

Symposia

Non-sensory Influences on Auditory Learning and Plasticity

MELISSA L. CARAS¹ , MAX F. K. HAPPEL^{2,3,4}, BHARATH CHANDRASEKARAN⁵, PABLO RIPOLLÉS⁶, SARAH M. KEESOM⁷, LAURA M. HURLEY⁸, LUKE REMAGE-HEALEY⁹, LORI L. HOLT¹⁰, AND BEVERLY A. WRIGHT¹¹

¹ Department of Biology, University of Maryland at College Park, College Park, MD 20742, USA

² Department of Systems Physiology of Learning, Leibniz Institute for Neurobiology, 39118 Magdeburg, Germany

³ The Center for Behavioral Brain Sciences (CBBS), 39106 Magdeburg, Germany

⁴ MSB Medical School Berlin, 14197 Berlin, Germany

⁵ Department of Communication Science and Disorders, University of Pittsburgh, Pittsburgh, PA 15260, USA

⁶ Department of Psychology, Music and Audio Research Lab (MARL) and Center for Language, Music and Emotion (CLaME), New York University, New York, NY 10003, USA

⁷ Department of Biology, Utica College, Utica, NY 13502, USA

⁸ Department of Biology, Indiana University Bloomington, Bloomington, IN 47405, USA

⁹ Department of Psychological and Brain Sciences and the Center for Neuroendocrine Studies, University of Massachusetts Amherst, Amherst, MA 01003, USA

¹⁰ Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

¹¹ Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL 60208, USA

Received: 10 August 2021; accepted: 30 December 2021; Online publication: 2 March 2022

ABSTRACT

Distinguishing between regular and irregular heartbeats, conversing with speakers of different accents, and tuning a guitar—all rely on some form of auditory learning. What drives these experience-dependent changes? A growing body of evidence suggests an important role for non-sensory influences, including reward, task engagement, and social or linguistic context. This review is a collection of contributions that highlight how these non-sensory factors shape auditory plasticity and learning at the molecular, physiological, and behavioral level. We begin by presenting evidence that reward signals from the dopaminergic midbrain act on cortico-subcortical networks to shape sound-evoked responses of auditory cortical neurons, facilitate auditory category learning, and modulate the long-term storage of new words and their meanings. We then discuss the role of task engagement

in auditory perceptual learning and suggest that plasticity in top-down cortical networks mediates learning-related improvements in auditory cortical and perceptual sensitivity. Finally, we present data that illustrates how social experience impacts sound-evoked activity in the auditory midbrain and forebrain and how the linguistic environment rapidly shapes speech perception. These findings, which are derived from both human and animal models, suggest that non-sensory influences are important regulators of auditory learning and plasticity and are often implemented by shared neural substrates. Application of these principles could improve clinical training strategies and inform the development of treatments that enhance auditory learning in individuals with communication disorders.

Keywords: Top-down, Experience, Reward, Task engagement, Social context, Language

Correspondence to: Melissa L. Caras · Department of Biology · University of Maryland at College Park · College Park, MD, 20742, USA. email: mcaras@umd.edu

INTRODUCTION

Auditory learning takes many forms. We routinely acquire new words, generate associations between specific sounds and behavioral outcomes, and adapt to new acoustic or linguistic environments. We can even (with practice) improve our ability to detect sounds or discriminate subtle differences in sound features. Auditory learning is the behavioral manifestation of neural plasticity—defined as experience-dependent changes in neural circuits. Of broad interest here are the inputs to the auditory system that enable auditory learning and plasticity to occur. One clear contributor is the bottom-up sensory input itself. This review instead focuses on the contribution of top-down/non-sensory inputs from outside the auditory system. We present a compilation of research summaries that highlights the contributions of three of these non-sensory factors: reward, task engagement, and social or linguistic context. These summaries are based on work presented at a symposium at the Annual Mid-Winter Meeting of the Association for Research in Otolaryngology in February 2021.

This collection opens with three contributions that focus on the role of reward on auditory learning and plasticity. Max Happel explores the influence of reward on auditory plasticity by probing the impact of dopamine—a neurotransmitter implicated in reward signaling—on auditory cortical physiology. *In vivo* recordings and optogenetic manipulations in the Mongolian gerbil reveal that dopamine modulates the activity of descending projections from the auditory cortex to the auditory thalamus, which in turn shapes bottom-up sensory representations and auditory perception. In a related vein, Bharath Chandrasekaran studies the role of reward provided by feedback on performance in auditory category learning. Functional magnetic resonance imaging (fMRI) in humans implicates descending projections from the auditory cortex to the striatum in this process. Pablo Ripollés examines the contribution of intrinsically generated rewards in new word learning. Pharmacological manipulations and fMRI in humans highlight the interaction of dopamine and descending inputs from the cortex to the hippocampus in facilitating the formation of long-term memories for newly acquired words.

The next two contributions examine the influence of task engagement on auditory learning and plasticity. Beverly Wright probes the contribution of task engagement to auditory perceptual learning in humans. Interleaving periods of practice (a process that engages both bottom-up and top-down inputs) and stimulus exposure alone (a process that primarily engages bottom-up inputs) reveal that top-down and bottom-up influences can interact across time to generate learning. On a similar note, Melissa Caras explores the contribution of task engagement to auditory perceptual learning in a non-human animal model. *In vivo* recordings of

auditory cortical activity during periods of practice and stimulus exposure alone suggest that practice strengthens the top-down modulation of auditory cortical activity, leading to gradual enhancements in signal detection and, ultimately, perceptual learning.

The final three summaries address the impact of social and linguistic context on auditory learning and plasticity. Laura Hurley and Sarah Keesom examine how early social isolation alters auditory processing by focusing on the influence of serotonin—a neurotransmitter implicated in signaling social context—on auditory midbrain function. Voltammetry and immunohistochemistry experiments reveal that early social isolation alters the dynamics of serotonin release and enhances the response of auditory midbrain neurons to serotonin. Luke Ramage-Healey examines the influence of estradiol—a hormone known to signal social and reproductive context—on auditory forebrain physiology and auditory learning. Electrophysiological, behavioral, and pharmacological experiments in zebra finches reveal that brain-derived estrogens act like classical neuromodulators to shape sound encoding and learning. Finally, Lori Holt explores how linguistic context affects speech-perception learning. Behavioral experiments in humans show that both long-term linguistic experience and short-term manipulations of speech statistics shape perception of accented speech.

In sum, this collection illustrates how three non-sensory contributions (reward, task engagement, and social or linguistic context) shape auditory plasticity and learning in a variety of organisms (rodents, birds, and humans), using a wide range of stimuli (tones and noise, words, and non-human animal vocalizations), at multiple levels (molecular, physiological, and behavioral). Increasing cross-talk among researchers from different disciplines is therefore likely to accelerate the development of novel pharmaceuticals, biotechnologies, or training strategies that restore and augment auditory skills.

ROLE OF REWARD IN AUDITORY LEARNING AND PLASTICITY

Dopamine and Auditory Cortical Physiology: Max Happel

The auditory cortex is located at a crossroad of ascending and descending brain circuits (Nelken 2020). The cortical circuits receiving direct bottom-up input from the auditory thalamus are the last station along the auditory hierarchy where tonotopy still plays a major organizational role. On the other hand, top-down, non-sensory information reflecting learning, motor commands, and behavioral choice can influence auditory cortical activity (Kuchibhotla and Bathellier 2018; Scheich et al. 2011; Zempeltzi et al. 2020). Many of these top-down,

non-sensory influences are associated with the cortical release of dopamine, a neurotransmitter implicated in reward signaling (Salamone et al. 2005; Schultz 2015; Vickery et al. 2011) and cortical learning-dependent plasticity (Happel 2016).

We investigated the impact of dopamine on layer-specific circuit processing in the primary auditory cortex (A1) of Mongolian gerbils by current-source-density analysis (Brunk et al. 2019; Deliano et al. 2020; Happel et al. 2014). Optogenetic stimulation of the ventral tegmental area, the primary source of dopaminergic input to A1 (Campbell et al. 1987), enhanced tone-evoked thalamocortical input to the infragranular layers Vb/VIa and boosted activity in the supragranular layers I/II (Fig. 1). Those cortical layers contain the densest expression of dopamine receptors in auditory cortex (Brunk et al. 2019; Campbell et al. 1987; Phillipson et al. 1987). Dopamine has distinct influences on those two laminar circuits. First, dopaminergic modulation in the infragranular layers (Vb/VIa) affects local recurrent corticothalamic feedback (Happel et al. 2014). This feedback loop affects the synchronization and amplitude of recurrent thalamocortical oscillations. For example, a time–frequency analysis

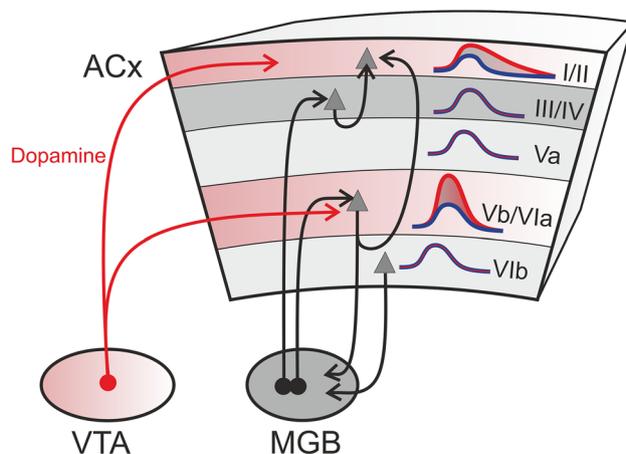


FIG. 1. Local and global impact of dopamine on cortical processing and learning. Auditory cortical (ACx) layers show a laminar distribution of dopamine receptor types with high levels of D1-like and D2-like receptors in infragranular layers Vb/VIa (shaded red) and higher levels of D1-receptors in supragranular layers I/II (also shaded red). Direct inputs of projection neurons from the ventral tegmental area (VTA) to the auditory cortex (red arrows) terminate in these layers and release dopamine when active. Dopamine release enhances early synaptic activity in infragranular layers (red versus blue evoked response curves in layer Vb/VIa). This enhancement has been linked to strengthening of the recurrent activity within the corticothalamic circuitry between ACx and the medial geniculate body (MGB) (black arrows). Furthermore, VTA activity prolongs corticocortical tone-evoked processing in the supragranular layers (red versus blue evoked response curves in layer I/II). Dopamine may therefore support prolonged local input processing of behaviorally relevant stimuli leading to potentially long-lasting plastic adaptations of more global corticocortical networks.

revealed that dopamine led to an increase in stimulus phase-locking in the high gamma band (75–110 Hz) in thalamocortical input layers (Deliano et al. 2020) leading to a gain of sensory inputs. Moreover, activation of these circuits by direct stimulation enhanced, while lesion by laser-induced apoptosis disturbed, salient auditory perception (Happel et al. 2014; Homma et al. 2017; Guo et al. 2017; Saldeitis et al. 2021), showing that infragranular activity is indeed an essential entry point for sensory representation. Second, after these local effects on sensory input, dopaminergic modulation in the supragranular layers (I/II) broadcasts local columnar activity via more long-range corticocortical circuits. This translaminar change of cortical activity persists over more than 30 min indicating that it is a long-lasting effect that transcends the signaling of a mere phasic dopaminergic reward-prediction error (Brunk et al. 2019). Such tonic dopamine release during auditory learning has been reported during early phases of acquisition learning, where correct task predictions are still poor (Stark and Scheich, 1997). Dopamine also effectuates protein biosynthesis-dependent plasticity over training days (Schicknick et al. 2008) and attenuates the rigid perisynaptic extracellular matrix (Mitlohner et al. 2020). In songbirds, it has been recently demonstrated that dopamine release in sensory-processing areas shapes the incentive salience of communication signals (Barr et al. 2021).

Taken together, our data suggest that dopaminergic modulation of recurrent cortico-thalamocortical processing via infragranular layers allows local sensory input to recruit long-range corticocortical supragranular networks and generate a salient sensory representation (Deliano et al. 2020). Thus, our work suggests that dopamine affects both local and long-range cortical circuits that integrate bottom-up sensory input with top-down information about behaviorally relevant non-sensory variables. By modulating corticothalamic feedback, dopamine may act as an ideal top-down regulator for active sensing and sensory grouping, decision making, and learning-dependent plasticity.

Cortico-striatal Networks in Auditory Category Learning: Bharath Chandrasekaran

Our program of research examines the role of non-sensory cortico-striatal networks in mediating speech and auditory category learning in adults. Anterograde tracing studies in animals demonstrate robust (many-to-one converging) connectivity between sensory association auditory cortex and the non-sensory striatum, which includes the putamen and the caudate nucleus (Yeterian and Pandya 1998). We posit that these connections provide a crucial infrastructure for sound-to-reward mapping and enable the acquisition of auditory categories in adulthood (Chandrasekaran et al. 2015; Feng et al. 2019). Our premise is as follows: during infancy, speech sound category representations

emerge within the association auditory cortex as a result of statistical learning without feedback (Vallabha et al. 2007). In contrast, during adulthood, auditory learning and plasticity require some amount of feedback. Feedback processing may be mediated by an implicit mapping between sound and reward in cortico-striatal circuitry. To test the premise that adult learners rely on non-sensory cortico-striatal circuitry for auditory category learning, we used functional magnetic resonance imaging (fMRI) to examine the neural mechanisms underlying non-native speech category acquisition (Yi et al. 2016). Monolingual American-English speakers learned to categorize non-native Mandarin phonemic tones (Fig. 2A, right side) using a training paradigm involving natural exemplars and trial-by-trial feedback (Fig. 2A, left side). On each trial, participants were instructed to classify a Mandarin phonemic tone as belonging to one of four tonal categories. Visual feedback immediately followed the categorization response, indicating whether the participant was correct or incorrect. Training engendered large individual differences in speech category learning (Fig. 2B). Tone category representations emerged within the auditory association cortex within a few hundred training trials (Feng et al. 2019). Feedback activated the non-sensory bilateral putamen, caudate nucleus, and nucleus accumbens (Fig. 2C). Participants who achieved greater categorization accuracy showed higher activation within the non-sensory putamen at the end of training. Moreover, the functional connectivity between regions demonstrating emergent representations (auditory association cortex) and the non-sensory putamen increased over the time course of training. Individual differences in structural connectivity between the superior temporal gyrus (STG) and the putamen, measured by diffusion tensor imaging, are associated with tone category learning accuracy (Fig. 2D). We posit that the functional and structural connectivity between the striatum and auditory association regions is crucial for emergent category

representations to ensure more accurate responses and therefore ensure more reliable rewards. With emerging expertise of the learner, the non-sensory cortico-striatal systems may “train” the sensory auditory temporal lobe networks to categorize information by validated rewards, thereby driving auditory learning (Feng et al. 2019).

A better understanding of the non-sensory influences on speech category learning has important implications for optimizing training paradigms for adults. While most adults can acquire novel speech categories, there are large-scale differences in learning speed and the extent of learning success, even in a relatively homogenous, neurotypical population (Fig. 2B) (Llanos et al. 2020). Only a small amount of this variability can be attributed to individual differences in the robustness of emergent category representations in the (sensory) associative auditory cortex. Much more can be attributed to individual differences in non-sensory striatal activity, structural and functional connectivity between sensory (auditory association cortex), and non-sensory striatal regions (Yi et al. 2016). Variability in performance may also reflect sub-optimal training regimens that lead to inconsistent engagement of the non-sensory cortico-striatal network across individuals (Chandrasekaran et al. 2014). One simple way to optimize training is to increase the engagement of the non-sensory cortico-striatal network by manipulating the nature of the feedback that is presented to the participant. Our prior work has shown that auditory category training paradigms in which feedback is provided immediately (within 500 ms after a button press) and contains minimum informational content (“wrong”) result in greater categorization accuracy for non-native Mandarin phonemic tones than training paradigms in which feedback is delayed (1000 ms after a button press) and contains full informational content (wrong, the correct category is “2”) (Chandrasekaran et al. 2014). Immediate feedback is a critical requirement for dopamine-mediated, non-sensory cortico-striatal learning

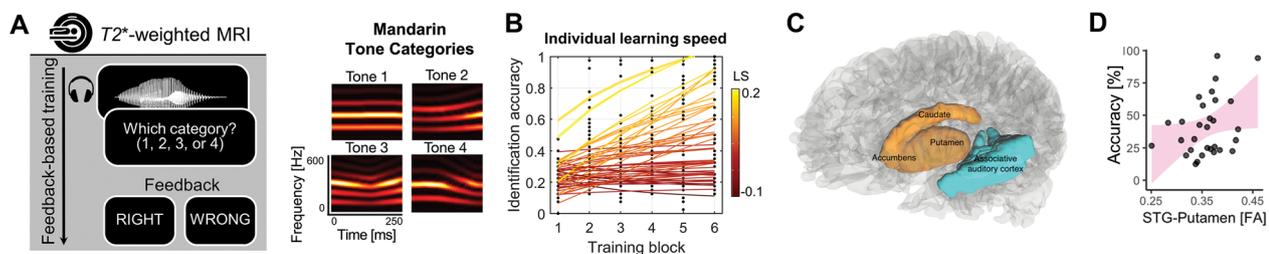


FIG. 2. Non-sensory contributions to non-native speech category acquisition. **A** To examine sensory and non-sensory influences on the learning of non-native Mandarin Chinese phonemic tones (tones 1–4), we leverage an auditory category training paradigm in the MR scanner that uses trial-by-trial feedback (left side) and natural exemplars of tone categories produced by Mandarin speakers (right side). **B** Despite the same amount of training, there are large individual differences in learning speed (LS) across participants. LS was estimated by fitting each learner’s block-by-block accuracies (black

dots) with a power function. Learning speed reflects initial learning gain based on the same amount of training as well as the changes in learning gain across subsequent training blocks. **C** Brain regions implicated during sound-to-category training (associative auditory cortex, blue; non-sensory striatum (caudate, putamen, nucleus accumbens), gold). **D** Structural connectivity between associative auditory cortex and putamen, measured via diffusion tensor imaging (DTI), is significantly associated with individual differences in tone category learning accuracy.

(Maddox and David 2005). Minimal-information feedback, relative to full-information feedback, allows for less effective hypothesis generation and testing, which in turn allows for less interference from cortical networks involved in an alternative learning process for learning speech categories: rule-based learning (Chandrasekaran et al. 2014). Thus, simple training manipulations in the timing and content of feedback can result in more robust engagement of the non-sensory cortico-striatal circuitry and have the potential to improve the efficiency of category training paradigms.

Intrinsically Generated Rewards in New Word Learning: Pablo Ripollés

One of the building blocks of language is the acquisition of new vocabulary, a learning process that starts early in childhood and is still present in later stages of life. In a series of studies, we tried to bridge the gap between language learning and reward, a cognitive function well-known to modulate learning, memory, and decision-making via top-down, non-sensory signals. To do so, we developed a paradigm that mimics natural new word learning from context (Mestres-Misse et al. 2007). Such learning occurs without the need for explicit reward, feedback or external guidance and is related to vocabulary growth across the lifespan (Nagy et al. 1985, 1987). In our task, participants learned, without any kind of explicit feedback, the meaning of new words that appeared at the end of two sentences (e.g., 1, “Every Sunday the grandmother went to the *jedin*”; 2, “The man was buried in the *jedin*”; *jedin* means graveyard). Importantly, research shows that *explicit* rewards (e.g., money) can modulate the entrance of new information into long-term memory by tapping into a dopaminergic and reward-related circuit formed by the substantia nigra/ventral tegmental area complex (SN/VTA, midbrain), the ventral striatum (VS, basal ganglia), and the hippocampus: the so-called SN/VTA-Hippocampal loop (Adcock et al. 2006; Lisman and Grace 2005). Based on the hypothesis that self-generating the solution to a problem can be inherently rewarding, we predicted that new word learning from context should elicit an intrinsic, top-down, non-sensory signal that—as is the case for explicit rewards—enhances learning and memory by tapping into the SN/VTA-Hippocampal loop. Our results showed that successful learning—in the absence of external feedback—increased: (i) objective physiological markers of arousal (i.e., electrodermal activity), (ii) subjective behavioral self-reports of reward (i.e., pleasure), and (iii), most importantly, brain activity in both cortical areas related to language processing and in the SN/VTA-Hippocampal loop (Ripollés et al. 2014, 2016). Moreover, our results showed that both increased activity in, and functional connectivity among, the areas of the SN/VTA-Hippocampal loop that were correlated with better memory for the newly learned words after a

consolidation period (24 h; Ripollés et al. 2016). Finally, by means of a double-blind, within-subject, pharmacological study, our results showed that dopamine does actually play a causal role in both the learning and consolidation of the new words (Ripollés et al. 2018).

Taken together, our results suggest that language learning from context can be its own intrinsic reward, generating a top-down, non-sensory signal that modulates the entrance of new words into long-term memory via dopaminergic and reward-related mechanisms. Combining the present results with previous literature, we put forward a neuroanatomical model showcasing how intrinsic, top-down, non-sensory signals can promote learning and induce plasticity in cortical regions related to auditory, speech, and/or language processing, bridging the gap between language models of semantic processing (Lau et al. 2008; Rodriguez-Fornells et al. 2009), hippocampal-neocortical accounts of memory formation (Davis and Gaskell 2009; McClelland et al. 1995; Ullman 2020), and the role that reward, dopamine, and the SN/VTA-Hippocampal loop play in long-term memory (Lisman and Grace 2005; Fig. 3).

First, contextual information (processed by visual and/or auditory regions; in yellow in Fig. 3A) is manipulated until the appropriate meaning for the new word is extracted (e.g., graveyard for *jedin*). This process is subserved by a series of cortical areas, more active during successful learning in our task (Ripollés et al. 2014), that are part of a semantic network (Lau et al. 2008; Rodriguez-Fornells et al. 2009). These regions are (see Fig. 3A) the posterior middle temporal gyrus (access and storage of lexical representations; light purple), the angular gyrus and anterior temporal cortex (integration and combination of the lexical representations into a higher semantical context; dark purple), and key regions within the prefrontal cortex (PFC): the anterior ventral (top-down retrieval of lexical representations; light blue) and posterior (selection among highly activated candidates; very light blue) inferior frontal gyrus.

Second, once a mapping between a new word and its meaning has been correctly assigned, this new information (*jedin* means graveyard) is sent to the hippocampus possibly via prefrontal-hippocampal connections for entry into long-term memory (Davis and Gaskell 2009; Preston and Eichenbaum 2013; Takahashi et al. 2007). The information then enters the SN/VTA-Hippocampal loop. In the downward arm of the loop (see Fig. 3B), an intrinsic, reward-related signal enhances VS activity which, in the upward arm of the loop, results in dopamine being released at the hippocampus. The release of dopamine ultimately enhances the probability that the new information learned (*jedin* means graveyard) enters long-term memory (Lisman et al. 2011).

Recall of the newly learned word (not shown) then follows hippocampal-neocortical accounts of memory formation, where memory recall for a new word relies on the

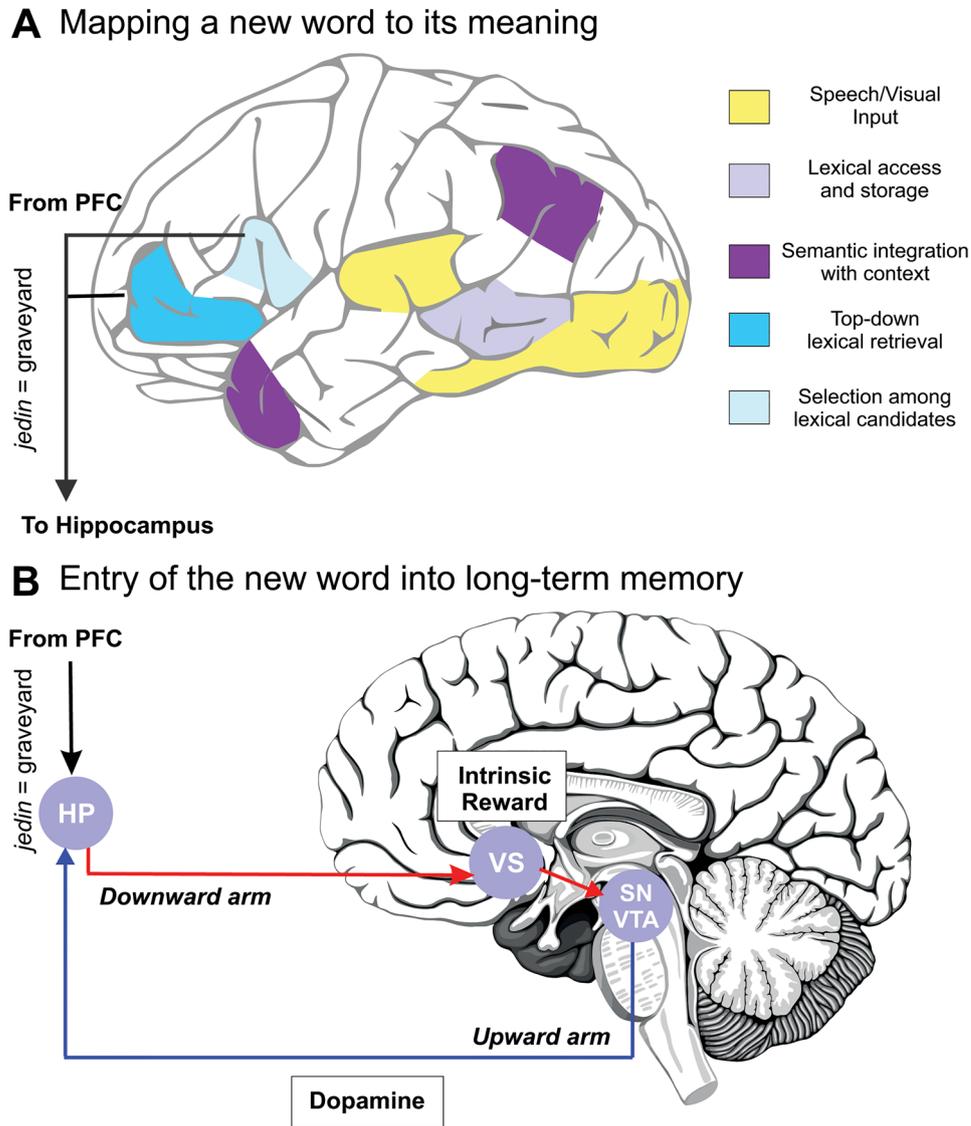


FIG. 3. A neuroanatomical model supporting cortical plasticity via intrinsic, non-sensory modulatory signals during word learning. For details, see text. **A** Contextual information is manipulated until the appropriate meaning is extracted via the activation of a semantic network. **B** Learning triggers a top-down, non-sensory, intrinsic

reward signal that aids the entrance of the new word into long-term memory via the activation of the SN/VTA-Hippocampal loop. HP, hippocampus; VS, ventral striatum; SN/VTA, substantia nigra/ventral tegmental area; PFC, prefrontal cortex.

hippocampal-dependent replay of the cortical patterns associated with that word-to-meaning mapping early after learning and becomes less hippocampal dependent after consolidation (Davis and Gaskell 2009; McClelland et al. 1995; Ullman 2020).

ROLE OF TASK ENGAGEMENT IN AUDITORY LEARNING AND PLASTICITY

Task Engagement and Auditory Perceptual Learning in Humans: Beverly Wright

Perceptual learning refers to the improvement in perceptual skills with practice. In audition, perceptual learning

contributes to the acquisition of foreign languages, musical skills, and specialized auditory expertise such as recognizing an irregular heartbeat. It also provides an avenue for the treatment of auditory disorders.

We investigate top-down non-sensory influences on auditory perceptual learning in humans. To do so, we examine the effects of two types of experience on learning: practice and stimulus exposure alone. By practice we mean performance of an auditory task that requires a decision, such as selecting which of two sounds has a higher frequency. By stimulus exposure alone, we mean exposure to the sounds that are used during practice, but without performance of the relevant auditory task.

We assume that practice provides top-down input—presumably related to arousal, attention, feedback or reward (Fig. 4A, left side)—that is not provided by stimulus exposure alone (Fig. 4A, right side).

In one line of work, we asked whether top-down input was necessary for auditory perceptual learning. If so, given our assumption, learning should occur from practice, but not from stimulus exposure alone. Consistent with this prediction, we documented that practice, but not stimulus exposure alone, induced learning on a non-native phonetic classification task (Wright et al. 2015) and a musical-interval discrimination task (Little et al. 2018) (Fig. 4B, top two rows).

In another line of work, we asked whether top-down input was necessary throughout the entire training period. In this case, we took advantage of the observation that sufficient practice, within a restricted time period, is typically required to induce an improvement in perceptual skills that persists across days (Little et al. 2017; Wright and Sabin 2007) (Fig. 4B, middle two rows). Knowing

that constraint, we provided too few trials of practice per day to yield learning and replaced the remaining required trials with stimulus exposure alone. The stimulus exposure alone was presented in the background while the participant was engaged in a written symbol-to-number matching task. If the top-down input from practice only affected trials on which the task was actually performed, the additional stimulus exposures—which were ineffective to drive learning on their own—should not contribute to learning. Contrary to this prediction, combining periods of practice and periods of stimulus exposure alone induced or enhanced learning on a variety of tasks including auditory frequency discrimination (Wright et al. 2010), musical-interval discrimination (Little et al. 2018), non-native phonetic discrimination (Wright et al. 2015), and adaptation to foreign accents (Wright et al. 2015) (Fig. 4B, bottom row). In related work, we reported that combining periods of practice on two different tasks—frequency discrimination and temporal-interval discrimination—induced learning on frequency discrimination, even though neither

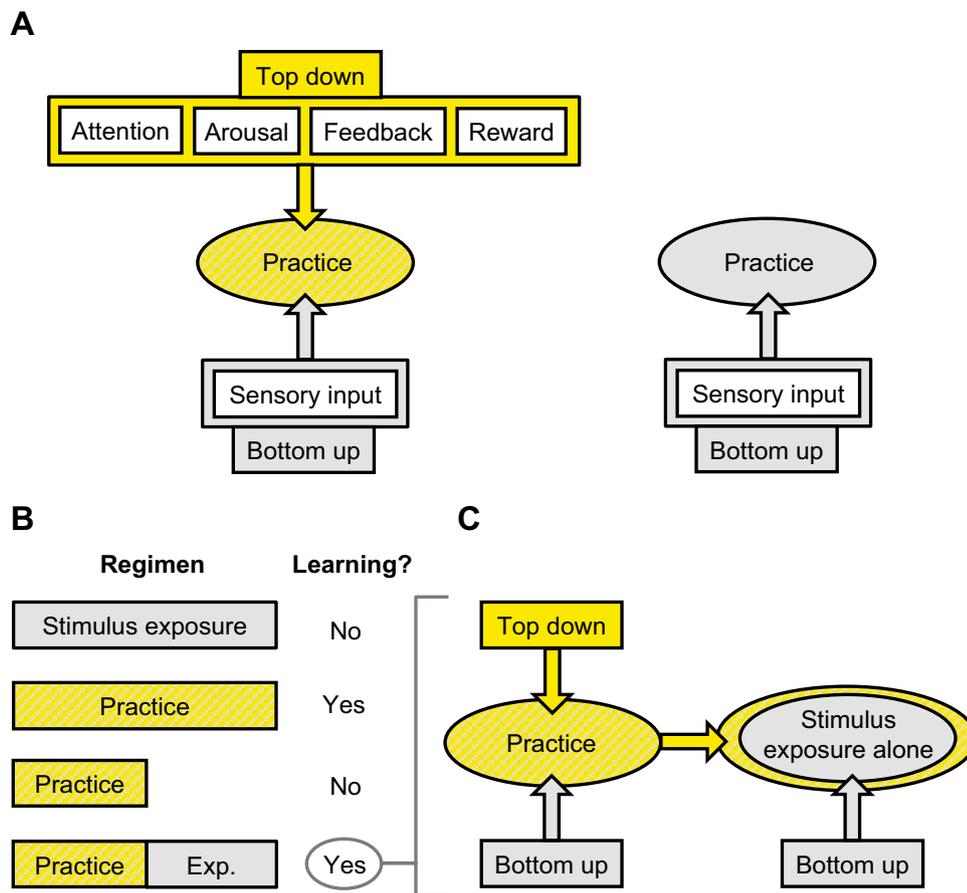


FIG. 4. Temporal spread of top-down influences during perceptual learning. **A** Practice on an auditory task (left side) involves both bottom-up sensory input (gray) and top-down input (yellow), while stimulus exposure alone (right side) primarily involves bottom-up sensory input. **B** Typically, stimulus exposure alone yields no auditory perceptual learning (top row); rather, learning requires

a sufficient amount of practice (middle two rows), but insufficient practice combined with stimulus exposure alone can yield learning (bottom row). **C** Thus, it appears that the top-down input required for learning can spread from a period of practice to infuse a period of stimulus exposure alone, making the stimulus exposures alone act as though they occurred during practice.

experience yielded learning on frequency discrimination on its own (Wright et al. 2010). We have also documented similar findings in visual perceptual learning in humans (Szpiro et al. 2014) as well as in odor learning in mice (Fleming et al. 2019).

Most recently, we reported another version of these training effects: semi-supervised learning (Wright et al. 2019). We showed, for a non-native phonetic classification task, that trial-by-trial feedback about performance (during which each response was labeled as correct or incorrect) was necessary for learning, but that the feedback was not required on every trial. Rather, the combination of practice trials with feedback (“supervised trials”) and practice trials without feedback (“unsupervised trials”) or the combination of practice trials with feedback and stimulus exposure alone, yielded learning, even though none of the individual experiences provided in these combined regimens generated learning on their own. Moreover, a relatively small but critical number of trials with feedback were required to trigger learning, but once learning was triggered, additional trials with feedback did not increase the amount of learning. Thus, feedback appears to engage an all-or-none process.

Overall, these results suggest that top-down input is necessary for many forms of auditory perceptual learning, but that the top-down input does not need to be actively engaged on every trial during the training period. Rather, it appears that the influences of top-down non-sensory input and bottom-up sensory input spread over time to promote learning (Fig. 4C). Knowledge of these dynamics could help constrain the search for potential neural mechanisms of learning, and markedly improve the efficiency and effectiveness of perceptual training regimens.

Task Engagement and Auditory Perceptual Learning in an Animal Model: Melissa Caras

As described in the preceding section, auditory perceptual learning is the process by which individuals improve their ability to hear subtle differences between sounds (Irvine 2018). Previous investigations into the neurobiological bases of auditory perceptual learning revealed that learning-related improvements in sound detection or discrimination may result from behaviorally relevant changes in sound-evoked responses within the bottom-up auditory pathway (Bao et al. 2004; Beitel et al. 2003; Recanzone et al. 1993). It is increasingly clear, however, that non-sensory processes, like attention, play an important role in shaping auditory perceptual learning (Amitay et al. 2014) and learning-related plasticity in the auditory cortex (Polley et al. 2006). These findings raise the possibility that learning-related changes may also occur within the top-down cortical networks that project to the auditory system and modulate its response properties.

We explored this possibility by wirelessly recording extracellular single- and multi-unit responses in the auditory cortex of freely moving Mongolian gerbils as they trained on an amplitude modulation (AM) detection task (Fig. 5A; Caras and Sanes 2017, 2019). We found that auditory cortical neurons were more sensitive to the target AM sound when animals performed the task (Fig. 5B, blue), compared to when they were exposed to the same AM sound in a passive, non-task context (Fig. 5B, green). Critically, the magnitude of this task-related enhancement increased over the course of training (compare orange arrows in Fig. 5B): When animals participated in the behavioral task, auditory cortical neurons showed a strong effect of training, such that neural AM detection thresholds improved across several days. In contrast, the *exact same* neurons displayed a weak effect of training when their responses were measured during passive sound exposure. As a result, the act of engaging in the task had a larger effect on auditory cortical responsiveness at the end of perceptual learning than it did at the beginning (Caras and Sanes 2017). This finding is consistent with the idea that task performance engages non-sensory, top-down brain networks that optimize auditory cortical responses to behaviorally relevant stimuli (David et al. 2012; Fritz et al. 2003, 2005; Yin et al. 2014). More importantly, these results suggest that training increases the strength of these top-down modulations, leading to gradual enhancements in auditory cortical sensitivity that underlie perceptual learning (Fig. 5C).

This hypothesis is consistent with data from human imaging experiments (Bartolucci and Smith 2011; Byers and Serences 2014; Mukai et al. 2007; Niu et al. 2014), electrophysiological recordings (Wang et al. 2016), and computational models (Schäfer et al. 2007), all of which indicate that functional changes in top-down cortical networks make an important contribution to visual perceptual learning. Plasticity in top-down circuits may therefore represent a general mechanism contributing to perceptual learning across sensory modalities. Future experiments that combine psychophysics, in vivo single-unit recordings, and projection-specific optical or chemical manipulations in animal models are needed to reveal the specific top-down brain regions and pathways involved.

ROLE OF SOCIAL AND LINGUISTIC CONTEXT IN AUDITORY LEARNING AND PLASTICITY

Early Social Isolation, Serotonin, and the Auditory Midbrain: Sarah Keesom and Laura Hurley

Neuromodulatory pathways are an important class of non-sensory inputs to the auditory system (Fig. 6). Multiple neuromodulatory systems project to both peripheral

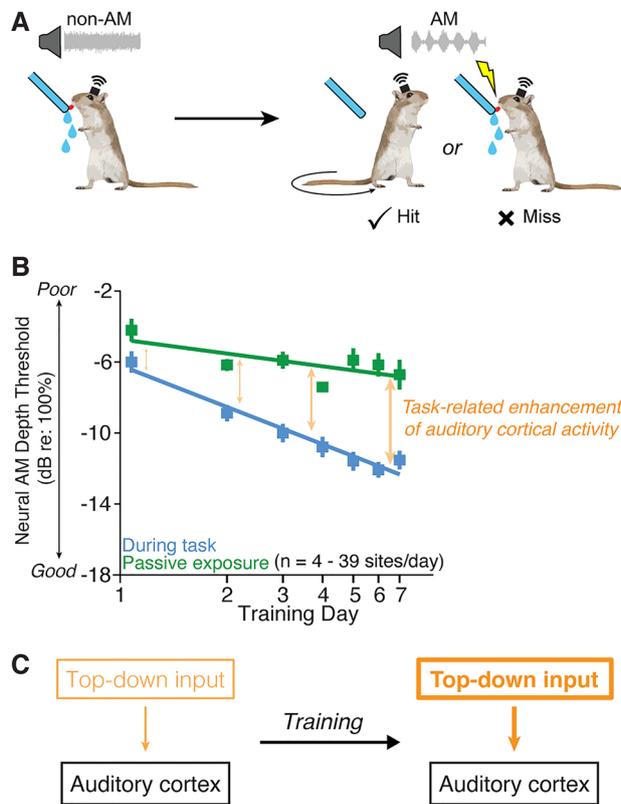


Fig. 5. Top-down modulation of auditory cortex during perceptual learning. **A** Mongolian gerbils were trained to drink from a spout while in the presence of continuous, unmodulated broadband noise and to cease drinking when the noise smoothly transitioned to an amplitude modulated (AM) noise. Single- and multi-unit recordings were obtained wirelessly from the auditory cortex of these animals as they trained with a range of AM depths. Over the course of several days of training, animals improve their ability to detect more subtle modulations, such that their psychometric thresholds gradually improve. **B** Amplitude modulation detection thresholds of auditory cortical neurons were lower (better) when animals engaged in the AM detection task (blue) compared to when they were passively exposed to sounds (green). This task-related enhancement of auditory cortical sensitivity (orange arrows) increased over the course of perceptual learning. **C** These data are consistent with the idea that behavioral performance engages higher order, top-down brain networks that enhance auditory cortical responses and suggests that this putative top-down enhancement grows larger over the course of perceptual learning. Adapted from Caras and Sanes (2017).

and central auditory structures, providing information on the external events and internal states that accompany acoustic signals (Kandler 2019; Schofield and Hurley 2018). One of these modulatory pathways, the serotonergic system, is closely involved in the brain's response to stressors and social stimuli. Information about these contexts is conveyed to central auditory regions through projections from serotonergic neurons in the dorsal raphe nucleus (DRN; Klepper and Herbert 1991; Niederkofler et al. 2016; Petersen et al. 2020). In the inferior colliculus (IC), a midbrain nucleus that is a hub for both ascending

and descending auditory pathways, serotonin release increases as mice interact with social partners (Hall et al. 2011). Across individuals, serotonin levels positively correlate with social behaviors (such as investigation of a social partner) and negatively correlate with antisocial behaviors (such as aggression) (Hall et al. 2011; Hanson and Hurley 2014; Keesom and Hurley 2016). Serotonin therefore represents aspects of the quality of a current social interaction.

In addition to responding to acute social contact, serotonin-auditory interactions are highly sensitive to manipulations of the social environment in early life. One such manipulation is a period of postweaning social isolation (Fone and Porkess 2008; Walker et al. 2019). In mice, postweaning social isolation for a period of a month alters both presynaptic and postsynaptic components of the serotonergic system within the adult IC (Fig. 6). Presynaptic targets of social isolation include the axons of serotonergic neurons expressing the serotonin transporter (Fig. 6(i)). Axon density is lower in mice that have been housed individually than in mice that have been housed socially (Keesom et al. 2018). A month of postweaning social isolation also alters the functional dynamics of serotonergic activity during subsequent social interactions (Fig. 6(ii)). When presented with novel social partners in an acute social encounter, mice that previously were housed individually show a longer latency to reach peak serotonergic activity in the IC relative to mice that were socially housed (Keesom et al. 2017). Additionally, serotonergic activity during acute social encounters in mice previously housed in isolation is less correlated with social investigation and overall activity than in socially housed mice (Keesom et al. 2017). This result suggests that serotonin release is less attuned to variation in the social environment following individual housing.

Postsynaptically, postweaning isolation influences the responses of IC neurons to the manipulation of serotonin (Fig. 6(iii)), as assessed by the number of IC neurons expressing cFos protein, an immediate early gene product that is a proxy marker for recent neural activity (Davis et al. 2021). The number of cFos-positive neurons in the IC is sensitive to an interaction between social isolation and serotonergic drugs. After the injection of fenfluramine, a drug that causes the release of endogenous stores of serotonin, individually housed mice have elevated numbers of cFos-positive cells in the IC relative to socially housed mice. This difference does not arise following injection of saline. The effect of social isolation on cFos expression is therefore observed when the serotonergic system is activated, as would naturally occur during salient contexts including acute social interaction. An additional aspect of this work that has not yet been fully explored in the auditory system is that the response of the serotonergic system to social isolation may depend on sex. For example, the effects of individual housing on the density of serotonergic axons in the IC are seen only

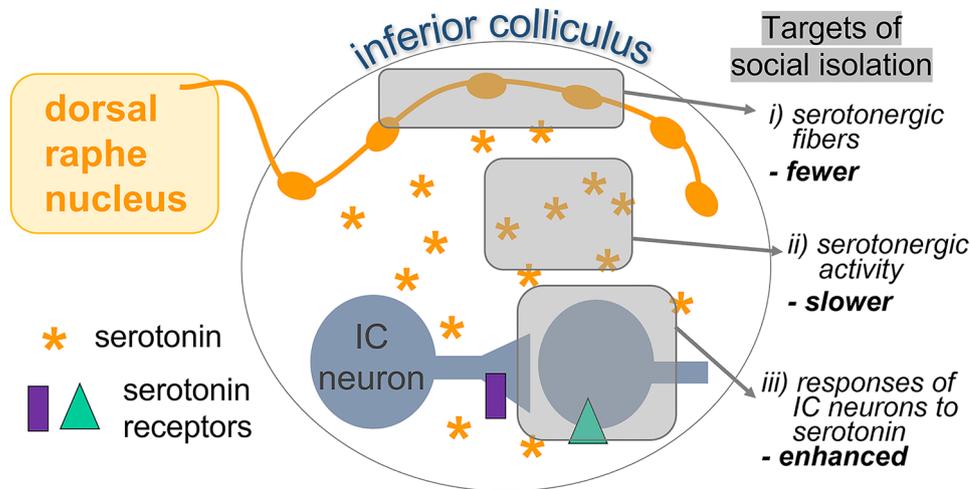


FIG. 6. Model of non-sensory serotonergic input to the auditory midbrain. Serotonergic axons from the dorsal raphe nucleus innervate the inferior colliculus (IC), altering the responses of IC neurons to sound through receptors expressed by IC neurons. Postweaning

social isolation affects serotonin-auditory interactions at the level of (i) the axons of serotonergic neurons, (ii) the dynamics of serotonergic activity, and (iii) the responses of IC neurons to the pharmacological manipulation of serotonin.

in females, not in males (Keesom et al. 2018). Overall, these findings demonstrate that the serotonergic system plays a key role in representing information on social experience in the auditory system.

Together with the growing evidence that social isolation influences the perception of vocal signals and that serotonin alters acoustically evoked responses (Schofield and Hurley 2018; Screven and Dent 2019; Sturdy et al. 2001), our findings suggest that the serotonergic system is one mechanism through which the experience of social isolation is transformed into a neural state capable of modulating auditory processing.

Estradiol and the Songbird Auditory Forebrain: Luke Ramage-Healey

Non-sensory influences on aural communication behaviors are mediated by sex steroid hormones (such as androgens and estrogens) in several species. For example, when a territorial boundary is breached (either by a vocalizing neighbor and/or an interloping human experimenter with a playback speaker and a decoy), the territory holder responds with a surge of sex steroid production by the gonads (Gleason et al. 2009; Wingfield 2005; Hirschenhauser and Oliveira 2006). These steroids, in turn, circulate throughout the body, and act on vocal-motor neural circuits, rapidly altering vocalization production (Fernández-Vargas 2017; Ramage-Healey and Bass 2006). Similarly, circulating sex steroids influence hearing by acting on both peripheral sensory organs and central auditory pathways (Caras 2013; Maney and Rodriguez-Saltos 2016; Sisneros 2009). Thus, by regulating the levels of sex steroid hormones in circulation, non-sensory environmental cues, like social or reproductive context,

shape both the production and the reception of vocal communication signals.

In addition to circulating hormones, which are synthesized in peripheral structures, the synthesis of some sex steroids, like estrogens, also occurs within the brain itself, including directly at neuronal synapses (Saldanha et al. 2011). There is increasing evidence that local, brain-derived estrogens also influence auditory physiology and perception (Fig. 7). The ascending auditory pathway in songbirds (yellow arrows) and descending song motor pathway (red arrows) are interconnected, and one node in the circuit (NCM; caudomedial nidopallium) is a hub for estrogen signaling. The expression and activity of aromatase (the enzyme responsible for converting androgens to estrogens) is particularly enriched in the circuits that regulate communication behaviors: the human temporal cortex (Azcoitia et al. 2011; Yague et al. 2006) and the songbird auditory forebrain (Saldanha et al. 2000). In addition, estrogen levels are rapidly elevated in the songbird auditory NCM when juveniles and adults hear new songs (Chao et al. 2014; de Bournonville et al. 2020; Ramage-Healey et al. 2008, 2012). Moreover, blockade of local estrogen synthesis within the NCM attenuates the expression of activity-dependent immediate early genes (Krentzel et al. 2020) and fMRI responses to song in NCM (De Groof et al. 2017). The same manipulation also disrupts some aspects of auditory learning in juveniles (Vahaba et al. 2020) and adults (Macedo-Lima and Ramage-Healey 2020). Brain-derived estrogens also have acute actions on the sensory coding of auditory neurons in NCM (Ramage-Healey and Joshi 2012; Ramage-Healey et al. 2010; Krentzel et al. 2018; Vahaba et al. 2017). These acute actions range from increased burst firing of auditory neurons, elevated evoked firing rates, and the propagation of modulated auditory representations into sensorimotor networks.

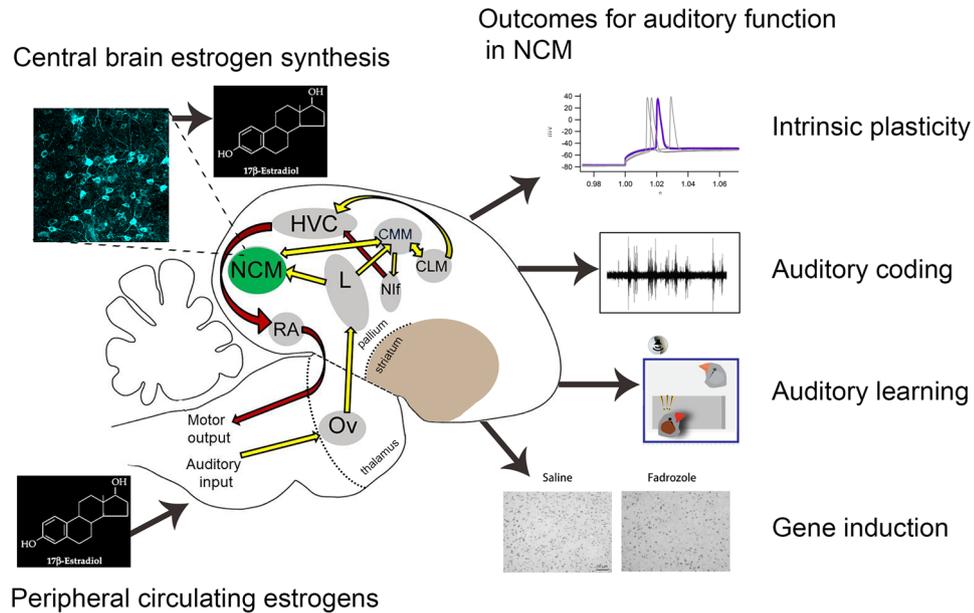


Fig. 7. Impact of estrogens on auditory processing and learning. Steroid hormones, such as the estrogen 17-beta-estradiol, can regulate the songbird ascending auditory pathway (yellow arrows) and descending motor pathway (red arrows; basal ganglia not pictured for clarity). Estrogenic effects are mediated by circulation from endocrine organs (lower left) and/or central brain synthesis (upper left) via the enzyme aromatase (neuronal somas and processes in cyan) in the caudomedial nidopallium (NCM, in green). The resultant outcomes on NCM function include intrinsic plasticity (top right; in vitro action potentials from NCM whole-cell slice recordings, LRH unpublished observations), auditory cod-

ing (upper middle right; in vivo auditory-evoked activity of NCM neurons; see Remage-Healey et al. 2010), auditory learning (lower middle right; operant task for auditory playback and learning; see Macedo-Lima and Remage-Healey 2020), and immediate-early gene induction (lower right; EGR1 positive nuclei for song-exposed males treated with saline or the aromatase inhibitor fadrozole; see Krentzel et al. 2020). Also shown is auditory thalamus Ov (ovoidalis), thalamorecipient field L (L), caudomedial mesopallium (CMM), caudolateral mesopallium (CLM), nucleus interface (Nif), sensorimotor HVC (proper name), and arcopallial song motor nucleus (RA).

Collectively, these results demonstrate that both circulating and brain-derived sex steroids fluctuate dynamically in response to non-sensory cues and that brain-derived estrogens can have minute-by-minute actions on auditory coding and learning, similar to traditional neuromodulators (Fig. 7). In the near term, it will be important to pin down potential interactions between neuroestrogen signaling systems and more “conventional” neuromodulators such as dopamine (Macedo-Lima and Remage-Healey 2021) or serotonin (Hurley and Sullivan 2012) to gain a more complete understanding of the neural mechanisms of auditory plasticity and learning.

Linguistic Context and Speech Perception Learning: Lori Holt

Speech perception is fundamentally shaped by experience. Listeners whose native language is English parse sensory acoustic speech input distinctly from, say, listeners whose native language is Korean or Swedish (Kuhl et al. 1992). Even more subtly, individuals who learn American English perceive speech somewhat differently than those who learn another dialect, like Scottish English (Escudero

2001). The detailed acoustic patterns of speech that differentiate languages and dialects mold speech perception; we hear speech through the lens of this experience. Inasmuch as these influences are language community-specific and acquired, they also are unambiguously non-sensory. This observation positions speech perception as an excellent testbed for examining the influence of non-sensory influences on auditory learning and plasticity (Guediche et al. 2014; Lim et al. 2014).

Imagine a chat with a stranger who hails from a different region. You each speak English with an accent shaped by your local language community norms. Although perhaps less apparent, you each *perceptually* parse speech through the lens of these language community norms as well. A rich research literature demonstrates that, at least initially, the mismatch between the accented speech input you hear and a perceptual system molded to your local language community will challenge your chat. Accented speech impairs speech comprehension. Nonetheless, there is abundant evidence that comprehension improves with exposure to accented speech (Bradlow and Bent 2008). There are rapid adjustments in how sensory input is mapped to speech perception as a function of experienced

short-term acoustic input regularities, like those arising in accented speech.

There is growing evidence that one contributor to these improvements derives from non-sensory influences of the acquired, language-specific molding of speech perception by long-term experience. In the laboratory, it is possible to manipulate short-term speech input regularities in a toy model of accented speech and its perception. Figure 8A shows an acoustic space defined by two dimensions, voice onset time (VOT) and fundamental frequency (F0), each of which contributes to the perception of American-English /b/ versus /p/ in an inter-dependent manner. Higher F0s and longer VOTs lead to more /p/ judgments; lower F0s and shorter VOTs lead to more /b/ judgments. Moreover, when VOT is held constant and perceptually ambiguous, F0 is sufficient to signal category membership as /b/ versus /p/ when the short-term input regularities mirror those typical of English (Fig. 8B, left top and bottom). This demonstrates that experience with the subtle correlation between F0 and VOT acoustics in the American English dialect has left its

fingerprint on native listeners' speech perception, as described above.

The right panels in Fig. 8B show what happens upon introduction of an “artificial accent” (Fig. 8B, right top) that reverses the canonical American-English correlation between VOT and F0. The influence of F0 in signaling /b/ versus /p/ is rapidly re-weighted such that this sensory evidence is no longer sufficient to signal a difference between /b/ and /p/ (Fig. 8B, right bottom). Importantly, the strength of activation of the native language-specific acquired non-sensory representations for American English /b/ and /p/ predicts the magnitude of the perceptual re-weighting (e.g., Idemaru and Holt 2011, 2014, 2020; Lehet and Holt 2020; Liu and Holt 2015; Zhang et al. 2021). Ultimately, this rapid adaptive plasticity in speech perception depends on the non-sensory language community-specific speech processing developed over long-term experience (Fig. 8C). The mapping from acoustic input to perception is not fixed, but rather flexibly adjusts to accommodate short-term input regularities in a manner that is guided by non-sensory influences driving rapid, online learning.

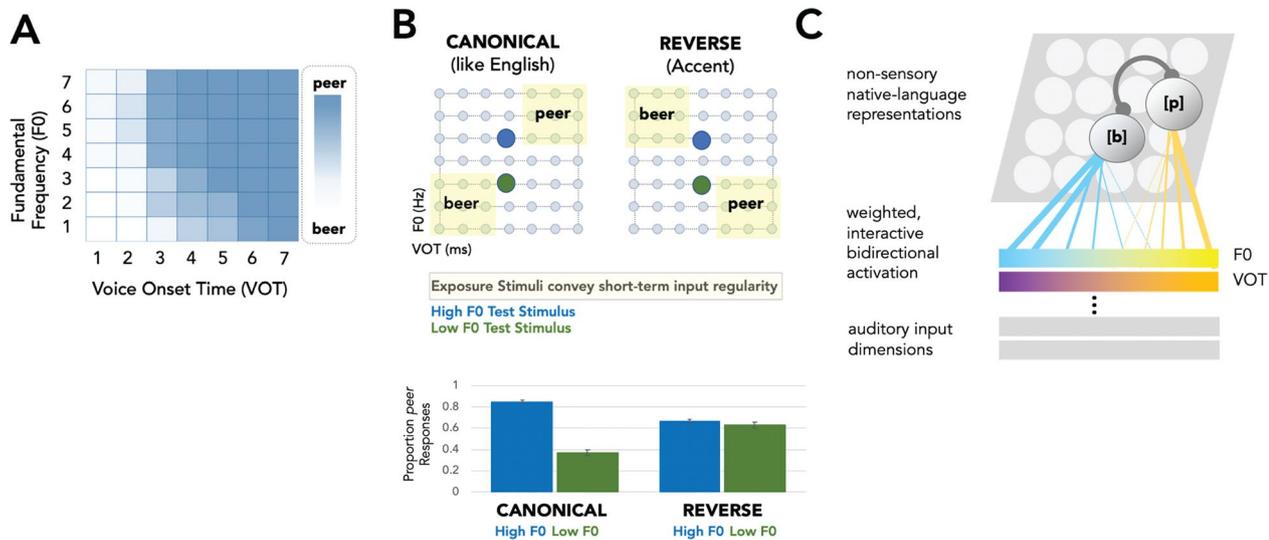


FIG. 8. Example of non-sensory influences on rapid plasticity in speech perception. **A** The color scale shows categorization of speech as *beer* or *peer* across an acoustic space defined by voice onset time (VOT) and fundamental frequency (F0). Each dimension contributes to categorization, and their influence covaries; longer VOT and higher F0 result in categorization as *peer*, whereas short VOT and lower F0 result in more *beer* responses. **B** This same acoustic space can be sampled selectively (yellow-shaded exposure stimuli, canonical) to convey the canonical English covariation of VOTxFO or an “accent” that reverses the typical English covariation of VOT and F0 (exposure stimuli in light yellow, reverse). Intermixed test stimuli with a perceptually ambiguous VOT and either high (blue) or low (green) F0 reveal how the short-term regularities conveyed by the canonical and reverse stimulus distributions affect the impact of F0 in signaling *beer* vs. *peer*. In the context of the

canonical English covariation, listeners rely on F0 to categorize the test stimuli as *beer* versus *peer* (**B**, bottom). In the context of the accent conveyed by the reverse covariation of VOTxFO, the influence of F0 is down-weighted such that it no longer signals category membership (**B**, bottom). **C** Schematic illustration of the origins of these effects shows sensory input dimensions that represent acoustic dimensions like F0 (illustrated at the top in blue-to-yellow gradient from low to high F0) providing weighted input (more effective inputs are illustrated as thicker lines) to non-sensory native language representations learned across long-term experience with a language community (here, /b/ versus /p/). These weighted inputs to the speech sound representations are interactive and bidirectional allowing for rapid online plasticity of the effectiveness of sensory information in informing speech perception (as in panel **B**).

DISCUSSION

It is increasingly accepted that top-down/non-sensory influences play an important role in auditory learning and plasticity. The present collection, drawn from studies of reward, task engagement, and social or linguistic context, reveals four overarching lines of evidence that are consistent with this idea:

1. Networks implicated in non-sensory signaling are recruited during auditory learning (Chandrasekaran: non-sensory striatum; Ripollés: SN/MTA-Hippocampal loop).
2. Signaling molecules associated with non-sensory factors are released during learning or as a result of experience (Hurley: serotonin; Remage-Healey: estradiol; Ripollés: dopamine).
3. Neuronal sound-evoked responses in the auditory system are influenced by non-sensory factors, and this influence changes over the course of learning (Caras/Happel/Remage-Healey: auditory cortical/forebrain responses; Hurley: auditory midbrain responses).
4. Several kinds of auditory learning are facilitated by non-sensory factors (Caras/Wright: auditory perceptual learning; Chandrasekaran/Holt/Wright: speech category learning; Ripollés: word learning).

The involvement of top-down/non-sensory factors in shaping auditory learning and plasticity has theoretical implications. For example, one fundamental question is what determines whether learning will occur when a listener encounters a new auditory experience. While a flexible system is important for the acquisition of new memories, learning entails an energetic cost, and change comes at the risk of eroding stable representations. The observation that top-down/non-sensory inputs are important for generating learning suggests that the presence or strength of these inputs may be one means by which the brain controls the balance of stability and plasticity in the auditory system. Another such question concerns the locus of neural plasticity during auditory learning. The studies presented here suggest that this plasticity may not be restricted to the auditory system. Rather, top-down/non-sensory inputs to the auditory system might themselves change as a result of practice or experience. This “top-down plasticity” might ultimately serve to shape bottom-up auditory processing and contribute to auditory learning.

The contribution of top-down/non-sensory inputs to learning and plasticity also has clinical and practical implications. Failures of learning could be attributed to the absence of top-down/non-sensory inputs, either because the training regimen did not properly recruit these inputs or because these inputs are in some way dysfunctional, being either absent or insufficiently engaged.

Failures of learning could also be attributed to an over-reliance on previously established top-down/non-sensory inputs, thereby preventing new activity patterns that are required for learning. Therefore, the restoration or augmentation of learning could potentially be aided by designing training regimens, social environments, neuro-active drugs, or brain computer interfaces that specifically target top-down function.

In sum, this collection highlights contributions of top-down/non-sensory influences on auditory learning and plasticity, providing an avenue for treating auditory communication disorders and optimizing the human perceptual experience.

Declarations

Conflict of Interest The authors declare no competing interests.

REFERENCES

- ADCOCK RA, THANGAVEL A, WHITFIELD-GABRIELI S, KNUTSON B, GABRIELI JD (2006) Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron* 50:507–517
- AMITAY S, ZHANG Y-X, JONES PR, MOORE DR (2014) Perceptual learning: top to bottom. *Vision Res* 99:69–77
- AZCOITIA I, YAGUE JG, GARCIA-SEGURA LM (2011) Estradiol synthesis within the human brain. *Neuroscience* 191:139–147
- BAO S, CHANG EF, WOODS J, MERZENICH MM (2004) Temporal plasticity in the primary auditory cortex induced by operant perceptual learning. *Nat Neurosci* 7(9):974–981
- BALTHAZART J, CHOLERIS E, REMAGE-HEALEY L (2018) Steroids and the brain: 50 years of research, conceptual shifts and the ascent of non-classical and membrane-initiated actions. *Horm Behav* 99:1–8
- BARR HJ, WALL EM, WOOLLEY SC (2021) Dopamine in the songbird auditory cortex shapes auditory preference. *Curr Biol* 31(20):4547–4559
- BARTOLUCCI M, SMITH AT (2011) Attentional modulation in visual cortex is modified during perceptual learning. *Neuropsychologia* 49(14):3898–3907
- BEITEL RE, SCHREINER CE, CHEUNG SW, WANG X, MERZENICH MM (2003) Reward-dependent plasticity in the primary auditory cortex of adult monkeys trained to discriminate temporally modulated signals. *Proc Natl Acad Sci USA* 100(19):11070–11075
- BRADLOW AR, BENT T (2008) Perceptual adaptation to non-native speech. *Cognition* 106(2):707–729
- BRUNK MGK, DEANE KE, KISSE M, DELIANO M, OHL FW, LIPPERT MT ET AL (2019) Ventral tegmental area modulates frequency-specific gain of thalamocortical input in deep layers of the auditory cortex. *Sci Rep* 7:20385
- BYERS A, SERENCES JT (2014) Enhanced attentional gain as a mechanism for generalized perceptual learning in human visual cortex. *J Neurophysiol* 112(5):1217–1227
- CAMPBELL MJ, LEWIS DA, FOOTE SL, MORRISON JH (1987) Distribution of choline acetyltransferase-, serotonin-, dopamine-beta-hydroxylase-, tyrosine hydroxylase-immunoreactive fibers in monkey primary auditory cortex. *J Comp Neurol* 261:209–220

- CARAS ML (2013) Estrogenic modulation of auditory processing: a vertebrate comparison. *Front Neuroendocrinol* 34:285–299
- CARAS ML, SANES DH (2017) Top-down modulation of sensory cortex gates perceptual learning. *Proc Natl Acad Sci USA* 114(37):9972–9977
- CARAS ML, SANES DH (2019) Neural variability limits adolescent skill learning. *J Neurosci* 39(15):2889–2902
- CHANDRASEKARAN B, YI HG, BLANCO NJ, MCGEARY JE, MADDOX WT (2015) Enhanced procedural learning of speech sound categories in a genetic variant of FOXP2. *J Neurosci* 35(20):7808–7812
- CHANDRASEKARAN B, YI HG, MADDOX WT (2014) Dual-learning systems during speech category learning. *Psychon Bull Rev* 21(2):488–495
- CHAO A, PAON A, REMAGE-HEALEY L (2014) Dynamic variation in forebrain estradiol levels during song learning. *Dev Neurobiol* 75(3):271–286
- DAVID SV, FRITZ JB, SHAMMA SA (2012) Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc Natl Acad Sci U S A* 109(6):2144–2149
- DAVIS MH, GASKELL MG (2009) A complementary systems account of word learning: neural and behavioural evidence. *Philos Trans R Soc Lond B Biol Sci* 364:3773–3800
- DAVIS SED, SANSONE J, HURLEY LM (2021) Postweaning isolation alters the responses of auditory neurons to serotonergic modulation. *Integr Comp Biol* 61:302–315
- DE BOURNONVILLE C, MCGRATH A, REMAGE-HEALEY L (2020) Testosterone synthesis in the female songbird brain. *Horm Behav* 121:104716
- DE GROOF G, BALTHAZART J, CORNIL CA, VAN DER LINDEN A (2017) Topography and lateralized effect of acute aromatase inhibition on auditory processing in a seasonal songbird. *J Neurosci* 37:4243–4254
- DELIANO M, BRUNK MGK, EL-TABBAL M, ZEMPELTZI MM, HAPPEL MFK, OHL FW (2020) Dopaminergic neuromodulation of high gamma stimulus phase-locking in gerbil primary auditory cortex mediated by D1/D5-receptors. *Eur J Neurosci* 51(5):1315–1327
- ESCUDEIRO P (2001) The role of the input in the development of L1 and L2 sound contrasts: Language-specific cue weighting for vowels. In AHJ Do, L Dominguez, & A Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press
- FENG G, YI HG, CHANDRASEKARAN B (2019) The role of the human auditory corticostriatal network in speech learning. *Cereb Cortex* 29(10):4077–4089
- FERNÁNDEZ-VARGAS M (2017) Rapid effects of estrogens and androgens on temporal and spectral features in ultrasonic vocalizations. *Horm Behav* 94:69–83
- FLEMING G, WRIGHT BA, WILSON DA (2019) The value of homework: exposure to odors in the home cage enhances odor discrimination learning in mice. *Chem Senses* 44:135–143
- FONE KCF, PORKESS MV (2008) Behavioural and neurochemical effects of post-weaning social isolation in rodents—relevance to developmental neuropsychiatric disorders. *Neurosci Biobehav Rev* 32:1087–1102
- FRITZ J, SHAMMA S, ELHILALI M, KLEIN D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 6(11):1216–1223
- FRITZ JB, ELHILALI M, SHAMMA SA (2005) Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *J Neurosci* 25(33):7623–7635
- GLEASON ED, FUXJAGER MJ, OYEBILE TO, MARLER CA (2009) Testosterone release and social context: when it occurs and why. *Front Neuroendocrinol* 30:460–469
- GUEDICHE S, BLUMSTEIN SE, FIEZ JA, HOLT LL (2014) Speech perception under adverse conditions: insights from behavioral, computational, and neuroscience research. *Front Syst Neurosci* 7:126
- GUO W, CLAUSE AR, BARTH-MARON A, POLLEY DB (2017) A corticothalamic circuit for dynamic switching between feature detection and discrimination. *Neuron* 95:180–194.e5
- HALL IC, SELL GL, HURLEY LM (2011) Social regulation of serotonin in the auditory midbrain. *Behav Neurosci* 125:501–511
- HANSON JL, HURLEY LM (2014) Context-dependent fluctuation of serotonin in the auditory midbrain: the influence of sex, reproductive state and experience. *J Exp Biol* 217:526–535
- HAPPEL MFK (2016) Dopaminergic impact on local and global cortical circuit processing during learning. *Behav Brain Res* 299:32–41
- HAPPEL MFK, DELIANO M, HANDSCHUH J, OHL FW (2014) Dopamine-modulated recurrent corticoefferent feedback in primary sensory cortex promotes detection of behaviorally relevant stimuli. *J Neurosci* 34:1234–1247
- HIRSCHENHAUSER K, OLIVEIRA R (2006) Social modulation of androgens in male vertebrates: meta-analyses of the challenge hypothesis. *Anim Behav* 71:265–277
- HOMMA N, HAPPEL MFK, NODAL FR, OHL FW, KING AJ, BAJO VM (2017) A role for auditory corticothalamic feedback in the perception of complex sounds. *J Neurosci* 37(25):6149–6161
- HURLEY LM, SULLIVAN MR (2012) From behavioral context to receptors: serotonergic modulatory pathways in the IC. *Front Neural Circuits* 6:58
- IDEMARU K, HOLT LL (2011) Word recognition reflects dimension-based statistical learning. *J Exp Psychol Hum Percept Perform* 37(6):1939–1956
- IDEMARU K, HOLT LL (2014) Specificity of dimension-based statistical learning in word recognition. *J Exp Psychol Hum Percept Perform* 40(3):1009–1021
- IDEMARU K (2020) Holt LL (2020) Generalization of dimension-based statistical learning. *Atten Percept Psychophys* 82(4):1744–1762
- IRVINE DRF (2018) Auditory perceptual learning and changes in the conceptualization of auditory cortex. *Hear Res* 366:3–16
- KANDLER K (ED) (2019) *The Oxford handbook of the auditory brainstem*. Oxford University Press, New York, NY
- KEESOM SM, HURLEY LM (2016) Socially induced serotonergic fluctuations in the male auditory midbrain correlate with female behavior during courtship. *J Neurophysiol* 115:1786–1796
- KEESOM SM, MORNINGSTAR MD, SANDLAIN R ET AL (2018) Social isolation reduces serotonergic fiber density in the inferior colliculus of female, but not male, mice. *Brain Res* 1694:94–103
- KEESOM SM, SLOSS BG, ERBOWOR-BECKSEN Z, HURLEY LM (2017) Social experience alters socially induced serotonergic fluctuations in the inferior colliculus. *J Neurophysiol* 118:3230–3241
- KLEPPER A, HERBERT H (1991) Distribution and origin of noradrenergic and serotonergic fibers in the cochlear nucleus and inferior colliculus of the rat. *Brain Res* 557:190–201
- KRENTZEL AA, IKEDA MZ, OLIVER TJ, KOROVESHI E, REMAGE-HEALEY L (2020) Acute neuroestrogen blockade attenuates song-induced immediate early gene expression in auditory regions of male and female zebra finches. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 206:15–31
- KRENTZEL AA, MACEDO-LIMA M, IKEDA MZ, REMAGE-HEALEY L (2018) A membrane g-protein-coupled estrogen receptor is necessary but not sufficient for sex differences in zebra finch auditory coding. *Endocrinology* 159:1360–1376
- KUCHIBHOTLA K, BATHELLIER B (2018) Neural encoding of sensory and behavioral complexity in the auditory cortex. *Curr Opin Neurobiol* 52:65–71
- KUHL PK, WILLIAMS KA, LACERDA F, STEVENS KN, LINDBLOM B (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255(5044):606–608
- LAU EF, PHILLIPS C, POEPEL D (2008) A cortical network for semantics: (de)constructing the N400. *Nat Rev Neurosci* 9:920–933
- LEHET M, HOLT LL (2020) Nevertheless, it persists: dimension-based statistical learning and normalization of speech impact different levels of perceptual processing. *Cognition* 202:104328
- LIM SJ, FIEZ JA, HOLT LL (2014) How may the basal ganglia contribute to auditory categorization and speech perception? *Front Neurosci* 8:230

- LISMAN J, GRACE AA, DUZEL E (2011) A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. *Trends Neurosci* 34:536–547
- LISMAN JE, GRACE AA (2005) The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron* 46:703–713
- LITTLE DF, CHENG HH, WRIGHT BA (2018) Inducing musical-interval learning by combining task practice with periods of stimulus exposure alone. *Attention, Perception, Psychophys* 81:344–357
- LITTLE DF, ZHANG YX, WRIGHT BA (2017) Disruption of perceptual learning by a brief practice break. *Curr Biol* 27:3699–3705.e3
- LIU R, HOLT LL (2015) Dimension-based statistical learning of vowels. *J Exp Psychol Hum Percept Perform* 41(6):1783–1798
- LLANOS F, MCHANEY JR, SCHUERMAN WL, YI HG, LEONARD MK, CHANDRASEKARAN B (2020) Non-invasive peripheral nerve stimulation selectively enhances speech category learning in adults. *NPJ Sci Learn* 5:12
- MACEDO-LIMA M, REMAGE-HEALEY L (2020) Auditory learning in an operant task with social reinforcement is dependent on neuroestrogen synthesis in the male songbird auditory cortex. *Horm Behav* 121:104713
- MACEDO-LIMA M, REMAGE-HEALEY L (2021) Dopamine modulation of motor and sensory cortical plasticity among vertebrates. *Integrative and Comparative Biology* Jul 23;61(1):316–336
- MADDOX WT, DAVID A (2005) Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. *J Exp Psychol Learn Mem Cogn* 31(1):100
- MANEY DL, RODRIGUEZ-SALTOS CA (2016) Hormones and the incentive salience of birdsong. In: Bass AH (ed) *Hearing and Hormones*. Springer, Heidelberg, pp 101–132
- MCCLELLAND JL, MCNAUGHTON BL, O'REILLY RC (1995) Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev* 102:419–457
- MESTRES-MISSE A, RODRIGUEZ-FORNELLS A, MUNTE TF (2007) Watching the brain during meaning acquisition. *Cereb Cortex* 17:1858–1866
- MITLÖHNER, J, KAUSHIK R, NIEKISCH H, BLONDIAUX A, GEE CE, HAPPEL MFK, GUNDELFINGER E, DITYATEV A, FRISCHKNECHT R, SEIDENBECHER C (2020) Dopamine receptor activation modulates the integrity of the perisynaptic extracellular matrix at excitatory synapses. *Cells* 9(260)
- MUKAI I, KIM D, FUKUNAGA M, JAPEE S, MARRETT S, UNGERLEIDER LG (2007) Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. *J Neurosci* 27(42):11401–11411
- NAGY WE, ANDERSON RC, HERMAN PA (1985) Learning words from context. *Read Res Q* 20:233–255
- NAGY WE, ANDERSON RC, HERMAN PA (1987) Learning word meanings from context during normal reading. *Am Educ Res J* 24:237–270
- NELKEN I (2020) From neurons to behavior: the view from auditory cortex. *Curr Opin Physiol* 18:37–41
- NIEDERKOFLER V, ASHER TE, OKATY BW ET AL (2016) Identification of serotonergic neuronal modules that affect aggressive behavior. *Cell Rep* 17:1934–1949
- NIU H, LI H, SUN L, SU Y, HUANG J, SONG Y (2014) Visual learning alters the spontaneous activity of the resting human brain: an fNIRS study. *Biomed Res Int* 2014:631425
- PETERSEN CL, KOO A, PATEL B, HURLEY LM (2020) Serotonergic innervation of the auditory midbrain: dorsal raphe subregions differentially project to the auditory midbrain in male and female mice. *Brain Struct Funct* 225:1855–1871
- PHILLIPSON OT, KILPATRICK IC, JONES MW (1987) Dopaminergic innervation of the primary visual cortex in the rat, and some correlations with human cortex. *Brain Res Bull* 18:621–633
- POLLEY DB, STEINBERG EE, MERZENICH MM (2006) Perceptual learning directs auditory cortical map reorganization through top-down influences. *J Neurosci* 26(18):4970–4982
- PRESTON AR, EICHENBAUM H (2013) Interplay of hippocampus and prefrontal cortex in memory. *Curr Biol* 23:R764–R773
- RECANZONE GH, SCHREINER CE, MERZENICH MM (1993) Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J Neurosci* 13(1):87–103
- REMAGE-HEALEY L, BASS AH (2006) From social behavior to neural circuitry: steroid hormones rapidly modulate advertisement calling via a vocal pattern generator. *Horm Behav* 50:432–441
- REMAGE-HEALEY L, COLEMAN MJ, OYAMA RK, SCHLINGER BA (2010) Brain estrogens rapidly strengthen auditory encoding and guide song preference in a songbird. *Proc Natl Acad Sci U S A* 107:3852–3857
- REMAGE-HEALEY L, DONG SM, CHAO A, SCHLINGER BA (2012) Sex-specific, rapid neuroestrogen fluctuations and neurophysiological actions in the songbird auditory forebrain. *J Neurophysiol* 107:1621–1631
- REMAGE-HEALEY L, JOSHI NR (2012) Changing neuroestrogens within the auditory forebrain rapidly transform stimulus selectivity in a downstream sensorimotor nucleus. *J Neurosci* 32:8231–8241
- REMAGE-HEALEY L, MAIDMENT NT, SCHLINGER BA (2008) Forebrain steroid levels fluctuate rapidly during social interactions. *Nat Neurosci* 11:1327–1334
- RIPOLLES P, FERRERI L, MAS-HERRERO E ET AL (2018) Intrinsically regulated learning is modulated by synaptic dopamine signaling. *Elife* 7:e38113
- RIPOLLES P, MARCO-PALLARES J, ALICART H, TEMPELMANN C, RODRIGUEZ-FORNELLS A, NOESSELT T (2016) Intrinsic monitoring of learning success facilitates memory encoding via the activation of the SN/VTA-Hippocampal loop. *Elife* 5:e17441
- RIPOLLES P, MARCO-PALLARES J, HIELSCHER U ET AL (2014) The role of reward in word learning and its implications for language acquisition. *Curr Biol* 24:2606–2611
- RODRIGUEZ-FORNELLS A, CUNILLERA T, MESTRES-MISSE A, DE DIEGO-BALAGUER R (2009) Neurophysiological mechanisms involved in language learning in adults. *Philos Trans R Soc Lond B Biol Sci* 364:3711–3735
- SALAMONE JD, CORREA M, MINGOTE SM, WEBER SM (2005) Beyond the reward hypothesis: alternative functions of nucleus accumbens dopamine. *Curr Opin Pharmacol* 5:34–41
- SALDANHA CJ, REMAGE-HEALEY L, SCHLINGER BA (2011) Synaptocrine signaling: steroid synthesis and action at the synapse. *Endocr Rev* 32:532–549
- SALDANHA CJ, TUEK MJ, KIM YH, FERNANDES AO, ARNOLD AP, SCHLINGER BA (2000) Distribution and regulation of telencephalic aromatase expression in the zebra finch revealed with a specific antibody. *J Comp Neurol* 423:619–630
- SALDEITIS K, JESCHKE M, BUDINGER E, OHL FW, HAPPEL MFK (2021) Laser-induced apoptosis of corticothalamic neurons in layer VI of auditory cortex impact on cortical frequency processing. *Front Neural Circuits* 15:659280
- SCHÄFER R, VASILAKI E, SENN W (2007) Perceptual learning via modification of cortical top-down signals. *PLoS Comput Biol* 3(8):e165
- SCHWEIG H, BRECHMANN A, BROSCHE M, BUDINGER E, OHL FW, SELEZNEVA E ET AL (2011) Behavioral semantics of learning and crossmodal processing in auditory cortex: the semantic processor concept. *Hear Res* 271:3–15
- SCHICKNICK H, SCHOTT BH, BUDINGER E, SMALLA K-H, RIEDEL A, SEIDENBECHER CI, SCHWEIG H, GUNDELFINGER E, TISCHMEYER W (2008) Dopaminergic modulation of auditory cortex-dependent memory consolidation through mTOR. *Cereb Cortex* 18(11):2646–2658
- SCHOFIELD BR, HURLEY L (2018) Circuits for modulation of auditory function. In: Oliver DL, Cant NB, Fay RR, Popper AN (eds) *The mammalian auditory pathways*. Springer International Publishing, Cham, pp 235–267

- SCHULTZ W (2015) Neuronal reward and decision signals: from theories to data. *Physiol Rev* 95:853–951
- SCREVEN LA, DENT ML (2019) Perception of ultrasonic vocalizations by socially housed and isolated mice. *eNeuro* 6:49–19
- SISNEROS JA (2009) Adaptive hearing in the vocal plainfin midshipman fish: getting in tune for the breeding season and implications for acoustic communication. *Integr Zool* 4:33–42
- STARK H, SCHEICH H (1997) Dopaminergic and serotonergic neurotransmission systems are differentially involved in auditory cortex learning: a long-term microdialysis study of metabolites. *J Neurochem* 68(2):691–697
- STURDY CB, PHILLMORE LS, SARTOR JJ, WEISMAN RG (2001) Reduced social contact causes auditory perceptual deficits in zebra finches, *Taeniopygia guttata*. *Anim Behav* 62:1207–1218
- SZPIRO SF, WRIGHT BA, CARRASCO M (2014) Learning one task by interleaving practice with another task. *Vision Res* 101:118–124
- TAKAHASHI E, OHKI K, KIM DS (2007) Diffusion tensor studies dissociated two fronto-temporal pathways in the human memory system. *Neuroimage* 34:827–838
- ULLMAN MT (2020) The declarative/procedural model: a neurobiologically motivated theory of first and second language. In: Van Paten B, Keating G, Wulff S, editors. *Theories in second language acquisition: an introduction*. Routledge p. 128–61
- VAHABA DM, HECSH A, REMAGE-HEALEY L (2020) Neuroestrogen synthesis modifies neural representations of learned song without altering vocal imitation in developing songbirds. *Sci Rep* 10:3602
- VAHABA DM, MACEDO-LIMA M, REMAGE-HEALEY L (2017) Sensory coding and sensitivity to local estrogens shift during critical period milestones in the auditory cortex of male songbirds. *eNeuro* 4
- VALLABHA GK, MCCLELLAND JL, PONS F, WERKER JF, AMANO S (2007) Unsupervised learning of vowel categories from infant-directed speech. *Proc Natl Acad Sci U S A* 104(33):13273–13278
- VICKERY TJ, CHUN MM, LEE D (2011) Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron* 72:166–177
- WALKER DM, CUNNINGHAM AM, GREGORY JK, NESTLER EJ (2019) Long-term behavioral effects of post-weaning social isolation in males and females. *Front Behav Neurosci* 13:66
- WANG F, HUANG J, LV Y, MA X, YANG B, WANG E, DU B, LI W, SONG Y (2016) Predicting perceptual learning from higher-order cortical processing. *Neuroimage* 124(Pt A):682–692
- WINGFIELD JC (2005) A continuing saga: the role of testosterone in aggression. *Horm Behav* 48:253–255
- WRIGHT BA, SABIN AT (2007) Perceptual learning: how much daily training is enough? *Exp Brain Res* 180:727–736
- WRIGHT BA, BAESE-BERK MM, MARRONE N, BRADLOW AR (2015) Enhancing speech learning by combining task practice with periods of stimulus exposure without practice. *J Acoust Soc Am* 138:928–937
- WRIGHT BA, LEBLANC EK, LITTLE DF, CONDERMAN JS, GLAVIN CC (2019) Semi-supervised learning of a nonnative phonetic contrast: how much feedback is enough? *Attention, Perception, Psychophys* 81:927–934
- WRIGHT BA, SABIN AT, ZHANG Y, MARRONE N, FITZGERALD MB (2010) Enhancing perceptual learning by combining practice with periods of additional sensory stimulation. *J Neurosci* 30:12868–12877
- YAGUE JG, MUÑOZ A, DE MONASTERIO-SCHRADER P, DEFELIPE J, GARCIA-SEGURA LM, AZCOITIA I (2006) Aromatase expression in the human temporal cortex. *Neuroscience* 138:389–401
- YETERIAN EH, PANDYA DN (1998) Corticostriatal connections of the superior temporal region in rhesus monkeys. *J Comp Neurol* 399(3):384–402
- YI HG, MADDOX WT, MUMFORD JA, CHANDRASEKARAN B (2016) The role of corticostriatal systems in speech category learning. *Cereb Cortex* 26(4):1409–1420
- YIN P, FRITZ JB, SHAMMA SA (2014) Rapid spectrotemporal plasticity in primary auditory cortex during behavior. *J Neurosci* 34(12):4396–4408
- ZEMPELTZI MM, KISSE M, BRUNK MGK, GLEMSER C, AKSIT S, DEANE E ET AL (2020) Task rule and choice are reflected by layer-specific processing in rodent auditory cortical microcircuits. *CommBiol* 345:1–21
- ZHANG X, WU YC, HOLT LL (2021) The learning signal in perceptual tuning of speech: bottom up versus top-down information. *Cogn Sci* 45(3):e12947

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.