



Long-term priors constrain category learning in the context of short-term statistical regularities

Casey L. Roark^{1,2} · Lori L. Holt^{1,2,3}

Accepted: 27 April 2022
© The Psychonomic Society, Inc. 2022

Abstract

Cognitive systems face a constant tension of maintaining existing representations that have been fine-tuned to long-term input regularities and adapting representations to meet the needs of short-term input that may deviate from long-term norms. Systems must balance the stability of long-term representations with plasticity to accommodate novel contexts. We investigated the interaction between perceptual biases or priors acquired across the long-term and sensitivity to statistical regularities introduced in the short-term. Participants were first passively exposed to short-term acoustic regularities and then learned categories in a supervised training task that either conflicted or aligned with long-term perceptual priors. We found that the long-term priors had robust and pervasive impact on categorization behavior. In contrast, behavior was not influenced by the nature of the short-term passive exposure. These results demonstrate that perceptual priors place strong constraints on the course of learning and that short-term passive exposure to acoustic regularities has limited impact on directing subsequent category learning.

Keywords Category learning · Statistical learning · Audition · Perceptual priors

Introduction

The natural world is structured – rain is nearly always accompanied by dark clouds; the words a speaker says are temporally aligned with their mouth movements. This structure is useful to learn because it keeps us from getting caught outside without an umbrella and helps us understand what someone is saying in a noisy restaurant. Perceptual systems are sensitive to input regularities at multiple levels, encoding both long-term regularities across a lifetime of experience and short-term regularities within individual contexts. This enables stability across the long term and flexibility in the short term when we encounter regularities that may deviate from long-term norms, such as when traveling to a location

with a novel climate or encountering a new speaker who has an accent.

Across the long term, sensory systems efficiently encode natural signal statistics in vision (Simoncelli, 2003; Simoncelli & Olshausen, 2001), audition (Kluender et al., 2013; Lewicki, 2002; Ming & Holt, 2009; Smith & Lewicki, 2006; Stilp & Lewicki, 2014; Wang, 2007), and across multiple modalities (Ernst & Banks, 2002). For example, auditory cochlear filters resemble filters optimized to code for the regularities of natural sounds (Smith & Lewicki, 2006) and human speech recognition is “efficient” in the sense that it is supported when signal regularities align with these filters (Ming & Holt, 2009).

The sensitivity to long-term sensory statistics, which might be understood as *priors*, introduces observable biases in perception. For example, the McGurk effect (McGurk & Macdonald, 1976) is observed when spoken instances of syllables (e.g., /ba/ or /ga/) are paired with mouth movements that conflict with expectations developed across long-term alignment of speech sounds and mouth position. The resulting effect is that the visual input biases speech perception toward the percept that better matches the prior (e.g., hearing /ba/ and seeing /ga/ leads to an intermediate percept of /da/). Thus, long-term statistics like the congruency of auditory

✉ Casey L. Roark
casey.l.roark@gmail.com

¹ Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA

² Center for the Neural Basis of Cognition, Pittsburgh, PA, USA

³ Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA

and visual speech come to be reflected in cognitive and neural representations of the sensory world.

People are also sensitive to short-term regularities in sensory environments. Many studies demonstrate sensitivity to evolving statistical regularities in the input across infants (Adriaans & Swingle, 2017; Aslin et al., 1998; Maye et al., 2002; McMurray et al., 2009; Toscano & McMurray, 2010), adults (Escudero & Williams, 2014; Wanrooij & Boersma, 2013), and non-human animals (Pons, 2006). Learning of short-term regularities in the sensory world involves the *rapid* learning of *novel* input regularities, which can occur even in a *passive* manner (Barlow & Földiák, 1989; Coen-Cagli et al., 2015; Lu et al., 2019; Stilp et al., 2010). Rapid, and putatively efficient, adaptation to short-term statistical regularities has been examined in the auditory modality where perception appears to be able to rapidly adapt to short-term statistical structure over as few as 2 min of passive exposure (Stilp et al., 2010; Stilp & Kluender, 2012, 2016).

However, it would not be adaptive for short-term experience to overwrite long-term representations: perception requires maintenance of long-term regularities to support *stable* representations, while also remaining *flexible* to short-term regularities that may deviate from the priors developed across the long term. Efficient and rapid processing of a complex sensory world thus requires balance across long-term and short-term regularities. However, how short-term exposure to novel statistical regularities interacts with long-term priors is not well understood.

Perceptual category learning provides an excellent testbed of the interaction of long-term priors and short-term statistical learning, as category learning is influenced by prior experience. For example, learning second language speech categories is more difficult when categories directly conflict with one's native language (Best et al., 2001; Kuhl et al., 2007). Outside of language contexts, existing representations of sensory dimensions affect how people learn categories based on those dimensions (Ell et al., 2012; Holt et al., 2004; Roark et al., 2022; Roark & Holt, 2019b; Scharinger et al., 2013). For example, over a lifetime of experience, we develop perceptual biases and priors in the representation of pairs of dimensions as integral or separable (Garner, 1974, 1976). Consequently, integral dimensions (e.g., saturation and brightness) are difficult to separate into their component dimensions, whereas separable dimensions (e.g., length and orientation of a line) are difficult to combine and may be separated automatically (Lockhead, 1972; Nelson, 1993). These priors influence behavior in short-term category-learning contexts – categories that require selective attention to dimensions are more difficult to learn when the dimensions are integral than separable, and categories that require integration across dimensions are more difficult

to learn when the dimensions are separable than integral (Ashby & Maddox, 1990; Ell et al., 2012).

In the current study, we investigate the influence of long-term priors and short-term statistical learning on perceptual category learning. Specifically, we use perceptual category learning to examine whether representations efficiently adapt to short-term regularities or whether long-term priors are stably maintained in the face of novel short-term input regularities. We do so by aligning (and misaligning) long-term priors with category exemplar distributions in a category-learning task.

With regard to long-term perceptual priors, we capitalize on the observation that spectral and temporal modulation are interdependent in auditory representations. Each thought to be a fundamental component of sound, *spectral modulation* reflects oscillations in power across the frequency spectrum at particular times and *temporal modulation* reflects oscillations in amplitude across time (Woolley et al., 2005). At some level, the neural populations encoding these dimensions are relatively separable (Depireux et al., 2001; Elliott & Theunissen, 2009; Langers et al., 2003; Schönwiesner & Zatorre, 2009; Visscher et al., 2007; Woolley et al., 2005), but their representations may be interdependent. Specifically, neurons that code *high* temporal modulation also code *low* spectral modulation (and vice versa; Allen et al., 2018; Hullett et al., 2016). As a result, the long-term representation of these dimensions comprises a prior wherein representations are stretched along the negative axis (e.g., *high* temporal modulation is associated with *low* spectral modulation) and shrunk along the positive axis relative to a naïve, untrained space (Fig. 1A). A perceptual prior like this may influence category learning such that categories aligned with the prior (i.e., the distinction between the categories is along the negative axis in Fig. 1A) may be easier to learn than categories that are misaligned with the prior (i.e., the distinction between the categories is along the positive axis, Fig. 1A). Another related possibility is that learners may demonstrate biases in how much they rely upon each dimension in making category decisions (Roark & Holt, 2019b). We assess this latter possibility using decision-bound models (Ashby, 1992).

There is limited research on the impact of perceptual priors on short-term statistical learning. It is also not well understood how short-term statistical learning may influence more overt behavior such as category learning. Here, we expose listeners to brief (~8 min) exposure to a statistical regularity prior to an overt category-learning task involving stimuli sampled from the same acoustic space as exposure stimuli, with category distributions aligned or misaligned with the long-term prior. This allows us to examine three hypotheses regarding the intersection of long-term priors, statistical learning, and category learning (Fig. 1B).

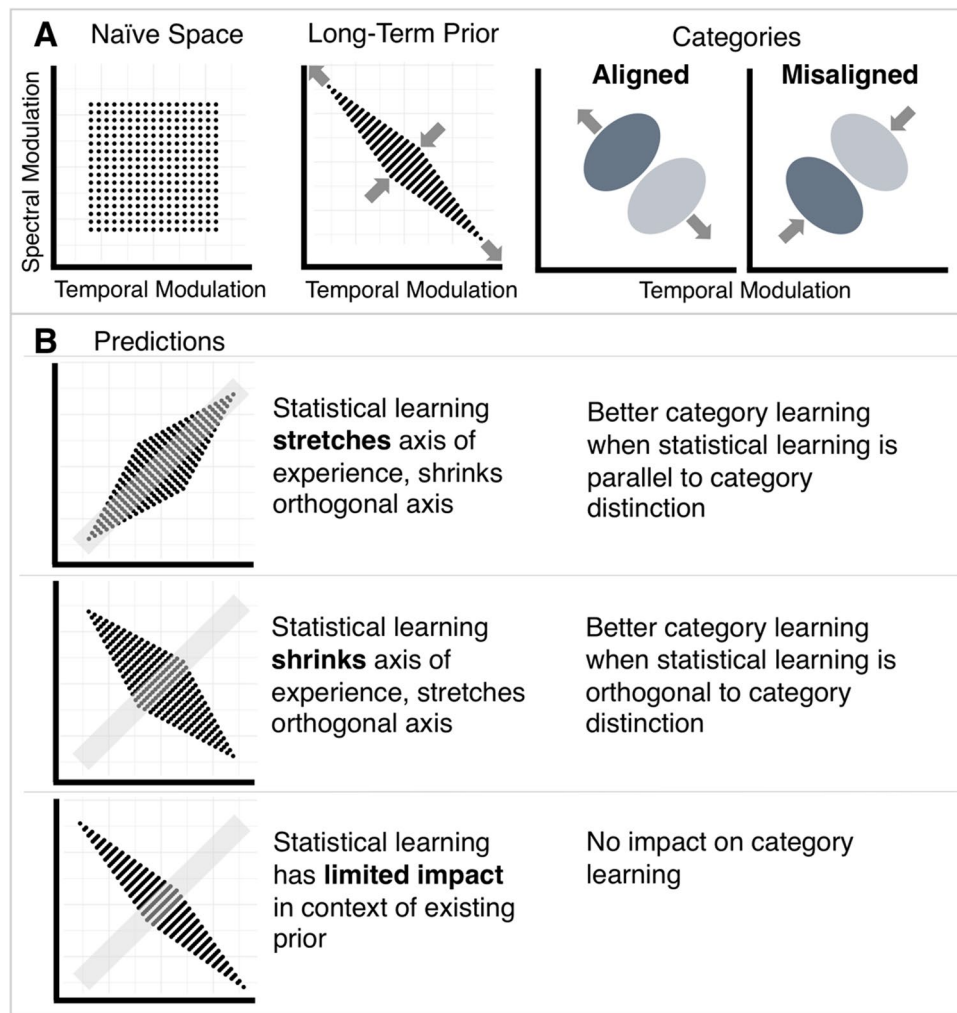


Fig. 1 Framework and predictions. *Note.* (A) Illustration of the interaction between long-term prior (relative to naïve physical space) and category learning with distributions that are aligned or misaligned with the prior. Categories that are Aligned with the prior are more distinguishable and should be easier to learn. Categories that are Mis-

aligned with the prior are less distinguishable and should be more difficult to learn. (B) Illustration of the different predictions for the influence of statistical learning on category learning. Example shows statistical learning distribution with positive correlation in gray

The first hypothesis is associated with an efficient coding perspective. Specifically, there are reasons to expect that short-term passive exposure will influence representations in a way that will influence behavior. Prior studies have shown that even as little as 2 min of passive exposure to a correlation between two acoustic dimensions can increase discriminability along that correlation (e.g., stretch axis in representations) and decrease discriminability orthogonal to that correlation (e.g., shrink axis in representations) (Stilp et al., 2010; Stilp & Kluender, 2012, 2016) and has been linked to efficient coding in single neurons in auditory cortex in animal studies (Lu et al., 2019). This is also consistent with studies that demonstrate that experience with variability along a relevant feature prior to category learning can improve learning (Antoniou & Wong, 2016; Holt & Lotto,

2006). In line with this prediction, we would expect that statistical learning experience will stretch whichever axis is being experienced and shrink the orthogonal axis. As a result of this, category learning will be better when the statistical learning distribution is parallel to the category distinction that needs to be learned.

In contrast, a second hypothesis predicts the opposite pattern – that statistical learning experience will shrink the axis of experience and stretch the orthogonal axis. This hypothesis stems from research that has shown that experience with variability along a dimension makes this dimension less reliable in subsequent category-learning contexts (Rost & McMurray, 2010). That is, the more variability one experiences along a specific dimension, the less informative the dimension for behavior. As a result, we would expect that

a statistical learning distribution that is orthogonal to the category distinction will support category learning.

Our final hypothesis is that long-term priors will override any influence of short-term statistical learning. This would support an interaction between long-term priors and short-term statistical learning. In Stilp and Kluender (2016), the effects of passive experience on discrimination diminished within 128 trials of discrimination testing, suggesting that even if statistical learning influences the representational space, the impact is quick to revert to alignment with existing long-term representations. This would predict that short-term statistical learning will not influence category learning.

In summary, long-term experience (such as native language experience) and short-term statistical learning have each been linked to perceptual warping of physical input space (Feldman et al., 2021; Kuhl, 2000; Kuhl et al., 2007; Maye et al., 2008). Yet, it is not yet clear how long-term priors and short-term statistical learning may independently or interactively influence novel category learning. Here, we test three competing hypotheses to understand how short-term statistical learning interacts with long-term priors and the behavioral demands of overt category learning.

Methods

This experiment examines differences in category learning in the same two-dimensional acoustic space as a function of (1) short-term statistical learning of a regularity between the two dimensions and (2) category distribution type is aligned or misaligned with long-term priors. We trained participants on one of two pairs of category distributions, which were identical in their statistical regularities and differed only in their orientation in the input space. These categories are multidimensional, in that the category identity cannot be determined by a single dimension. We refer to the categories based on whether the category distinction is Aligned or Misaligned with long-term priors. Stimuli and data are available at osf.io/qyg7z/ (Roark & Holt, 2022).

Participants

Participants were 305 (102 male, 201 female, two prefer not to answer) Carnegie Mellon University undergraduates ages 18–29 years and were given \$10 or course credit for participating. All participants gave informed consent and the experimental protocols were approved by the Institutional Review Board at Carnegie Mellon University. Participants were randomly assigned to one of five statistical learning conditions (Naïve, Positive, Negative, Spectral, or Temporal) and one of two category types (Aligned, Misaligned). In the statistical learning phase, with the exception of the Naïve condition, participants passively experienced

Table 1 Number of participants in each condition

Exposure condition	Category type	<i>N</i>
Spectral	Aligned	29
Temporal	Aligned	32
Positive	Aligned	29
Negative	Aligned	29
Naïve	Aligned	30
Spectral	Misaligned	32
Temporal	Misaligned	31
Positive	Misaligned	31
Negative	Misaligned	32
Naïve	Misaligned	30

specific statistical regularities in the acoustic space – variability along either one dimension (Temporal or Spectral modulation) or along both dimensions (Positive or Negative correlation). A power analysis was conducted with the *WebPower* package in R (Zhang & Mai, 2018) and indicated that to detect an interaction between statistical regularity type and category type with a medium effect size ($f = .25$), a sample of at least 26 participants would be needed in each group to obtain statistical power at a .90 level with an alpha of .05. We exceeded this target recruitment for each group (Table 1), with approximately 30 participants in each of ten conditions. Nine additional participants were run, but not included due to experimenter or software error.

Stimuli

The stimuli were complex static acoustic ripples varying on spectral modulation and temporal modulation. The stimuli were generated using a custom MATLAB script. Stimuli were defined with the following parameters based on prior work (Yi & Chandrasekaran, 2016): duration = 1 s; phase = 0°; $F_0 = 200$ Hz; spectral bandwidth = -3.18; amplitude modulation depth = 0 dB; sampling rate = 44.1 kHz.¹ Stimuli were then root mean square (RMS) amplitude matched at 70 Hz in Praat (Boersma & Weenink, 2021). Stimuli could take on temporal modulation values from 4–12 Hz and spectral modulation values from 0.1 oct/cyc to 2 oct/cyc. Stimuli and scripts are available via the Open Science Framework. Spectrograms are shown in Fig. 2 and were created using the *phonTools* in R (Barreda, 2015).

Results of a pilot experiment indicated that discriminability was equivalent in these ranges across the two dimensions and along a perfect positive and negative correlation between

¹ Phase of 0 degrees assures that all elements are positionally aligned with one another. Spectral bandwidth is the range of spectral information around the median and is related to perception of timbre. Amplitude modulation depth reflects the variability in amplitude modulation, reflecting that amplitude modulation does not change within a stimulus.

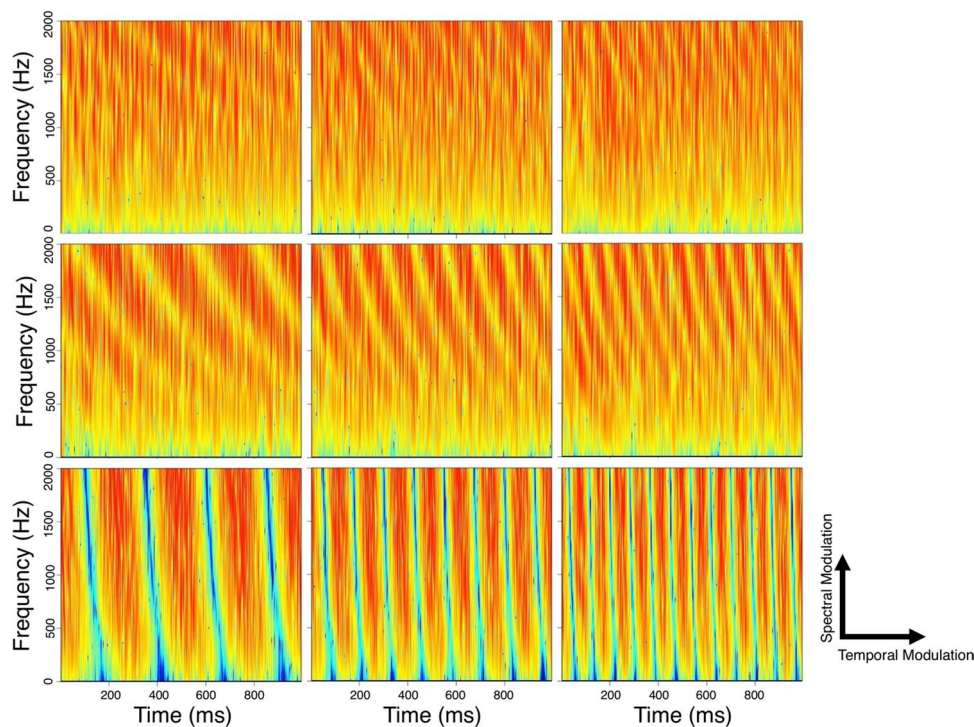


Fig. 2 Stimuli spectrograms. *Note.* Spectrograms for stimuli across grid of temporal and spectral modulation space

the dimensions. In the pilot, participants were 80 (25 male, 53 female, two prefer not to answer) Carnegie Mellon University undergraduates ages 18–25 years and were given \$10 or course credit for participating. Participants were randomly assigned to one of the four distributions (Spectral, Temporal, Positive, Negative; 20 participants per condition) and made same-different discrimination judgments of pairs of stimuli along an 18-step continuum (Fig. 3A). Participants made judgments across 496 trials (248 same, 248 different), with each “different” pair repeated twice. We calculated d' values across all stimuli for each participant using hit and false alarm rates using the *dprime* function in the *Psycho* R package (Makowski, 2018). Discriminability was equivalent across the four dimensions, indicated by the fact that d' values for the four dimensions were not statistically different, according to a one-way ANOVA ($F(3, 76) = 1.08, p = 0.36, \eta^2 = 0.041$; Fig. S2, Online Supplementary Material, OSM).

Stimulus distributions

Statistical learning distributions During the statistical learning phase participants passively listened to one of four distributions of sounds, according to condition (Positive, Negative, Temporal, Spectral). As shown in Fig. 3A, two conditions involved variation across both dimensions, with either a positive or a negative distribution reflecting a perfect ($r = 1.0, r = -1.0$) correlation between the two dimensions. The other two

conditions involved variance across only one of the acoustic dimensions. Eighteen equidistant stimuli defined each distribution. For the positive and negative distributions, one step between each of the stimuli varied 0.47 Hz along the temporal modulation dimension and 0.11 cyc/oct along the spectral modulation dimension. Temporal stimuli had a constant mean spectral modulation value of 1.05 cyc/oct, with 0.47 Hz per step. Spectral stimuli had a constant mean temporal modulation value of 8 Hz, with 0.11 cyc/oct per step.

Category learning distributions Participants learned one of two category pairs: Aligned or Misaligned (Fig. 2B). Two category pairs were created by sampling a bivariate Gaussian distribution using the *mvnorm* function in the *MASS* R package (Venables & Ripley, 2002). We sampled for a single category (100 exemplars) using normalized coordinates and then rotated and mirrored that distribution to create all other categories. Thus, both category types possess identical variance and covariance of exemplars, and the relationship between the categories is equal in terms of overlap (Table 2; Fig. S1, OSM). The categories differ in how they are aligned or misaligned with the long-term representational prior. Separate test distributions (50 exemplars/category) were sampled using the same parameters and due to random sampling have slightly different means, variance, and covariance than the training distributions (Table 2).

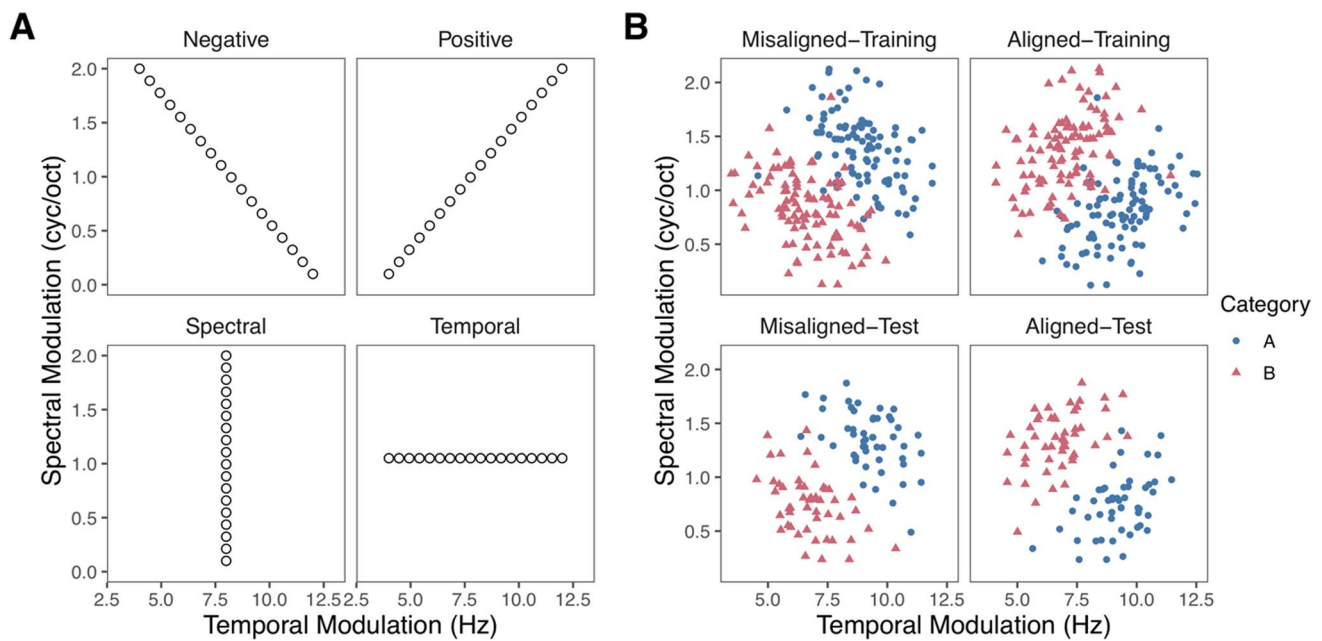


Fig. 3 Statistical learning and category learning distributions. *Note:* Stimulus distributions for the (A) statistical learning and (B) supervised category learning phases (separately for training and generalization test)

Table 2 Category distribution information

Category	M (temporal, spectral)	Variance (temporal, spectral)	Covariance
Training distributions			
Aligned: Category A	9.35, 0.83	2.01, 0.10	0.190
Aligned: Category B	7.08, 1.37	1.80, 0.11	0.190
Misaligned: Category A	8.92, 1.37	1.80, 0.11	-0.190
Misaligned: Category B	6.65, 0.83	2.01, 0.10	-0.190
Test distributions			
Aligned: Category A	9.23, 0.75	1.34, 0.077	0.15
Aligned: Category B	6.73, 1.34	1.37, 0.076	0.15
Misaligned: Category A	9.27, 1.34	1.37, 0.076	-0.15
Misaligned: Category B	6.77, 0.75	1.34, 0.077	-0.15

Procedure

During the statistical learning phase, all participants except those in the Naïve conditions passively listened to a stream of sounds with a particular statistical regularity (Positive, Negative, Temporal, Spectral) for approximately 8 min. They heard 450 presentations of sounds (25 repetitions each of 18 sounds), a repetition number that has been shown in another stimulus space to affect perceptual discriminability (Stilp et al., 2010). Each sound (1 s) was followed by a 50-ms silent intertrial interval (ITI). Participants were given markers and blank pieces of paper and told to draw whatever they wanted.

Participants next learned the categories in a supervised categorization task across eight blocks of training with 48 trials per block for a total of 384 training trials. On each trial,

participants heard a single exemplar selected randomly without replacement followed by a screen on which they were prompted about whether they believed the sound belonged to Category A or Category B. Participants indicated their category response with a key-press (*u* or *i*), with response keys for each category counterbalanced across participants. After a response was made there was a 500-ms pause after which participants were given feedback about the correctness of their response (“Correct!” or “Incorrect!”). Participants also saw boxes on the screen that were associated with the individual categories. In addition to the written feedback, a red X appeared in the box associated with the correct category. This red X was presented regardless of the correctness of the response. Feedback was displayed for 500 ms before a 1-s ITI preceding the next category exemplar. Participants

were told to use feedback to inform future category decisions. Finally, participants completed a test without feedback to assess generalization of learning to novel category exemplars.

Decision strategies

To understand how participants used the underlying dimensions in category decisions, we used decision bound computation models to assess their decision strategies (Ashby, 1992; Maddox & Ashby, 1993). These models are derived from General Recognition Theory (Ashby & Townsend, 1986) and applied widely to understand decision strategies during category learning (Ashby & Maddox, 1992; Reetzke et al., 2016; Roark & Holt, 2019a, 2019b; Yi & Chandrasekaran, 2016).

We fit several classes of decision bound models. Each model assumes participants create decision boundaries to separate the stimuli into two categories. The four classes of models that we fit were: two unidimensional rule-based models (one along the temporal modulation dimension and another along the spectral modulation dimension), an information-integration model in which both dimensions contribute to decisions, and a random responder model.

The two unidimensional models instantiate a linear decisional bound along one of the two dimensions – temporal modulation or spectral modulation. Unidimensional models have two free parameters – the decision boundary and the variance of noise (both perceptual and criterial).

The information-integration model employs a general linear classifier that assumes a linear decision boundary but, in contrast to the unidimensional models, uses both dimensions. This model is optimal for both kinds of categories in the current study. For the Positive condition, the optimal decision boundary has a positive slope whereas for the Negative condition, the optimal decision boundary has a negative slope. Both training and test distributions were subjected to decision bound modeling to ensure that the true optimal model was the one idealized by the experimenter. The integration model has three free parameters: the slope and intercept of the decision boundary and the variance of noise (perceptual and criterial).

To understand if participants were just randomly guessing, we fit a random responder model that assumes equal response probability across categories on each trial.

We fit the models separately to each participant's data for each of the training blocks and the generalization test. Model parameters were estimated using a maximum likelihood procedure (Wickens, 1982) and model selection used the Bayesian Information Criterion ($BIC = r \cdot \ln N - 2 \ln L$, where r is the number of free parameters, N is the number of trials in a given block, and L is the likelihood of the model given the data (Schwarz, 1978)). BIC applies penalties for

extra free parameters and the best-fit model was defined as the model with the lowest BIC value.

Results

To understand the interaction between priors and statistical learning, we examined how statistical learning of different acoustic regularities influenced category learning performance and decision strategies while learning categories that align or misalign with long-term perceptual priors. We tested three competing hypotheses: an efficient coding hypothesis that suggests that statistical learning experience stretches the axis of experience and stretches the orthogonal axis, improving category learning for categories that make distinctions along the axis of experience; a variability hypothesis that suggests that experience shrinks the axis of experience and stretches the orthogonal axis, improving category learning for categories that make distinctions along the orthogonal axis; and a long-term prior bias hypothesis that suggests that short-term statistical learning experience has limited impact on representations and will not impact category learning. Instead, according to the long-term prior bias hypothesis, the long-term prior may have a substantial and stable impact on learning that does not interact with statistical learning experience.

Behavioral results

To confirm that the expected long-term bias was present for these categories, we examined how participants with no exposure prior to categorization (Naïve) learned the categories (Fig. 4). Naïve participants who learned the Aligned categories had significantly better Block 1 accuracy than participants who learned the Misaligned categories (Naïve-Aligned: $M = 65\%$; Naïve-Misaligned: $M = 55\%$; $t(49.1) = 3.23$, $p = .0022$, $d = 0.83$, 95% CI [3.97, 17.0]). This finding supports the assumption that there is a long-term bias across learners for better learning of Aligned relative to Misaligned categories.

We next examined the influence of short-term statistical learning on category learning performance. To minimize potential washout effects due to experience in the categorization task (e.g., Stilp & Kluender, 2016), we examined the group differences within the first block (Fig. 4B). Using a two-way ANOVA, we examined effects of the statistical regularity (Naïve, Positive, Negative, Spectral, Temporal) and category type (Aligned, Misaligned). In line with the perceptual prior, we found an overall advantage for the Aligned categories over Misaligned categories ($F(1, 295) = 47.3$, $p < 0.0005$, $\eta_p^2 = 0.14$), such that accuracy for Aligned categories was 9.1% (95% CI: [6.5, 11.7]) higher than Misaligned categories in Block 1.

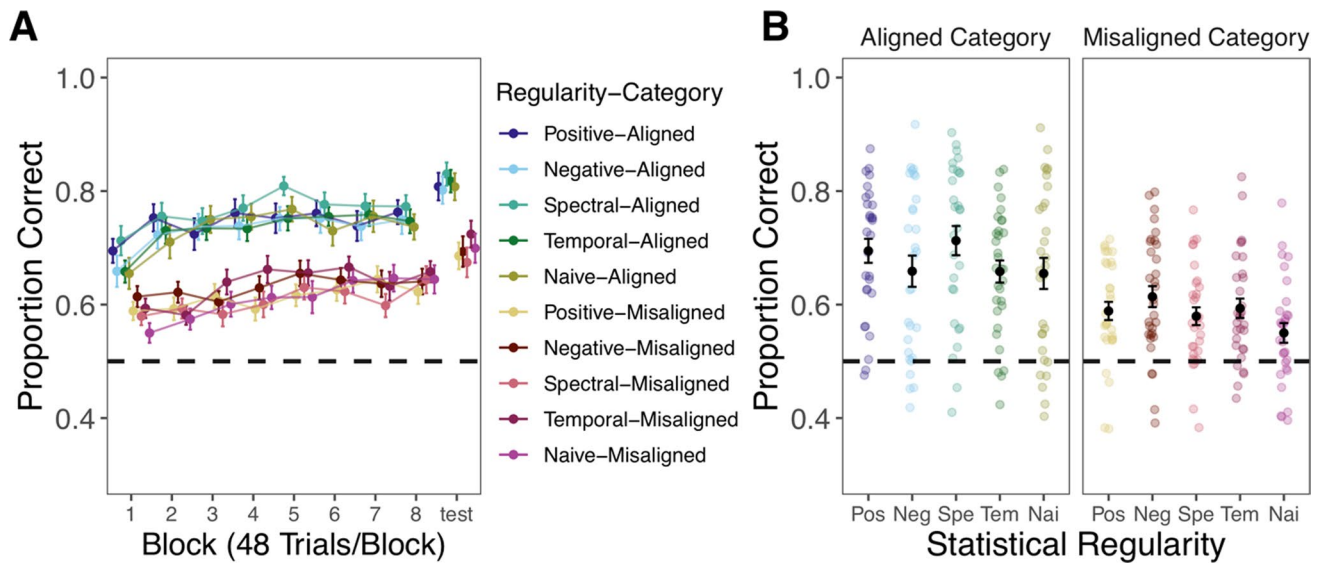


Fig. 4 Category learning performance. *Note.* (A) Mean accuracy across the eight training blocks and generalization test with chance performance (50%) denoted by a dashed line. Error bars reflect *SEM*.

Short-term statistical learning did not influence category learning performance. There was neither an effect of the type of statistical regularity ($F(4, 295) = 1.38, p = 0.24, \eta_p^2 = 0.018$) nor an interaction between regularity and category type ($F(4, 295) = 1.45, p = 0.22, \eta_p^2 = 0.019$).

We also compared learning across all training blocks (Fig. 4A), using a mixed-model ANOVA to examine the effects of statistical regularity (Naïve, Positive, Negative, Spectral, Temporal), training block (1–8), and category type (Aligned, Misaligned).² The effects observed in the first block were persistent across all blocks – there was an overall advantage for Aligned over Misaligned categories across all blocks ($F(1, 295) = 126.9, p < 0.0005, \eta_p^2 = 0.30$), exposure to different regularities in the statistical learning phase did not affect learning across blocks ($F(4, 295) = 0.22, p = 0.93, \eta_p^2 = 0.003$), and there was no interaction between statistical regularity and category type across blocks ($F(4, 295) = 1.07, p = 0.37, \eta_p^2 = 0.014$).

Participants' accuracy improved across blocks, indicated by a main effect of block ($F(5.7, 1666.9) = 25.6, p < 0.0005, \eta_p^2 = 0.080$), which was driven by a significant improvement from the first to the second block (Bonferroni-corrected $p < 0.0005$), with no other subsequent differences among adjacent blocks ($ps > 0.26$). The improvement across blocks also had a distinct pattern for those learning Aligned and Misaligned categories ($F(5.7, 1666.9) = 2.91, p = 0.009, \eta_p^2 = 0.10$). Aligned categories had more drastic improvement

(B) Accuracy in the first block with mean and *SEM* shown in black and individual subject variability shown in color for each condition

from the first to second block, whereas Misaligned categories had more gradual improvement across blocks. Critically, the type of statistical regularity did not impact the pattern of learning across blocks ($F(22.6, 1666.9) = 0.93, p = 0.55, \eta_p^2 = 0.12$), and there was no interaction between block, regularity, and category type ($F(22.6, 1666.9) = 1.045, p = 0.40, \eta_p^2 = 0.014$).

Finally, the pattern of results in the generalization test was identical to training – generalization of learning was better for Aligned than Misaligned categories ($F(1, 295) = 60.0, p < .001, \eta_p^2 = 0.17$), there was no effect of statistical regularity type ($F(4, 295) = 0.44, p = .78, \eta_p^2 = 0.0045$), or an interaction between regularity and category type ($F(4, 295) = 0.49, p = .74, \eta_p^2 = 0.066$).

To summarize, performance during the category-learning task was not impacted at any stage (even the earliest stages of learning) by short-term statistical learning of acoustic regularities via passive exposure. However, there was a persistent effect of long-term perceptual priors such that Aligned categories requiring distinctions between categories across the negative axis in spectral-temporal modulation space exhibited a learning advantage over Misaligned categories defined by the inverse relationship, even among Naïve participants.

Decision strategy results

We restrict our discussion of strategy use to Block 1, as we were primarily interested in behavior before participants received extensive feedback. Results for the other blocks can be found in the OSM. Participants used similar decision

² Mauchly's test of sphericity was significant ($p < 0.0005$), so we report the Huynh-Feldt corrected values.

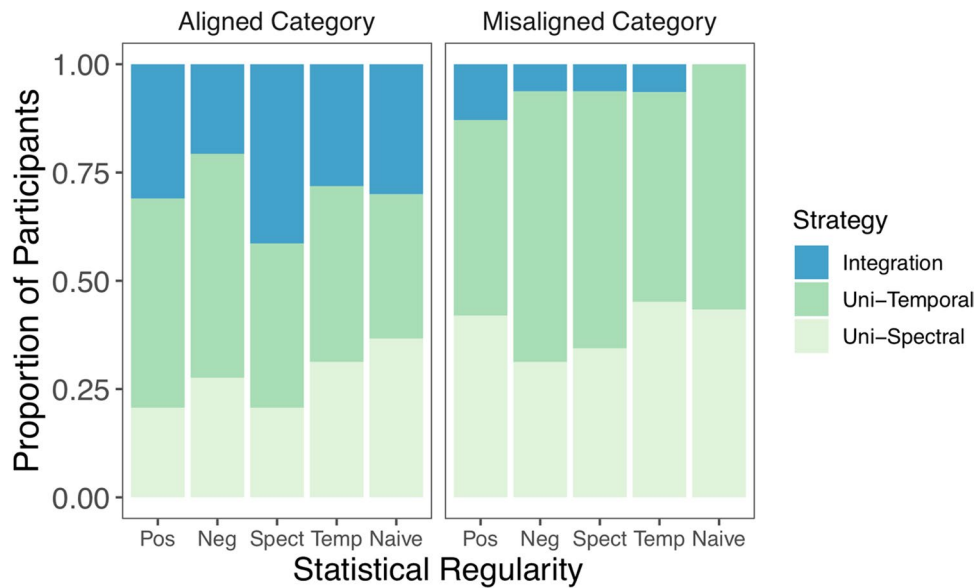


Fig. 5 Proportion of participants best-fit by each strategy in Block 1. None of the participants were best fit by a random responder model, so it is not shown

strategies, regardless of the type of statistical regularities they experienced (Fig. 5). According to Fisher's exact tests, in the first block, there were no significant differences in the strategies participants used across the five regularity conditions for the Aligned categories ($p = 0.70$) or the Misaligned categories ($p = 0.61$). Participants who learned the Aligned categories had a roughly even mix between optimal integration (30%), unidimensional-temporal (42%), and unidimensional-spectral (28%) strategies. Participants who learned the Misaligned categories primarily used unidimensional-temporal (54%) and unidimensional-spectral (39%) strategies. Only 6% of Misaligned category participants used the optimal integration strategy.³ No participants were best fit by a random responder model.

While short-term statistical learning did not influence participants' strategies, long-term priors did. According to Fisher's exact tests, strategies were significantly different for the Aligned and Misaligned categories ($p < .001$). More participants learning the Aligned categories used the optimal integration strategy than participants learning the Misaligned categories. These differences in strategies across categories persisted throughout the rest of the task (see Fig. S3, OSM). These results complement the behavioral accuracy data: accuracy was higher for the Aligned than the Misaligned categories because individuals learning the Aligned categories used optimal integration strategies while individuals learning the Misaligned categories used suboptimal unidimensional strategies.

Discussion

We investigated the interaction of long-term perceptual priors and short-term statistical learning in a category-learning task. Passive statistical learning had no impact on decision strategies or overall performance. However, there were large and persistent differences between the two statistically identical category types (Aligned, Misaligned), indicating that perceptual priors can place strong constraints on learning. Our study extends prior work on the influence of short-term statistical learning by examining the influence of this experience on relevant overt learning behavior. This study is also the first to examine the interaction of perceptual priors, statistical learning, and category learning.

Interaction between short-term and long-term regularities

Perceptual systems are sensitive to long-term (Ernst & Banks, 2002; Lewicki, 2002; Simoncelli & Olshausen, 2001; Wang, 2007) and short-term regularities (Aslin et al., 1998; Barlow & Földiák, 1989; Pons, 2006; Wanrooij & Boersma, 2013), which enables stable yet flexible perception in a complex sensory world. The current results suggest that long-term representations may be robust in the face of short-term regularities. These results are consistent with findings from the speech category learning literature that suggest that non-native speech categories that conflict with long-term native language representations are much more difficult to learn than categories that do not conflict with the native language (Best et al., 2001; Kuhl et al., 2007).

³ The values do not sum to 100% due to rounding of percentages.

Regardless of the nature of the short-term statistical learning experience, we observed large and persistent differences in the ability to learn *statistically identical* categories that differed only in the arbitrary assignment of stimuli to categories based on the rotation of the categories in the acoustic space. Participants learning Misaligned categories performed worse throughout training and used more suboptimal decision strategies than participants learning Aligned categories. These persistent differences indicate that priors reflected in the representations of these dimensions may not be shifted, moved, or otherwise substantially influenced by short-term passive experience. The bias observed in this spectral-temporal modulation space may directly relate to the long-term representations of these dimensions in auditory cortex (Allen et al., 2018; Hullett et al., 2016). Neurons in these regions may encode a joint representation of spectral-temporal modulation that may relate to enhanced efficient processing of natural sounds, such as speech.

It is informative to compare the findings of our pilot study – which examined discrimination behavior across the positive and negative axes – and the main category learning study. We found no differences in the discriminability of sounds varying across the negative and positive axes in the pilot study but found that the Aligned categories were persistently learned better than the Misaligned categories. We believe this difference stems from the nature of these two tasks. Specifically, in the pilot study, participants reported only whether the sounds were the same or different from one another. The positive and negative stimuli differed on both the temporal and spectral modulation dimensions. As a result, participants could detect differences across *either* of the dimensions. We can contrast this with the requirements in the category learning study in which participants learned arbitrary labels for categories through feedback. While participants were able to *detect* differences based on temporal and spectral modulation along the positive and negative axes, as evidenced by the pilot study, they were impaired in their ability to assign the stimuli to arbitrary categories when the category differences were misaligned with a long-term prior. It is possible that more graded measures of behavior, such as similarity judgments, may reveal differences across the positive and negative axes. Future work should address this directly.

The bias in learning category learning distributions rotated differently in space was also present in other studies with different dimensions in both auditory (Roark & Holt, 2019b) and visual modalities (Markant, 2018). Specifically, across our study and this prior work, categories that can be distinguished across the negative axis (Aligned) were learned better than categories distinguished across the positive axis (Misaligned). While these prior studies did not address this possibility, these directional biases may reflect constraints of existing representations, such that if

categories do not align with existing representations, learners will encounter more difficulty than if they align (Holt et al., 2004; Roark et al., 2022). Our results suggest that other long-term priors should also influence category learning in predictable ways. For instance, there is an association between amplitude modulation (i.e., change in rate of modulation over time) and changes in carrier frequency (i.e., changes in pitch over time), such that sounds that increase in frequency are more likely to be perceived as getting faster over time and sounds that decrease in frequency are more likely to be perceived as getting slower (Bond & Feldstein, 1982; Feldstein & Bond, 1981; Henry & McAuley, 2009; Herrmann & Johnsrude, 2018). Our results suggest that any long-term prior or bias may influence category learning with categories that are aligned with the bias being easier to learn than categories that are misaligned with the bias.

Regardless of direction, it is possible that the source of these biases could be based in hardwired functionality of the neural systems or based on the physics of the dimensions themselves (e.g., faster temporal modulations can naturally accommodate more spectral modulations). The effect could also be learned – it is possible that long-term experience with distributions in the sensory world that accentuate certain distinctions contributes to these long-term priors (Roark et al., 2022). To understand the source of these biases, more will need to be understood about the distributions along these dimensions in the natural sensory world and the nature of the neural representations.

Cognitive systems face the tension of maintaining existing representations that have been fine-tuned to the long-term input regularities and adapting representations to meet the unique needs of short-term input that may deviate from long-term norms. It would be extremely costly for a system to fundamentally change representations that do a good job of reflecting stable aspects of the environment when presented with novel information. To facilitate speech perception, listeners can rapidly adapt to the novel regularities in foreign or artificially accented speech, without overwriting their long-term representations (Clarke & Garrett, 2004; Idemaru & Holt, 2014; Liu & Holt, 2015; Norris et al., 2003; Skoruppa & Peperkamp, 2011). Even years of experience with a second language may not substantially change stable representations developed across the long-term (Idemaru et al., 2012). It is sometimes adaptive for the system not to adapt. This experiment demonstrates the robustness of some representations in response to short-term structured experience.

Nature of the short-term experience

There are several components regarding the nature of the statistical learning phase that impact the interpretation of the findings.

The experience was passive

The statistical learning phase was completely passive. It is possible that perceptual systems are sensitive to these regularities but that changes to representations or generalizability to broader cognitive behavior, such as during category learning, is not possible with passive exposure alone. Prior research on rapid efficient coding of regularities in sensory systems suggests some representational change can occur through passive exposure (Lu et al., 2019; Stilp et al., 2010, 2018; Stilp & Kluender, 2012, 2016). It could be that to see impacts in a categorization context or changes to representations when there is a strong prior, more active engagement or feedback may be needed.

Supporting this view on the limits of passive exposure, both computational modeling and behavioral work have demonstrated that general sensitivity to passive exposure to statistical regularities may not be sufficient to drive learning of complex categories and, instead, feedback or prediction mechanisms might play a more substantial role (Emberson et al., 2013; Feldman et al., 2013; McMurray et al., 2009; Nixon, 2020; Roark et al., 2021; Wade & Holt, 2005). Using hybrid passive plus supervised paradigms, researchers have demonstrated enhanced perceptual learning, relative to passive exposure alone (Wright et al., 2010). Future research should address the extent to which representation change might occur with passive, active, or hybrid short-term experience.

The experience was brief

Relative to the lifetime of acoustic experience that participants had before the experiment, the 8 min of exposure to 450 stimuli is extremely brief. This length of exposure was chosen based on prior work that suggested that even short-lived representational change may occur with as little as 2 min of exposure (Stilp et al., 2010). It is possible that this amount of exposure is not enough to substantially change representations or impact behavior, but longer exposure times might. When representational changes occur (if they do) with further experience is an open question for future research.

Statistical learning conflicted with category learning

Finally, it is possible that we were simply unable to see any impact of the statistical learning phase because of the way that we measured the impact. After the statistical learning phase, participants immediately entered a testing environment with no relationship between the dimensions. During passive exposure, participants experienced a regularity, and during categorization, they experienced a different regularity – a lack of a correlation between the dimensions. When measuring the effect of passive exposure, researchers have found that effects rapidly disappear in a transfer task (Stilp & Kluender, 2016; after 128 trials). Even across the first 48

trials, we did not see any effects of short-term passive experience on categorization. We are unable to conclude whether statistical learning failed to occur at all or, alternatively, statistical learning occurred but effects disappeared rapidly during the categorization task. To disentangle these possibilities, future studies could examine trial-wise behavioral or neural representations to examine change at a finer level.

Conclusion

Although organisms are sensitive to the statistical structure in the world, the interaction between short-term statistical learning and long-term perceptual biases, or priors, is not yet well understood. We found that passive statistical learning had limited effects on subsequent category learning in an acoustic environment with strong perceptual priors. These findings highlight the limits of short-term passive exposure on restructuring of perceptual representations that influence learning and decision-making processes, such as those involved in category learning. The mind does not rapidly adapt to *all* regularities in an environment and the generalizable effects of passive exposure to regularities on subsequent behavior are limited. Long-term priors can be quite rigid in the face of short-term experience and statistically identical categories can be learned very differently based on existing representations.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13423-022-02114-z>.

References

- Adriaans, F., & Swingle, D. (2017). Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *The Journal of the Acoustical Society of America*, *141*(5), 3070–3078.
- Allen, E. J., Moerel, M., Lage-Castellanos, A., Martino, F. D., Formisano, E., & Oxenham, A. J. (2018). Encoding of natural timbre dimensions in human auditory cortex. *NeuroImage*, *166*(March 2017), 0–70.
- Antoniou, M., & Wong, P. C. M. (2016). Varying irrelevant phonetic features hinders learning of the feature being trained. *The Journal of the Acoustical Society of America*, *139*(1), 271–278.
- Ashby, F. G. (1992). *Multidimensional models of categorization* (F. G. Ashby, Ed.; pp. 449–483). Lawrence Erlbaum. <http://psycnet.apa.org/psycinfo/1992-98026-016>
- Ashby, F. G., & Maddox, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 598–612.
- Ashby, F. G., & Maddox, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 50–71.
- Ashby, F. G., & Townsend, J. T. (1986). Varieties of Perceptual Independence. *Psychological Review*, *93*(2), 154–179.
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by human infants. *Psychological Science*, *9*(4), 321–324.

- Barlow, H., & Földiák, P. (1989). Adaptation and Decorrelation in the Cortex. In R. Durbin, C. Miall, & G. Mitchison (Eds.), *The Computing Neuron* (pp. 54–72). Addison-Wesley.
- Barreda, S. (2015). *phonTools: Functions for phonetics in R*. (Version 0.2-2.1) [Computer software].
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of Acoustical Society of America*, *109*(2), 775–794.
- Boersma, P., & Weenink, D. (2021). *Praat: doing phonetics by computer* (Version 6.1.51) [Computer software]. <http://www.praat.org>
- Bond, R. N., & Feldstein, S. (1982). Acoustical correlates of the perception of speech rate: An experimental investigation. *Journal of Psycholinguistic Research*, *11*(6), 539–557.
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658.
- Coen-Cagli, R., Kohn, A., & Schwartz, O. (2015). Flexible gating of contextual influences in natural vision. *Nature Neuroscience*, *18*(11), 1648–1655.
- Depireux, D. A., Simon, J. Z., Klein, D. J., & Shamma, S. A. (2001). Spectro-Temporal Response Field Characterization With Dynamic Ripples in Ferret Primary Auditory Cortex. *Journal of Neurophysiology*, *85*(3), 1220–1234.
- Ell, S. W., Ashby, F. G., & Hutchinson, S. (2012). Unsupervised category learning with integral-dimension stimuli. *The Quarterly Journal of Experimental Psychology*, *65*(8), 1537–1562.
- Elliott, T. M., & Theunissen, F. E. (2009). The Modulation Transfer Function for Speech Intelligibility. *PLoS Computational Biology*, *5*(3), e1000302.
- Emberson, L. L., Liu, R., & Zevin, J. D. (2013). Is statistical learning constrained by lower level perceptual organization? *Cognition*, *128*(1), 82–102.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.
- Escudero, P., & Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition*, *133*(2), 408–413.
- Feldman, N. H., Goldwater, S., Dupoux, E., & Schatz, T. (2021). Do Infants Really Learn Phonetic Categories? *Open Mind*, *5*, 113–131.
- Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). A Role for the Developing Lexicon in Phonetic Category Acquisition. *Psychological Review*, *120*(4), 751–778.
- Feldstein, S., & Bond, R. N. (1981). Perception of Speech Rate as a Function of Vocal Intensity and Frequency. *Language and Speech*, *24*(4), 387–394.
- Garner, W. R. (1974). *The Processing of Information and Structure*. Erlbaum.
- Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, *8*(1), 98–123.
- Henry, M. J., & McAuley, J. D. (2009). Evaluation of an Imputed Pitch Velocity Model of the Auditory Kappa Effect. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 551–564.
- Herrmann, B., & Johnsrude, I. S. (2018). Attentional State Modulates the Effect of an Irrelevant Stimulus Dimension on Perception. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 89–105.
- Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, *119*(5), 3059–3059.
- Holt, L. L., Lotto, A. J., & Diehl, R. L. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *The Journal of the Acoustical Society of America*, *116*(3), 1763–1773.
- Hullett, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E., & Chang, E. F. (2016). Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. *Journal of Neuroscience*, *36*(6), 2014–2026.
- Idemaru, K., & Holt, L. L. (2014). Specificity of Dimension-Based Statistical Learning in Word Recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1009–1021.
- Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *Journal of Acoustical Society of America*, *132*(6), 3950–3964.
- Kluender, K. R., Stilp, C. E., & Kiefte, M. (2013). Perception of vowel sounds within a biological realistic model of efficient coding. In ["G. Morrison" & P. Assmann (Eds.), *Vowel Inherent Spectral Change, Modern Acoustics and Signal Processing* (pp. 117–151). Springer-Verlag. https://doi.org/10.1007/978-3-642-14209-3_6
- Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850–11857.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2007). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979–1000.
- Langers, D. R. M., Backes, W. H., & van Dijk, P. (2003). Spectrotemporal features of the auditory cortex: the activation in response to dynamic ripples. *NeuroImage*, *20*(1), 265–275.
- Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature Neuroscience*, *5*(4), 356–363.
- Liu, R., & Holt, L. L. (2015). Dimension-based statistical learning of vowels. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1783–1798.
- Lockhead, G. R. (1972). Processing dimensional stimuli: A note. *Psychological Review*, *79*(5), 410–419.
- Lu, K., Liu, W., Dutta, K., Zan, P., Fritz, J. B., & Shamma, S. A. (2019). Adaptive Efficient Coding of Correlated Acoustic Properties. *Journal of Neuroscience*, *39*(44), 8664–8678.
- Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*(1), 49–70.
- Makowski, D. (2018). The Psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science. *Journal of Open Source Software*, *3*(22), 470.
- Markant, D. B. (2018). Effects of Biased Hypothesis Generation on Self-Directed Category Learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *45*(9), 1552–1568.
- Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), 101–111.
- McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.
- McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science*, *12*(3), 369–378.
- Ming, V. L., & Holt, L. L. (2009). Efficient coding in human auditory perception. *The Journal of the Acoustical Society of America*, *126*(3), 1312–1320.
- Nelson, D. G. K. (1993). Processing Integral Dimensions: The Whole View. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(5), 1105–1113.

- Nixon, J. S. (2020). Of mice and men: Speech sound acquisition as discriminative learning from prediction error, not just statistical tracking. *Cognition*, *197*, 104081.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.
- Nystrom, N. A., Levine, M. J., Roskies, R. Z., and Scott, J. R. (2015). Bridges: A Uniquely Flexible HPC Resource for New Communities and Data Analytics. In *Proceedings of the 2015 Annual Conference on Extreme Science and Engineering Discovery Environment* (St. Louis, MO, July 26–30, 2015). XSEDE15. ACM, New York, NY, USA. <https://doi.org/10.1145/2792745.2792775>.
- Pons, F. (2006). The Effects of Distributional Learning on Rats' Sensitivity to Phonetic Information. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(1), 97–101.
- Reetzke, R., Maddox, W. T., & Chandrasekaran, B. (2016). The role of age and executive function in auditory category learning. *Journal of Experimental Child Psychology*, *142*, 48–65.
- Roark, C. L., & Holt, L. L. (2019a). Auditory information-integration category learning in young children and adults. *Journal of Experimental Child Psychology*, *188*, 104673.
- Roark, C. L., & Holt, L. L. (2019b). Perceptual dimensions influence auditory category learning. *Attention, Perception, and Psychophysics*, *81*(4), 912–926.
- Roark, C. L., & Holt, L. L. (2022). Statistical learning does not overrule perceptual priors during category learning. <https://doi.org/10.17605/OSF.IO/QYG7Z>
- Roark, C. L., Lehet, M. I., Dick, F., & Holt, L. L. (2021). The representational glue for incidental category learning is alignment with task-relevant behavior. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. <https://doi.org/10.1037/xlm0001078>
- Roark, C. L., Plaut, D. C., & Holt, L. L. (2022). A neural network model of the effect of prior experience with regularities on subsequent category learning. *Cognition*, *222*(104997), 104997.
- Rost, G. C., & McMurray, B. (2010). Finding the Signal by Adding Noise: The Role of Noncontrastive Phonetic Variability in Early Word Learning. *Infancy*, *15*(6), 608–635.
- Scharinger, M., Henry, M. J., & Obleser, J. (2013). Prior experience with negative spectral correlations promotes information integration during auditory category learning. *Memory & Cognition*, *41*(5), 752–768.
- Schönwiesner, M., & Zatorre, R. J. (2009). Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proceedings of the National Academy of Sciences*, *106*(34), 14611–14616.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464.
- Simoncelli, E. P. (2003). Vision and the statistics of the visual environment. *Current Opinion in Neurobiology*, *13*(2), 144–149.
- Simoncelli, E. P., & Olshausen, B. A. (2001). Natural Image Statistics and Neural Representation. *Annual Review of Neuroscience*, *24*, 1193–1216.
- Skoruppa, K., & Peperkamp, S. (2011). Adaptation to Novel Accents: Feature-Based Learning of Context-Sensitive Phonological Regularities. *Cognitive Science*, *35*(2), 348–366.
- Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, *439*(7079), 978–982.
- Stilp, C. E., Kiefte, M., & Kluender, K. R. (2018). Discovering acoustic structure of novel sounds. *The Journal of the Acoustical Society of America*, *143*(4), 2460–2473.
- Stilp, C. E., & Kluender, K. R. (2012). Efficient Coding and Statistically Optimal Weighting of Covariance among Acoustic Attributes in Novel Sounds. *PLoS ONE*, *7*(1), e30845.
- Stilp, C. E., & Kluender, K. R. (2016). Stimulus Statistics Change Sounds from Near-Indiscriminable to Hyperdiscriminable. *PLOS ONE*, *11*(8), e0161001.
- Stilp, C. E., & Lewicki, M. S. (2014). Statistical structure of speech sound classes is congruent with cochlear nucleus response properties. *Proceedings of Meetings on Acoustics*, *20*(2014), 050001.
- Stilp, C. E., Rogers, T. T., & Kluender, K. R. (2010). Rapid efficient coding of correlated complex acoustic properties. *Proceedings of the National Academy of Sciences*, *107*(50), 21914–21919.
- Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*(3), 434–464.
- Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G. D., Roskies, R., Scott, J. R., & Wilkens-Diehr, N. (2014). XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering*, *16*(5), 62–74.
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer.
- Visscher, K. M., Kaplan, E., Kahana, M. J., & Sekuler, R. (2007). Auditory Short-Term Memory Behaves Like Visual Short-Term Memory. *PLoS Biology*, *5*(3), e56.
- Wade, T., & Holt, L. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *The Journal of the Acoustical Society of America*, *118*(4), 2618–2633.
- Wang, X. (2007). Neural coding strategies in auditory cortex. *Hearing Research*, *229*(1–2), 81–93.
- Wanrooij, K., & Boersma, P. (2013). Distributional training of speech sounds can be done with continuous distributions. *The Journal of the Acoustical Society of America*, *133*(5), EL398–EL404.
- Wickens, T. D. (1982). *Models for Behavior: Stochastic Processes in Psychology*. Freeman.
- Woolley, S. M. N., Fremouw, T. E., Hsu, A., & Theunissen, F. E. (2005). Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nature Neuroscience*, *8*(10), 1371–1379.
- Wright, B. A., Sabin, A. T., Zhang, Y., Marrone, N., & Fitzgerald, M. B. (2010). Enhancing perceptual learning by combining practice with periods of additional sensory stimulation. *Journal of Neuroscience*, *30*(38), 12868–12877.
- Yi, H.-G., & Chandrasekaran, B. (2016). Auditory categories with separable decision boundaries are learned faster with full feedback than with minimal feedback. *The Journal of the Acoustical Society of America*, *140*(2), 1332–1335.
- Zhang, Z., & Mai, Y. (2018). *WebPower: Basic and Advanced Statistical Power Analysis*. <https://CRAN.R-project.org/package=WebPower>

Open Practices Statement: The data and materials for the experiment are available via the Open Science Framework at <https://doi.org/10.17605/OSF.IO/QYG7Z>, and the experiment was not preregistered.

Author Note: C. L. R. is now at the University of Pittsburgh, Department of Communication Science and Disorders. This work was supported by the National Institutes of Health (T32-GM0081760 & F32DC018979 to C. L. R.), the National Science Foundation (BCS1950054 to L. L. H.) and the Carnegie Mellon University Department of Psychology Centennial Fund Fellowship. This work used the Extreme Science and Engineering Discovery Environment (XSEDE, Towns et al., 2014), which is supported by NSF (ACI-1548562). Specifically, it used the Bridges system (Nystrom, Levine, Roskies, & Scott, 2015), which is supported by NSF (ACI-1445606), at the Pittsburgh Supercomputing Center (PSC).

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