$), suggesting that the learning rate for the P65 condition was not different from the no-pressure control condition. Next, we looked at the condition simple effects, which would indicate the degrees to which accuracies in the P70 and P65 conditions differed from those in the no-pressure condition. Because the trial numbers were mean-centered, the inference would apply to the 240th trial, which is the midpoint in the course of learning. The P70 condition effect was not significant ($b = 0.136$, $SE = 0.281$, $z = 0.483$, 95% CI $= -0.415, 0.686$; $p = .629$), and the P65 condition effect was also not significant ($b = 0.23$, $SE = 0.285$, $z = 0.81$, 95% CI $= -0.327, 0.788$; $p = .418$), suggesting that learning performance in the two pressure groups did not significantly differ with the no-pressure group at the midpoint of the learning task. The intercept, which coded how much the logit odds of producing a correct response in the no-pressure condition at the midpoint in learning differed from 0 (equivalent to 50% probability), was not significant ($b = -0.35$, $SE = 0.201$, $z = -1.74$, 95% CI $= -0.743, 0.043$; $p = .081$). Because the chance level for each trial is 25%, we did not consider the implications of the intercept to be important.

Because the previous analysis did not provide a direct comparison between the two pressure conditions, an additional analysis was performed without the

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

no-pressure condition. The design was identical with trial as a fixed effect and condition as another (P70 vs. P65: reference level). The trial effect was significant ($b = 0.003$, $SE = 1.8 \times 10^{-4}$, $z = 19.22$, 95% CI = 0.003, 0.003; $p < 2 \times 10^{-16}$), replicating the positive learning rate in the P65 condition. The Trial × P70 Condition interaction was marginally significant ($b = 4.4 \times 10^{-4}$, $SE = 2.5 \times 10^{-4}$, $z = 1.77$, 95% CI = -4.6×10^{-5}, 9.2×10^{-4}; $p = .076$), presenting a trend in which the learning rate was higher for the P70 condition than for the P65 condition. The P70 condition effect was not significant ($b = -0.095$, $SE = 0.291$, $z = -0.327$, 95% CI = –0.665, 0.475; $p = .743$), suggesting that the performance did not differ in the midpoint in learning. The intercept was not significant ($b = -0.119$, $SE = 0.221$, $z = -0.537$, 95% CI= –0.551, 0.314; $p = .591$).

Computational modeling

The accuracy-based analyses suggested a performance advantage for individuals under pressure. To investigate the processing locus of this performance pressure advantage, we applied computational modeling techniques that provide the necessary window onto cognitive processing. We predicted that the performance advantage was due to a faster shift from reflective to reflexive strategy use in the pressure conditions.

Computational modeling details. We fitted a series of decision-bound models on a block-by-block basis at the individual participant level because of problems with interpreting fits to aggregate data (Maddox, 1999). We assumed that the two-dimensional space (pitch height vs. pitch direction) displayed in Figure 2a accurately describes the perceptual representation of the stimuli. Based on the results from our earlier work (Chandrasekaran, Koslov, et al., 2014) that showed that participants can categorize by the sex of the speaker (male/female) with greater than 95% accuracy with no feedback, we also assumed that participants applied category learning strategies separately to the male (Figure 2b) and female (Figure 2c) perceptual spaces. Note that as long as the major dimensions are known, these modeling procedures can be applied to any type of speech category structure, so this offers an exciting new approach to the study of speech category learning. As discussed earlier, there are several secondary cues to tone categorization that are not examined here. However, with the goal of understanding the locus of the pressure effect on speech learning, we use the two-dimensional space as a starting point in estimating the perceptual space in nonnative listeners. Several prior studies have shown that this two-dimensional space (pitch height and direction) is capable of explaining significant variance in native English speakers discriminating tone categories (Chandrasekaran et al., 2007, 2010; Francis et al., 2008).

Here we provide a brief description of each model, as well as an interpretation of the model results. Details are available in numerous previous publications (e.g., Chandrasekaran, Koslov, et al., 2014). Each model assumes that decision bounds were used to classify stimuli into each of the four Mandarin tone categories (T1, T2, T3, or T4). We applied three classes of models: reflexive, reflective, and random responding. The first class is computational models of the reflexive

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

procedural learning system. This is instantiated with the Striatal Pattern Classifier (SPC; Ashby & Waldron, 1999; Maddox, Molis, & Diehl, 2002). The SPC is a computational model whose processing is consistent with what is known about the neurobiology of the procedural-based category learning system thought to underlie reflexive-optimal classification performance (Ashby et al., 1998; Ashby & Ennis, 2006; Maddox et al., 2002; Nomura et al., 2007; Seger & Cincotta, 2005; Yi et al., 2014). The second class is reflective, rule-based, and instantiate hypothesis-testing strategies, such as the application of unidimensional or conjunctive rules. These are verbalizable strategies. The third model is a random responder model that assumes that the participant guesses on each trial. The models were fit to the Mandarin tone category learning data from each trial by maximizing negative log-likelihood. In brief, the goal of maximum likelihood parameter estimation is to identify the set of model parameter for which the observed probability of responding tone categories T1–T4 across a set of trials is as similar as possible to the predicted probability of responding tone categories T1–T4 from the model. To determine the best fitting model, the maximum negative log-likelihood value was converted into an Akaike information criterion (AIC) value (Akaike, 1974). For each model i, the AIC is defined as

$$\mathrm{AIC}_i = 2\ln L_i + 2V_i,$$

where L_i is the maximum likelihood for model i and V_i is the number of free parameters in the model. Note that AIC penalizes models with more free parameters. Smaller AIC values indicate a better fit to the data. The best fitting model was defined as the model with the smallest AIC value. Another goodness of fit measure that is based on maximum likelihood is called the Bayesian information criterion. We replicated all of the model fits with the Bayesian information criterion, and none of the patterns in the data changed. Thus, we focus the presentation below on AIC. We provide the specifics of each model in the next section.

SPC. The SPC assumes that stimuli are represented perceptually in higher level auditory areas, such as the superior temporal gyrus. Because of the massive many-to-one (~10,000:1) convergence of afferents from the primary and secondary sensory cortices to the striatum (Ashby & Ennis, 2006; Wilson, 1995), a low-resolution map of the perceptual space is represented among the *striatal units*. Within the auditory domain, it is well known that there are direct projections from secondary auditory areas such as the superior temporal gyrus and supratemporal plane to the caudate (Arnauld, Jeantet, Arsaut, & Desmotes-Mainard, 1996; Hikosaka, Sakamoto, & Usui, 1989; Yeterian & Pandya, 1998). During feedback-based learning, the striatal units become associated with one of the category labels so that, after learning is complete, a category response label is associated with each of a number of different regions of perceptual space. In effect, the striatum learns to associate a response with clumps of cells in the auditory cortex. In fitting behavioral data we utilize a computational version of the SPC that is inspired by what is known about the neurobiology of the striatum. The version of the SPC that we fit to data assumes that there is one striatal "unit" in the pitch height–pitch

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

direction space for each category, and a single "noise" parameter that represents the noise associated with the placement of the striatal units. Stimuli are assigned probabilistically (because of the noise in the striatal unit placement) into the category with the nearest striatal unit. Responses from a hypothetical participant using the SPC are displayed in Figure 2d.

Conjunctive rule-based model. Rule-based models assume that the participant generates hypotheses regarding category assignment. Rule generation, selection, storage, and application recruit a network of brain regions including the anterior cingulate, head of the caudate nucleus, and the prefrontal cortex (Ashby et al., 1998; Chandrasekaran, Koslov, et al., 2014). A conjunctive rule-based model that assumes that the participant sets two criteria along the pitch direction dimension and one criterion along the pitch height dimension was applied to the data. The model assumes that the two criteria along the pitch direction dimension are used to separate the stimuli into those that are of low, medium, or high pitch direction. Low pitch direction items are classified into tone category 4 (T4), and high pitch direction items are classified into tone category 2 (T2). If an item is classified as having a medium pitch direction, then the pitch height dimension is examined. The single criterion along the pitch height dimension is used to separate the stimuli into low and high pitch height. Stimuli that have medium pitch direction and low pitch height are classified into tone category 3 (T3), and medium pitch direction items of high pitch height are classified into tone category 1 (T1). Responses from a hypothetical participant using a conjunctive strategy are displayed in Figure 2e.

Unidimensional rule-based model. A unidimensional height rule-based model that assumes that the participant sets three criteria along the pitch height dimension was also applied to the data. The model assumes that the three criteria along the pitch height dimension are used to separate the stimuli into those that are of low, medium-low, medium-high, or high pitch height, with each of these being associated with one of the four tone categories. Note that this model completely ignores the pitch direction dimension. Although 24 versions of the model are possible given four category labels, some are highly unrealistic (e.g., a model that assumes that tone category 1 [T1] was the lowest in pitch height). We examined the 8 most reasonable variants of the model. A unidimensional direction rule-based model that assumes that the participant sets three criteria along the pitch direction dimension was also applied to the data. The model assumes that the three criteria along the pitch direction dimension are used to separate the stimuli into those that are of low, medium-low, medium-high, or high pitch direction with each of these being associated with one of the tone categories. Note that this model completely ignores the pitch height dimension. We examined the two most reasonable variants of the model. Responses from a hypothetical participant using a unidimensional strategy along pitch height are displayed in Figure 2f, and responses from a hypothetical participant using a unidimensional strategy along pitch direction are displayed in Figure 2g.

Random responder model. The random responder model assumes a fixed probability of responding Tone 1, Tone 2, Tone 3, and Tone 4 but allows for response

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

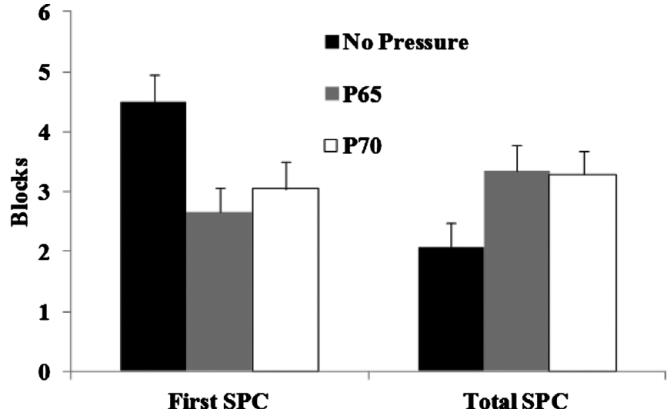


Figure 4. (Left) Block during which no-pressure control, P65, and P70 participants first used a strategy consistent with reflexive processing (i.e., the SPC). (Right) Total number of blocks during which no-pressure control, P65, and P70 participants used a strategy consistent with reflexive processing. Standard error bars included. Confidence intervals (95%): first SPC block: no-pressure control = 3.681, 5.319; P65 = 1.716, 3.584; and P70 = 2.136, 3.959. Confidence intervals (95%): total SPC block: no-pressure control = 1.308, 2.846; P65 = 2.473, 4.227; and P70 = 2.430, 4.142.

biases. The model has three free parameters to denote the predicted probability of responding "1," "2," or "3" with the probability of responding "4" equal to one minus the sum for the other three categories.

Computational modeling results. The accuracy-based analyses provide support for the prediction that performance pressure leads to enhanced speech learning. We hypothesize that this is because pressure leads to enhanced reflexive processing. We argue that this can come in two forms. First, pressure should enhance reflexive processing by speeding the transition from reflective to reflexive processing. As a direct test of this hypothesis, we determined the first block of trials for which the SPC (a model of reflexive processing) provided the best fit of the data. If performance pressure speeds the transition from reflective to reflexive processing, then the first block of trials for which the SPC fits best should occur sooner in learning (i.e., during an earlier block of trials) than in the no-pressure control condition. Second, once the transition from reflective to reflexive processing occurs, reflexive processing should remain stable. As a direct test of this hypothesis, we determined the total number of block of trials for which the SPC provided the best fit of the data. If individuals under performance pressure use reflexive strategies sooner and with more regularity, then they should use reflexive strategies in more blocks of trials than individuals in the no-pressure control condition. Data from these two measures are displayed in Figure 4 (with 95% confidence intervals).

Comparing across all three conditions, a one-way analysis of variance on the first block of trials best fit by the SPC revealed a significant effect of condition, $F(2, 64) = 5.114$, $p = .009$, partial $\eta^2 = 0.138$. Post hoc tests using a Bonferroni

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

correction for multiple comparisons suggested that a reflexive strategy: (a) was used significantly sooner ($p = .012$) in the P65 condition (mean = 2.65) than in the no-pressure control condition (mean = 4.50), $t\,(44) = 2.954$, $p = .005$, partial $\eta^2 = 0.166$; (b) was used marginally significantly sooner ($p = .063$) in the P70 condition (mean = 3.05) than in the no-pressure control condition; and (c) showed no difference between the P65 and P70 conditions in first block best fit by the SPC ($p = 1.00$). A comparison of the no-pressure control with the aggregate of the two pressure conditions was significant, $F\,(1, 65) = 9.953$, $p = .002$, partial $\eta^2 = 0.133$. Next, we conducted a one-way analysis of variance on the total number of blocks best fit by the SPC and found a significant effect of condition, $F\,(2, 64) = 3.180$, $p = .048$, partial $\eta^2 = 0.090$. Post hoc tests using a Bonferroni correction for multiple comparisons suggested that a reflexive strategy: (a) was used marginally significantly more often ($p = .099$) in the P65 condition (mean = 3.35) than in the no-pressure control condition (mean = 2.08); (b) was not used significantly more often ($p = .119$) in the P70 condition (mean = 3.29) than in the no-pressure control condition; and (c) showed no difference between the P65 and P70 conditions in total number of SPC blocks ($p = 1.00$). A comparison of the no-pressure control with the aggregate of the two pressure conditions was significant, $F\,(1, 65) = 6.447$, $p = .014$, partial $\eta^2 = 0.090$. These findings establish that individuals under performance pressure use reflexive strategies sooner and with greater regularity than individuals under no-pressure conditions but that there are no performance differences between the two pressure conditions.

GENERAL DISCUSSION

This is the first study to examine the effects of performance pressure on nonnative speech category learning. Within a DLS framework, and based on a growing body of evidence (Chandrasekaran, Koslov, et al., 2014; Chandrasekaran, Yi, et al., 2014; Maddox & Chandrasekaran, 2014; Maddox et al., 2013, 2014; Yi et al., 2014), we posited that nonnative speech (Mandarin tone) category learning is mediated by an automatic, procedural reflexive system. A successful theory of the effects of pressure on cognitive processing is the distraction hypothesis (Beilock et al., 2004; Beilock & Carr, 2005; Markman et al., 2006). The distraction hypothesis suggests that performance pressure shifts the balance between reflective and reflexive processing toward reflexive processing by reducing available executive function processes, and thus the effectiveness of reflective processing. The distraction hypothesis predicts that pressure should attenuate reflective-optimal (rule-based) learning but should enhance reflexive-optimal (information-integration) learning. This pattern has been observed with experimenter-defined category structures and simple perceptual stimuli presented in the visual domain (e.g., Gabor patches that vary in frequency and orientation; DeCaro et al., 2011; Markman et al., 2006; McCoy et al., 2014). Support for the distraction hypothesis in the visual category learning literature, along with recent evidence to suggest that nonnative speech learning is mediated by the reflexive system, led us to predict that performance pressure would *enhance* nonnative speech (Mandarin tone) category learning through early and more frequent use of reflexive strategies. As predicted, we found an advantage in Mandarin tone category learning for individuals under

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

performance pressure. Neurobiologically motivated computational model-based analyses suggested that the locus of this performance advantage was due to faster, and more frequent, use of reflexive-processing strategies. The implications of this work are many. We briefly discuss some of these below.

Individual differences in learning success

Laboratory speech training research often attributes the large individual differences in learning success to perceptual abilities, genetics, cognitive factors, and listening experience (Chandrasekaran et al., 2015; Perrachione et al., 2011). We demonstrate that situational factors, such as performance pressure, can also significantly alter learning strategies and, consequently, performance, contributing to the individual differences in speech learning success. The proposed role of situational factors may explain how a videogame paradigm could achieve robust nonnative speech learning comparable with a much longer multiweek laboratory-based training paradigm (Lim & Holt, 2011), which may reflect a combination of greater performance pressure, incentive to learn, mediating faster shift to optimal procedural-based learning strategies.

Although we found robust effects of pressure on performance, we found few performance differences between the P65 and P70 conditions. This was not unexpected. We chose these two values because average performance in our previous work reached asymptote at 65%–70% accuracy (Chandrasekaran, Koslov, et al., 2014). Even so, performance in the P70 condition was slightly worse than that in the P65 condition. Performance pressure influences behavior when the target performance level is not too easy and not too hard (Maddox & Markman, 2010; Markman et al., 2006). It is likely that the P65 condition satisfies these conditions slightly better than the P70 condition, leading to slightly better performance in the P65 condition. More work is needed, though, that explores a broader range of performance levels.

Other manipulations of performance pressure and performance

Although the present study is the first to examine the effects of pressure on speech category learning, it represents one of several that have explored the effects of pressure and stress on the balance of processing in the reflexive and reflective systems (DeCaro et al., 2011; Markman et al., 2006; McCoy et al., 2014). Recent work by DeCaro et al. (2011) and McCoy et al. (2014) suggests that the nature of the performance pressure manipulation has a large effect on the balance of reflective and reflexive processing. In particular, whereas they found that the performance pressure manipulation used here does lead to enhanced reflexive processing, they found that other pressure manipulations, in particular those that enhance explicit monitoring, were more likely to enhance reflective processing. Thus, it is important to be clear that the effects of pressure are complex and depend upon the specifics of the pressure manipulation as well as the nature of the learning system that optimizes performance. Future work should examine a broader range of pressure manipulations and explore their effects on Mandarin tone learning as well as on other speech category learning domains.

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

Implications for education

Although the results presented in this research report are intriguing, it is highly premature at this stage to make any strong claims about the implications of this work for educational settings. The present study used a well-vetted laboratory manipulation of pressure and a highly studied form of nonnative speech learning (Mandarin tone learning) for which strong evidence exists to suggest that learning is optimally mediated by the reflexive system. Other pressure manipulations can have different effects on system-level processing, and other learning tasks (including other forms of speech learning) may not be so simply characterized as reflexive-optimal or reflective-optimal. Even so, with more basic and applied research, we may be able (in the future) to identify conditions under which certain forms of pressure (even high stakes testing) are advantageous.

Individual differences in trait variables

The research presented here completely ignores personality and trait factors of the individual. Clearly these factors influence how individuals approach a task as well as their inherent balance of reflexive and reflective processing. We have begun to explore this in our speech learning work with some interesting consequences. As just one example, we argued that the presence of elevated depressive symptoms leads to a shift in balance toward the reflexive system. Thus, just as we predicted for performance pressure, we predicted, and found support for the claim, that elevated depressive symptoms should lead to enhanced speech learning (Maddox et al., 2014). Future research should extend this to other populations that differ in personality traits and should incorporate situational manipulations such as performance pressure (Maddox & Markman, 2010).

Pressure and motivation

Early in this paper we stated that performance pressure affects "motivation" but that we wanted to avoid that term and instead focus on performance pressure. In this section, we want to briefly discuss the implications of our work for motivation. Motivation, broadly defined, is critical to nearly all behavior. It follows that motivation is highly complex and as such is not well understood. We argue that pressure is a form of "global" situational motivation. Previous research suggests that a person's global motivational focus interacts with the local trial-by-trial motivation (Higgins, 2000; Maddox & Markman, 2010). For example, when the global motivation is to approach something positive (such as a monetary bonus) and the trial-by-trial motivation is to obtain positive things (such as points), there is a regulatory match. Previous research suggests that this leads to enhanced processing in the reflective system (Maddox & Markman, 2010). However, if the global motivation is to avoid losing something desirable (such as an already obtained bonus, or disappointing a partner who did their job) and the trial-by-trial motivation is to obtain points, there is a regulatory mismatch that leads to enhanced processing in the reflexive system (Maddox & Markman, 2010). Maddox and Markman argue that pressure represents a mixture of a global situational approach motivation

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

(to obtain a monetary bonus) with a global social avoidance motivation (not to disappoint the partner). Future work should disentangle these global motivational effects and examine how they interact with local motivational effects. This would lead to a more complete understanding of the broader motivational framework and its effects on speech learning.

Motivation and second language learning

Seminal work by Gardner (1985, 2007) and others (Dornyei, 1994; Dornyei & Otto, 1998) have examined the relationship between motivation and second language learning. For example, Gardner (2007) posits four stages of language acquisition (elemental, consolidation, conscious expression, and automaticity) and two types of motivational processes (language learning motivation and classroom motivation). Although a detailed discussion of this important work is beyond the scope of this paper, some integration of our results into this broader framework is possible. Gardner argues for the importance of motivation in second language learning and states explicitly that its effect is more complex than simply *wanting* to learn the language. Our work examining the effects of pressure, and the broader motivation framework incorporating global and local aspects of motivation outlined above, clearly supports this claim. In our view, "wanting to learn" something implies the goal of "trying harder." Trying harder most likely involves conscious effort and utilization of working memory and executive attention; in other words, reflective learning processes. Our study suggests that the form of social pressure utilized here shifts the balance of processing *away* from reflective toward reflexive processing. Future work should explore how pressure in general, and social pressure in particular, influences the various stages of language acquisition. It is possible that the effects are quite different if pressure is introduced at the elemental stage versus the conscious expression stage.

Conclusions

This represents the first study to examine the effects of performance pressure on speech category learning. As predicted from our application of DLS theory to speech category learning and from the extensive literature suggesting that pressure enhances reflexive processing, we predicted that performance pressure would enhance speech category learning. In support of our prediction, we found that pressure enhanced speech category learning with the computational modeling, suggesting that this performance advantage was due to faster and more frequent use of reflexive category learning strategies.

NOTES

1. Recent evidence suggests that a third system, referred to as the perceptual-representation system, can also mediate category learning under certain conditions (Casale & Ashby, 2008; Zeithamova et al., 2008).
2. Although the control and pressure procedures outlined here are standard in the field, it is the case that the pressure condition includes a monetary bonus for good performance

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

that is not included in the control condition. In a small-scale pilot study, we collected additional data from the control condition ($N = 10$) and a control condition that included a monetary bonus for obtaining 70% correct ($N = 21$). We found a small 1% performance advantage in the control with bonus condition, whereas the P70 condition advantage was 6% (presented in the Results). This suggests that the presence of an incentive is not driving the performance advantage in the pressure conditions.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *Transactions on Automatic Control, 19*, 716–723.

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120*, 3–19.

Arnauld, E., Jeantet, Y., Arsaut, J., & Desmotes-Mainard, J. (1996). Involvement of the caudal striatum in auditory processing: c-fos response to cortical application of picrotoxin and to auditory stimulation. *Molecular Brain Research, 41*, 27–35.

Aron, A. R., Shohamy, D., Clark, J., Myers, C., Gluck, M. A., & Poldrack, R. A. (2004). Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *Journal of Neurophysiology, 92*, 1144–1152.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442–481.

Ashby, F. G., & Crossley, M. J. (2011). A computational model of how cholinergic interneurons protect striatal-dependent learning. *Journal of Cognitive Neuroscience, 23*, 1549–1566.

Ashby, F. G., Ell, S. W., Valentin, V. V., & Casale, M. B. (2005). FROST: A distributed neurocomputational model of working memory maintenance. *Journal of Cognitive Neuroscience, 17*, 1728–1743.

Ashby, F. G., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. In B. H. Ross & A. Markman (Eds.), *Psychology of learning and motivation: Vol. 47. Categories in use* (pp. 1–36). New York: Elsevier.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56*, 149–178.

Ashby, F. G., & Maddox, W. T. (2010). Human category learning 2.0. *Annals of the New York Academy of Sciences, 1224*, 147–161.

Ashby, F. G., Paul, E., & Maddox, W. T. (2011). COVIS 2.0. In E. P. A. Wills (Ed.), *Formal approaches in categorization*. Cambridge: Cambridge University Press.

Ashby, F. G., & Valentin, V. V. (2005). Multiple systems of perceptual category learning: Theory and cognitive tests. In H. Cohen & C. Lefebvre (Eds.), *Categorization in cognitive science*. New York: Elsevier.

Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin and Review, 6*, 363–378.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes [Computer software]. Retrieved from http://cran.R-project.org/package=lme4

Beilock, S. L., Bertenthal, B. I., McCoy, A. M., & Carr, T. H. (2004). Haste does not always make waste: Expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills. *Psychonomic Bulletin and Review, 11*, 373–379.

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Science, 16*, 101–105.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn & M. J. Munro (Eds.), *Language experience in second*

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

language speech learning: In honor of James Emil Flege (pp. 13–34). Amsterdam: John Benjamins.

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer (Version 5.2 18) [Computer software]. Retrieved from http://www.fon.hum.uva.nl/praat/

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*, 977–985.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*, 707–729.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*, 2299–2310.

Brooks, L. (1978). *Nonanalytic concept formation and memory for instances*. Hillsdale, NJ: Erlbaum.

Casale, M. B., & Ashby, F. G. (2008). A role for the perceptual representation memory system in category learning. *Perception & Psychophysics, 70*, 983–999.

Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience, 25*, 195–210.

Chandrasekaran, B., Koslov, S. R., & Maddox, W. T. (2014). Toward a dual-learning systems model of speech category learning. *Frontiers in Psychology, 5*, 825.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America, 128*, 456–465.

Chandrasekaran, B., Yi, H. G., Blanco, N. J., McGeary, J. E., & Maddox, W. T. (2015). Enhanced procedural learning of speech sound categories in a genetic variant of FOXP2. *Journal of Neuroscience, 35*, 7808–7812.

Chandrasekaran, B., Yi, H. G., & Maddox, W. T. (2014). Dual-learning systems during speech category learning. *Psychonomic Bulletin and Review, 21*, 488–495.

Cooper, J. A., Worthy, D. A., Gorlick, M. A., & Maddox, W. T. (2013). Scaffolding across the lifespan in history-dependent decision-making. *Psychology of Aging, 28*, 505–514.

DeCaro, M. S., Thomas, R. D., Albert, N. B., & Beilock, S. L. (2011). Choking under pressure: Multiple routes to skill failure. *Journal of Experimental Psychology: General, 140*, 390–406.

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition, 107*, 284–294.

Dornyei, Z. (1994). Motivation and motivating in the foreign language classroom. *Modern Language Journal, 78*, 273–284.

Dornyei, Z., & Otto, I. (1998). Motivation in action: A process model of L2 motivation. *Working Papers in Applied Linguistics, 4*, 43–69.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 127*, 107–140.

Filoteo, J. V., Lauritzen, J. S., & Maddox, W. T. (2010). Removing the frontal lobes: The effects of engaging executive functions on perceptual category learning. *Psychological Science, 21*, 415–423.

Filoteo, J. V., Maddox, W. T., Salmon, D. P., & Song, D. D. (2005). Information-integration category learning in patients with striatal dysfunction. *Neuropsychology, 19*, 212–222.

Filoteo, J. V., Maddox, W. T., Simmons, A. N., Ing, A. D., Cagigas, X. E., Matthews, S., et al. (2005). Cortical and subcortical brain regions involved in rule-based category learning. *NeuroReport, 16*, 111–115.

Folstein, J. R., & Van Petten, C. (2004). Multidimensional rule, unidimensional rule, and similarity strategies in categorization: Event-related brain potential correlates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1026–1044.

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36*, 268–294.

Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech, 21*, 1–33.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.

Gardner, R. C. (2007). Motivation and second language acquisition. *Porta Linguarum, 8*, 9–20.

Grimm, L. R., Markman, A. B., & Maddox, W. T. (2012). End-of-semester syndrome: How situational regulatory fit affects test performance over an academic semester. *Basic and Applied Social Psychology, 34*, 376–385.

Grimm, L. R., Markman, A. B., Maddox, W. T., & Baldwin, G. C. (2009). Stereotype threat reinterpreted as a regulatory mismatch. *Journal of Personality and Social Psychology, 96*, 288–304.

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist, 52*, 1280–1300.

Higgins, E. T. (2000). Making a good decision: Value from fit. *American Psychologist, 55*, 1217–1230.

Hikosaka, O., Sakamoto, Y., & Usui, S. (1989). Functional properties of monkey caudate neruons: III. Activities related to expectation of target and reward. *Journal of Neurophysiology, 61*, 814–832.

Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science, 17*, 42–46.

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, and Psychophysics, 72*, 1218–1227.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*, B47–B57.

Keri, S. (2003). The cognitive neuroscience of category learning. *Brain Research Reviews, 43*, 85–109.

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category level knowledge. *Science, 262*, 1747–1749.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning and Memory, 1*, 106–120.

Lee, C. Y., & Hung, T. H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *Journal of the Acoustical Society of America, 124*, 3235–3248.

Lim, S. J., & Holt, L. L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science, 35*, 1390–1405.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96*, 2076–2087.

Lombardi, W. J., Andreason, P. J., Sirocco, K. Y., Rio, D. E., Gross, R. E., Umhau, J. C., et al. (1999). Wisconsin Card Sorting Test performance following head injury: Dorsolateral fronto-striatal circuit activity predicts perseveration. *Journal of Clinical and Experimental Neuropsychology, 21*, 2–16.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*, 309–332.

Maddox, W. T. (1999). On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization. *Perception & Psychophysics, 61*, 354–375.

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes, 66*, 309–332.

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cogniion, 29*, 650–662.

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory and Cognition, 32*, 582–591.

Maddox, W. T., & Chandrasekaran, B. (2014). Tests of a dual-systems model of speech category learning. *Bilingualism (Cambridge, England), 17*, 709–728.

Maddox, W. T., Chandrasekaran, B., Smayda, K., & Yi, H. G. (2013). Dual systems of speech category learning across the lifespan. *Psychology and Aging, 28*, 1042–1056.

Maddox, W. T., Chandrasekaran, B., Smayda, K., Yi, H. G., Koslov, S., & Beevers, C. G. (2014). Elevated depressive symptoms enhance reflexive but not reflective auditory category learning. *Cortex, 58*, 186–198.

Maddox, W. T., & Filoteo, J. V. (2001). Striatal contributions to category learning: Quantitative modeling of simple linear and complex nonlinear rule learning in patients with Parkinson's disease. *Journal of the International Neuropsychological Society, 7*, 710–727.

Maddox, W. T., & Filoteo, J. V. (2005). The neuropsychology of perceptual category learning. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 573–599). New York: Elsevier.

Maddox, W. T., & Ing, A. D. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 100–107.

Maddox, W. T., Ing, A. D., & Lauritzen, J. S. (2006). Stimulus modality interacts with category structure in perceptual category learning. *Perception & Psychophysics, 68*, 1176–1190.

Maddox, W. T., Love, B. C., Glass, B. D., & Filoteo, J. V. (2008). When more is less: Feedback effects in perceptual category learning. *Cognition, 108*, 578–589.

Maddox, W. T., & Markman, A. B. (2010). The motivation–cognition interface in learning and decision-making. *Current Directions in Psychological Science, 19*, 106–110.

Maddox, W. T., Molis, M. R., & Diehl, R. L. (2002). Generalizing a neuropsychological model of visual categorization to auditory categorization of vowels. *Perception & Psychophysics, 64*, 584–597.

Markman, A. B., Maddox, W. T., & Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science, 17*, 944–948.

McCoy, S. K., Hutchinson, S., Hawthorne, L., Cosley, B. J., & Ell, S. W. (2014). Is pressure stressful? The impact of pressure on the stress response and category learning. *Cognitive, Affective, and Behavioral Neuroscience, 14*, 769–781.

Monchi, O., Petrides, M., Petre, V., Worsley, K., & Dagher, A. (2001). Wisconsin Card Sorting revisited: Distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *Journal of Neuroscience, 21*, 7733–7741.

Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., et al. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex, 17*, 37–43.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). A rule-plus-exception model of classification learning. *Psychological Review, 101*, 53–79.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America, 130*, 461–472.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature, 414*, 546–550.

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience & Biobehavioral Reviews, 32*, 197–205.

Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2003). Dissociating explicit and implicit category knowledge with fMRI. *Journal of Cognitive Neuroscience, 15*, 574–583.

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

Regehr, G., & Brooks, L. R. (1993). Perceptual manifestations of an analytic structure: The priority of holistic individuation. *Journal of Experimental Psychology: General, 122*, 92–114.

Schnyer, D. M., Maddox, W. T., Ell, S., Davis, S., Pacheco, J., & Verfaellie, M. (2009). Prefrontal contributions to rule-based and information-integration category learning. *Neuropsychologia, 47*, 2995–3006.

Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neuroscience & Biobehavioral Reviews, 32*, 265–278.

Seger, C. A., & Cincotta, C. M. (2002). Striatal activity in concept learning. *Cognitive, Affective, and Behavioral Neuroscience, 2*, 149–161.

Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *Journal of Neuroscience, 25*, 2941–2951.

Seger, C. A., & Cincotta, C. M. (2006). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cerebral Cortex, 16*, 1546–1555.

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience, 33*, 203–219.

Shohamy, D., Myers, C. E., Grossman, S., Sage, J., Gluck, M. A., & Poldrack, R. A. (2004). Cortico-striatal contributions to feedback-based learning: Converging data from neuroimaging and neuropsychology. *Brain: A Journal of Neurology, 127*, 851–859.

Shohamy, D., Myers, C. E., Onlaor, S., & Gluck, M. A. (2004). Role of the basal ganglia in category learning: How do patients eith Parkinson's disease learn? *Behavioral Neuroscience, 118*, 676–686.

Smayda, K. E., Chandrasekaran, B., & Maddox, W. T. (2015). Enhanced cognitive and perceptual processing: A computational basis for the musician advantage in speech learning. *Frontiers in Psychology, 6*, 682.

Smith, J. D., Berg, M. E., Cook, R. G., Murphy, M. S., Crossley, M. J., Boomer, J., et al. (2012). Implicit and explicit categorization: A tale of four species. *Neuroscience & Biobehavioral Reviews, 36*, 2355–2369.

Smith, J. D., Johnston, J. J., Musgrave, R. D., Zakrzewski, A. C., Boomer, J., Church, B. A., et al. (2014). Cross-modal information integration in category learning. *Attention, Perception, and Psychophysics, 76*, 1473–1484.

Tharp, I. J., & Pickering, A. D. (2008). A note on DeCaro, Thomas, and Beilock (2008): Further data demonstrate complexities in the assessment of information-integration category learning. *Cognition, 111*, 411–415.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin and Review, 8*, 168–176.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America, 113*, 1033–1043.

Wang, Y., Sereno, J. A., Jongman, A., & Hirsch, J. (2003). fMRI evidence for cortical modification during learning of Mandarin lexical tone. *Journal of Cognitive Neuroscience, 15*, 1019–1027.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106*, 3649–3658.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49*, 25–47.

Wilson, C. J. (1995). *The contribution of cortical neurons to the firing pattern of striatal spiny neurons*. Cambridge, MA: MIT Press.

Wong, P., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics, 28*, 565–585.

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press

Worthy, D. A., Markman, A. B., & Maddox, W. T. (2009a). Choking and excelling under pressure in experienced classifiers. *Attention, Perception, and Psychophysics, 71*, 924–935.

Worthy, D. A., Markman, A. B., & Maddox, W. T. (2009b). What is pressure? Evidence for social pressure as a type of regulatory focus. *Psychonomic Bulletin and Review, 16*, 344–349.

Worthy, D. A., Markman, A. B., & Maddox, W. T. (2013). Feedback and stimulus-offset timing effects in perceptual category learning. *Brain and Cognition, 81*, 283–293.

Yeterian, E. H., & Pandya, D. N. (1998). Corticostriatal connections of the superior temporal region in rhesus monkeys. *Journal of Comparative Neurology, 399*, 384–402.

Yi, H. G., Maddox, W. T., Mumford, J. A., & Chandrasekaran, B. (2014). The role of corticostriatal systems in speech category learning. *Cerebral Cortex*. Advance online publication.

Zeithamova, D., & Maddox, W. T. (2006). Dual task interference in perceptual category learning. *Memory and Cognition, 34*, 387–398.

Zeithamova, D., & Maddox, W. T. (2007). The role of visuo-spatial and verbal working memory in perceptual category learning. *Memory and Cognition, 35*, 1380–1398.

Zeithamova, D., Maddox, W. T., & Schnyer, D. M. (2008). Dissociable prototype learning systems: Evidence from brain imaging and behavior. *Journal of Neuroscience, 28*, 13194–13201.

https://doi.org/10.1017/S0142716415000600 Published online by Cambridge University Press