Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-021-01878-0

BRIEF REPORT

Comparing perceptual category learning across modalities in the same individuals

Casey L. Roark[1,2] · Giorgio Paulon[3] · Abhra Sarkar[3] · Bharath Chandrasekaran[1,2]

Accepted: 4 January 2021

Abstract

Category learning is a fundamental process in human cognition that spans the senses. However, much still remains unknown about the mechanisms supporting learning in different modalities. In the current study, we directly compared auditory and visual category learning in the same individuals. Thirty participants (22 F; 18–32 years old) completed two unidimensional rule-based category learning tasks in a single day – one with auditory stimuli and another with visual stimuli. We replicated the results in a second experiment with a larger online sample ($N = 99$, 45 F, 18–35 years old). The categories were identically structured in the two modalities to facilitate comparison. We compared categorization accuracy, decision processes as assessed through drift-diffusion models, and the generalizability of resulting category representation through a generalization test. We found that individuals learned auditory and visual categories to similar extents and that accuracies were highly correlated across the two tasks. Participants had similar evidence accumulation rates in later learning, but early on had slower rates for visual than auditory learning. Participants also demonstrated differences in the decision thresholds across modalities. Participants had more categorical generalizable representations for visual than auditory categories. These results suggest that some modality-general cognitive processes support category learning but also suggest that the modality of the stimuli may also affect category learning behavior and outcomes.

Keywords Category learning · Modality effects · Audition · Vision

Introduction

Categorization is critical to all sensory systems, enabling complex behaviors such as object recognition and speech perception. However, much still remains unknown about the mechanisms supporting learning across different modalities.

Theories of perceptual category learning typically focus on a single modality with relatively little generalization across modalities. While cross-modal category learning is a large area of research, understanding the commonalities and differences of single-modality category learning is an unexplored area. Much has been gained by studying auditory and visual category learning separately. Theories focused on visual learning have deepened understanding about the nature of the neural and cognitive mechanisms underlying category learning (Ashby, 1992a; Goldstone, 1994; Love, Medin, & Gureckis, 2004; Nosofsky, 1986). Theories focused on auditory and speech category learning have demonstrated the influences of native language experience and training methods on the outcomes of second language learning (Best, 1995; Logan, Lively, & Pisoni, 1991; McClelland, Fiez, & McCandliss, 2002). These complimentary yet distinct approaches have resulted in two sets of literature that have remained relatively separated by modality.

There are reasons to suspect that generic cognitive processes support perceptual category learning, regardless of modality. Recently, theories originally developed in the visual modality have been applied to learning in the auditory modality (Francis & Nusbaum, 2002; Goudbeek, Swingley, & Smits, 2009; Maddox, Chandrasekaran, Smayda, & Yi, 2013; Yi, Maddox, Mumford, & Chandrasekaran, 2014). Generally,

✉ Casey L. Roark
croark@pitt.edu

✉ Bharath Chandrasekaran
b.chandra@pitt.edu

[1] Department of Communication Science and Disorders, University of Pittsburgh, Pittsburgh, PA, USA

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

[3] Department of Statistics and Data Sciences, The University of Texas at Austin, Austin, TX, USA

Published online: 02 February 2021

the application of these theories of visual learning to the auditory modality has proved fruitful for understanding cognitive mechanisms that support auditory learning, demonstrating similar overall learning patterns to the visual modality.

When learning a relatively simple unidimensional category distinction, the learner's task is essentially the same regardless of modality: based on their sensory experience, learners need to create abstract rules, make categorization decisions, and incorporate feedback to test and update these rules to optimize their decisions. Beyond perception, learning involves complex cognitive processes including attention, working memory, hypothesis generation and testing, and feedback processing. In support of a modality-general perspective, these basic components of cognition operate similarly regardless of modality (Lehnert & Zimmer, 2006; Visscher, Kaplan, Kahana, & Sekuler, 2007; Zvyagintsev et al., 2013). Additionally, the same fronto-striatal brain networks support working memory, attention, and feedback processing for both auditory and visual tasks, with domain-general response properties (Crittenden & Duncan, 2014; Duncan & Owen, 2000; Fedorenko, Duncan, & Kanwisher, 2013; McNab & Klingberg, 2008; Myers, 2014).

Even if category learning is supported by domain-general cognitive processes, stimulus modality may still play a substantial role in learning. To effectively learn perceptual categories that can be described by a rule along one stimulus dimension, learners must perceive the stimulus, separate the stimulus into components, identify which components are relevant or irrelevant to category identity, and specify the decision criterion along the relevant dimension. For instance, in a unidimensional auditory categorization task with sounds varying in duration and pitch, learners must hear the sound, separate the duration and pitch components, identify which of those components is relevant, and, for example, specify that sounds belong to category A when the duration is longer than 0.5 s and category B when the duration is shorter than 0.5 s. The rules and processes are not necessarily so explicit, but the steps are required for accurate rule-based category learning.

The expansion of theories of category learning that originated in the visual modality into the auditory modality has highlighted some important differences between vision and audition (Heffner, Idsardi, & Newman, 2019; Roark & Holt, 2019; Scharinger, Henry, & Obleser, 2013). Specifically, auditory objects may be more difficult to separate into their individual components than visual objects, meaning that selective attention to dimensions may be more difficult for auditory than visual dimensions (Garner, 1974**;** Roark & Holt, 2019). Auditory dimensions may also be more difficult to describe than visual dimensions, which may make hypothesizing about category rules more difficult (Brashears & Minda, 2020; Zettersten & Lupyan, 2020). Together, these are reasons to suspect that modality-specific processes support perceptual category learning.

In the current study, we examine unidimensional auditory and visual category learning in the same individuals. In two experiments (a smaller in-person experiment and a large-scale online replication), we employed identical category training tasks that differ only in the modality of the stimuli. We use a diverse set of tools to compare perceptual and cognitive processes of auditory and visual category learning. We assess category learning by examining accuracy during training. We examine decision processes during learning by comparing evidence accumulation rates and decision thresholds assessed with drift-diffusion models. Finally, we assess learners' category representations in a generalization test with novel stimuli.

If domain-general cognitive processes support learning across modalities, then we predict that learners will show similarities in performance and decision processes regardless of modality. If instead, modality plays a substantial role in learning, we predict that learners will show differences across modalities. These differences might manifest in one or more behaviors, such as the performance, decisional processes, or representation of categories.

Methods

We first ran Experiment 1, an in-person first pass of the methodology with 30 participants. After this initial study, we ran Experiment 2 as a replication in a larger online sample. To accommodate online testing, we made two small changes to the overall methodology, outlined below. We ran the same analyses across both experiments.

Participants

In Experiment 1, participants were 30 (22 F, 8 M) members of the Pittsburgh community, age 18–32 years. All participants passed a hearing screening indicating that their pure-tone thresholds were 25 dB HL or better at octave and half-octave frequencies from 125 Hz to 6,000 Hz. All participants were native speakers of English. A power analysis examining the main effect of modality on accuracy showed that 24 participants would provide greater than 80% power ($\alpha = .05$) to detect a medium-large effect ($d = 0.6$). In Experiment 2, participants were 99 (45 F, 54 M) individuals recruited through Prolific (www.prolific.co), age 18–35 years and were tested using the Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2019). An additional participant was run but excluded because they only completed one task (1F). Both experiments were approved by the University of Pittsburgh Institutional Review Board and participants were paid $10/h.

Stimuli

The auditory stimuli were dynamic ripples varying in spectral and temporal modulation. The visual stimuli were Gabor patches varying in spatial frequency and orientation. The pairs of dimensions were chosen because they are thought to be similarly complex, have independent neural representations (Schönwiesner & Zatorre, 2009), and are processed similarly in working memory (Visscher et al., 2007). Further, categorization with these perceptual dimensions has been examined in separate individuals (Maddox, Ashby, & Bohil, 2003; Rabi & Minda, 2014; Yi & Chandrasekaran, 2016).

Category distributions

The category distributions were identical for the auditory and visual tasks (Fig. 1, circles). To create the stimulus distributions for category training, a single category was sampled from a bivariate Gaussian distribution in normalized space. This category was mirrored across the space to form the other category, ensuring that the categories have identical variability. The stimuli were then transformed separately into the auditory and visual spaces. Optimal performance during learning could be obtained by selectively attending to temporal modulation in the auditory task and spatial frequency in the visual task.

A separate set of generalization stimuli was generated, sampling a grid across the same two-dimensional space (Fig. 1, black Xs). These generalization stimuli provide the opportunity to understand how general the representations learned during the training phase are (broad sampling of the space) and how specific the learning of the distinction between the categories is (dense sampling around the optimal category boundary).

Procedure

Participants performed the auditory and visual tasks in the same session with the order counterbalanced across participants.

Auditory and visual category training

The auditory and visual tasks were identical (Fig. 1). Participants were instructed to categorize the stimuli into one of two categories by pressing buttons on the keyboard (1, 2), with category-response association counterbalanced across subjects. Across six blocks of 50 trials (300 trials total), the

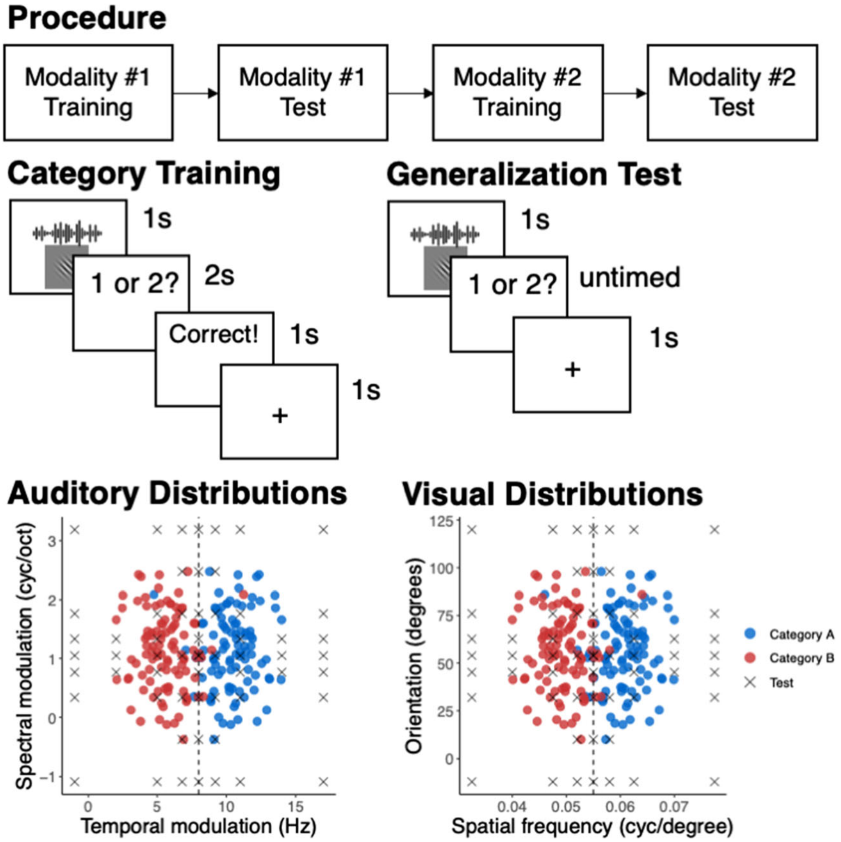

Fig. 1 Task procedures for category training and generalization test, across the two modalities. Stimulus distributions for auditory (*left*) and visual (*right*) for the category training (circle) and generalization test (x), with optimal distinction between the categories shown as a dashed line

stimulus was presented for 1 s, followed by a 2-s delay during which participants were able to respond, corrective feedback (Correct/Incorrect) for 1 s, and a 1-s intertrial interval (ITI). Participants were required to respond within 2 s after the onset of the stimulus or else no response was recorded, and they were told to respond faster. Trials for which there was no response (1.2% of all trials) were coded as incorrect.

The method for Experiment 2 was nearly identical to Experiment 1, with two differences to accommodate testing online. First, we shortened the number of training blocks from six to five blocks to enhance the quality of data collected online. Second, we changed the time allowed for response from 2 s to infinite time to ensure that participants made a response on every trial. This change also forced one difference in our analysis method. We removed the top and bottom 1% of reaction time responses in the drift-diffusion modeling to account for cases where participants had impossibly short reaction times (< 3 ms) or unreasonably long reaction times (> 7 s) that would alter the ability of the models to fit the data appropriately.

Generalization test

A hallmark of category learning is the ability to generalize category knowledge to novel exemplars. After category training, participants were given a generalization test where they encountered novel exemplars and no longer received any feedback. Participants were instructed to use the same category-to-response mapping as during training. In each of 58 trials, the stimulus was presented for 1 s, followed by the participant's response, and a 1-s ITI. There was no limit to participants' response time during the test.

Drift-diffusion modeling

Drift-diffusion models (DDMs; Nosofsky & Palmeri, 1997; Ratcliff, 1978; Smith & Vickers, 1988) take into account accuracies and response times to understand processes underlying perceptual decision making. Theoretically, drift-diffusion models assume that during decision making, evidence for multiple decision options (in our case, categories) is accumulated at varying rates in a single accumulator (Nosofsky & Palmeri, 1997) and a decision is made when this evidence reaches a particular threshold. These models have recently been extended to a multi-alternative, longitudinal, mixed-model setting specifically in the context of category learning by considering multiple simultaneous accumulators of evidence (Paulon, Llanos, Chandrasekaran, & Sarkar, 2020).

We apply the same methods of parameterization of the DDM as Paulon et al. (2020). Specifically, the model fits an offset parameter (δ_s) for each category and, for every combination of decision response (d) and stimulus (s) category, an evidence accumulation rate parameter ($\mu_{d,\ s}$) and decision threshold parameter ($b_{d,\ s}$; see Fig. 2 for an illustration).

The offset represents the time taken by all the actions that are not directly relevant to the actual decision making (e.g., stimulus encoding, motor response, etc.). Evidence accumulation rate reflects extraction of information from the stimulus relevant for the categorization decision. Learners have lower evidence accumulation rates for more difficult tasks, when it is more difficult to extract what is relevant. Decision threshold reflects response caution in the speed-accuracy tradeoff (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010). Learners have higher decision thresholds when they are more cautious in their responses (i.e., slower reaction times in favor of more accurate responses). Importantly, the DDM of Paulon et al. (2020) allows the rates and thresholds to evolve longitudinally as the participants become more experienced in their decision tasks. Moreover, drifts and the boundaries are allowed to differ between individuals, capturing the heterogeneity in category learning performance across different participants. We will examine two parameters: evidence accumulation rate and decision thresholds.

Results

The main goal of the current study is to understand how the *same individuals* learn identically structured auditory and visual categories. We assess several components of this process: learning performance, dynamics of decision processes (evidence accumulation rates and decision thresholds), and generalizability of category representations. We present the results from the in-person and online experiments together.

Category learning

Participants learned the categories in the two modalities fairly well, with substantial variability across participants (Fig. 3a). This kind of inter-subject variability is common in learning tasks and demonstrates the necessity of examining learning in the same individuals to illuminate the common sources of this variability.

We compared performance during category learning across the two tasks using repeated-measures ANOVAs with task (auditory, visual) and block (1–6) as factors. In both experiments, participants had similar overall accuracies in the two modalities (Experiment 1: $F(1,29) = 1.33$, $p = 0.26$, $\eta_p^2 = .04$; Experiment 2: $F(1,305.2) = 0.29$, $p = 0.60$, $\eta_p^2 = .003$), but had different patterns of accuracy across blocks (Experiment 1: $F(5,145) = 2.83$, $p = 0.018$, $\eta_p^2 = .09$; Experiment 2: $F(4,305.2) = 5.89$, $p = 0.001$, $\eta_p^2 = .057$). Specifically, the slope of the learning curves across blocks was steeper in the visual task (Experiment 1: block 6 – block 1: $M = 19\%$; Experiment 2: block 5 – block 1: $M = 16\%$) than the auditory

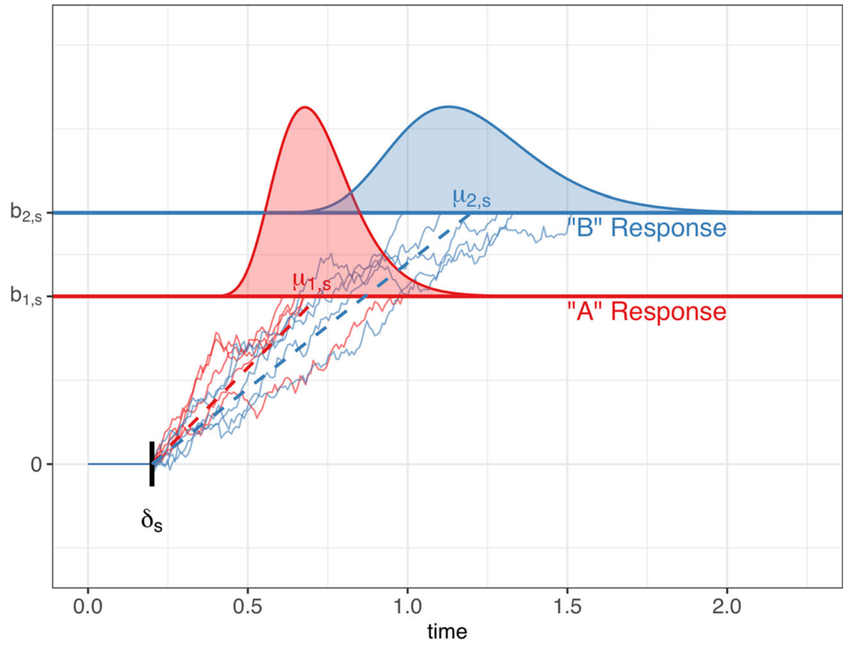

Fig. 2 Drift-diffusion model for perceptual decision making. After an initial amount of time δ_s required to encode an input signal s, the evidence in favor of a response category d accumulates according to a Wiener diffusion process with drift $\mu_{d,\ s}$. The decision d is eventually taken if the underlying process is the first to reach its decision threshold $b_{d,\ s}$. The curves represent the probability distributions of the response times corresponding to the two possible decisions $d \in \{1, 2\}$

task (Experiment 1: block 6 – block 1: $M = 9\%$; Experiment 2: block 5 – block 1: $M = 10.5\%$). The order in which participants completed the two tasks did not affect performance (Experiment 1: $F(1,28) = 0.011$, $p = 0.92$, $\eta_p^2 = 0.0$; Experiment 2: $F(1,97) = 0.56$, $p = 0.46$, $\eta_p^2 = 0.006$).

To understand how the outcomes of learning in the two tasks were related, we examined the correlation between accuracies in the final block. Because the accuracies are bounded by 0 and 1, we applied a logit transform to accuracies (Lesaffre, Rizopoulos, & Tsonaka, 2007), which did not change any of the statistical patterns in the results. Final block accuracy in the auditory and visual tasks was significantly positively correlated in both experiments (Fig. 3b, Experiment 1: $r(28) = 0.61$, p = 0.00034; Experiment 2: $r(97) = 0.54$, p = 0.000000009281), indicating a common source of variability underlying auditory and visual category learning. Performance in the first block was also positively correlated in Experiment 1 ($r(28) = 0.57$, $p = 0.0011$); however, this did not replicate in Experiment 2 ($r(97) = 0.16$, $p = 0.12$).

Dynamics of decision processes

We compared the dynamics of the decision processes in the two modalities by examining evidence accumulation rates and decision thresholds using DDMs. Across both experiments and for both tasks, the population-level estimates for correct trials' rate and threshold parameters did not differ across the two category responses (e.g., A and B). Thus, without loss of generality, we restrict our comparison across modalities to correct response category A trials.

We compared evidence accumulation rates and decision thresholds across the two tasks by determining where the 95% credible intervals (95% posterior probability that the parameter estimate lies within this interval) do not overlap for auditory and visual tasks (Fig. 3d and e). In both experiments, evidence accumulation rates differed across auditory and visual tasks in a way that depended on block (Fig. 3d). Specifically, the visual task had lower evidence accumulation rates than the auditory task in block 1 and higher evidence accumulation rates in later blocks (Experiment 1: blocks 5 and 6, Experiment 2: block 4). In both experiments, the rates at which participants accumulate evidence towards a categorization decision were stable across blocks in the auditory task and stabilized in the second block in the visual task. Final block auditory and visual evidence accumulation rates were positively correlated in Experiment 1 ($r(28) = 0.54$, $p = 0.0020$) and Experiment 2 ($r(97) = 0.57$, $p = 0.00000000098$).

In both experiments, decision thresholds also differed across auditory and visual tasks in a way that depended on block (Fig. 3e). For both experiments, the auditory task had higher decision thresholds than visual in block 1 and the visual task had higher decision thresholds in blocks 3–5. For identical categorization decisions in two different modalities, there were differences in decision thresholds in the majority of blocks. For the auditory modality, there was a decrease in decision thresholds with more training, meaning that with more experience in the task, participants were less cautious in their responses and were willing to respond faster even if it meant being incorrect more often. For the visual modality, response caution was relatively stable; there were no differences in decision thresholds across blocks except in the final block. Final block auditory and visual decision thresholds were positively correlated in Experiment 1 ($r(28) = 0.34$, $p = 0.064$) and Experiment 2 ($r(97) = 0.37$, $p = 0.00020$).

In sum, across both experiments in block 1, participants accumulated evidence for visual categories at a slower rate than auditory categories. However, this disappeared by the second block and by the end of training, evidence accumulation rates were faster for visual categories than auditory categories. Participants' decision thresholds were also different across modalities with initially higher thresholds in the auditory task (i.e., more response caution—bending towards accuracy in speed-accuracy tradeoff) and higher thresholds in the visual task later in training.[1]

Generalization test

To understand how participants generalized their category knowledge, we computed the accuracy of their responses to novel stimuli when they did not receive feedback. We defined accuracy based on the optimal response, with items to the right of the boundary belonging to category A and items to the left belonging to category B. We removed trials for stimuli that lay directly on the category boundary as performance for these stimuli would have been at chance for an optimal observer. In Experiment 1, participants were better able to generalize their knowledge about visual categories than auditory categories (Fig. 4a, $M_{\text{diff}} = 7.57\%$, $t(29) = 2.86$, $p = 0.0078$, $d = 0.52$). However, this effect did not replicate in Experiment 2; there was no difference in generalization performance across modalities ($M_{\text{diff}} = 2.18\%$, $t(98) = 1.22$, $p = 0.22$, $d = 0.12$). In both experiments, as with training accuracy, the generalization test accuracy for the auditory and visual tasks (with a logit transform) was significantly positively correlated (Fig. 4c, Experiment 1: $r(28) = 0.51$, $p = 0.0043$; Experiment 2: $r(97) = 0.34$, $p = 0.00056$).

To understand the nature of the category representations, we computed proportion of category A responses as a function of a stimulus's placement on the category-relevant dimension (Fig. 4b). Across the nine steps in the generalization grid, values near 0 represent a consistent "Category B" response and values near 1.0 represent a consistent "Category A" response. In both modalities, participants were more accurate further from the category boundary and were less able to determine the category identity for stimuli closer to the category boundary. We compared the categorization response curves by examining relevant-dimension value (9) × modality (auditory, visual) repeated-measures ANOVAs. In both experiments, participants demonstrated different patterns of generalization across the stimulus space for the two modalities, as indicated by significant interactions between dimension value and modality (Experiment 1: $F(8, 232) = 3.51$, $p = 0.001$ $\eta_p^2 = .11$; Experiment 2: $F(8, 424.1) = 6.65$, $p < 0.0005$ $\eta_p^2 = .063$). In Experiment 1, this was driven by more polarized category representations for visual than auditory categories at the second, sixth, and seventh steps in the generalization grid (step 2 ($t(29) = 1.759$, $p = 0.089$), step 6 $t(29) = -3.161$, $p = 0.004$, step 7 $t(29) = -2.819$, $p = 0.009$). All other steps were not significantly different (ps > 0.077). In Experiment 2, there were significant differences at steps 1 ($t(98) = 2.98$, $p = 0.004$), 4 ($t(98) = -5.08$, $p < 0.0005$), 6 ($t(98) = -3.94$, $p < .0005$) and 7 ($t(98) = -3.46$, $p = 0.001$). All other steps were not significantly different (ps > 0.10). The effect at steps 6 and 7 was consistent across both experiments. These generalization results indicate that participants had more "categorical" representations of the visual stimuli than the auditory stimuli, especially in the regions of space near to the boundary on the right side of the stimulus space.

Discussion

We compared the cognitive and perceptual processes during auditory and visual category learning in the same individuals in an in-person experiment and a large-scale online replication. Participants had similar performance but demonstrated differences in decision processes across modalities, especially early in learning. Participants also demonstrated more categorical generalizable representations for visual than auditory categories. These results suggest that modality plays a substantial role in perceptual category learning.

We found that learning outcomes and decisional processes were correlated across modalities. However, evidence of a correlation does not provide clear information about why this correlation exists. Future studies should focus on identifying

[1] We allowed the evidence accumulation and decision threshold parameters to vary flexibly across learning. As a supplementary analysis, we also compared this flexible model of Paulon et al. (2020) with other sub-cases in which one parameter is fixed but the other is allowed to vary to understand the relative importance of each parameter for auditory and visual learning. We compared the models using the Watanabe-Akaike information criterion (WAIC), a popular approach for assessing predictive performances of competing methods. WAIC is obtained by computing the log point-wise predictive density and then adding a correction reflecting the effective number of degrees of freedom to prevent from overfitting (see Gelman, Hwang, & Vehtari, 2014, for details). The results of this comparison indicated that the model that best describes the data for both auditory and visual tasks is a fully flexible one, allowing both evidence accumulation and decision threshold parameters to vary across learning (WAIC for Visual: flexible: −6690.71, constant accumulation: −7247.56, constant threshold: −7084.33; WAIC for Auditory: flexible: −4404.38, constant accumulation: −4443.75, constant threshold: −4590.62). These model comparisons also help pinpoint which parameter is most relevant for a particular modality. In the visual task, the flexibility for the accumulation parameter is more important than for the threshold parameter, as measured by the relative increase in WAIC. The converse is true for the auditory task. This is consistent with our results and interpretation of the accumulation and drift results from the fully flexible model, discussed in the manuscript.

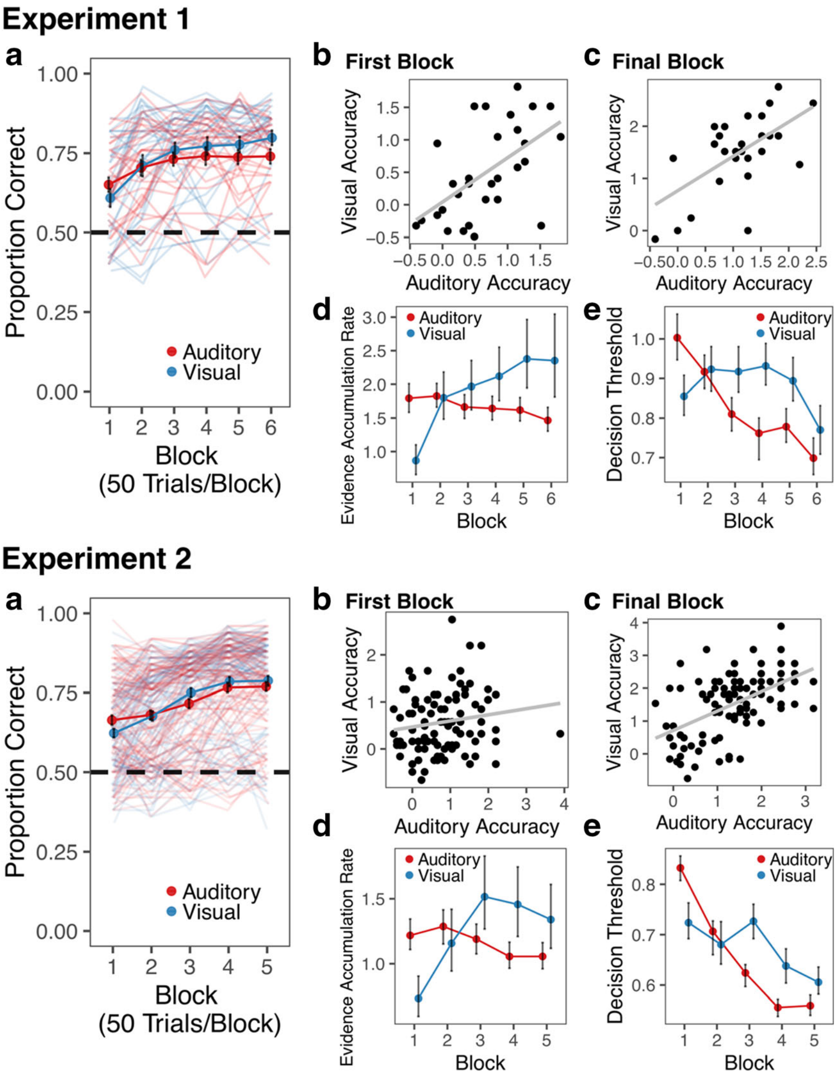

Fig. 3 **a** Accuracy during category training for the auditory and visual tasks relative to chance (dashed line at 50% accuracy) in Experiment 1 (top) and Experiment 2 (bottom). Lighter lines show individual performance and darker lines show the mean. Error bars reflect standard error of the mean. **b** and c Correlations between accuracy in the auditory task and visual task, with a logit transform **b** in the first block and **c** in the final block. **d** and **e** Population estimates of **d** evidence accumulation rates and **e** decision thresholds from the drift-diffusion models in the auditory and visual tasks. Error bars reflect 95% credible intervals around population estimates

the underlying source the similar patterns of behavior across modalities. For instance, to identify what underlying skills or abilities support modality-general or modality-specific learning, a study might include a large battery of sensory, perceptual, and cognitive assessments to discover which measures reliably account for variance in auditory and visual category learning.

While there were similar learning outcomes across modalities, we found differences in the decisional processes supporting learning (evidence accumulation and decision thresholds). Evidence accumulation reflects extraction of information about a stimulus to guide categorization decisions and has been linked to attention and motivation processes in presumably domain-general cognitive brain regions (i.e., dorsolateral prefrontal cortex, inferior frontal gyrus), which support accumulation regardless of sensory modality (Noppeney, Ostwald, & Werner, 2010). In the current study, evidence accumulation differed across modalities in the earliest stage of learning. In the first block, participants had lower accumulation rates in the visual than the auditory task, indicating that learners were slower to get the information they needed from the stimulus. Later when learners were more accurate and discovered the relevant dimension for category identity, the accumulation rate increased and stabilized. In the auditory modality, the accumulation rate was stable across blocks. The differences in the patterns across modalities in the same individuals suggests that modality affects how learners extract information from stimuli to guide categorization decisions.

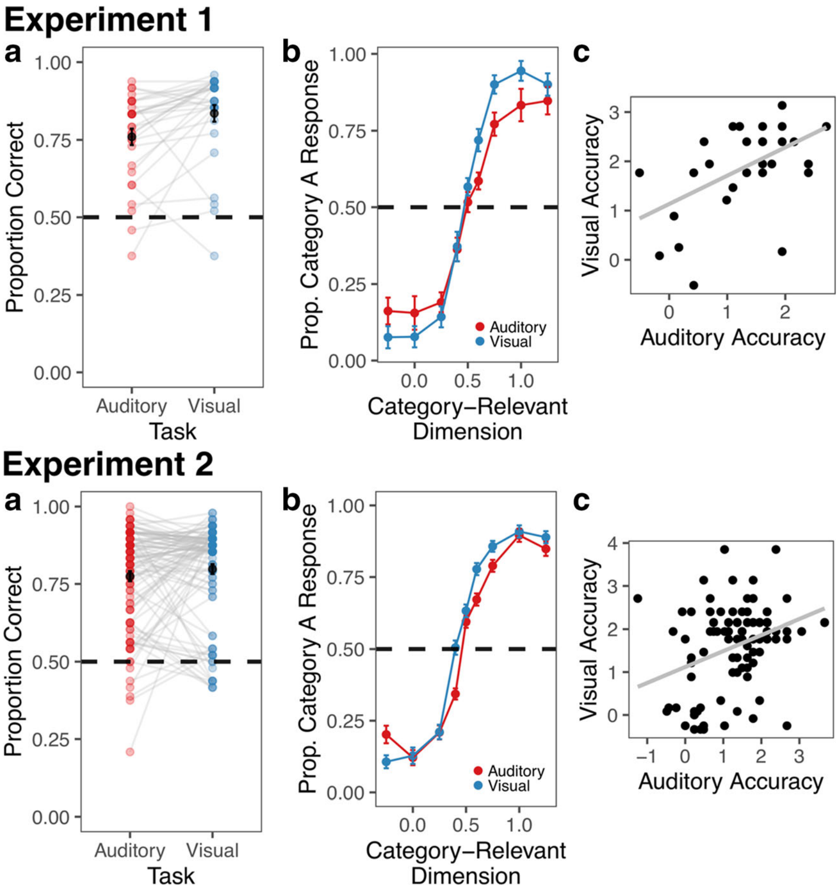

Fig. 4 **a** Accuracy in the generalization test for the auditory and visual tasks relative to chance (dashed line at 50% accuracy), with performance for individuals connected with gray lines with Experiment 1 on top and Experiment 2 at the bottom. **b** Categorization curve as a function of the placement of the stimulus along the category-relevant dimension for the auditory and visual tasks, where the dashed line at 50% reflects an equal response of Category A and Category B for a given category-relevant dimension value. **c** Correlation between generalization accuracy in the auditory task and visual task, with a logit transform

Another component of the decision process involves the balance between accuracy and speed in responses. This "response caution" is reflected in learners' decision thresholds. Participants were told to prioritize both speed and accuracy and had equal time to respond across auditory and visual tasks. Even so, the patterns of decision thresholds differed across modalities. At the beginning of learning, participants had less cautious responses in the visual than the auditory modality, and were thus willing to accept faster, more inaccurate responses.

These results demonstrate that within identical categorization tasks, decisional processes differed across modalities even when accuracies were the same. With training, participants became more efficient in gathering relevant information from the stimulus in the visual, but not the auditory task. This could be due to differences in sensory processing. With auditory stimuli, information unfolds across time. With visual stimuli, information is constant in time; participants can learn to quickly direct attention to a visual feature and evidence accumulation rates increase with training.

For auditory stimuli, participants initially had higher decision thresholds, meaning they collected more evidence before making a decision. Thresholds decreased markedly with training, suggesting that participants refined their understanding of the categories. For visual stimuli, participants relied more on sensory information (faster evidence accumulation) to guide their decisions rather than lowering their decision thresholds, which decreased less dramatically across blocks relative to auditory categories.

Finally, participants demonstrated more categorical responses in the generalization test for visual than auditory categories. Critically, during the generalization test, participants did not receive feedback. One interpretation of this result is that without the reference of consistent feedback as a reminder, participants were less able to remember and apply the relevant rule in the auditory task but were able to apply the visual rule accurately (see Appendix for supporting evidence from post hoc decision-bound modeling). It is also possible that these categorization differences in this region of the stimulus space could be due to subtle differences in perceptual discriminability across modalities that we cannot rule out in the current experiment. However, we note that categorization accuracy was not significantly different across modalities in the higher-powered Experiment 2.

We designed these categories to be as equal as possible. However, when comparing across modalities, it is extremely difficult if not impossible to exactly equate dimensions. We chose these pairs of dimensions because they have been shown to be equally independent in representations and behave similarly in working memory (Visscher et al., 2007). However, there could still be differences in separability (Garner, 1974) or verbalizability of the dimensions that may have affected learning (Brashears & Minda, 2020; Zettersten & Lupyan, 2020). Because direct comparisons across modalities are difficult, these persistent differences further highlight the necessity of considering modality in theories of perceptual category learning.

Importantly, these categories reflect one specific and relatively simple kind of category – unidimensional rule-based categories. This study is not meant to reflect the full spectrum of category learning problems. Indeed, it is possible that different kinds of categories (multidimensional, information-integration, similarity-based, etc.) may demonstrate different patterns across modalities. Comparison across different kinds of categories is an open area for future research and the current study is an important first step. These experiments were also not designed to distinguish among different models of category learning (i.e., prototype, exemplar, decision bound, etc.). Instead, we tested a fundamental assumption common across models: that the processes supporting category learning are the same across modalities. Future studies should test predictions from specific models of category learning, while including categorization problems across different modalities.

Conclusion

Though category learning is ubiquitous across modalities, it is rare to see comparisons across modalities, and instead researchers focus their investigations on a single modality. Theories of perceptual category learning have traditionally focused on the visual modality. The more recent expansion of some popular theories of visual category learning into the auditory modality has highlighted potential issues in the generalization of these theories across modalities. The extent to which mechanisms supporting perceptual category learning are shared across the senses is an open question. The current study demonstrates that even when the ultimate task of category learning is the same (map some sensory experience onto discrete category responses through feedback), modality also plays an important role, even in the same individuals and identical category tasks. Understanding the mechanisms driving perceptual category learning will require understanding both *perceptual* and *learning* components.

Author note This work was supported by grants from the National Institute On Deafness and Other Communication Disorders of the National Institutes of Health (R01DC013315A1 to B.C. and F32DC018979 to C.L.R.) and the National Science Foundation (NSF-1953712 to Co-PIs B.C. and A.S.). 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).

Open practices statement The data and materials for the experiment and replication are available at osf.io/msnq2 and neither the experiment nor replication were preregistered.

Appendix

Decision-bound modeling methods

To address the post-hoc question of whether participants used more rule-based processing during visual generalization than auditory generalization, we applied a series of decision-bound models to participants' response data from the test blocks of Experiment 1 and Experiment 2. As a comparison, we also applied the models to participants' response data from the final blocks of Experiment 1 (block 6) and Experiment 2 (block 5), where participants still received feedback.

Rationale

Decision-bound models (Ashby, 1992a; Maddox & Ashby, 1993) give information about the strategies participants use to separate perceptual stimuli into categories. We used decision-bound models to understand how individuals use rule-based processes during auditory and visual rule-based categorization and generalization.

Decision-bound models assume that participants separate stimuli into categories with a decision boundary. This boundary can be based on a single dimension or multiple dimensions. Additionally, participants can separate the categories using rules, which are thought to reflect overt decisional processes and hypothesis testing, or they can separate the categories with a boundary that reflects more implicit, procedural processes (Ashby, Alfonso-Reese, Turken, & Waldron, 1998). We fit a series of rule-based models that assume that participants separate the categories based either on the dimension that is relevant for categorization or the dimension that is irrelevant. We also fit an integration model that assumes that participants use both dimensions to separate the categories, in a manner that reflects implicit, rather than rule-based processing. Finally, we fit a random responder model that assumes that participants are randomly guessing.

Rule-based models

The rule-based models assume that participants draw a decision boundary along one of the two stimulus dimensions. We fit separate models assuming participants used a rule-based strategy along the category-relevant and category-irrelevant dimensions. The rule-based models have two free parameters – the location of the decision boundary along the dimension and a perceptual/criterial noise parameter. Rule-based models assume that participants are using hypothesis testing and overt rules to separate the stimuli into categories. For instance, while learning the auditory categories, a specific rule a participant could use would be to categorize all stimuli that have temporal modulation faster than 8 Hz into Category B and all stimuli slower than 8 Hz into Category A. A rule-based strategy is the optimal strategy to separate the categories in the current experiments.

Integration model

In contrast to rule-based models, the integration model assumes that participants use both stimulus dimensions to separate the categories. Integration strategies are thought to reflect more implicit, procedural learning processes, separating categories by a boundary that is not easily verbalizable (Ashby et al., 1998). The integration model assumes a linear decision boundary and has three free parameters: the slope and intercept of the decision boundary and a perceptual/criterial noise parameter. If a participant is using an integration strategy, it means they are using both dimensions to separate the categories, which is suboptimal in this case.

Random responder model

The random responder model assumes that participants guess on each trial.

Model fitting and selection

For each participant (30 in Experiment 1, 99 in Experiment 2) and each block (final categorization block, generalization test block), we fit rule-based, integration, and random responder models. For each model type, the model parameters were estimated using a maximum likelihood procedure (Ashby, 1992b; Wickens, 1982). Model selection used the Bayesian Information Criterion (BIC): BIC = r*ln$N$ - 2ln$L$, where r is the number of free parameters, N is the number of trials in a given block for a given subject, and L is the likelihood of the model given the data (Schwarz, 1978). The BIC allows for comparison of model fits because it penalizes models for extra free parameters such that the smaller the BIC, the closer the model is to the "true" model.

The model fitting and selection procedure produces the best-fitting model for each participant and each block (final categorization block, generalization test block). We grouped the models by whether they reflected rule-based or integration processing. No participants were best fit by the random responder model. Below, we report the percentage of participants best fit by the rule-based models.

Decision-bound modeling results

In the final block of training, there is no evidence that participants used more rule-based strategies in the auditory or visual task. There were no significant differences in strategy use between auditory and visual tasks (Experiment 1: McNemar's χ^2= 2.78, p = 0.096; Experiment 2: McNemar's χ^2 = 1.67, p = 0.20). In the final block of Experiment 1, 73% (22/30) of participants in the auditory task and 90% (27/30) of participants in the visual task used rule-based strategies. In Experiment 2, 89% (88/99) of participants in the auditory task and 94% (93/99) of participants in the visual task used rule-based strategies.

In contrast, during the generalization test, significantly more participants used rule-based strategies in the visual task than in the auditory task (Experiment 1: McNemar's χ^2= 4.0, p = 0.046; Experiment 2: McNemar's χ^2= 13.37, p = 0.00026). In the generalization block of Experiment 1, 83% (25/30) of participants in the auditory task and 97% (29/30) of participants in the visual task used rule-based strategies. In Experiment 2, 77% (76/99) of participants in the auditory task and 96% (95/99) of participants in the visual task used rule-based strategies.

Overall, these results demonstrate that there are no differences in auditory and visual rule-based processing during categorization, but during generalization, when there is no longer any feedback, more participants rely on rule-based processing for visual than auditory stimuli. These results also align with our other measures of performance in the generalization test. While there were not differences in overall accuracy (in the highly powered Experiment 2), there were differences in the pattern of responses. Visual category representations were more categorical than auditory category representations and participants found it easier to consistently apply a unidimensional rule to separate the visual categories even in the absence of feedback.

References

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2019). Gorilla in our Midst: An online behavioral experiment builder. *Behavior Research Methods*, *438242*. https://doi.org/10.3758/s13428-019-01237-x

Ashby, F. G. (1992a). *Multidimensional models of categorization* In F. G. Ashby (Ed.), Multidimensional Models of Perception and Cognition (pp. 449–483). Lawrence Erlbaum. Retrieved from http://psycnet.apa.org/psycinfo/1992-98026-016

Ashby, F. G. (1992b). Multivariate Probability Distributions. In F. G. Ashby (Ed.), Multidimensional Models of Perception and Cognition (pp. 1–34). Lawrence Erlbaum.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481. https://doi.org/10.1037/0033-295x.105.3.442

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), Speech Perception and Linguistic Experience: Issues in Cross-Language Research (pp.171-204). Timonium, MD: York Press.

Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed–accuracy tradeoff. *Trends in Neurosciences*, *33*(1), 10–16. https://doi.org/10.1016/j.tins.2009.09.002

Brashears, B. N., & Minda, J. P. (2020). The effects of feature verbalizability on category learning. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference on the Cognitive Science Society* (pp. 655–660). Austin, TX: Cognitive Science Society.

Crittenden, B. M., & Duncan, J. (2014). Task difficulty manipulation reveals multiple demand activity but no frontal lobe hierarchy. Cerebral Cortex, 24(2), 532–540. https://doi.org/10.1093/cercor/bhs333

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends in Neurosciences, 23(10), 475–483. https://doi.org/10.1016/s0166-2236(00)01633-7

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. Proceedings of the National Academy of Sciences, 110(41), 16616–16621. https://doi.org/10.1073/pnas.1315235110

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. Journal of Experimental Psychology: Human Perception and Performance, 28(2), 349–366. https://doi.org/10.1037//0096-1523.28.2.349

Garner, W. R. (1974). *The processing of information and structure*. Hillsdale, NJ: Erlbaum.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016. https://doi.org/10.1007/s11222-013-9416-2

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. Journal of Experimental Psychology: General, 123(2), 178–200.

Goudbeek, M., Swingley, D., & Smits, R. (2009). Supervised and unsupervised learning of multidimensional acoustic categories. Journal of Experimental Psychology: Human Perception and Performance, 35(6), 1913–1933. https://doi.org/10.1037/a0015781

Heffner, C. C., Idsardi, W. J., & Newman, R. S. (2019). Constraints on learning disjunctive, unidimensional auditory and phonetic categories. Attention, Perception & Psychophysics, 81(4), 958–980. https://doi.org/10.3758/s13414-019-01683-x

Lehnert, G., & Zimmer, H. D. (2006). Auditory and visual spatial working memory. *Memory & Cognition*, *34*(5), 1080–1090. https://doi.org/10.3758/bf03193254

Lesaffre, E., Rizopoulos, D., & Tsonaka, R. (2007). The logistic transform for bounded outcome scores. Biostatistics, 8(1), 72–85. https://doi.org/10.1093/biostatistics/kxj034

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. Journal of the Acoustical Society of America, 89(2), 874–886. https://doi.org/10.1016/j.biotechadv.2011.08.021.secreted

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. Psychological Review, 111, 309–332. https://doi.org/10.1037/0033-295x.111.2.309

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*(1), 49–70. https://doi.org/10.3758/bf03211715

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29(4), 650–662. https://doi.org/10.1037/0278-7393.29.4.650

Maddox, W. T., Chandrasekaran, B., Smayda, K., & Yi, H.-G. (2013). Dual systems of speech category learning across the lifespan. Psychology and Aging, 28(4), 1042–1056. https://doi.org/10.1037/a0034969

McClelland, J. L., Fiez, J. A., & McCandliss, B. D. (2002). Teaching the /r/–/l/ discrimination to Japanese adults: Behavioral and neural aspects. *Physiology & Behavior*, *77*, 657–662. Retrieved from file:///Users/devans/Documents/Papers2/Articles/2003/Unknown/2003R8705.pdf%5Cnpapers2://publication/uuid/D9D9D273-E580-4543-BB39-F6DA81E6B21F

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. Nature Neuroscience, 11(1), 103–107. https://doi.org/10.1038/nn2024

Myers, E. B. (2014). Emergence of category-level sensitivities in non-native speech sound learning. Frontiers in Neuroscience, 8, 1–11. https://doi.org/10.3389/fnins.2014.00238

Noppeney, U., Ostwald, D., & Werner, S. (2010). Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. *The Journal of Neuroscience*, *30*(21), 7434–7446. https://doi.org/10.1523/jneurosci.0455-10.2010

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General, 115(1), 39–57.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300. https://doi.org/10.1037/0033-295x.104.2.266

Nystrom, N. A., Levine, M. J., Roskies, R. Z., & 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. https://doi.org/10.1145/2792745.2792775.

Paulon, G., Llanos, F., Chandrasekaran, B., & Sarkar, A. (2020). Bayesian semiparametric longitudinal drift-diffusion mixed models for tone learning in adults. *Journal of the American Statistical Association*, 1–14. https://doi.org/10.1080/01621459.2020.1801448

Rabi, R., & Minda, J. P. (2014). Rule-based category learning in children: The role of age and executive functioning. PLoS ONE, 9(1), e85316. https://doi.org/10.1371/journal.pone.0085316

Ratcliff R, (1978) A theory of memory retrieval.. Psychological Review 85 (2):59-108

Roark, C. L., & Holt, L. L. (2019). Perceptual dimensions influence auditory category learning. Attention, Perception & Psychophysics, 81(4), 912–926. https://doi.org/10.3758/s13414-019-01688-6

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. https://doi.org/10.3758/s13421-013-0294-9

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. https://doi.org/10.1073/pnas.0907682106

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. Journal of Mathematical Psychology, 32(2), 135–168. https://doi.org/10.1016/0022-2496(88)90043-0

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., … Wilkens-Diehr, N. (2014). XSEDE: Accelerating scientific discovery. *Computing in Science & Engineering*, 16(5):62-74. https://doi.org/10.1109/MCSE.2014.80.

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. https://doi.org/10.1371/journal.pbio.0050056

Wickens, T. D. (1982). *Models for behavior: stochastic processes in psychology*. San Francisco, CA: W. H. Freeman.

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. https://doi.org/10.1121/1.4961163

Yi, H.-G., Maddox, W. T., Mumford, J. A., & Chandrasekaran, B. (2014). The role of corticostriatal systems in speech category learning. *Cerebral Cortex*, 1–12. https://doi.org/10.1093/cercor/bhu236

Zettersten, M., & Lupyan, G. (2020). Finding categories through words: more nameable features improve category learning. *Cognition*, *196*, 539–547. https://doi.org/10.17605/osf.io/uz2m9

Zvyagintsev, M., Clemens, B., Chechko, N., Mathiak, K. A., Sack, A. T., & Mathiak, K. (2013). Brain networks underlying mental imagery of auditory and visual information. *European Journal of Neuroscience*, *37*(9), 1421–1434. https://doi.org/10.1111/ejn.12140

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