nature neuroscience

Article

https://doi.org/10.1038/s41593-025-01899-1

Expectation-driven sensory adaptations support enhanced acuity during categorical perception

Received: 7 February 2023

Accepted: 21 January 2025

Published online: 13 March 2025

Check for updates

Tim Sainburg[1,2,10] ✉, Trevor S. McPherson[3,10], Ezequiel M. Arneodo[1,4], Srihita Rudraraju[1], Michael Turvey[1], Bradley H. Theilman[3], Pablo Tostado Marcos[5,6,7], Marvin Thielk[3] & Timothy Q. Gentner[1,3,8,9] ✉

Expectations can influence perception in seemingly contradictory ways, either by directing attention to expected stimuli and enhancing perceptual acuity or by stabilizing perception and diminishing acuity within expected stimulus categories. The neural mechanisms supporting these dual roles of expectation are not well understood. Here, we trained European starlings to classify ambiguous song syllables in both expected and unexpected acoustic contexts. We show that birds employ probabilistic, Bayesian integration to classify syllables, leveraging their expectations to stabilize their perceptual behavior. However, auditory sensory neural populations do not reflect this integration. Instead, expectation enhances the acuity of auditory sensory neurons in high-probability regions of the stimulus space. This modulation diverges from patterns typically observed in motor areas, where Bayesian integration of sensory inputs and expectations predominates. Our results suggest that peripheral sensory systems use expectation to improve sensory representations and maintain high-fidelity representations of the world, allowing downstream circuits to flexibly integrate this information with expectations to drive behavior.

Categorical perception groups smoothly varying signals into discrete classes, affording generalization at the expense of acuity. In many settings, categorical perception is driven by expectation. For example, in speech, as contexts change, perception is biased toward likely sounds, words and phrases[1–3], reflecting a shift in prior expectations. This warping of perception toward expected categories is called the 'perceptual magnet effect' (refs. 4,5) and can be formally described as a process of Bayesian inference over acoustic distributions[3,6–9]. Under this framework, an optimal perceiver resolves sensory ambiguity by integrating noisy or imperfect sensory information with information about their prior expectations in a particular context. By contrast, prior expectations also play a pivotal role in enhancing sensory acuity by refocusing attention toward expected regions of stimulus space[10–12]; this ability of priors to bias our sensory systems enables more accurate discrimination of closely related signals[13–15]. In the domain of language, this phenomenon is exemplified by listeners'

[1]Department of Psychology, University of California, San Diego, San Diego, CA, USA. [2]Center for Academic Research and Training in Anthropogeny, University of California, San Diego, San Diego, CA, USA. [3]Neurosciences Graduate Program, University of California, San Diego, San Diego, CA, USA. [4]Departamento de Física, Universidad Nacional de La Plata, La Plata, Argentina. [5]Department of Bioengineering, University of California, San Diego, San Diego, CA, USA. [6]Department of Electrical and Computer Engineering, University of California, San Diego, San Diego, CA, USA. [7]Institute for Neural Computation, University of California, San Diego, San Diego, CA, USA. [8]Neurobiology Section, Division of Biological Sciences, University of California, San Diego, San Diego, CA, USA. [9]Kavli Institute for Brain and Mind, University of California, San Diego, San Diego, CA, USA. [10]These authors contributed equally: Tim Sainburg, Trevor S. McPherson. ✉e-mail: tim_sainburg@hms.harvard.edu; tgentner@ucsd.edu

https://doi.org/10.1038/s41593-025-01899-1

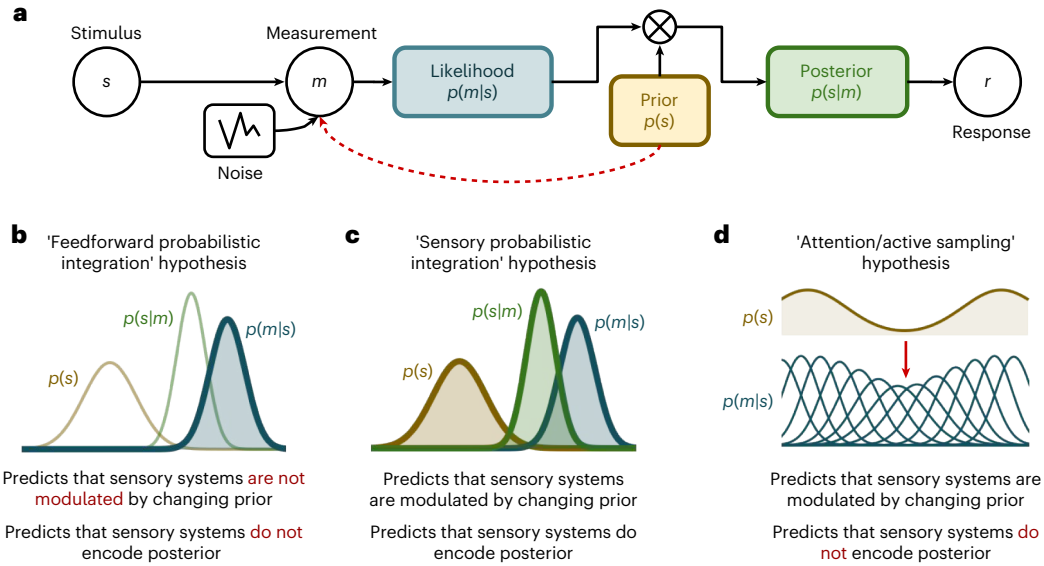


Fig. 1 | Models of expectation-driven sensory modulation. a, Bayesian models of perception predict that physical stimuli, *s*, received by sensory systems are corrupted by noise in measurement, *m*, representing the likelihood $p(m|s)$, which is integrated with prior expectations, $p(s)$, to form a posterior probability, $p(s|m)$, on which decisions are made[21]. **b**, In a modular implementation of Bayesian integration, sensory populations represent only the likelihood (denoted by shading). **c**, In a distributed implementation, sensory systems may reflect all components of the Bayesian integration model. **d**, Alternatively, attention and active sampling may modulate how sensory systems measure real-world stimuli (dashed red arrow in **a**), altering the likelihood (acuity) on the basis of expectation, without reflecting the posterior.

ability to fine-tune speech recognition based on the characteristics of the speaker's voice[16]. It remains unknown how these dual contrasting functions of prior expectations (generalization underlying categorization and sharpened sensory acuity) are implemented neurally. Specifically, the extent to which early sensory representations are influenced by prior expectation remains unclear. This ambiguity is highlighted in speech sound perception, which can manifest as either categorical or continuous depending upon the specific task at hand[17,18]. The task-dependent nature of perception fuels ongoing debates about the inherent nature of categorical perception: whether it is an intrinsic aspect of the sensory systems responsible for perception or a product of downstream decision-making processes[19–22].

One possibility is that sensory codes are not impacted by expectation. Most empirical evidence for probabilistic integration derives from work on regions of the brain associated with motor and decision making[23–28] and suggests that sensory populations encode a likelihood distribution[29] integrated with expectation by downstream circuits. Alternatively, expectation and context could be integrated in sensory systems themselves, which then are used as a substrate for decision making; such modulation could occur via feedback loops with higher-order systems (for example, decision making and multisensory integration[30]) or through local network dynamics[21,31–33]. Both theoretical work[19,21,31,34] and recent brain-wide analyses[22,35] raise the possibility that even early sensory brain regions engage in probabilistic integration of information about sensation and expectation. However, it is also possible that expectation impacts sensory codes but does so in a manner that is inconsistent with classical Bayesian integration. Instead, sensory modulation may be related to the active role that prior expectations play in shaping our interaction with the environment. For example, prior expectations drive attention and active sensing[36,37], which modulates sensory encoding[11,15,38,39] and improves sensory acuity[12–14,40,41]; these findings suggest that expectation may flexibly focus neural resources on regions of sensory space where signals are expected, thus improving perceptual acuity rather than pulling percepts toward expected categorical representations.

These three hypotheses can be restated in the language of Bayes' rule. Noisy measurements of physical signals received by sensory systems represent the likelihood of a particular stimulus. The likelihood is then integrated with prior expectations to form a posterior probability on which decisions are made (Fig. 1a). In the 'feedforward probabilistic integration' account, sensory systems represent only the likelihood and prior probabilities and the posterior distribution are not reflected in the sensory brain (Fig. 1b). By contrast, the 'sensory probabilistic integration' account predicts that sensory populations will reflect the Bayesian integration of prior and likelihood (Fig. 1c). Finally, in the 'attention and active sampling' account, expectations sharpen the stimulus likelihood without directly integrating the prior probability distribution (Fig. 1d).

Here, we analyze the activity patterns of sensory neuron populations in an auditory perceptual decision-making task to discern which computational model (feedforward probabilistic integration, sensory probabilistic integration or attention and active sampling) most accurately explains neural processing in sensory systems. Songbirds provide an important opportunity to study mechanisms of categorical vocal perception in neurobiological detail, as they perceive some elements of song categorically[2,42] and those elements are biased by expectation[43]. We developed methods to explicitly impose probabilistic predictive information in a sequence of birdsong syllables and trained European starlings, a songbird with complex vocal repertoires, to classify smoothly varying syllables while controlling sequential predictive song structure. To begin, we show that categorical perception of vocal elements is explained well by Bayesian probabilistic integration and that sensory neural responses capture the perceptual uncertainty (that is, the likelihood) of a Bayesian model representing the birds' behavior. We observe that sensory neural responses are directly modulated by the predictive structure of vocal sequences, thereby ruling out the feedforward probabilistic integration hypothesis. These response biases do not align, however, with a Bayesian integration of prior and likelihood (that is, the posterior) as predicted by the sensory probabilistic integration hypothesis. Instead, we find that the bias is consistent with dynamic changes in sensory acuity (that

https://doi.org/10.1038/s41593-025-01899-1

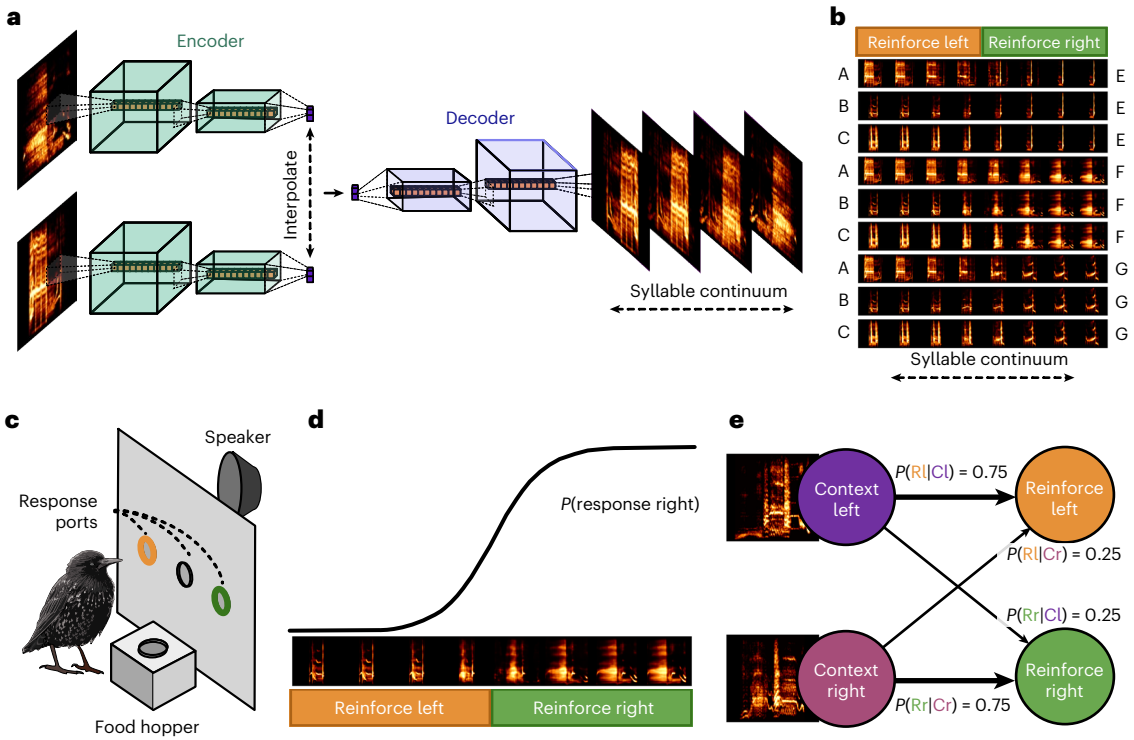


Fig. 2 | Context-dependent categorical perception paradigm. a, Syllable morphs are generated as interpolated projections between two song syllables in the latent space of a neural network. **b**, Example syllables from the nine syllable continua (rows) used for behavioral training. The reinforced category is shown on the top, and the endpoint syllables are labeled on the left and right sides. **c**, Operant apparatus used for this experiment. Green and orange response ports correspond to the syllable classes in **b**. **d**, A psychometric curve depicting syllable classification over one continuum. **e**, Two example context cue syllables precede the target syllables, holding predictive information about the class to which each belongs. Cl, context left; Cr, context right; Rl, reinforce left; Rr, reinforce right.

is, the likelihood of the Bayesian model) in predicted regions of acoustic space. As a consequence, an unbiased representation of the sensory signal is left available for flexible use in behavior.

Results

Categorical decision making is modulated by expectation

Sensory neuroscience and psychophysics have long, productive histories founded on the idea of relating parametric change in a stimulus to quantifiable changes in both neural activity[44] and behavior[45]. Implicit in this approach is the strong assumption that sensory inputs can be discretized into stimulus events parametrically varying along one or two continuous dimensions. This approach is ideally suited to investigate how simple, nonnatural and easily controllable signals are perceived behaviorally or encoded neurally but neglects the natural history of sensory systems, which are adapted to complex ethologically relevant signals like birdsong[46,47]. Attempts to apply the same kind of parametric stimulus control to natural stimuli are rare because natural signals tend to vary along multiple dimensions simultaneously[48,49].

To address this challenge, we developed a behavioral paradigm to control context-dependent categorical perception in a natural stimulus space by synthesizing smoothly varying starling song syllables using a neural network. We captured the complex spectrotemporal statistics of song acoustics using a deep convolutional variational autoencoder (VAE; Fig. 2a)[50] trained on a large library of conspecific song. From the latent space of this network, we synthesized acoustic continua ($N = 9$), each comprising 128 synthetic syllables (morphs) that smoothly vary between two acoustically distinct syllable endpoints. We trained starlings using a two-alternative choice category learning task to classify the naturalistic syllable morphs that lie along these continua (Fig. 2b). We divided each continuum at the midpoint and reinforced one half with food for pecks into the left response port and the other half for pecks into the right response port (Fig. 2c, bottom). We trained birds ($n = 20$) on the syllable classification task to obtain psychometric functions for each syllable continuum (Fig. 2d) and then introduced cue syllables preceding the target (to-be-classified) syllable. Each cue syllable provided predictive information about the likely response category of the target syllable (Fig. 2e). All subjects learned the task to at least 75% accuracy (Supplementary Table 1), performing a total of 4.8 million behavioral trials.

We fit a psychometric model (Fig. 3a) to each subject's classification behavior for each syllable continuum (Fig. 3b) and then used the parameters of the fit psychometric model to understand how the cue affected behavior. Under the Bayesian integration hypothesis (Fig. 2f–h), categorical perceptual decision making (that is, syllable classification) is modulated by integrating the likelihood imposed by the stimulus (that is, the target syllable) with the prior imposed by its sequential context (that is, the preceding cue syllable). As a result, the decision boundary (Fig. 3a, inflection point) shifts in the direction predicted by the cue (Fig. 3c, top). We also examined whether information from the cue and target syllables are treated independently, in a non-Bayesian manner, as evidenced by an overall shift in the probability of a left or right response but not a shift in the decision boundary (Fig. 3c, bottom).

Across each syllable continuum and for each bird, we observed robust shifts in the decision boundary (Fig. 3b,d–f; linear mixed effects; psychometric inflection ~ cue probability + (1|subject) + (cue probability|subject); cue probability: $\beta = -11.12$, s.e.m. = 1.158, $z = -9.6$, $P < 0.001$), consistent with Bayesian integration underlying context-dependent categorical perception. To examine this shift more closely, we fit the Bayesian model (in particular, the likelihood) to each bird's behavioral data in uncued trials (for each continuum) and predicted the inflection point shift given each cue probability. The red

https://doi.org/10.1038/s41593-025-01899-1

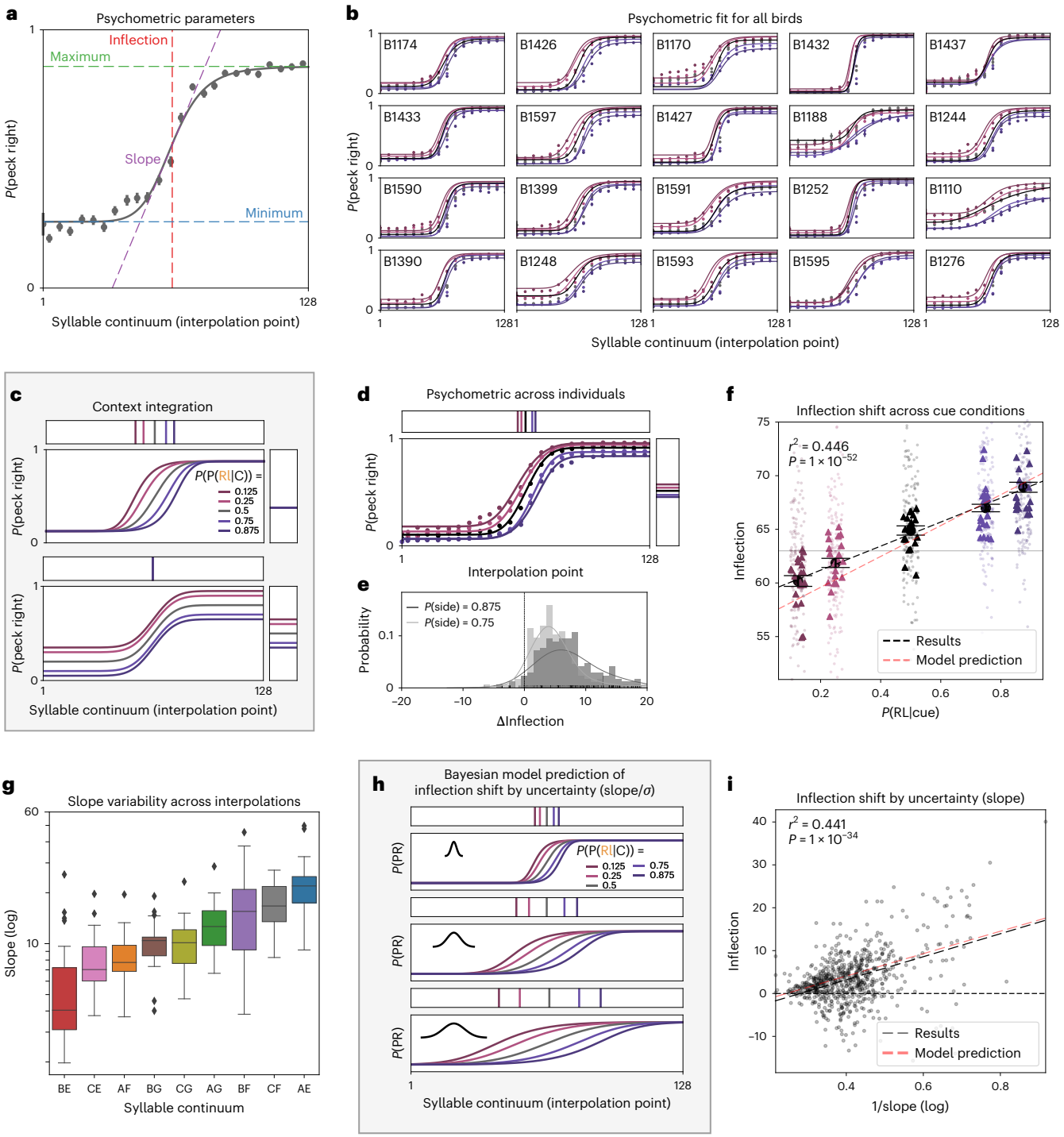


Fig. 3 | Decision-making behavior reflects Bayesian integration. a, An example psychometric fit with parameters. **b**, Psychometric fits for cued conditions for each of the subjects. **c**, Top, an example of the context-dependent category shift as a function of the cue, as predicted by the Bayesian model. **c**, Bottom, an example of an alternative hypothesis, in which decisions are made either using the cue or the categorical stimuli, without integration of the two sources of information results in no category boundary shift. The corresponding lines in the connected horizontal and vertical boxes indicate the shift in the inflection point (vertical lines) as well as the midpoint between mid and maximum in the psychometric function (horizontal lines). Colors indicate the cue probabilities given in Fig. 2. **d**, The results across birds and morphs indicate that both strategies from **c** are present in behavior. **e**, Cue shift between left and right cues for each morph and bird at $P = 0.875$ and 0.75. **f**, The categorical boundary (inflection point) shifts as a function of the strength of the cue (Pearson's correlation; $n = 1{,}050$ over each bird, morph or cue). RL, reinforce left. The Bayesian model predicts a similar shift from the uncued data (red line). Points correspond to each morph for each bird, and triangles correspond to each bird, averaged over morphs. Error bars represent s.e.m. for each cue. **g**, Morphs (interpolations) vary on the slope of the fit psychometric function, indicating variation in uncertainty in decision making by morph ($n = 1{,}050$). For box plots, the center line represents the median, the bounds of the box span from the 25th to the 75th percentile (the interquartile range), and the whiskers extend to the furthest data points within 1.5 times the interquartile range. **h**, The Bayesian model predicts a greater shift in categorical boundary as a function of the uncertainty of the categorical stimulus (σ of the likelihood and slope of the psychometric model). PR, peck right. **i**, As predicted by the Bayesian model, the shift in the categorical boundary increases as a function of uncertainty (Pearson's correlation; $n = 700$ over each bird or morph). Points correspond to the shift in inflection between left and right cues (averaged across cue strength conditions) for each morph for each bird.

dashed line in Fig. 3f depicts a linear regression showing the close correspondence between the observed shift in inflection point and that predicted by the Bayesian model. In addition, we observed an overall shift in decision probability (Fig. 3d), suggesting that, in a subset of trials, subjects responded independently to the cue or the target syllable alone, which aligns with previous findings that animals alternate between decision-making strategies from trial to trial[51].

We observed substantial variation in the slope of the psychometric functions fit to each bird's behavior. Some individuals showed a much sharper categorical boundary than others (for example, B1432 versus B1110 in Fig. 3b), and the mean slope (averaged across individuals) also varied between syllable continua (Fig. 3g). The slope of the psychometric curve reflects uncertainty in the Bayesian model. Under greater uncertainty about the target syllable, the Bayesian model predicts that integration with the cue stimulus will result in a greater shift in categorical perception (that is, the inflection point; Fig. 3h ref. 52). Consistent with this, we observed a smaller inflection point shift in the direction of the cue as the slope of the psychometric curve steepened (Fig. 3i; linear mixed effects; inflection shift ~ psychometric slope + (1|subject) + (psychometric slope|subject); cue probability: $\beta = -0.401$, s.e.m. = 0.099, $z = -4.03$, $P < 0.001$), which again matches the quantitative prediction of the model (Fig. 3i, red dashed line).

Both the likelihood of a given stimulus and its prior probability were reflected in the response times of birds. Response times were longer in incorrect trials than in correct trials (Extended Data Fig. 1a; linear mixed effects; response time ~ correct response + (1|subject); correct response: $\beta = -0.298$, s.e.m. < 0.001, $z = -339.448$, $P < 0.001$), suggesting that challenging decisions take longer to make. When looking only at trials where the bird was correct and controlling for side bias (Methods), we found that response times decreased proportionally to the prior probability imposed by the cue (Extended Data Fig. 1b) and that response times near the center of the morph increased following the bird's psychometric slopes for each morph (that is, the likelihood; Extended Data Fig. 1c,d).

Sensory neural responses reflect behavioral likelihood

These behavioral results indicate that, in our task, birds are probabilistically integrating expectations with sensory experiences to categorize song syllables. To investigate whether sensory forebrain neural populations reflect Bayesian integration, we recorded extracellular neural spiking activity using one to two (unilaterally or bilaterally) implanted 32–64-channel 1–8 shank silicon electrode arrays in freely behaving subjects ($N = 10$) while they completed trials on the syllable categorization task and passively listened to the same stimuli (during both sleeping and waking states). We targeted electrode arrays broadly across the auditory forebrain, including the primary auditory region field L (Extended Data Fig. 2) and secondary regions caudal mesopallium, caudomedial nidopallium and medial caudolateral nidopallium (NCL). We recorded from a total of 13,854 putative single neurons (Spike sorting and merging over long-term chronic recordings).

We analyzed spike train data as spike vectors over the different syllable continua by convolving the time histogram (bin width = 10 ms) of the stimulus-aligned spike train for each trial with a Gaussian kernel ($\sigma = 25$ ms; Fig. 4a–e). Fig. 4e,f shows sample spike trains and trial-averaged spike vectors for a sample unit for each syllable continuum. From the trial-averaged spike vectors, we computed a cosine similarity matrix between spike vectors for each syllable on each continuum (Fig. 4i) from which we then computed a neurometric function (Methods and Fig. 4j). We also used the cosine similarity matrix to compute a metric for each unit's task relevance (Fig. 4k,l and Assessing task relevance for units), reflecting the similarity of the unit's response within versus between syllable categories. Importantly, this analysis is not meant to suggest that these neurons reflect learned categories, only that they show response variance over the task-relevant stimulus space. Of the 13,854 units recorded, 7,994 had task-relevant responses to the syllable continua (Subsetting task-relevant units). On average, the spike vector responses for these task-relevant units changed smoothly across the syllable continua, but the degree of this smoothness varied (Fig. 4m,n).

To assess whether neural responses reflected behavior, we compared the slope of the neurometric function to the slope of the psychometric function for each bird and syllable continuum. The neurometric slopes were predicted well by the psychometric function (Fig. 4o,p; linear mixed effects; log (neurometric slope) ~ log (psychometric slope) + (1|unit) + (log (psychometric slope)|unit); log (psychometric slope): $\beta = 0.172$, s.e.m. < 0.005, $z = -37.006$, $P < 0.001$). Because, in the Bayesian decision-making model, the slope of the psychometric function is modulated by the likelihood, that is, stimulus uncertainty, it follows that these neural responses carry information about the stimulus uncertainty. Additionally, we assessed whether individual bird-level variation in psychometric slopes was reflected in the unit neurometrics and found that it was not (Methods), indicating a possible downstream role for estimates of stimulus uncertainty in decision making.

Expectation suppresses sensory spike rate

Prior work has established that expectation modulates neural activity in sensory and decision-making brain regions, with activity in decision-making regions increasing with predictability and activity in sensory regions decreasing[11]. The reduction in activity in sensory areas during predictable stimuli may reflect the dampened coding of task-irrelevant features, which could be useful for improving acuity[12]. To assess whether cue syllables modulate neural responses to target syllables, we quantified how much the overall spike rate in each unit changed as a function of the predictive cue syllable in trials in which birds behaviorally responded to stimuli. The presence of a cue syllable significantly suppressed the spike rate evoked by the target syllable when controlling for spike rate variability across units (Fig. 5a; ANOVA for linear mixed-effects model comparison: $\chi^2(4, N = 857{,}301) = 15{,}196$, $P < 1 \times 10^{-5}$; see Methods for details). This suppression was consistent across the motif continuum, stronger in active trials than in passive playbacks (Fig. 5b), and was most prominent early and continued throughout much of the target stimulus playback (Fig. 5c). Moreover, the magnitude of the cue-dependent suppression was consistent within each cue condition and persisted throughout stimulus playback. In passive playback trials, any cue-dependent effects quickly diminished (Fig. 5d).

Because cue syllables are differentially informative (that is, they establish different priors) for upcoming target syllables, we reasoned that the magnitude of cue-specific response suppression might covary with the strength of the predictive information. We therefore measured the impact of the cue's predictive probability on spike rates while controlling for differences in each unit's response between syllable continua (Supplementary Fig. 13). We found that, as the predictive strength of the cue syllable increased, the associated spike rates decreased (Fig. 5e; ANOVA for linear mixed-effects model comparison: $\chi^2(1, N = 857{,}301) = 399$, $P < 1 \times 10^{-5}$; see Methods for details). This effect was abolished during passive playback (Fig. 5f) and when cue labels were shuffled (Supplementary Fig. 13). In sum, these results are consistent with previous observations that signal predictability decreases responses to stimuli (that is, expectation suppression[11]), which is believed to reflect a dampening of task-irrelevant noise[12].

Expectation drives sensory likelihood modulation

As Fig. 5 shows, prior expectations modulate sensory neural responses. This rules out the feedforward probabilistic integration hypothesis that sensory populations representing the stimulus are unmodulated by expectation (Fig. 1b). The remaining two hypotheses make opposing predictions about how sensory populations should be modulated by prior expectations. The attention and active sampling hypothesis predicts that sensory representations become increasingly discriminable

https://doi.org/10.1038/s41593-025-01899-1

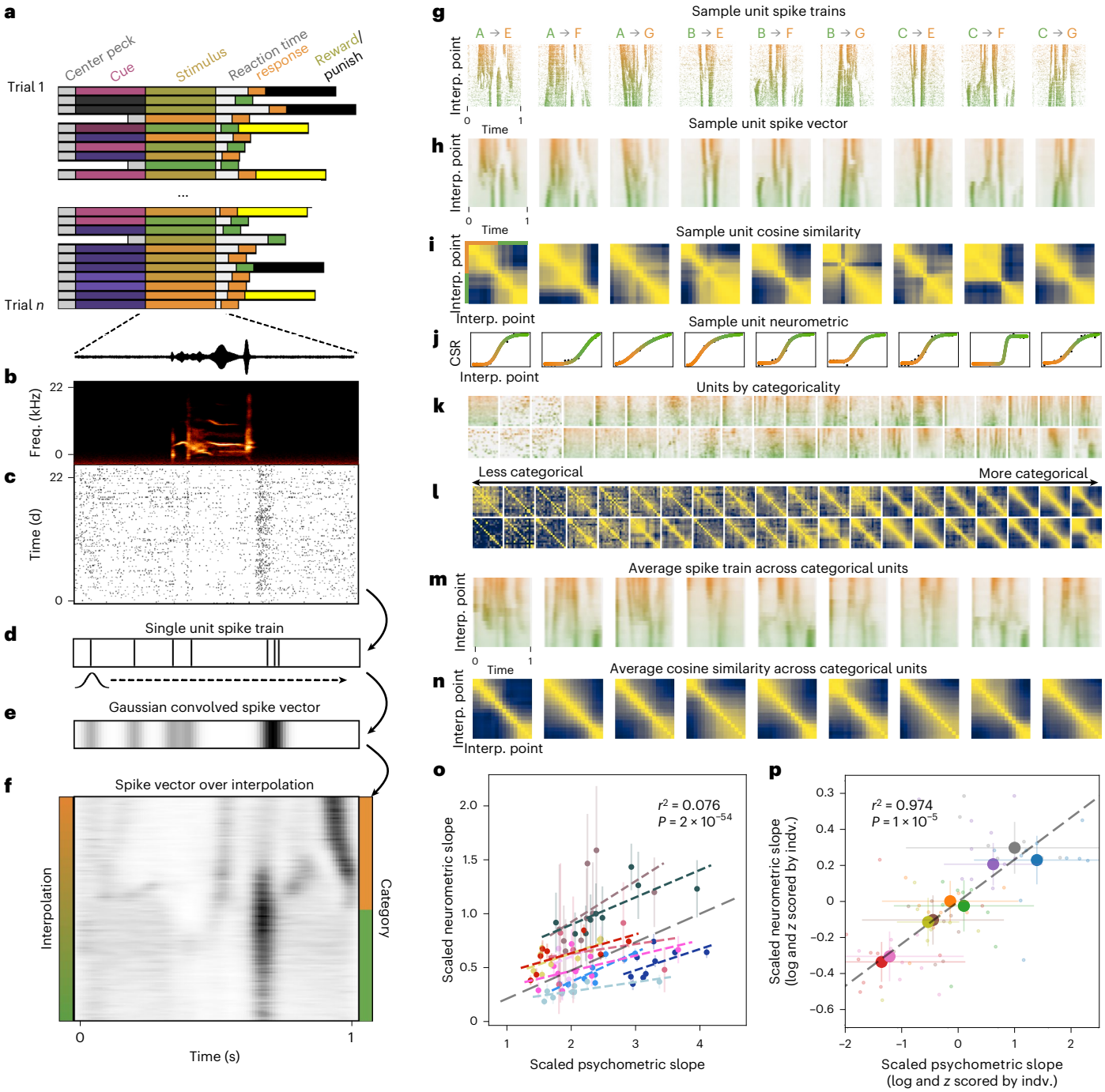


Fig. 4 | Neurometric functions of single units reflect perceptual likelihood. **a**, Trial-by-trial behavioral data structure. **b**, Spectrogram of a categorical (morph) stimulus for a single trial. Freq., frequency. **c**, Spike raster for a single unit across trials. **d**, A single example spike train from **c**. **e**, A spike vector is computed as the spike train from **d** convolved with a Gaussian kernel. **f**, The average spike vector for the unit in **c**,**d** for a single morph (A → E). **g**, Sample spike trains for one unit across nine morphs. Interp., interpolation. **h**, Spike vector representations of the spike trains from **g**. **i**, Cosine similarity matrices computed from the spike trains in **h**. **j**, Neurometric functions are computed from the similarity matrices in **i**. CSR, categorical similarity ratio (Estimating a neurometric function from the similarity matrix). **k**, Sample morph spike vectors (as in **h** for units, sorted by unit categoricality (rows show two examples of the same categoricality). **l**, Similarity matrices for the units in **k**. **m**, Average spike trains across each task-relevant unit for morphs. **n**, Average cosine similarity matrices across all task-relevant units for each morph. **o**, Psychometric slope (log transformed and scaled by the psychometric range) versus neurometric slope (log transformed and scaled by the psychometric range) for each subject and morph. Each subject is shown with a unique color and regression line. Points are means across subjects or morphs, and error bars correspond to the mean ± 3 × s.e.m. Pearson's correlation is computed between the psychometric and neurometric slopes across units and morphs ($n = 41{,}181$). A regression line for pooled data is shown in gray. **p**, The same data as in **o** *z* scored by subject, where color corresponds to syllable continua (as in Fig. 3). Large points correspond to the average psychometric and neurometric slopes across subjects for each continuum, and small points correspond to individual (indv.) continua or birds. Error bars correspond to the mean ± 3 × s.e.m. Pearson's correlation between mean psychometric and neurometric is plotted ($n = 9$ morphs).

https://doi.org/10.1038/s41593-025-01899-1

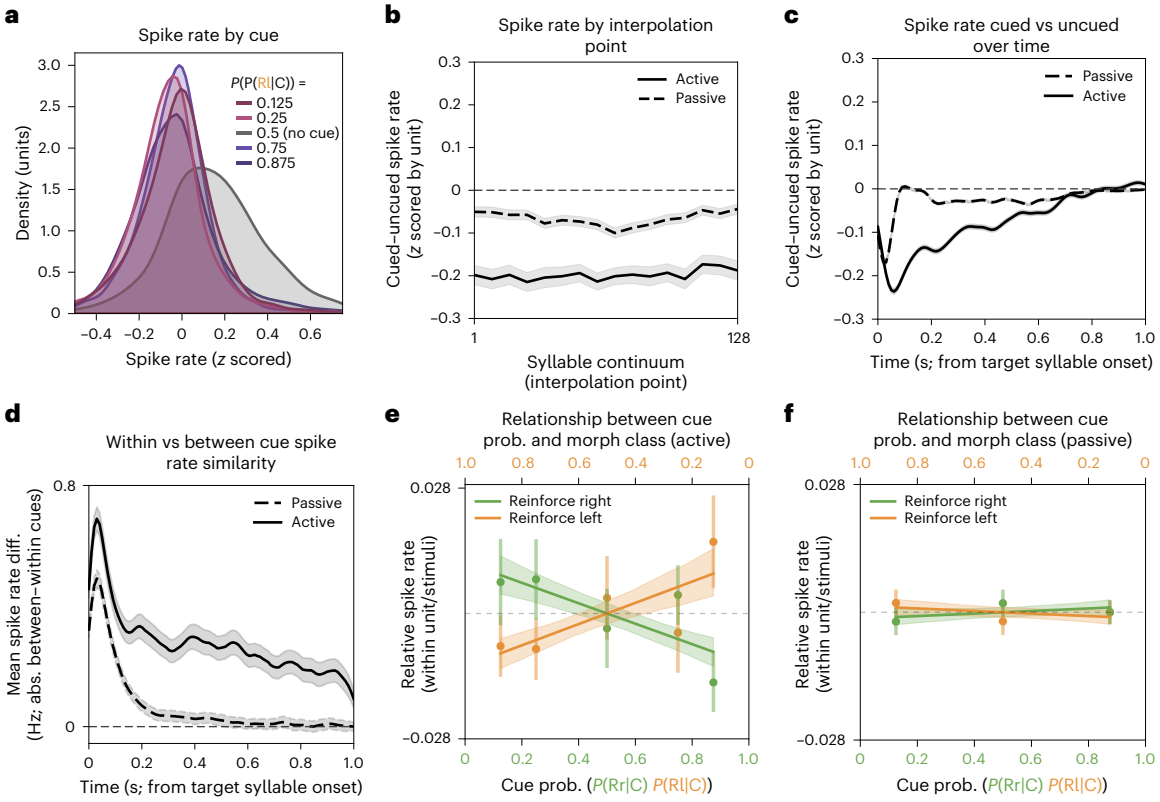


Fig. 5 | Predictive syllables suppress spike rate. a, Differences in spike rate for each predictive cue syllable for active behavior playbacks, where uncued trials yield the highest spike rate. Spike rate is *z* scored within each unit, and the difference is shown as the difference within stimulus across cueing conditions for each unit. **b**, Spike rate suppression across units for cued trials by interpolation point. Confidence intervals reflect the mean ± 3 × s.e.m. **c**, Spike rate suppression across units for cued trials by time. Confidence intervals reflect the mean ± 3 × s.e.m. **d**, Within-cue versus between-cue spike rate similarity over time. Confidence intervals reflect the mean ± 3 × s.e.m. Mean difference (diff.) in spike rate between cue conditions minus mean difference within cue conditions across time for passive and active trials. A value above zero indicates greater similarity within cues than between cues. Abs., absolute. **e**, Relationship between cue probability (prob.) and morph class, within unit and stimulus, measured through spike rate across morph playback. For both the left and right morph stimuli, a regression line is shown with a 95% bootstrapped confidence interval (n = 857,301 unit–stimuli pairs). Points reflect the mean, and error bars show the mean ± 3 × s.e.m. **f**, Same as **e** for passive trials (n = 540,033 unit–stimuli pairs).

in high-probability regions of stimulus space (Fig. 6b,d). In our model of attention and active sampling (Extended Data Fig. 3), we assume that an increase in sensory acuity in one region of sensory space (here, the cued section of the syllable continuum) comes at the expense of acuity in other sensory dimensions (that is, acoustic features irrelevant to our categorization task; see Methods for details)[53,54]. Under this model, if neural responses reflect the stimulus likelihood, their similarity should decrease as a function of predictive cue strength (Fig. 6f,i), both because signals become more discriminable along the task-relevant dimension (here, the syllable continuum) and because representational noise along task-irrelevant dimensions increases. Alternatively, Bayesian integration mirrors an effect in categorical phoneme perception called the perceptual magnet effect[4] in which speech perception is warped around categorical boundaries to reduce discriminability of within-category sounds (Fig. 6c,e). In the Bayesian model, this perceptual warping results from the integration of prior distributional information with a noisy representation of the acoustic signal, yielding a shift in the posterior toward higher-probability regions of acoustic space[6] (Fig. 6c). As a result, similarity within high-probability regions of stimulus space increases, compressing within-category representations together (that is, perceptual magnetism). In the context of our task, increasing predictive probability toward one side of the syllable continuum (that is, in the context of a predictive cue) leads to two outcomes: the within-category similarity of the posterior on the predicted side of the continuum will increase and the within-category similarity of the low-probability side of the continuum will decrease (Fig. 6g). Under this model, if neural responses reflect this posterior distribution, their similarity should also increase as a function of predictive cue strength (Fig. 6f,i).

To determine which model best aligns with our neural data, we employed two methodologies. First, we directly compared similarity of neural responses to syllables across the different continua as a function of cue condition. Second, we used a decoder model to estimate the accuracy of stimulus and stimulus class predictions from neural data in different cue conditions.

Comparing the trial-to-trial cosine similarity of the spike vector response across syllable continua revealed that, in the presence of a predictive cue, the within-category similarity was higher in the nonpredicted (cue-invalid) class than in the predicted (cue-valid) class (Fig. 6h; linear mixed effects, Methods; $\beta = 0.009$, s.e.m. < 0.001, $P < 0.001$). Moreover, the within-category similarity across units and continua decreased significantly as a function of the probability of the cue class (Fig. 6j; linear mixed effects, Methods; $\beta = -0.01$, s.e.m. < 0.001, $P < 0.001$). This effect was abolished when cue labels were shuffled (Supplementary Fig. 1), suggesting that perceptual acuity is selectively sharpened over predicted regions of acoustic space, which likely decreases the overall representational noise. To test this directly, we compared the variance in spike rates as a function of cue validity. On average, spike rate variance in the cue-valid condition was slightly but significantly lower than that in the cue-invalid condition (s.d. ~ cue

https://doi.org/10.1038/s41593-025-01899-1

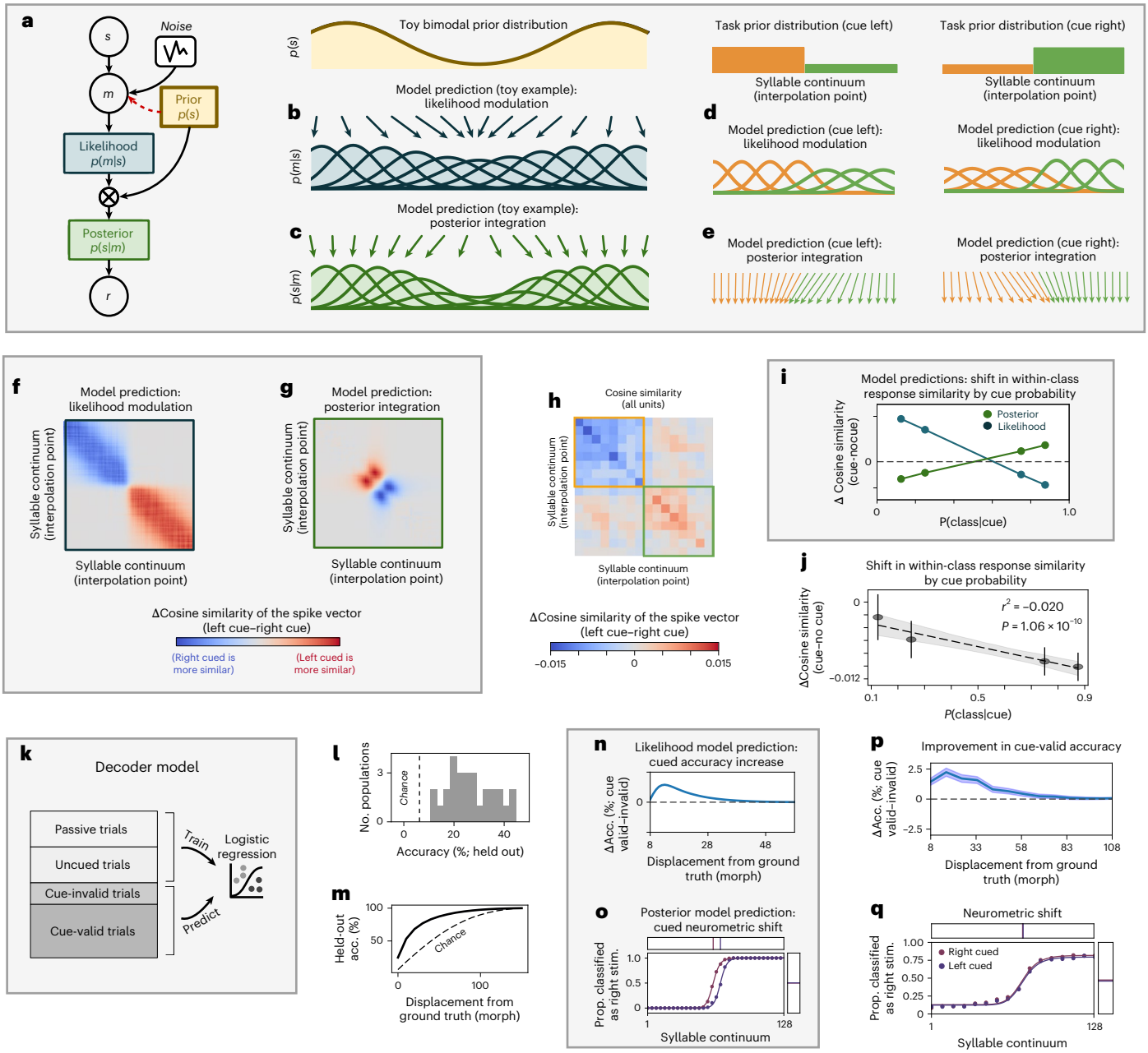


Fig. 6 | Context-dependent spike train modulation reflects change in perceptual acuity and not Bayesian integration. a, Bayesian integration model. **b**, Attention and active sampling model predicts decreased similarity in high-probability regions due to noise modulation, shown on a toy bimodal prior. Arrows denote a one-dimensional (1D) graph embedding with cosine similarity weights. **c**, Sensory probabilistic integration model reflects the perceptual magnet effect, compressing high-probability vocalizations. Arrows indicate shifts from true stimulus positions to posterior distribution peaks. **d**, Likelihood modulation model for our task, where acuity increases with predictability, as in **b**. **e**, Posterior integration model for our task, where boundary distribution shifts based on cue probability, as in **c**. **f**, Shift in representational similarity predicted by the likelihood model between left- and right-cued stimuli. **g**, Shift in representational similarity predicted by the posterior integration model between left- and right-cued stimuli. **h**, Observed shift in spike train vector cosine similarity for left-cued versus right-cued trials, averaged across units and morphs; compare to models in **f**,**g**. **i**, Model predictions for similarity changes: decrease in sensory modulation and increase in Bayesian integration relative to stimulus probability. **j**, Relationship between stimulus class probability and similarity shift from baseline (uncued); compare to model predictions in **i** (Pearson's correlation; $n = 108{,}153$ unit, cue and stimulus samples). Points correspond to mean ± 3 × s.e.m. Line plot shows bootstrapped 95% confidence intervals. **k**, Neural population decoder model overview. A logistic regression is trained on passive and uncued trials, evaluating accuracy on held-out cued data. **l**, Decoder accuracy histogram for held-out cued data, showing above chance performance. **m**, Mean decoder accuracy (acc.) by classifier specificity. The *x* axis reflects maximum interpolation point deviations from ground truth, called a correct prediction. **n**, The likelihood modulation model predicts higher classification accuracy for cue-valid conditions than for cue-invalid conditions. **o**, The posterior shift model predicts that classifier predictions will shift between left- and right-cued stimuli (stim.). prop., proportion. **p**, Improvement in accuracy for cue-valid trials over cue-invalid trials across all populations. Compare to **n**. The confidence interval shows mean improvement in accuracy ± 3 × s.e.m. **q**, Neurometric shift of the decoder for left- and right-cued trials across all populations. Compare to **o**.

https://doi.org/10.1038/s41593-025-01899-1

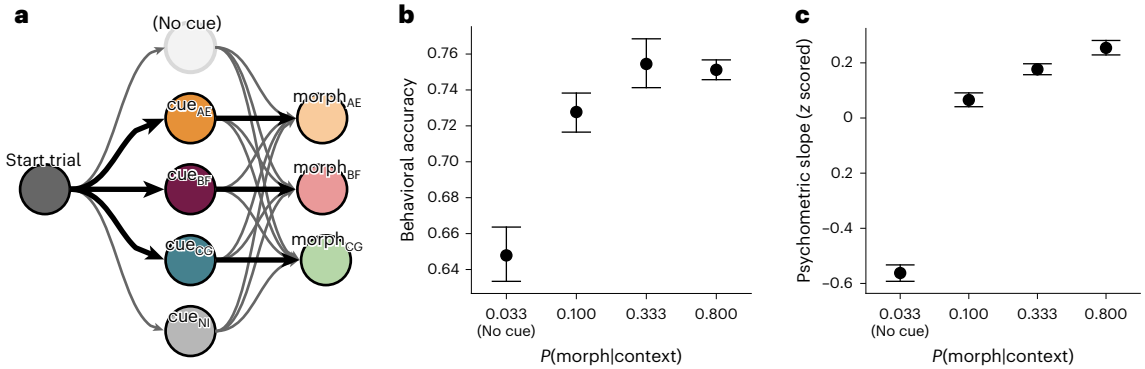


Fig. 7 | Expectation improves perceptual acuity. a, Schematic of the behavioral design in which cues predict full morph continua rather than response classes. One cue predicts each morph continuum at p(morph|cue) = 0.8 and the other two at p(morph|cue) = 0.1. A fourth cue (cue_{NI}, NI stands for 'not informative') predicts all morphs equally (p(morph|cue) = 0.33). In 10% of trials, no cue is given and all three morph continua are equally likely, yielding p(morph|peck) = 0.033. **b**, Mean classification accuracy under the cue conditions in **a**, averaged across subjects and morphs with bootstrapped 95% confidence intervals (n = 37,993 trials). **c**, Mean z-score normalized slopes from psychometric functions fit to responses across morph continua under the different cue conditions, averaged over subjects and morphs with bootstrapped 95% confidence intervals (n = 37,993 trials).

condition + (1|stimulus:unit); $\beta = -0.515$ Hz, $P < 0.001$), consistent with the idea that valid cues reduce representation noise.

We next asked whether prediction accuracy increases and whether stimulus predictions are shifted when stimuli are expected. To this end, for each morph and bird, we trained a logistic regression on principal-component analysis (PCA) projections of population activity fit to passive playbacks and uncued behavioral trials (Fig. 6k). We then applied the PCA projection and predicted morph positions of the held-out cued trials. Each decoder model performed well above chance, validating that stimulus identity can be decoded from sensory populations (Fig. 6l,m).

Both models make dichotomous predictions about the decoded responses. The attention and active sampling model predicts that the decoder accuracy of cue-valid (for example, cue left, morph left) trials will be higher than that of cue-invalid (for example, cue left, morph right) trials (Fig. 6n). The Bayesian integration model predicts instead that decoder predictions will shift in the direction of the cue (Fig. 6q). We analyzed our population decoder results on the basis of these predictions and found that decoding accuracy improves for the validly cued stimulus (Fig. 6p; linear mixed effects; correct ~ cue valid + (1|population) + (cue valid|population); cue valid: $\beta = 0.007$, s.e.m. = 0.002, $z = -3.78$, $P < 0.001$) and that the inflection point fit to model predictions did not shift toward the cue (Fig. 6q; z-test between psychometric model fits; $z = -0.0555$, $P = 0.478$). Additionally, we confirmed that the neurometric curve does not shift over single units by computing the neurometric curve directly on the response similarity matrices, as in our initial comparisons between neurometric and psychometric curves (Supplementary Fig. 2). We measured the change in the inflection point between cue-valid and cue-invalid trials. Across units, we did not find a significant shift in inflection point between high-probability and low-probability cues (linear mixed effects; neurometric shift ~ 1 + (1|unit); $\beta = 0.13$, s.e.m. = 0.82, $z = 1.64$, $P = 0.10$). These results support the sensory modulation model over the Bayesian integration model, demonstrating that sensory acuity is enhanced at the population level and that sensory representations do not shift toward expected stimulus classes.

Receptive field remapping underlies sensory modulation

The preceding results show that expectation modulates sensory responses and supports a model of sensory modulation in which expectation drives changes in sensory acuity. We observed that, on average, spike rates are suppressed as expectation increases (Fig. 5e). One possible mechanism for the spike rate reduction is that expectation modulates the gain of an otherwise static stimulus–response relationship (that is, receptive field). Alternatively, expectations may drive a more dynamic remapping of receptive fields. To differentiate between these hypotheses, we fit a maximum noise entropy (MNE) composite receptive field model[55] to stimulus-evoked single-neuron activity in a subset of trials in which the cue provided a valid prediction of the upcoming target syllable. If the cue has no effect on the receptive field, then model performance (correlation between model-predicted and empirical response) should be similar for the same target syllable presented on held-out cue-valid and cue-invalid trials (Extended Data Fig. 4a,b). Across all cue strengths, however, the MNE receptive field models provided significantly better (more accurate) predictions of responses to target syllables in cue-valid trials than in cue-invalid trials (Extended Data Fig. 4c; linear mixed effects, trial correlation ~ cue validity + (1|unit) + (1|day); $\beta = 0.015$, s.e.m. < 0.001, $z = 41.074$, $P < 0.001$). Thus, contextual cues rapidly reorganize receptive fields to better encode predicted stimuli. This reorganization is produced by cue-dependent changes in both the gain and stimulus feature tuning of linear and nonlinear components of the receptive fields (Methods).

To directly link changes in receptive fields to sensory likelihood modulation, we reproduced the similarity analysis from Fig. 6h,j with the output of the MNE encoder model. We fit MNE receptive fields to target syllables for each cue condition separately and passed all syllables through the model to generate predicted spiking probabilities for the duration of each stimulus. Cue-driven information is then encoded in the variability of the neural response to a given stimulus across cue conditions and hence will produce distinct spiking probability vectors for each cue condition. We computed the similarity of the spiking probability vectors across the different continua as a function of cued direction, taking the difference of the resulting similarity matrices derived for left and right cue conditions, as we did for the empirical responses. Paralleling our empirical results in Fig. 6h–j, we see that within-category similarity is higher in the nonpredicted (cue-invalid) class than in the predicted (cue-valid) class (Supplementary Fig. 3a; linear mixed effects, cosine $\text{similarity}_{\text{empirical – shuffled; cue left – cue right}}$ ~ validity + (1|unit) + (1|day): $\beta = -0.013$, s.e.m. < 0.001, $z = -154.841$, $P < 0.001$; see 'Maximum noise entropy receptive fields') and that within-category similarity decreases as a function of the probability of the cue class (Supplementary Fig. 4; linear mixed effects, cosine $\text{similarity}_{\text{empirical; cued – no cue}}$ ~ validity + (1|unit) + (1|day): $\beta = -0.07$, s.e.m. = 0.001, $z = -12.970$, $P < 0.001$; see 'Maximum noise entropy receptive fields'). These results further support the notion that neuronal responses are dynamically restructured to optimize the differentiation of expected stimuli.

Expectation improves perceptual acuity

The preceding physiological evidence supports a model of expectation-dependent sensory modulation in which sensory representations are flexibly reorganized to improve acuity in expected regions of stimulus space (that is, the attention and active sampling model). These neural changes should also lead to improved behavioral acuity in expected regions of stimulus space. Testing this prediction in the original behavioral task is not possible, however, because the cued portion of the stimulus space is tied to the response class (peck left or peck right), and therefore changes in perceptual acuity cannot be dissociated from the behavioral decision.

To directly test whether expectation modulates sensory acuity, we designed a modified behavioral task in which cues predict individual syllable continua rather than a response class (Fig. 7a). By presenting the same syllable continua under differing levels of expectation, we can assess how perceptual acuity is modulated by expectation.

In the modified task, we paired each of three syllable continua (A–E, B–F, C–G) with a cue syllable. Each cue preceded its paired syllable continuum with 80% probability (for example, $p(\text{morph}_{AE}|\text{cue}_{AE}) = 0.8$) and the other two syllable continua with 10% probability (for example, $p(\text{morph}_{BF}|\text{cue}_{AE}) = 0.1$). These cued trials accounted for 80% of trials. On the remaining 20% of trials, we presented either an uninformative cue (for example, $p(\text{morph}_{AE}|\text{cue}_{NI}) = 0.33$) or no cue ($p(\text{morph}_{AE}|\text{peck}) = 0.033$).

Given our physiological results, we predicted that, in trials where syllable continua are more expected, perceptual acuity would increase and the birds would perform better when discriminating the stimulus. We computed psychometric functions for each continuum and cue condition and took the slope of each as an estimate of perceptual sensitivity across the stimulus space (syllable continuum).

As predicted by our model and consistent with our physiological results, sensitivity improved in the presence of predictive cues (Fig. 7c; permutation test controlling for subject and morph, Methods; $r = 0.22$, $P < 0.001$), coinciding with an improvement in behavioral accuracy (Fig. 7b; linear mixed effects; accuracy ~ cue probability + (1|subject) + (cue probability|subject); cue probability: $\beta = 0.068$, s.e.m. = 0.014, $z(37{,}993) = 5.02$, $P < 0.001$). These behavioral effects corroborate our observations of improved sensory acuity at the neural level.

Discussion

Expectation plays a varied, yet fundamental role in perception. It can facilitate generalization through probabilistic integration, and it can improve acuity through attention. How these diverse outcomes of expectation-driven categorization and acuity are balanced in the course of real-world perception has not been clear. Here, we find that early sensory processing reflects prior information, thereby improving sensory acuity while relegating probabilistic integration of these expectations to downstream circuits involved in decision making and behavior.

To disambiguate models for how expectation might influence sensory representations, we trained songbirds on a categorical perceptual decision-making task and manipulated the predictive contextual information in sequences of vocal elements. Songbirds exploit this information, biasing the categorical perception of their vocalizations. A Bayesian model of perceptual decision making captures both qualitative and quantitative aspects of this behavioral bias (Fig. 3), reflecting the integration of predictive contextual information with uncertainty over natural stimulus dimensions. This model is similar to that which has been proposed for human context-dependent categorical speech perception[3,6]. Neural recordings revealed that many sensory neuronal responses throughout the auditory forebrain are broadly responsive across the natural stimulus space dimensions in which our task was embedded, mirroring the animal's perceptual behavior (Fig. 4). Syllable sequence predictability influenced these sensory representations by suppressing spike rates and modulating syllable encoding and decoding (Figs. 5 and 6 and Extended Data Fig. 4). Contrary to the explicit predictions of the Bayesian model, these neural responses do not directly represent the integration of prior information in these sensory regions. Instead, the context-dependent modulation more closely reflects an increase in perceptual acuity in predicted regions of stimulus space (Fig. 6), facilitating an increase in behavioral performance (Fig. 7). Our results indicate that the coordinated variability of sensory forebrain neuronal populations dynamically shifts in the face of predictions, facilitating optimal encoding along (anticipated) stimulus-relevant dimensions. This restructuring of stimulus–response mapping is suggestive of a top–down predictive model reshaping the stimulus likelihood within sensory regions in anticipation of upcoming stimuli[12]. This conceptualization has the potential to explain how internal models can reduce spiking variance when predictions are valid and casts 'noise' when predictions are invalid as predictive error.

Current speech research aims to uncover neural systems involved in processing predictive information related to lexical and pre-lexical feedback[3]. Many have proposed that a Bayesian framework provides a mechanistic explanation for speech categorization and comprehension more broadly[3,6,8]. However, human studies have methodological limitations, leaving gaps in our understanding of neural representations of speech category information and their modulation under various comprehension-related conditions. Our results support Bayesian integration as a mechanism for categorical perception but leave open the possibility that biases imposed through probabilistic integration, such as categorical perception, are at least in part the product of task-dependent decision making, rather than early sensory and perceptual processes. This is reminiscent of behavioral observations in speech, where the degree to which speech is categorically perceived is task dependent[3]. Our observations suggest a functional segregation of Bayesian integration processes that is adaptive for communication in the sense that it preserves a veridical sensory representation of the stimulus that can be used flexibly in the service of multiple task demands.

Although our findings suggest that sensory populations do not reflect Bayesian integration, it remains possible that some perceptual biases are encoded in sensory systems. In speech, some secondary sensory populations have been found to exhibit categorical responses[56]. Some aspects of categorical speech perception also appear to be less task dependent. For example, native Japanese speakers often have trouble distinguishing between the English phonemes /r/ and /l/ (as in 'rake' versus 'lake') because there is no distinction between /r/ and /l/ in Japanese[4,5], a perceptual bias that can extend for years after exposure to a second language[57]. Reflections of prior expectations have been observed throughout the sensory hierarchy in mice[22]; it may be that the role of expectation in the brain differs along more dimensions than sensory versus decision-making systems. For example, language-related sensory systems may adapt to phonetic categories during critical periods of language acquisition, imposing immutable biases to perception. However, our observations contextualize the broad qualitative differences observed between sensory and decision-making brain regions, where suppression in the sensory brain is linked to dampening noise in task-irrelevant dimensions[12] and increased activity is associated with the integration of expectation and sensory experience[11]. Our results suggest that, even while animals perform Bayesian inference at the behavioral level, sensory populations reflect expectation in a manner wholly unrelated to Bayesian integration. Simultaneous and large-scale recordings across the perceptual and decision-making hierarchy will be crucial for understanding how expectation is used broadly across the brain. Studying these hierarchies remains a challenge in nonmodel systems such as the European starling (and more broadly, songbirds) in which neural substrates for motor control and cognition outside of song production are not well characterized. For instance, the lateral-most region of the NCL is a promising candidate for probabilistic integration in songbirds, paralleling the premotor and

cognitive functions of the primate frontal cortex. However, the role of the NCL has been primarily described in visual processing and multisensory integration, with no evidence yet found in auditory cognition[58], highlighting an important area for future research.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41593-025-01899-1.

References

1. Ganong, W. F. Phonetic categorization in auditory word perception. *J. Exp. Psychol. Hum. Percept. Perform.* **6**, 110–125 (1980).
2. Marslen-Wilson, W. D. & Welsh, A. Processing interactions and lexical access during word recognition in continuous speech. *Cogn. Psychol.* **10**, 29–63 (1978).
3. Norris, D., McQueen, J. M. & Cutler, A. Prediction, Bayesian inference and feedback in speech recognition. *Lang. Cogn. Neurosci.* **31**, 4–18 (2016).
4. Kuhl, P. K. Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Percept. Psychophys.* **50**, 93–107 (1991).
5. Kuhl, P. K. Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* **5**, 831–843 (2004).
6. Feldman, N. H., Griffiths, T. L. & Morgan, J. L. The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychol. Rev.* **116**, 752–782 (2009).
7. Knill, D. C. & Richards, W. *Perception as Bayesian Inference* (Cambridge University Press, 1996).
8. Kuperberg, G. R. & Jaeger, T. F. What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* **31**, 32–59 (2016).
9. Stocker, A. A. and Simoncelli, E. A Bayesian model of conditioned perception. *Adv. Neural Inf. Process. Syst.* **2007**, 1409–1416 (2007).
10. Yon, D., Gilbert, S. J., de Lange, F. P., & Press, C. Action sharpens sensory representations of expected outcomes. *Nat. Commun.* **9**, 4288 (2018).
11. Summerfield, C. & De Lange, F. P. Expectation in perceptual decision making: neural and computational mechanisms. *Nat. Rev. Neurosci.* **15**, 745–756 (2014).
12. Kok, P., Jehee, J. F. M. & De Lange, F. P. Less is more: expectation sharpens representations in the primary visual cortex. *Neuron* **75**, 265–270 (2012).
13. Rohenkohl, G., Cravo, A. M., Wyart, V. & Nobre, A. C. Temporal expectation improves the quality of sensory information. *J. Neurosci.* **32**, 8424–8428 (2012).
14. Correa, Ángel, Lupiáñez, J. & Tudela, P. Í. O. Attentional preparation based on temporal expectancy modulates processing at the perceptual level. *Psychon. Bull. Rev.* **12**, 328–334 (2005).
15. Xin, Y. et al. Sensory-to-category transformation via dynamic reorganization of ensemble structures in mouse auditory cortex. *Neuron* **103**, 909–921 (2019).
16. Johnson, K. & Sjerps, M. J. Speaker normalization in speech perception. In *the Handbook of Speech Perception* (eds Pardo, J. S. et al.) 145–176 (Wiley, 2021).
17. Gerrits, E. & Schouten, M. E. H. Categorical perception depends on the discrimination task. *Percept. Psychophys.* **66**, 363–376 (2004).
18. McMurray, B. The myth of categorical perception. *J. Acoust. Soc. Am.* **152**, 3819–3842 (2022).
19. Ganguli, D. & Simoncelli, E. P. Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. *Neural Comput.* **26**, 2103–2134 (2014).
20. Echeveste, R., Aitchison, L., Hennequin, G. & Lengyel, M. Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference. *Nat. Neurosci.* **23**, 1138–1149 (2020).
21. Sohn, H. & Narain, D. Neural implementations of Bayesian inference. *Curr. Opin. Neurobiol.* **70**, 121–129 (2021).
22. Findling, C. et al. Brain-wide representations of prior information in mouse decision-making. Preprint at *bioRxiv* https://doi.org/10.1101/2023.07.04.547684 (2023).
23. Vilares, I. & Kording, K. Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Ann. N. Y. Acad. Sci.* **1224**, 22–39 (2011).
24. Darlington, T. R., Beck, J. M. & Lisberger, S. G. Neural implementation of Bayesian inference in a sensorimotor behavior. *Nat. Neurosci.* **21**, 1442–1451 (2018).
25. Jazayeri, M. & Movshon, J. A. Optimal representation of sensory information by neural populations. *Nat. Neurosci.* **9**, 690–696 (2006).
26. Funamizu, A., Kuhn, B. & Doya, K. Neural substrate of dynamic Bayesian inference in the cerebral cortex. *Nat. Neurosci.* **19**, 1682–1689 (2016).
27. Akrami, A., Kopec, C. D., Diamond, M. E. & Brody, C. D. Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature* **554**, 368–372 (2018).
28. Hou, H., Zheng, Q., Zhao, Y., Pouget, A. & Gu, Y. Neural correlates of optimal multisensory decision making under time-varying reliabilities with an invariant linear probabilistic population code. *Neuron* **104**, 1010–1021 (2019).
29. Walker, E. Y., Cotton, R. J., Ma, WeiJi & Tolias, A. S. A neural basis of probabilistic computation in visual cortex. *Nat. Neurosci.* **23**, 122–129 (2020).
30. Yin, P., Strait, D. L., Radtke-Schuller, S., Fritz, J. B. & Shamma, S. A. Dynamics and hierarchical encoding of non-compact acoustic categories in auditory and frontal cortex. *Curr. Biol.* **30**, 1649–1663 (2020).
31. Sohn, H., Narain, D., Meirhaeghe, N. & Jazayeri, M. Bayesian computation through cortical latent dynamics. *Neuron* **103**, 934–947 (2019).
32. Berkes, P., Orbán, G., Lengyel, M. & Fiser, J. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* **331**, 83–87 (2011).
33. Hiratani, N. & Latham, P. E. Rapid Bayesian learning in the mammalian olfactory system. *Nat. Commun.* **11**, 3845 (2020).
34. Lange, R. D. & Haefner, R. M. Task-induced neural covariability as a signature of approximate Bayesian learning and inference. *PLoS Comput. Biol.* **18**, e1009557 (2022).
35. International Brain Laboratory et al. A brain-wide map of neural activity during complex behaviour. Preprint at *bioRxiv* https://doi.org/10.1101/2023.07.04.547681 (2023).
36. Mirza, M. B., Adams, R. A., Friston, K. & Parr, T. Introducing a Bayesian model of selective attention based on active inference. *Sci. Rep.* **9**, 13915 (2019).
37. Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P. & Pezzulo, G. Active inference: a process theory. *Neural Comput.* **29**, 1–49 (2017).
38. Todorovic, A., van Ede, F., Maris, E. & de Lange, F. P. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *J. Neurosci.* **31**, 9118–9123 (2011).
39. Todorovic, A., Schoffelen, J.-M., Van Ede, F., Maris, E. & De Lange, F. P. Temporal expectation and attention jointly modulate auditory oscillatory activity in the beta band. *PLoS ONE* **10**, e0120288 (2015).

40. Block, N. The puzzle of perceptual precision. *Open Mind* (2014).
41. Anton-Erxleben, K. & Carrasco, M. Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. *Nat. Rev. Neurosci.* **14**, 188–200 (2013).
42. Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S. & Mooney, R. Neural correlates of categorical perception in learned vocal communication. *Nat. Neurosci.* **12**, 221–228 (2009).
43. Lachlan, R. F. & Nowicki, S. Context-dependent categorical perception in a songbird. *Proc. Natl Acad. Sci. USA* **112**, 1892–1897 (2015).
44. Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–154 (1962).
45. Fechner, G. T. *Elements of Psychophysics, 1860* (Appleton-Century-Crofts, 1948).
46. Singh, N. C. & Theunissen, F. E. Modulation spectra of natural sounds and ethological theories of auditory processing. *J. Acoust. Soc. Am.* **114**, 3394–3411 (2003).
47. Hauber, M. E., Cassey, P., Woolley, S. & Theunissen, F. E. Neurophysiological response selectivity for conspecific songs over synthetic sounds in the auditory forebrain of non-singing female songbirds. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **193**, 765–774 (2007).
48. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).
49. Smith, E. C. & Lewicki, M. S. Efficient auditory coding. *Nature* **439**, 978–982 (2006).
50. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. Preprint at arxiv.org/abs/1312.6114 (2013).
51. Ashwood, Z. C. et al. Mice alternate between discrete strategies during perceptual decision-making. *Nat. Neurosci.* **25**, 201–212 (2022).
52. Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* **113**, 700–765 (2006).
53. Raymond, J. E., Shapiro, K. L. & Arnell, K. M. Temporary suppression of visual processing in an RSVP task: an attentional blink? *J. Exp. Psychol. Hum. Percept. Perform.* **18**, 849–860 (1992).
54. Treisman, A. M. & Gelade, G. A feature-integration theory of attention. *Cogn. Psychol.* **12**, 97–136 (1980).
55. Kozlov, A. S. & Gentner, T. Q. Central auditory neurons have composite receptive fields. *Proc. Natl Acad. Sci. USA* **113**, 1441–1446 (2016).
56. Chang, E. F. et al. Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* **13**, 1428–1432 (2010).
57. Best, C. T. & Tyler, M. D. Nonnative and second-language speech perception. In *Language Experience in Second Language Speech Learning* (Bohn, O.-S. & Munro, M. J.) 13–34 (John Benjamins, 2007).
58. Sainburg, T. & Gentner, T. Q. Toward a computational neuroethology of vocal communication: from bioacoustics to neurophysiology, emerging tools and future directions. *Front. Behav. Neurosci.* **15**, 811737 (2021).

Methods

Ethical note

All procedures were approved by the Institutional Animal Care and Use Committee of the University of California (S05383).

Summary

Experiments consisted of a behavioral component and a chronic physiology component. The experimental protocol for the behavioral component was kept constant by using the same software and hardware in both conditions, with the addition of chronic electrophysiological recording in the physiology component.

Subjects

Behavioral data were collected from 20 wild-caught European starlings of unknown sex. Before beginning experimental training, subjects were housed in a large mixed-sex aviary. Of the 20 starlings used for behavior experiments, ten individuals were used for chronic physiology.

Datasets

Our final behavioral dataset was composed of 4.8 million behavioral trials from 20 birds. Our final chronic neural dataset was composed of 402,797 behavioral trials, with 365,360 responses, a total of 1,594,257 audio playbacks, occurring over 5,345 h (222 d) of recording, across ten birds.

Stimulus generation

Stimuli were syllables of European starling song synthesized from a VAE trained on syllables extracted from a library of European starling song[59].

Training dataset. Syllables were segmented from the full songs of starlings with the dynamic thresholding approach outlined in ref. 60 and available in the vocalization segmentation Python package (https://github.com/timsainb/vocalization-segmentation). Syllables were zero-padded symmetrically at their beginning and end to be 1 s long. Spectrograms of each syllable were computed with 128 frequency bands spaced between 50 and 22,050 Hz and downsampled to 128 time bins (128 Hz), resulting in a 128 × 128 spectrogram of each syllable, used to train the VAE.

Neural network. The neural network architecture we used followed those in our AVGN jhk fmhtcg bn. We used a convolutional VAE architecture with a 16-dimensional latent space. The network was trained on batches of 32 syllables at a time. Artificial neurons used a leaky ReLu nonlinearity. The network was trained with the ADAM optimizer in TensorFlow.

Sampling and synthesis. Each syllable stimulus (used for cues and endpoints) was sampled from the original dataset (Supplementary Fig. 5) and passed through the VAE. The stimuli were chosen to be diverse, well reconstructed in the VAE and roughly equidistant both in spectrogram space (both input and reconstruction) as well as the latent space of the VAE. It is not expected that distances in spectral or neural network latent space would have a 1:1 relationship with an animal's perception of similarity. Morph syllables were sampled from 126 evenly spaced points along the linear interpolation between the latent (16D) representations of a pair of endpoint syllables and passed through the decoder of the VAE, producing the final 128-syllable continuum including the two endpoint syllables (Extended Data Fig. 6). Waveform stimuli were then generated from the spectrogram output of the decoder of the VAE using the Griffin–Lim algorithm. These waveforms were the stimuli used for playback to the birds.

Behavioral training paradigm

Birds were initially trained to differentiate between the two syllable endpoints for a single continuum. After several days of above chance accuracy with one pair of syllable endpoints, the number of endpoints was increased until the birds showed above accuracy classification of the endpoints of all nine continua. After learning the correct response for all endpoints, birds were transferred to the full stimulus set, which included all 128 syllables (linearly sampled and equally spaced in latent space) spanning each of the nine continua (1,152 syllables in total). After the birds were performing reliably above chance on each full syllable continuum for several days, we added cue syllables preceding the target syllables to provide context-dependent information at $P = 0.125$, $P = 0.25$, $P = 0.5$, $P = 0.75$ and $P = 0.875$.

Training parameters

Several behavioral parameters were used in behavioral training, given here for reproducibility. Trials were reinforced on a variable ratio schedule between two and four responses, manually set for each bird to maximize the number of trials each day without the loss of more than 10 g of weight from baseline when in the restricted feeding condition. Punishment was set at a 5-s lights-off period, during which new behavioral trials could not be initiated. A minimum of 1 s between trials, regardless of response, was imposed. Birds could not respond during stimulus playback. Birds were given a 5-s window to respond after stimulus playback. Lighting conditions were set to match seasonal sunrise and sunset times in the experimental location (San Diego, California).

Cue stimulus

Like the morph syllables, the cue syllables are 1 s long, synthesized by reconstruction from the VAE. Behavioral trials were presented with one of six cue conditions: no cue $P(\text{L}|\text{no cue}) = 0.5$ (NC), cue with no predictive information $P(\text{L}|\text{cue}) = 0.5$ (CN), cue left at $p = 0.875\%$ $P(\text{L}|\text{cue}) = 0.875$ (CL1), cue left at $p = 0.75\%$ $P(\text{L}|\text{cue}) = 0.75$ (CL0), cue right at $p = 0.875\%$ $P(\text{L}|\text{cue}) = 0.125$ (CR1), cue right at $p = 0.75\%$ $P(\text{L}|\text{cue}) = 0.25$ (CR0). Sixteen percent of trials were presented in the no-cue condition (NC). Four percent of trials were presented with the uninformative cue condition (CN). The remaining 80% of trials were evenly split between the cue right and cue left conditions. Because the CN condition was sampled with a substantially lower probability than the other conditions, resulting in a low number of total trials in comparison to each other cue condition, it was not included in physiological analyses. In passive physiology playback conditions, due to time constraints in playing back the full stimulus set of 128 interpolation points for each of nine morphs and six cue conditions, we played back only the 87.5% predictive cue conditions in the AE and BF morphs.

Psychometric fit

To assess the shift in categorical perception, in each of the birds ($n = 20$), we fit a psychometric (four-parameter logistic) function both to the overall responses to stimuli in the left and right categories of the morph as well as to each individual morph. The fit psychometric across all morphs is given in Supplementary Fig. 6, that across all birds is given in Supplementary Fig. 7 and that broken out across all birds and morphs is given in Supplementary Fig. 8:

$$\text{logistic}(x) = maximum + \frac{minimum - maximum}{1 + \left(\frac{x}{inflection}\right)^{slope}}.$$

Bayesian integration model

To formalize our hypothesis, when a stimulus varies upon a single dimension x, the perceived value of x as a function of the true value of x and contextual cue information can be described by Bayes' rule:

$$\underbrace{P(x_{\text{true}}|x_{\text{sensed}}, \text{cue})}_{\text{posterior}} \propto \underbrace{P(x_{\text{sensed}}|x_{\text{true}}, \text{cue})}_{\text{likelihood}} \underbrace{P(x_{\text{true}}|\text{cue})}_{\text{prior}} \quad (1)$$

By modulating the prior distribution of the categorical stimuli (x) with a cue, we predict that the perception of x will shift.

Preceding each to-be-categorized target stimulus (x), the cue stimulus provides predictive information about the category of the target stimulus. By treating this cue stimulus as a prior probability over x, we predicted that the determined posterior probability of x given sensory information and the cue stimulus would shift the classification of stimuli near the boundary between the two classes in the direction predicted by the cue stimulus.

Explicitly, we treat the likelihood of a target being sensed $P(x_{\text{sensed}}|x_{\text{true}}, \text{cue})$ as a Gaussian probability distribution around the true target x_{true} (refs. 6,61):

$$P(x_{\text{sensed}}|x_{\text{true}}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_{\text{true}}-x_{\text{sensed}}}{\sigma_{\text{sensed}}}\right)^2} \quad (2)$$

and set the prior probability as a function of the cue

$$P(x_{\text{true}} > \text{ categorical boundary } | \text{cue}) = \text{cue}_{prob}, \quad (3)$$

where cue_{prob} represents the predictive probability of the cue stimulus. In our case then,

$$P(x_{\text{true}}|\text{cue}) = \begin{cases} cue_{prob}/64 & x_{\text{true}} > \text{ categorical boundary} \\ (1-\text{cue}_{prob})/64 & x_{\text{true}} < \text{ categorical boundary} \end{cases}. \quad (4)$$

We predict that birds will make a categorical decision on the basis of the posterior,

$$P(\text{ right peck}|x_{\text{sensed}}, \text{cue}) = P(x_{true} > \text{ categorical boundary } |x_{\text{sensed}}, cue). \quad (5)$$

Under this model, the categorical decision of the bird is modulated by the prior cue information, resulting in a shift in the categorical decision point along the stimulus dimension in the direction predicted by the cue (Fig. 2i).

Bayesian fit. In addition to fitting a psychometric function capturing the shape of the behavioral responses, we fit a Bayesian model reflecting our probabilistic hypothesis described above. This model used five parameters: the shape of the Gaussian of the likelihood (σ_{sensed}), a parameter corresponding to side bias in the apparatus (γ) and parameters representing inattention to the cue stimulus (δ), the target stimulus (β) and overall inattention to the task (α).

$$\text{bias}_{\text{side}}(\gamma) = \text{category}(x_{\text{true}})(1-2(1-\gamma)) + 1 - \gamma$$

$$\text{likelihood} = P(x_{\text{sensed}}|x_{\text{true}}, \text{cue})(1-\beta) + \text{bias}_{\text{side}}(\gamma)\beta$$

$$\text{posterior} \propto P(x_{\text{true}}|x_{\text{sensed}}, \text{cue})(1-\alpha) + \text{bias}_{\text{side}}(\gamma)\alpha$$

$$\text{prior} = P(x_{\text{true}}|\text{cue})(1-\delta) + \text{bias}_{\text{side}}(\gamma)\delta$$

To fit the model, we used the lmfit Python package[62] (see Supplementary Information for additional information).

Response time

For each behavioral trial, we measured the time between the end of a stimulus presentation and the time that a subject's beak was detected in a behavioral response port. In Extended Data Fig. 1, we found that the response time varied based on stimuli and cue conditions.

We used a linear mixed-effects model to statistically test the relationship between response time and cue probability plotted in Extended Data Fig. 1b. We predict response time from the stimulus probability (given the cue), controlling for side bias and overall subject differences in response time (response time ~ stimulus probability + (1 + stimulus class|subject)). We observe that, as expectations increase, response times decrease ($\beta = -0.151$, s.e.m. < 0.001, $z = -188.011$, $P < 0.001$).

To parameterize the decay in response time as a function of the distance in the morph from the decision or class boundary, we fit the decay in response time (controlling for side bias) to an exponential decay function (Supplementary Fig. 9). To account for side biases in decision making (for example, the bird having a position preference when engaging with the behavioral apparatus that positions them further toward the left or right peck port), for each analysis, we z scored response time for each bird's responses to each class (over a 500-trial block, to account for changes in side bias over time). In addition, in the analyses examining response time relative to the morph interpolation point, we discounted the average bias of the cue for each trial. To account for variability in response time related to correctness, we ran all response time analyses only on correct trials.

We excluded two birds from analysis (B1426, B1170) that we observed did not exhibit the same decay in response time as a function of distance from the decision boundary (Extended Data Fig. 1c). For syllable continua where a decay was observed (set at $r^2 > 0.001$ and decay range > 0.1 s.d.), we found a strong relationship between the exponential decay constant and the psychometric slope (Extended Data Fig. 1d; $r^2 = 0.421$, $P = 6 \times 10^{-8}$, $n = 153$).

Chronic electrophysiology

We used 32- or 64-channel NeuroNexus Si probes (A4x2-tet-7mm-150-200-121, Buzsaki32, Buzsaki64, A1x2-Edge-5mm-20-177) implanted either unilaterally or bilaterally. Probes were coated with PEDOT using an Intan RHD Electroplating Board no more than 1 week before implant. Probes were mounted on 3D-printed drives (described in Microdrives and head caps), which were stereotactically implanted using the procedure outlined in Electrode implant procedure. Extracellular voltages were amplified and digitized at 30 kHz using an Intan RHD recording headstage, output through an SPI cable through an electrically assisted commutator to an Open Ephys recording system.

Behavioral neural acquisition interfacing with PiOperant

Behavioral and physiology data were synced using a custom-designed Raspberry Pi-based system (PiOperant) for automating our behavioral paradigm and interfacing with the Open Ephys neural acquisition device (Supplementary Fig. 10). PiOperant interfaces with our behavioral panel using the Python software pyoperant (https://github.com/gentnerlab/pyoperant). Behavioral states and audio signals were input and synced with Open Ephys over two HDMI inputs (digital and analog) and a ZMQ interface containing additional information about behavioral trials.

Microdrives and head caps

Microdrives and head caps (Supplementary Fig. 11) were custom designed over the course of this experiment and were printed using a Formlabs Form 3 3D printer using Formlabs standard gray resin printed at a resolution of 25–50 µm. Microdrives were composed of a drive, a shuttle and a MiniTaps 6/16" 00-90 gold screw, hand-tapped and fastened to the drive with a brass nut. The screw was used to raise and lower the shuttle manually, at a depth of 282 µm per full rotation. Head caps were designed to be removable and enable moving probes further down as well as easy to explant, allowing reuse of probes.

Electrode implant procedure

Subjects were given analgesia by means of a dose of carprofen (5 mg per kg) (Rimadyl) administered intramuscularly. Animals were then anesthetized with a gaseous mixture of isoflurane and oxygen (1–2.5%, liters per minute). The scalp and feathers around the scalp were then removed, and part of the skull over the y sinus (the stereotactic reference sinus between the cerebellum and the two hemispheres of the brain) was visible. A craniotomy was opened above the recording site. A second craniotomy for the ground was then performed several millimeters away from the primary craniotomy. A platinum–iridium

ground wire was then inserted in the craniotomy above the dura and glued to the skull. The baseplate for the head cap was then cemented (Metabond) to the skull. The durotomy was then performed in the original craniotomy, and the electrode, attached to the microdrive, was stereotactically lowered at a rate of no more than 100 µm per minute. Once the final site was reached, the microdrive was then cemented to the skull, and a silicone base was applied above the craniotomy to prevent infection. The head cap was then screwed into the baseplate, protecting the recording site and probe. The headstage was then attached to the outside of the head cap.

In some individuals, multiple implants were performed in serial when one probe failed by explanting and removing the first probe and microdrive, creating a new craniotomy in the opposite hemisphere and durotomy, and implanting a new probe and microdrive. In one individual, two probes and drives were implanted simultaneously, one in each hemisphere.

Recordings and behavior blocks

Recordings were performed 24 h per day to track individual neurons over days. Recordings consisted of (1) behavior blocks, in which subjects freely interacted with the behavioral apparatus, (2) a free feeding period, in which the behavioral apparatus presented food to the bird without requiring the bird to perform trials, (3) a passive playback block, in which lights were turned off and the birds were passively presented with stimuli and (4) a sleep block, in which the lights were left off and no stimuli were played back.

We recorded from ten subjects over a total of 222 d (5,317 h) of recordings. Chronically implanted subjects performed over 400,000 behavioral trials during recording. In addition, during the evening after the birds had completed their behavioral trials for the day, we turned the lights out in the behavior boxes and passively played back the same morph stimuli to the birds, totaling 1.2 million passive playbacks while recording.

Chronic behavior blocks. Chronic behavior blocks were matched to behavior blocks without physiology. The behavioral apparatus was left on throughout the day, allowing subjects to initiate trials with a peck in the central peck port. Trials were intermittently reinforced with a food reward and punished with the lights briefly turning off in incorrect trials. Using this paradigm, subjects performed several thousand trials per day.

Chronic passive playback blocks. At a set time at the end of each day, we turned the lights off in the bird's operant conditioning block and passively played back the morph stimuli to the bird. The bird's activity and sleep state during this time were not monitored. The silence interval between stimuli was randomly sampled between 1.1 and 1.5 s.

Spike sorting and merging over long-term chronic recordings

Spike sorting was performed over each 12-h block of recording using KiloSort 2–2.5 (ref. 63) and SpikeInterface[64]. LFP was bandpass filtered between 300 and 6,000 Hz and further normalized using common median referencing. To retain units across days or sorts, we additionally used an overlapping procedure to merge each neighboring pair of recordings together. To do so, we took the last 30 min of the previous recording and the first 30 min of the following recording and separately sorted that hour-long recording, which overlapped with the two larger recordings. We then computed the overlap between units in the overlapping recording and each of the two full recordings. Units were then considered to be the same unit if their 'agreement' score (SpikeInterface; the spike coincidence of the two units) was above a set threshold (set at 0.5). Units from each of the larger recordings that were merged with the same unit in the overlapped recording were then merged, allowing the same unit to be tracked over multiple days (Extended Data Fig. 5).

Stimulus alignment

Stimulus playback was aligned to neural data using a 1-kHz sine wave sent from the MagPi behavioral control device to the Open Ephys acquisition board collected simultaneously with neural data, alongside a binary switch indicating the onset and offset of playback. An additional message giving information about the specific trial was sent over the local network via ZMQ.

Localizing units

Unit locations were defined as the location of the peak recording channel on which the unit was present. The recording channel was determined from its position within the shank and the shank's position relative to the stereotactic implant. Stereotactic implant locations were recorded relative to the y sinus between the cerebellum and two hemispheres of the brain, and the depth relative to the surface of the brain. Implant locations relative to nuclei were then determined relative to voxel mapping of the European starling brain atlas[65], as shown in Extended Data Fig. 2a,b. We recorded from units in the primary auditory forebrain region field L, two secondary auditory forebrain regions (the caudal mesopallium and the caudal medial nidopallium) and the NCL. Note that, while NCL is a higher-order forebrain region implicated in visual and multimodal working memory[66–68], our recordings were performed only on the most medial regions of the NCL (Extended Data Fig. 2b), which have been less well characterized. Sample unit spike trains for each nucleus are shown in Extended Data Fig. 2c.

Neural feature representation and response similarity

We represented spike trains as vectors using the methods outlined in Fig. 4a–f. In particular, a PSTH of spike trains was computed with 10-ms time bins, which was then smoothed with a Gaussian kernel with a σ of 25 ms. Morphs were sampled at a resolution of 128 points. For physiological analyses, we reduced the sampling resolution, binning the 128 interpolation points into 16 points along the morph; thus, the neural response vectors and similarity matrices are 100 time bins by 16 interpolation bins and 16 interpolation bins by 16 interpolation bins, respectively.

We computed neural response similarity as the cosine similarity of the Gaussian convolved spike vectors, which has been effectively used to find similarity in spike trains in the past[69]. A number of different similarity metrics could have been used in its place, for example, correlation coefficients[70,71] and Euclidean distance between Gaussian convolved spike trains. We compared the cosine similarity to several other similarity metrics used in neural analyses including the correlation coefficient, the Euclidean distance and the Manhattan distance and found broadly similar results (Supplementary Fig. 12).

Estimating a neurometric function from the similarity matrix

The neurometric function is computed on the basis of the similarity matrix and is detailed in Extended Data Fig. 7. For each interpolation point, we took the average of the within- and between-category similarity (S_{C1} and S_{C2}) and took the ratio ($\frac{S_{C1}}{S_{C1}+S_{C2}}$) as the categorical similarity ratio. We then fit the same four-parameter logistic function as used in the psychometric function to the categorical similarity ratio as a function of the interpolation point.

Between-subject neurometric versus psychometric variability. To determine whether between-subject variability in the slope of the psychometric was reflected in the neurometric, we used a linear mixed-effects model in the Python statsmodels package: neurometric slope ~ morph + psychometric slope + (1|unit). Controlling for the morph and random effects of the individual unit, the relationship between the individual variance in the psychometric slope and the neurometric slope is not significant and slightly negative ($t(41{,}171) = -5.75$, $P > 0.999$, $\beta = -0.04$) and only explained an additional 0.03% of the variance ($r^2 = 0.778$).

Assessing task relevance for units

Task relevance was measured as the categoricality of the neural response. Unit categoricality was computed using the similarity matrix (as seen in Fig. 4). The similarity matrix used to compute a unit's categoricality was the mean cosine similarity matrix across interpolation responses, where the cosine similarity matrix was computed over average response vectors for each interpolation point.

Similarity matrices were divided into four quadrants, corresponding to the within-category similarities for each category and the between-category similarities. Categoricality was computed as the mean similarity in the within-category quadrants of the similarity matrix (that is, the top left and the bottom right) minus the between-category similarities.

Subsetting task-relevant units

We operationalized task-relevant, categorical units on the basis of their response characteristics to the morph stimuli. Task-relevant units were determined by a threshold set in the categoricality metric. This threshold was set at a categoricality metric value above 0.1. These thresholds were set based on visual assessment of unit responses (Supplementary Fig. 14) and similarity matrices. For reference, figures showing units sorted by the categoricality metric are provided for subject (Extended Data Fig. 8), morph (Extended Data Fig. 9) and region (Supplementary Fig. 15).

Comparing spike rate across units, cues and morphs

The physiological analyses performed in the main text were performed over unit spike rates in response to the morph stimuli, where spike rate was *z* scored over the unit's spike rates across all stimuli. Supplementary Fig. 13 visualizes the main effect of cue and interactions between cue probability and stimulus class. In addition, we shuffled the cue labels to ensure that our results were not due to inherent sampling biases present in the data (for example, a left cue is more predictive of a left morph point; thus, more cue left to left morph point samples exist in the dataset).

Cue suppresses spike rate model comparison. An analysis of variance compared a baseline mixed-effects model with only a random intercept for the unit or stimulus to a model including both a random intercept for the unit stimulus and the fixed effect of cue:

$$\texttt{model0 : spike rate} \sim \texttt{(1|unit\_stimulus)}$$

$$\texttt{model1 : spike rate} \sim \texttt{cue + (1|unit\_stimulus)},$$

where unit_stimulus is a variable representing a combination of the unit and the stimulus (for example, neuron 8, stimulus BF, interpolation point 7).

Spike rate suppression increases with cue strength. We next tested the relationship between the cue's predictive strength and the spike rate, again using an ANOVA between linear mixed-effects models.

$$\texttt{model1: spike rate} \sim \texttt{cue + (1|unit\_stimulus)}$$

$$\texttt{model2: spike rate} \sim \texttt{cue + cue\_p\_right:side + (1|unit\_stimulus)},$$

where cue_p_right is the cue probability and side is the stimulus class.

Differences in spike rate as a function of time

In the main text, we compared differences in spike rate as a function of their cue (that is, within versus between cue).

To ensure that the effects of spike rate modulation occur between cue conditions and not only between cue conditions and the uncued condition (where the main effect of cue on spike rate is greatest), we did not include the uncued condition in the spike rate differences between cue conditions in Fig. 5b. We only included units and stimuli for which we had active and passive behavioral trials (n units = 4,722). We then, for each unit and stimulus, took the average absolute difference in response vectors between trials for trials with the same cue and trials with different cues. The difference between the average absolute difference between cues, minus within cues, will equal zero when there is no difference between cue conditions. To ensure that no factors exogenous to between-cue differences are causing this effect, in Supplementary Fig. 16, we show the same analysis where cue labels have been shuffled within stimuli.

Within-cue response similarity

For each unit and cue, we computed the cosine similarity matrix across each morph. Cosine similarity matrices were computed by taking the average cosine similarity across trials for each interpolation point (16) in the morph. Analyses were only performed over active behavioral trials in which the subject provided a response. We then contrasted the cosine similarity matrices across different cue conditions. Supplementary Fig. 6h shows the average cosine similarity across left cues subtracted from the average cosine similarity across right cues. Blue in the top left of the plot (the orange bounding box) depicts less similarity in the predicted left class in left-cued trials. The reverse is true for the red at the bottom right. We statistically test this with a linear mixed-effects model ($\text{cosine similarity}_{\text{cue left - cue right}} \sim \text{side} + (1|\text{unit}) + (\text{side}|\text{unit})$; $\beta = 0.009$, s.e.m. < 0.001, $P < 0.001$). We measured this relationship, showing that predicted morph classes are less similar within class in Supplementary Fig. 6j. Each point and confidence interval consists of the within-class similarity relative to the same unit's response to uncued stimuli across trials. The negative relationship shows that higher-probability cued trials exhibit less similar responses. We statistically test this with a linear mixed-effects model ($\text{cosine similarity}_{\text{rel. cue none}} \sim \text{probability class} + (1|\text{unit}) + (\text{probability class}|\text{unit})$; probability class: $\beta = -0.01$, s.e.m. < 0.001, $P < 0.001$). In a similar manner as in Supplementary Figs. 13 and 16, we repeated this analysis over the same data in which cue labels had been shuffled within unit or interpolation. In the shuffled condition, we observed that the effect was removed.

Acuity trade-off model

Central to the acuity trade-off model is the idea that focusing on task-relevant stimulus dimensions improves the precision of their representation but at the expense of less accurate representations of irrelevant dimensions. Thus, a key feature of this model is that noise in neural measurement and resulting representations decreases for stimuli in expected regions of stimulus space. This decrease in noise yields neural responses that are more easily separable. For example, consider two simple, 1D Gaussian distributions over our signal dimension (that is, the likelihood for two similar stimuli in our Bayesian model) separated by a fixed distance. Reducing measurement error (σ) decreases the overlap between the two distributions and thus decreases the similarity (on average) between points sampled from those distributions (Extended Data Fig. 3a). Somewhat paradoxically, when the means of the distributions are sufficiently close in 1D space, reducing σ results in an increase in similarity between points sampled from those distributions (Extended Data Fig. 3b). For example, for a 1D Gaussian distribution when the mean is equal and only the variance is changed, the average difference between two points sampled is $2\sigma/\sqrt{(\pi)}$. The average difference between two points sampled from a Gaussian distribution with a standard deviation of 1 ($N(0,1)$) is 1.13 and with a standard deviation of 2 ($N(0,2)$) is 2.26. This effect is illustrated in the context of a single dimension from our current stimulus set (that is, along a single syllable continuum) in Extended Data Fig. 3c. To generate Extended Data Fig. 3c, we sampled from Gaussian distributions with standard deviation (σ) along the syllable continuum and took the average difference

between samples at each point along the syllable continuum as similarity. Extended Data Fig. 3c then shows the difference between two similarity matrices when sigma is varied to mimic a cue that predicts the left or right stimulus class along the continuum. In this scenario, the predicted increase in similarity along the diagonal contrasts with our empirical observations of a decrease in similarity, particularly along the diagonal within the cued stimulus class (Extended Data Fig. 3d).

To account for this discrepancy, we can expand the 1D model to account for stimulus dimensions that are irrelevant to the task immediately at hand (that is, on a given trial). This includes, for example, acoustic features that might be relevant for other syllable continua, response classes and estimates of acoustic characteristics exogenous to the task-relevant stimuli such as background acoustics or other characteristics not relevant to classification but that are also represented by neural responses.

Under the assumption that the capacity to measure all possible characteristics of the sensory environment is limited by resources, improved accuracy in representing a task-relevant stimulus dimension (here, the syllable continuum) comes at a cost of decreased representation accuracy on task-irrelevant stimulus dimensions. We refer to this model as the 'acuity trade-off model', referring to the trade-off that exists in measuring and representing all features of our environment, requiring us to selectively attend to and represent only task-relevant features with high fidelity. Evidence for this model comes from both behavior, where limited attentional resources are available to keep track of different feature dimensions simultaneously[53,54], as well as theories and empirical observations of neural computation, where neural coding has been observed to shift to decrease noise in stimulus-relevant dimensions at the expense of increasing noise in stimulus-relevant dimensions[72,73].

By accounting for the task-irrelevant dimensions (Extended Data Fig. 3e), we observe, as task-irrelevant dimensions are added, that the diagonal line (corresponding to the change in the similarity of nearby signals on the morph dimension) increases. This reflects expectation, reducing the similarity of neural response along the diagonal in this model. The plots in Extended Data Fig. 3e are generated in the same manner as in Extended Data Fig. 3c, except where additional task-irrelevant dimensions are added, and, in contrast to the sharpening of the task-relevant dimension, the noise in measurement of the task-relevant dimension increases. A sample of points sampled from the one task-irrelevant dimension version of this model is provided in Extended Data Fig. 3f, where noise in sampling in the task-relevant dimension is decreased for the cued signal, whereas noise in sampling is increased for the behaviorally irrelevant dimension when the signal is cued. This decrease in similarity when additional behaviorally irrelevant noise is present (here through the addition of behaviorally irrelevant dimensions) occurs because the most similar signals (near the diagonal) are most susceptible to becoming more distant in representation when task-irrelevant noise is increased.

A self-contained Jupyter notebook (Google Colab link: https://colab.research.google.com/drive/1ZqOWgfPhaBOuoSYOr7UvR_wwDHYIdTsU) is available to reproduce this figure and aid in understanding the model through interaction.

Maximum noise entropy receptive fields

We used the MNE model to calculate receptive fields for all task-relevant units[55,74] (Extended Data Fig. 4). MNE models were computed separately for each unit and predictive cue condition. Model fitting used a jackknife procedure, averaging estimates from four subsets of the training data to yield the final parameters. Model-predicted spiking probabilities were correlated with empirical spike trains on held-out trials to evaluate receptive field model performance (Extended Data Fig. 4). The correlation values were then averaged across trials within a given cue condition.

To assess whether expectation modulates receptive fields, we trained MNEs for each neuron using the target syllable-evoked responses from trials with a valid cue, that is, trials where the cue accurately predicted the correct response to the subsequent target stimulus. We held out a subset of cue-valid trials equal to the number of cue-invalid trials (trials where the cue predicted the incorrect response) collected during the same recording session. We then used the model to predict responses on these held-out trials and used the correlation coefficient between predicted and actual responses (averaged across trials for a given predictive cue) as a measure of model performance (Extended Data Fig. 4c). To ensure consistency, we repeated these prediction tests using four different random samples of held-out cue-valid trials and averaged performance across these four samples. If the expectation does not alter the receptive field, then the models should fit responses in both the valid and invalid conditions equally well. We used a linear mixed-effects model with a fixed effect for cue validity and random effects for each unit's identity and recording day to predict trial correlation values. We used paired *t*-tests for post hoc comparisons.

To assess cue-dependent gain and tuning changes in the receptive fields, we first fit an MNE to the responses of a neuron in the NC condition and then refit responses for that same neuron to held-out data from the 'cued' and 'NC' conditions, using the no-cue refit as the null hypothesis for change across conditions. As proxies for feature tuning and gain of the receptive field, we examined changes in the orientation and magnitude of the MNE feature vectors (the h and J terms), respectively.

In the context of the MNE model, the orientation of the feature vector in high-dimensional stimulus space defines the features to which the neuron is tuned. If those features change, then the vector orientation changes. To measure this change, we computed the cosine distance between feature vectors before and after refitting. Because the cosine distance is invariant to scaling, a change in gain alone will not alter the orientation of the feature vector. By contrast, the orientation of the full feature vectors does change significantly between the 'cued' and the 'NC' conditions (linear mixed effects, cos-diff ~ cue present + (1|unit) + (1|day); $\beta = 0.048$, s.e.m. = 0.003, $z = 17.838$, $P < 0.001$). We also examined the orientation of the linear and nonlinear components (h and J) separately. In both cases, the feature vector orientation changes significantly more for the 'cued' condition than for the 'NC' condition (linear mixed effects, cos-diff-h ~ cue present + (1|unit) + (1|day): $\beta = 0.068$, s.e.m. = 0.007, $z = 10.369$, $P < 0.001$; cos-diff-J ~ cue present + (1|unit) + (1|day): $\beta = 0.048$, s.e.m. = 0.003, $z = 18.042$, $P < 0.001$). To examine the nonlinear changes in more detail, we tried to look at changes in matched sets of nonlinear features across conditions. This is difficult to do with strict assurance, but, as a proxy to feature similarity, we restricted the analysis to only the closest pairs of eigenvectors (that is, those with the minimum pairwise cosine distance from initial and refit MNE J matrices). Even here, the minimum change in these nearest eigenvectors is larger for the 'cued' condition than for the 'NC' condition (linear mixed effects, eig-min-cos-dist ~ cue present + (1|unit) + (1|day): $\beta = 0.019$, s.e.m. = 0.002, $z = 8.974$, $P < 0.001$). Collectively, these results consistently support our claim that expectation modulates the tuning of receptive fields to explicit stimulus features.

The foregoing feature vector orientation analyses rule out the possibility that cue-dependent response modulation is explained entirely by changes in the receptive field gain, but changes in gain may still contribute to the modulation. To assay the change in receptive gain directly, we first compared the change in the magnitude of the linear feature vector (h), taken as the change in the L_2 norm of the linear feature vector, after refitting to the cue and NC conditions. We found that the magnitude of the linear feature vector was significantly greater for the 'cued' condition than for the 'NC' condition (linear mixed effects, mag-diff ~ cue present + (1|unit) + (1|day); $\beta = 0.004$, s.e.m. = 0.001, $z = 4.752$, $P < 0.001$). We used a similar logic to examine the change in magnitude of nonlinear features across cue conditions by summing the absolute values of all eigenvalues for the J matrix. As

with the linear feature, this proxy for the gain of the nonlinear features was significantly greater in the 'cue' condition than in the 'NC' condition (linear mixed effects, eig-diff ~ cue present + (1|unit) + (1|day); $\beta = 0.186$, s.e.m. = 0.031, $z = 6.052$, $P < 0.001$). We note that these cue-dependent shifts in the magnitudes of the linear and quadratic MNE terms correspond directly to the width and the height of the tuning curves modeled in high-dimensional stimulus space, consistent with our conclusion that expectations enhance the selectivity of receptive fields. These more selective models produce fewer spikes with higher confidence, aligning with the expectation-induced firing rate suppression we observe in our empirical data.

To better understand how changes in MNE receptive fields are related to expectation-driven changes in the sharpness of the likelihood function, we attempted to replicate the empirical shifts in similarity (Fig. 6h) using MNE models. To do this, we fit MNE models with target syllable data for each cue condition and then passed all target syllables through the model to generate predicted spiking probabilities for each stimulus and cue condition. We pooled the produced spiking probability vectors across models and separated them by unit, interpolation and cued response ('left' versus 'right'). We then binned the spiking probability vectors for each unit, interpolation and cued response into 16 bins spanning the target syllable continuum. We computed the cosine similarity between pairs of probability vectors within and between bins to generate a similarity matrix for each unit, interpolation and cued response. We subtracted the 'left' and 'right' similarity matrices for each unit and interpolation and averaged the resulting difference matrices across all units and interpolations. We performed the analysis for both the linear (only h term included) and full (h and J terms included) MNE (Supplementary Figs. 3a,b and 4). As a control, we completed the same analyses using a version of each MNE feature vector that was shuffled before producing spiking probability vectors for each stimulus and then subtracted the resulting similarity matrix from that for the unshuffled MNEs. We used a linear mixed-effects model with a fixed effect for cue validity and random effects for each unit's identity and recording day to predict similarity values.

Stimulus decoder

For each population and syllable continuum, we trained a logistic regression with L_2 regularization and balanced class weighting using scikit-learn. Models were trained to predict the bin in the syllable continuum where the stimulus occurred, where the stimulus continua were split into 16 bins (from the total of 128 interpolation points along the syllable continuum). Because individual neurons were tracked longitudinally rather than in sessions, neural population representations are sparse; not all neurons are present during any given trial.

Activity for each neuron was represented as a histogram aligned to the 1-s playback of the target syllable (we used 20 time bins, each corresponding to 50 ms). Time-varying spike rates were then z scored for each neuron and clipped between −4 and 4 s.d. to weight units equally in the subsequent PCA projection. Thus, at this stage in processing, population activity is represented as a matrix of shape (number of trials, number of units, 20). Neural populations were projected into PCA space (256 dimensions), fit to the training data (passive and uncued trials). The decoder was trained on all trials that were either uncued (that is, with no cue or with the uninformative cue) or passive playback trials where the bird was not performing the task. Analyses were then performed on the held-out cued data, specifically looking at prediction accuracy between cue-valid and cue-valid trials. Any population without at least one cue-invalid sample for each syllable continuum bin was discarded.

Neurometric across cue conditions

For each unit and syllable continuum, we fit a neurometric model to spike response vectors (as described earlier). To compare changes in the neurometric as a function of cue probability, we subsetted units or morphs where there existed enough trials for each morph and cue (at least one per stimulus) to compute a similarity matrix. We additionally excluded any model fits where the fit inflection point did not converge. In total, this yielded 1,762 units in which we had well-fit neurometrics.

Morph prediction behavior

The primary task uses cue syllables to predict the likely correct response class associated with the target stimulus on the current trial (that is, the left and right stimulus classes, corresponding to the left or right half of each morph), biasing perceptual decision making toward a stimulus class. The morph prediction task instead uses cue syllables to predict the morph, independent of the stimulus class. We used this behavior to assess whether the accuracy increases and the psychometric slope sharpens when a morph is expected.

We used three of the morphs from the original dataset (AE, BF, CG). The trial structure is as follows. At the beginning of each trial, initiated by a peck in the center peck port, one of three things would happen. (1) In 10% of trials, a morph would be played without any cue. (2) In another 10% of trials, an uninformative cue would play, which equally predicts all three morphs. (3) In the remaining 80% of trials, an informative cue would play. The informative cue predicted a single morph (that is, morph_{AE} follows cue_{AE}) 80% of the time, and the remaining 20% of trials following cue_{AE} are divided between morph_{BF} and morph_{CG}. This yields trials with four sets of expectations. In the uncued trial, the probability that any given morph will play after a center peck is only 1/30 (10% of trials have no cue, and there is a one-in-three chance of hearing each morph). After the uninformative cue plays, the probability that any given morph will play is 1/3 (equal chances for each morph). After an informative cue plays, the probability that the predicted morph will play is 0.8, and each of the nonpredicted morphs is 0.1. To focus on challenging trials that define the psychometric slope, we sampled only the 32 points surrounding the midpoint of the original 128 points in the morph. In addition, we applied white noise over the stimulus (at 25% of the maximum amplitude) to keep the accuracy around 70% and avoid overtraining.

We retrained two birds on the modified task (B1590 and B1591) over approximately 7 weeks. B1590 engaged in 17,069 trials, while B1591 engaged in 21,879 trials. We compared the accuracy across cue conditions as well as the psychometric slope. To compare accuracy, we fit a linear mixed-effects model, predicting correctness for each trial by the cue probability, controlling for the subject as a random intercept and slope. To compare psychometric fits, we used a bootstrapping approach. We estimated the psychometric slope by sampling 1,000 trials (with replacement) from each morph for each subject and fitting the psychometric function to those trials. We repeated this 1,000 times and took the mean psychometric slope parameter. This method accounts for differing numbers of trials across cue conditions and stochastic error in model fitting. To then statistically test the relationship between cue probability and psychometric slope, we z scored the psychometric slope within subject and morph. We then computed the Pearson correlation between cue probability and z-scored slope. Finally, we compared our observed correlation between cue probability and slope to a distribution generated by shuffling the cue probabilities (shuffled within subject and morph) 1,000 times.

Statistics and reproducibility

No statistical method was used to predetermine sample size. No data were excluded from the analyses. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data are available at https://zenodo.org/records/7363595 (ref. 75).

Code availability

Code and code documentation are available at https://github.com/timsainb/cdcp_paper.

References

59. Arneodo, Z., Sainburg, T., Jeanne, J. & Gentner, T. An acoustically isolated European starling song library. *Zenodo* https://doi.org/10.5281/zenodo.3237217 (2019).
60. Sainburg, T., Thielk, M. & Gentner, T. Q. Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLoS Comput. Biol.* **16**, e1008228 (2020).
61. Körding, K. P. & Wolpert, D. M. Bayesian integration in sensorimotor learning. *Nature* **427**, 244–247 (2004).
62. Newville, M. et al. lmfit/lmfit-py: 1.0.3. *Zenodo* https://doi.org/10.5281/zenodo.598352 (2021).
63. Pachitariu, M., Steinmetz, N. A., Kadir, S. N., Carandini, M. & Harris, K. D. Fast and accurate spike sorting of high-channel count probes with KiloSort. *Adv. Neural Inf. Process. Syst.* **29**, 4448–4456 (2016).
64. Buccino, A. P. et al. SpikeInterface, a unified framework for spike sorting. *eLife* **9**, e61834 (2020).
65. De Groof, G. et al. A three-dimensional digital atlas of the starling brain. *Brain Struct. Funct.* **221**, 1899–1909 (2016).
66. Güntürkün, O. The avian 'prefrontal cortex' and cognition. *Curr. Opin. Neurobiol.* **15**, 686–693 (2005).
67. Kröner, S. & Güntürkün, O. Afferent and efferent connections of the caudolateral neostriatum in the pigeon (*Columba livia*): a retro- and anterograde pathway tracing study. *J. Comp. Neurol.* **407**, 228–260 (1999).
68. Nieder, A. Inside the corvid brain—probing the physiology of cognition in crows. *Curr. Opin. Behav. Sci.* **16**, 8–14 (2017).
69. Fellous, J.-M., Tiesinga, P. H. E., Thomas, P. J. & Sejnowski, T. J. Discovering spike patterns in neuronal responses. *J. Neurosci.* **24**, 2989–3001 (2004).
70. Schreiber, S., Fellous, J.-M., Whitmer, D., Tiesinga, P. & Sejnowski, T. J. A new correlation-based measure of spike timing reliability. *Neurocomputing* **52**, 925–931 (2003).
71. Theilman, B., Perks, K. & Gentner, T. Q. Spike train coactivity encodes learned natural stimulus invariances in songbird auditory cortex. *J. Neurosci.* **41**, 73–88 (2021).
72. Jeanne, J. M., Sharpee, T. O. & Gentner, T. Q. Associative learning enhances population coding by inverting interneuronal correlation patterns. *Neuron* **78**, 352–363 (2013).
73. Panzeri, S., Moroni, M., Safaai, H. & Harvey, C. D. The structures and functions of correlations in neural population codes. *Nat. Rev. Neurosci.* **23**, 551–567 (2022).
74. Fitzgerald, J. D., Rowekamp, R. J., Sincich, L. C. & Sharpee, T. O. Second order dimensionality reduction using minimum and maximum mutual information models. *PLoS Comput. Biol.* **7**, e1002249 (2011).
75. Sainburg, T. European starling categorical perception chronic ephys and behavior dataset. *Zenodo* https://doi.org/10.5281/zenodo.7363594 (2022).

Acknowledgements

T.S. acknowledges support from a CARTA Fellowship to T.S. and NIH 5T32MH020002-20 to T.S. T.Q.G. acknowledges support from NIH 5R01DC018055-02. PTM acknowledges support from the Kavli Institute for Brain and Mind (IRG no. 2021-1759), 'La Caixa' Foundation and an IIE Fulbright Fellowship. E.M.A. acknowledges support from a Pew Latin American Fellowship in the Biomedical Sciences and the Kavli Institute for the Brain and Mind (IRG no. 2021-1759). We thank B. Datta, J. Pearl, A. Pouget and C. Findling for valuable feedback on the manuscript.

Author contributions

T.S., T.S.M. and T.Q.G. designed experiments. T.S. and T.S.M. carried out experiments. E.M.A., S.R., M. Turvey, B.H.T., P.T.M. and M. Thielk aided in carrying out experiments and provided advice on study design. T.S.M. performed all analyses related to MNE receptive fields. T.S. performed all other analyses. T.S., T.S.M. and T.Q.G. wrote the paper; all other authors provided feedback.

Competing interests

The authors declare no competing interests.

Additional information

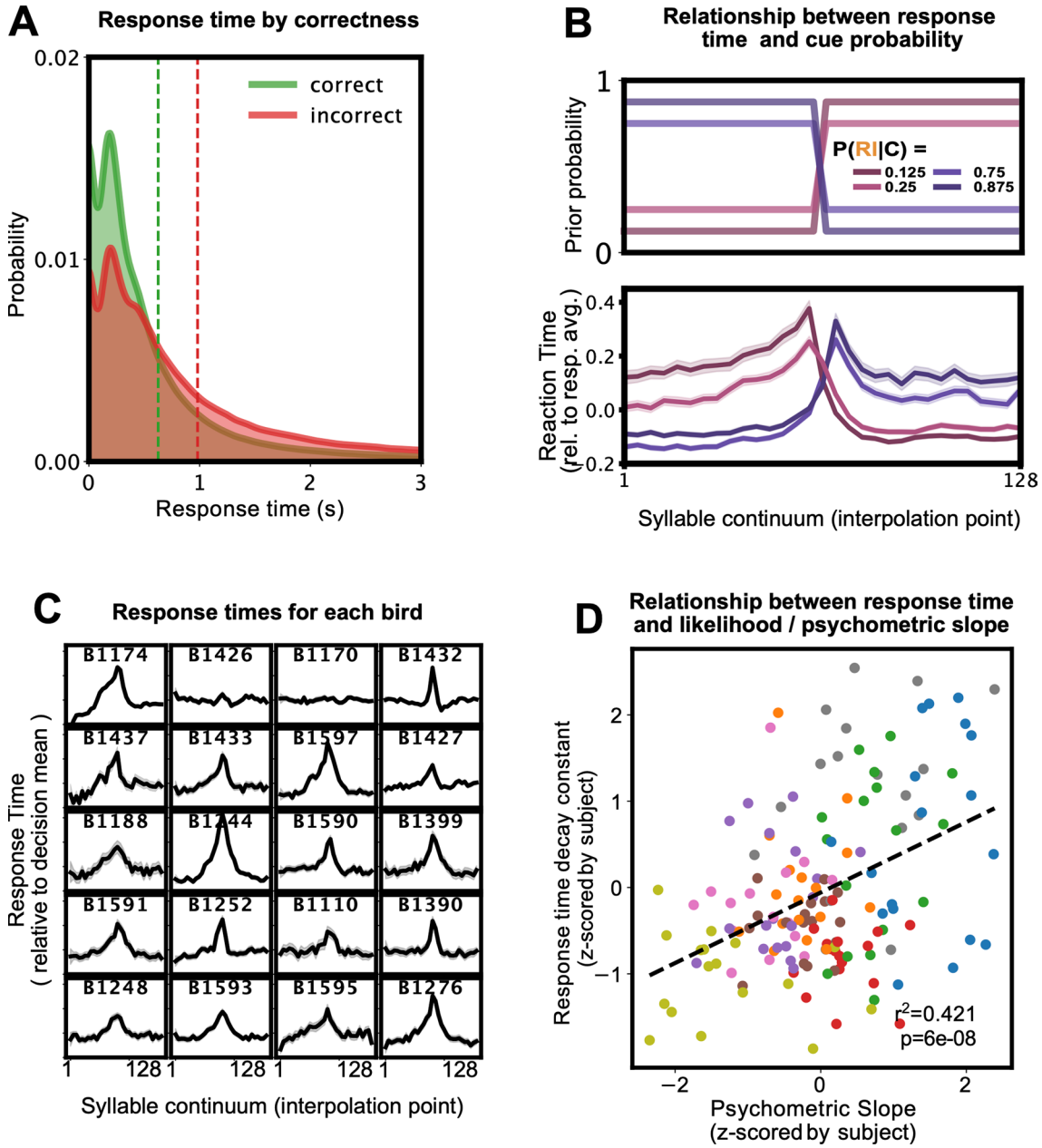
Extended data is available for this paper at https://doi.org/10.1038/s41593-025-01899-1.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41593-025-01899-1.

Correspondence and requests for materials should be addressed to Tim Sainburg or Timothy Q. Gentner.

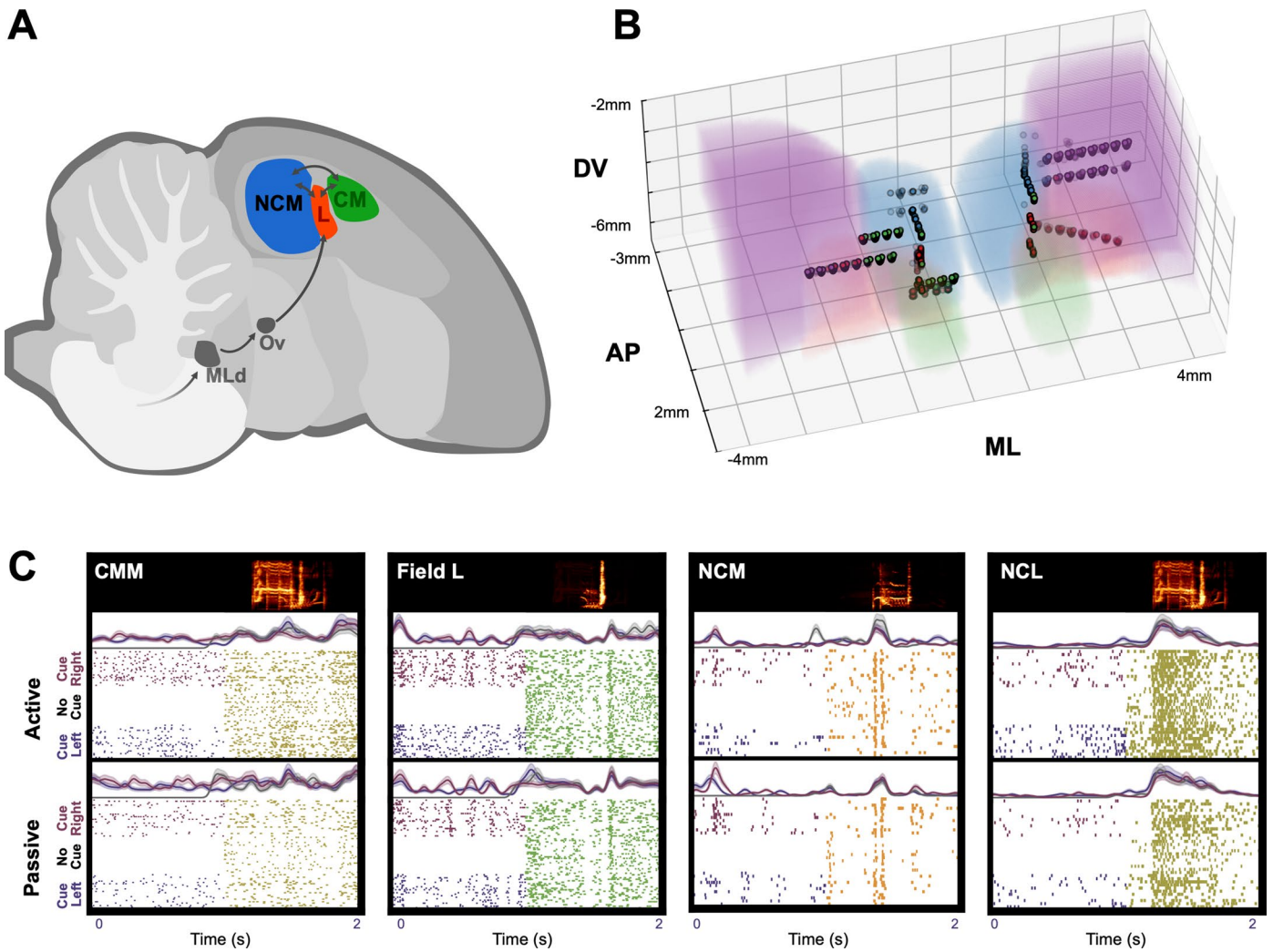
Peer review information *Nature Neuroscience* thanks the anonymous reviewers for their contribution to the peer review of this work.

https://doi.org/10.1038/s41593-025-01899-1



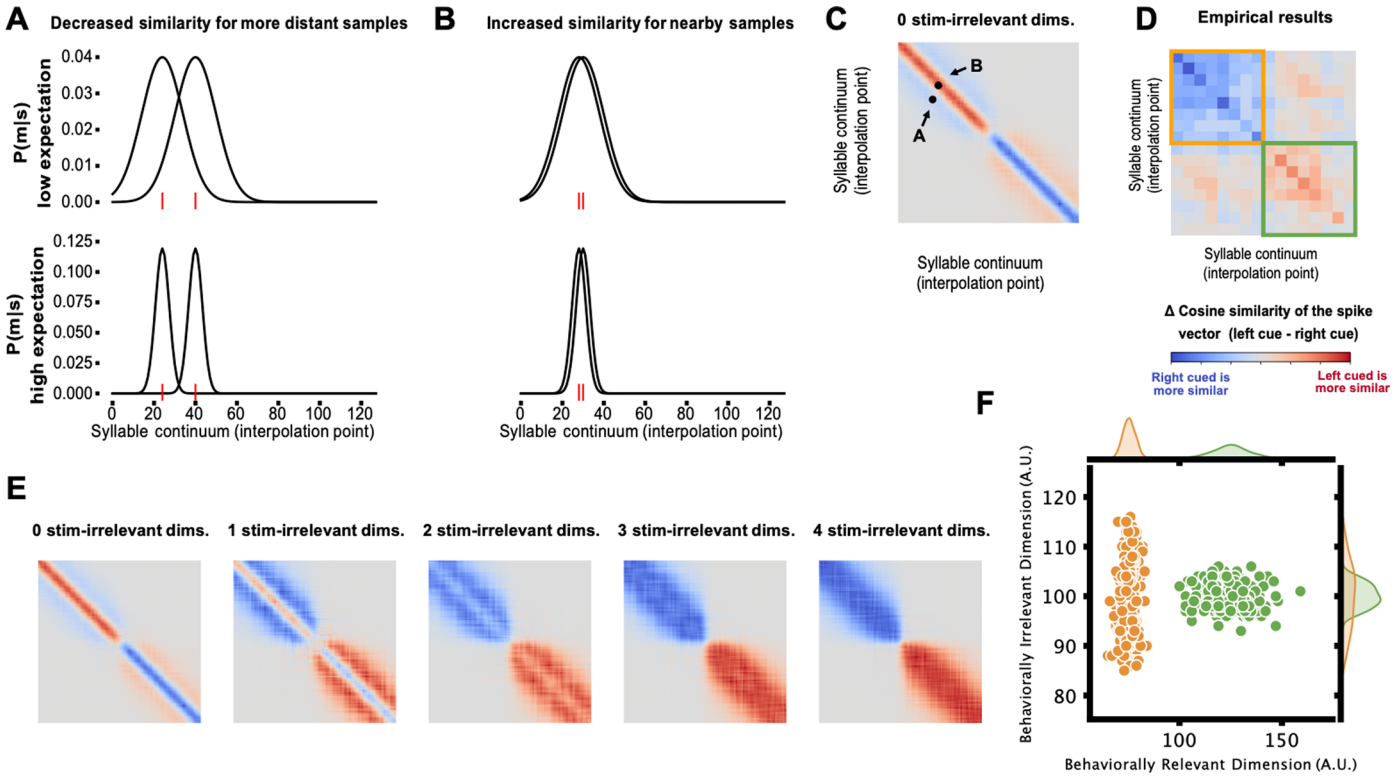
Extended Data Fig. 1 | Response times reflect Bayesian integration. (**A**) Response time across birds for correct versus incorrect trials. (**B**; top) The imposed prior probability in the task for each condition. (**B**; bottom) Average response time over morph for each cue condition (mean and 95% bootstrapped CI). (**C**) Response time over the morph for each bird (mean and 95% bootstrapped CI). (**D**) Decay constants of exponential decay fit to reaction time as a function of distance from decision boundary, in relation to the slope of the fit psychometric function, for each bird and morph. Point colors reflect the morph categories (as in Fig. 3G) (Pearson's correlation, n=121).

https://doi.org/10.1038/s41593-025-01899-1



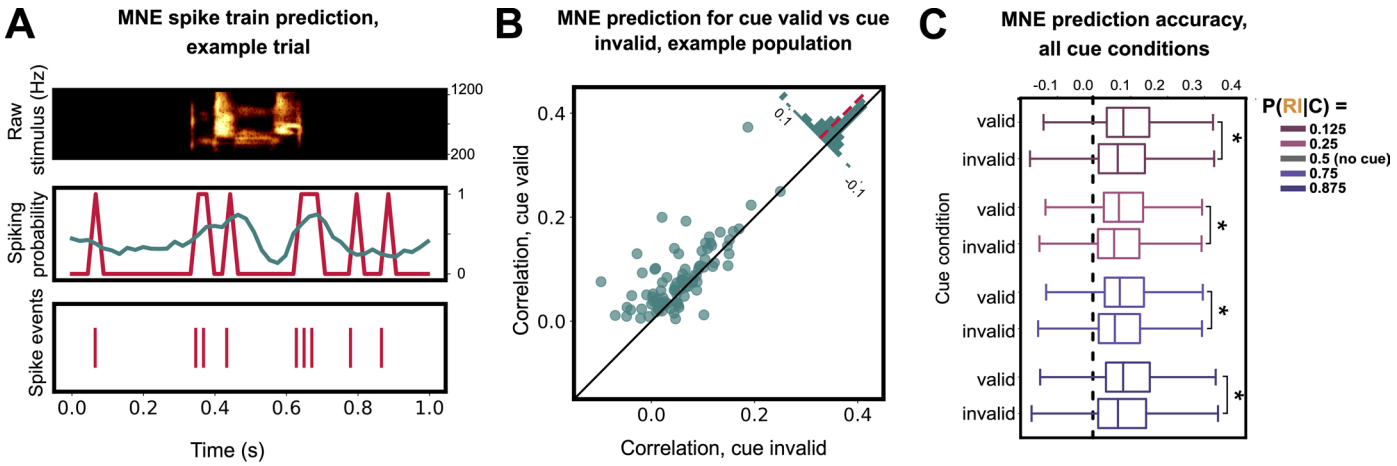
Extended Data Fig. 2 | Recording sites. (**A**) Diagram of auditory input to the songbird brain. Nuclei OV projects to the primary auditory region Field L, which has bidirectionally projections with NCM and CMM. NCL (not pictured), lateral to NCM, additionally exhibits bilateral projections with Field L. (**B**) A visualization of recording sites, shown over top of the starling brain atlas[65]. Colors are consistent with panel A, with NCL being shown in purple. (**C**) The top of each panel shows a spectrogram of the morph stimulus played back. Below, a trace is shown for three cue conditions (No cue, P(Rl|C) = 0.125, and P(Rl|C) = 0.875) corresponding to the average Gaussian convolved spike vector and 95% CI for active trials. Below the trace are sample spike rasters for each cue condition, where each row is a trial. Below the rasters, the sample trace and raster plots are repeated for the same unit in the passive trial condition.

https://doi.org/10.1038/s41593-025-01899-1



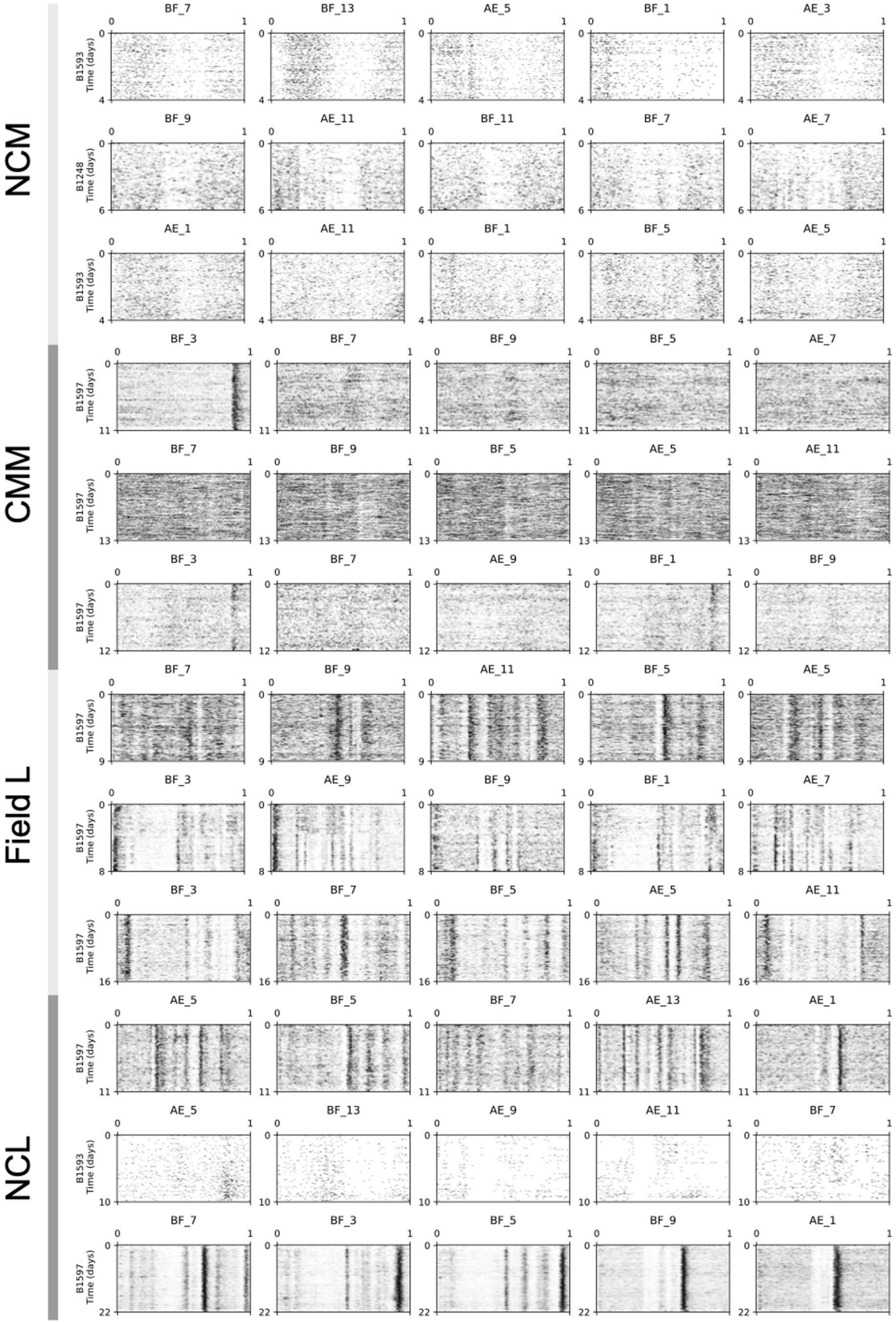
Extended Data Fig. 3 | An outline of the acuity trade-off model. (**A**) A decrease in measurement/representational noise reduces similarity and improves discriminability between stimuli. (**B**) When stimuli are sampled from regions of stimulus space that are sufficiently close to one another, similarity increases in the task-relevant dimension. (**C**) The difference between similarity matrices for the left-cued and right-cued syllables, based upon the 1D task-relevant model. The example from (**A**) and (**B**) are marked as dots with arrows pointing towards them. (**D**) Empirical results from our study. The observed shift in spike train vector cosine similarity for left-cued minus right-cued trials. The shift is depicted here is averaged across units and morphs. Compare to (**C**), where the diagonal does not match the predictions from the 1D model. (**E**) Predictions of the acuity trade-off model. If there are 0 task-irrelevant dimensions, points that are close to each other in stimulus space will become more similar because noise in measurement is reduced. As more task-relevant dimensions are added, the similarity of close points decreases. (**F**) A scatterplot of the noise in measurement for task-relevant and irrelevant dimensions under the acuity trade-off model. When a stimulus is cued, the noise in measurement is reduced in a task-relevant dimension (here the morph dimension) and noise is increased in another dimension.

https://doi.org/10.1038/s41593-025-01899-1



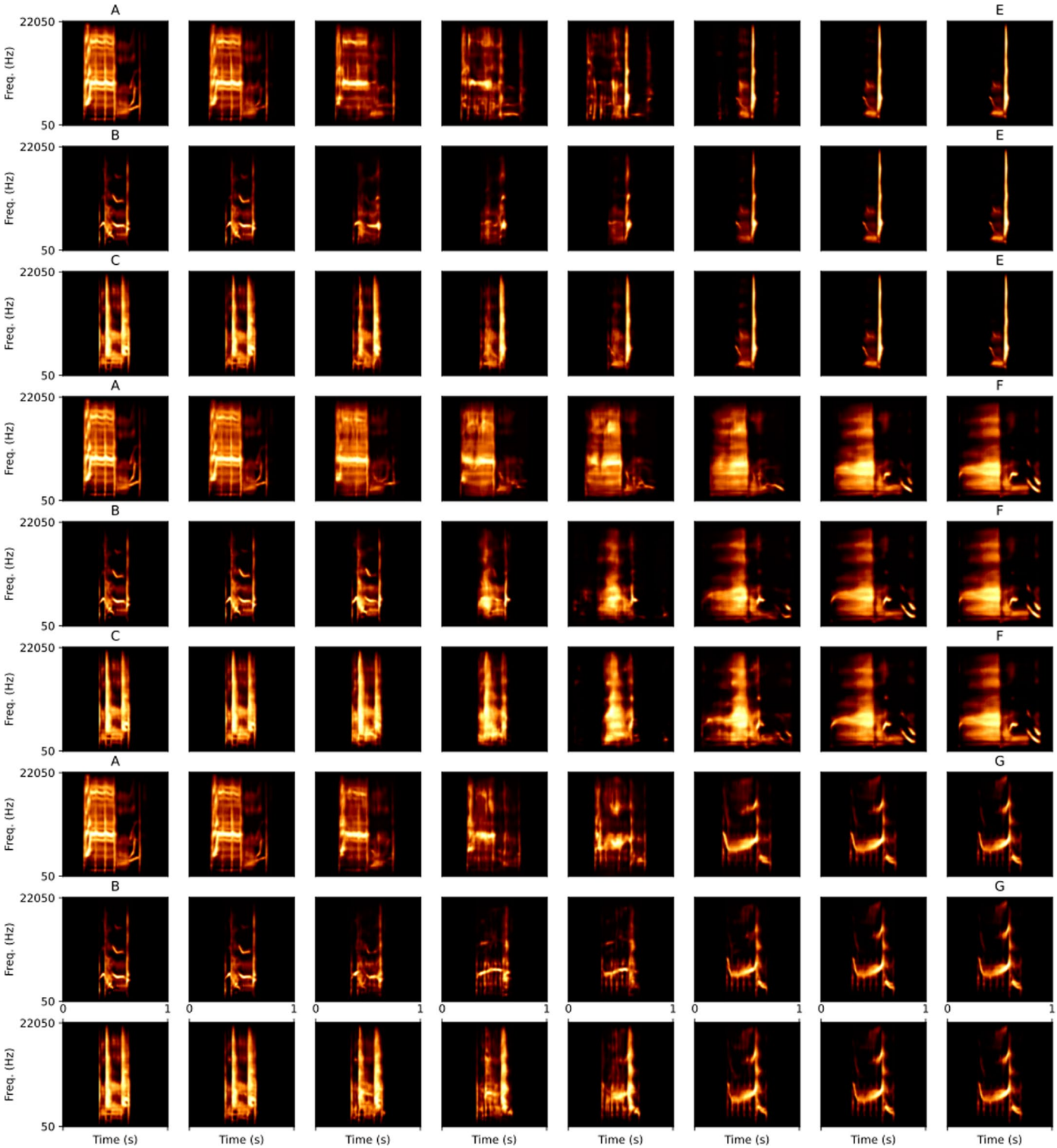
Extended Data Fig. 4 | Maximum Noise Entropy encoder model fit to neural data. (**A**) A sample MNE receptive field prediction. (top) Raw spectrogram of the target syllable on an individual trial. (middle) Actual (red) and receptive field model predicted (teal) spiking probability (same trial). (bottom) Raster plot of spiking events (same trial). (**B**) Correlation values between actual and predicted spiking for cue-valid vs. cue-invalid trials. Trial correlation values were averaged across valid or invalid trials for each unit on an example recording day (N = 98 units). (**C**) Box plots for the distribution of trial averaged correlation values (as in H) for all units broken down by cue-validity and strength. (* indicates significantly increased correlation value for valid verses invalid trials, post-hoc t-test, Cue 0.125, $t(9078) = 19.5$, $p < 0.001$; Cue 0.25, $t(9377) = 18.2$, $p < 0.001$; Cue 0.75, $t(9379) = 18.6$, $p < 0.001$; Cue 0.875, $t(9101) = 17.0$, $p < 0.001$).

https://doi.org/10.1038/s41593-025-01899-1



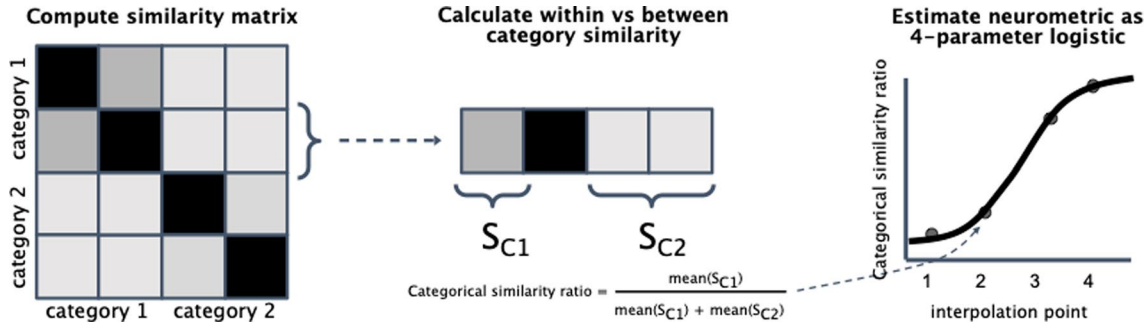
Extended Data Fig. 5 | Example units (rows) for each brain region, showing stability in response profiles to example stimuli (columns) across days/weeks. The units shown are the 3 longest-held units for each brain region. PSTHs are shown for the 1-second reinforced stimuli.

https://doi.org/10.1038/s41593-025-01899-1



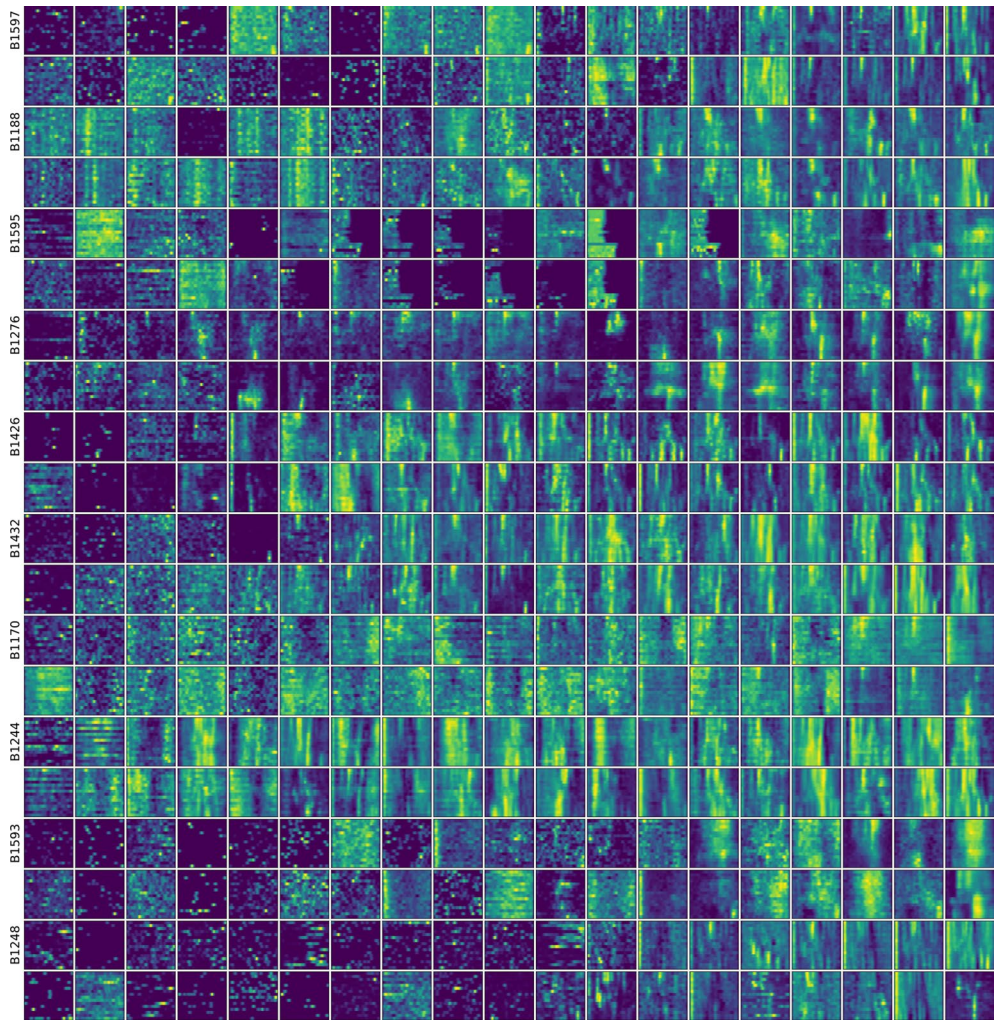
Extended Data Fig. 6 | Spectrograms of 8 sample morph points (of 128 total) from each morph used in the experiment. The starting morph points are written above the left and rightmost syllables.

https://doi.org/10.1038/s41593-025-01899-1



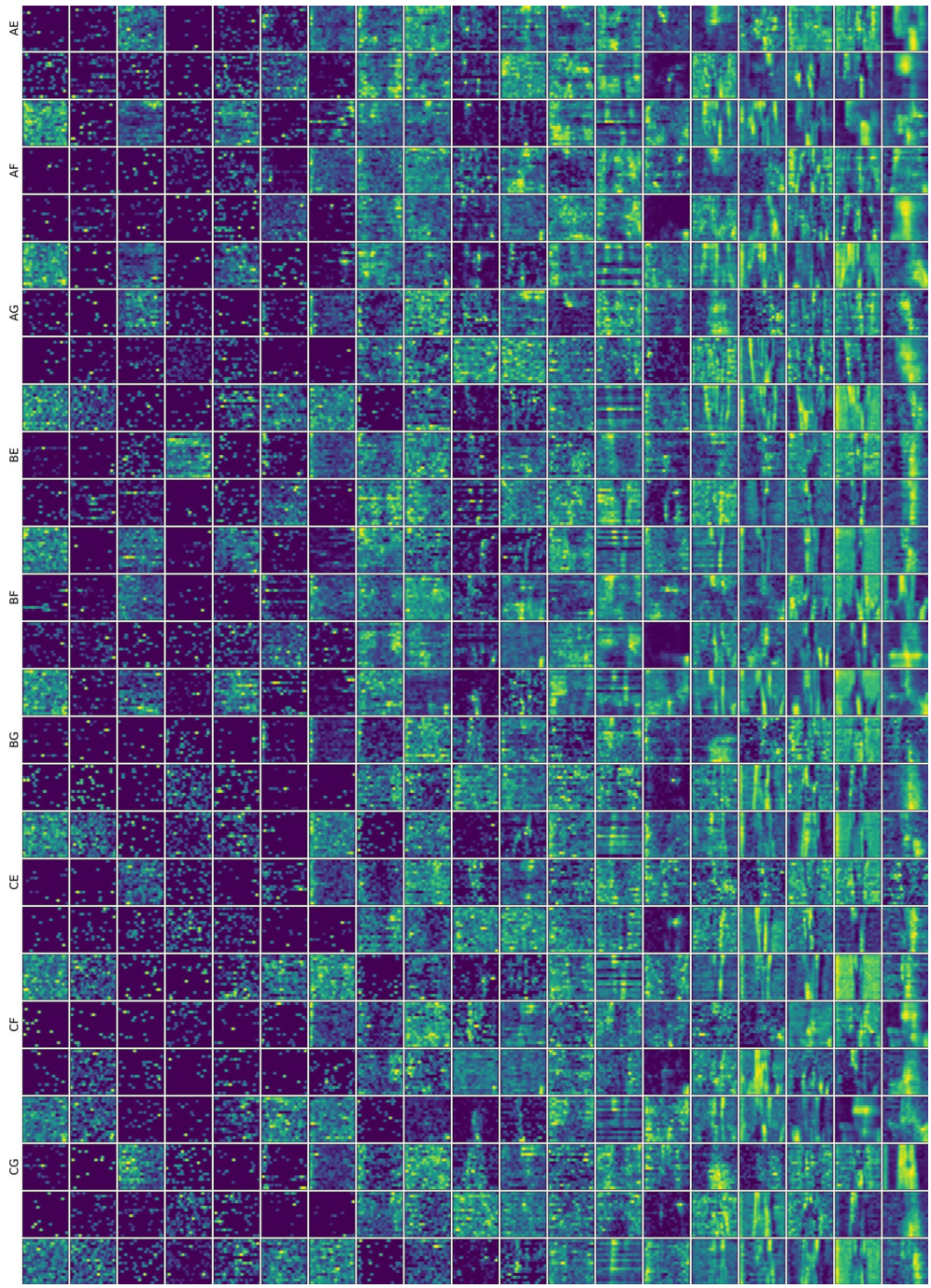
Extended Data Fig. 7 | Method for computing a neurometric function from a similarity matrix. S_{C1} (Similarity to Category 1) and S_{C1} (Similarity to Category 2) represent the within and between category similarities.

https://doi.org/10.1038/s41593-025-01899-1



Extended Data Fig. 8 | Sample units for each subject sorted by the categoricality metric. Each plot depicts the average firing rate across a randomly sampled unit, sorted by the categoricality metric, with time on the X-axis and morph position on the Y-axis. Rows correspond to the subject written on the left.

https://doi.org/10.1038/s41593-025-01899-1



Extended Data Fig. 9 | Sample units for each morph sorted by the categoricality metric. Each plot depicts the average firing rate across a randomly sampled unit, sorted by the categoricality metric, with time on the X-axis and morph position on the Y-axis. Rows correspond to the morph written to the left.

nature portfolio

Corresponding author(s): Tim Sainburg

Last updated by author(s): Nov 28, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Electrophysiology data were collected using Open Ephys. Behavioral data were collected using the pyoperant package for Python.
Data analysis	Data analyses are all in Python, and are available on GitHub (https://github.com/timsainb/cdcp_paper). Software versions: Tensorflow 2.0.0 (for VAE) Librosa 0.10.2 (audio processing) scipy 1.10.1 (signal processing) numpy 1.23.5 scikit-learn 1.2.1 (model fitting) statsmodels 0.13.5 (model fitting) lmfit 1.1.0 (model fitting)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All code for analysis are available on GitHub upon publication (https://github.com/timsainb/cdcp_paper).
Ephys and behavioral data have been deposited to Zenodo (https://zenodo.org/records/7363595).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	20 European starlings were used (10 for behavior only). This number was chosen to provide sufficient behavioral data across subjects, as well as a large sampling of neural populations Sample size was chosen to match prior sample sizes used in our lab
Data exclusions	No data were excluded.
Replication	A description and scripts for the behavior are avaliable with the analysis code. Ephys coordinates are provided. A single dataset was created and analyzed, for which most analyses, except where noted in the manuscript (e.g. bootstrapping), were deterministic.
Randomization	The same conditions were presented to each subject in our experiment, with only the neural recording location differing.
Blinding	Conditions were constant across subjects.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	No laboratory animals were used
Wild animals	European starlings, wild-caught. Sex unknown, age unknown. Starlings were caught by a 3rd party airport groundskeeper and transferred to UCSD. They were then kept in an outdoor aviary until used in our experiment. Animals used in physiology experiments were sacrificed after the experiment using approved university protocols. Behavioral subjects were kept to be used in additional experiments.
Reporting on sex	We were not able to sex the subjects.
Field-collected samples	No field collected samples were used.
Ethics oversight	All procedures were approved by the Institutional Animal Care and Use Committee of the University of California (S05383)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*