Recurrent Dynamics Give Rise to Individualized Category Representations

Maya Taliaferro[1] (mjt10029@nyu.edu) Julien Diriani[3] (jdirani@andrew.cmu.edu)

Esti Blanco-Elorrieta[1,2] (eb134@nyu.edu)

[1] Department of Psychology, New York University
[2] Department of Neural Science, New York University
[3] Neuroscience Institute, Carnegie Mellon University

Abstract

Visual object categorization unfolds on the timescale of milliseconds as sensory representations are transformed into decisions through recurrent processing, yet it remains unclear whether this processing merely sharpens category structure shared across individuals or gives rise to individualized category representations. Using MEG, we examined object categorization along continua spanning basic-level category pairs in which category membership was supported either by graded, weakly diagnostic features (*Experimental Continua*) or by discrete, highly diagnostic features (*Control Continua*). Time-resolved multivariate decoding showed that early neural activity reflected shared perceptual structure for all continua, whereas later, response-locked activity encoded individual-specific category boundaries. Temporal generalization analyses further demonstrated that recurrent dynamics stabilized these individual category representations. Finally, a distance-to-bound analysis revealed that neural category representations predicted reaction times for *Experimental Continua.* These findings show that recurrent dynamics do not merely refine universal category axes but reorganize neural representations to reflect each observer's learned category structure.

Keywords: Object categorization; Category representation; Individual differences; MEG; Multivariate pattern analysis (MVPA).

Introduction

Visual object categorization is not a single-step process but a temporally unfolding transformation from sensory input to decision. Neurophysiological and neuroimaging work has shown that early neural responses reflect coarse, largely stimulus-driven distinctions (e.g., animate vs. inanimate), whereas later activity patterns encode increasingly fine-grained and task-relevant category structure (Carlson et al., 2013; Cichy et al., 2014; Isik et al., 2014; Grootswagers et al., 2019; Contini et al., 2017). Recent MEG work by Gwilliams & King (2020) further demonstrated that this evolution is not purely feedforward: temporal generalization analyses revealed that later stages of object representations depend on earlier ones through recurrent interactions, yielding a cascade of increasingly specific categorical decisions over time.

A growing body of work suggests that these recurrent dynamics support the gradual formation of categorical decisions from ambiguous input. For example, prior studies using parametrically morphed stimuli have shown that neural activity evolves from encoding graded sensory evidence to reflecting categorical percepts, consistent with a process of evidence accumulation and sequential decision formation (O'Connell et al., 2012; Mohsenzadeh et al., 2018; Rajaei et al., 2019). In this framework, recurrent processing stabilizes representations over time and supports the transition from continuous sensory input to discrete behavioral reports. However, these studies have primarily examined category structures that are shared across observers, such that the relevant decision boundaries are largely determined by the stimulus and common across individuals (e.g., at the superordinate category level as either inanimate or animate).

Decades of behavioral work in categorization challenge this assumption and also reveal that not all category boundaries are alike. For some domains, boundaries are strongly constrained by visual structure and are largely shared across observers. For others, particularly basic-level categories, which are privileged for recognition and naming relative to more coarse distinctions (Rosch et al., 1976; Rosch, 1978), category boundaries are learned, probabilistic, and shaped by repeated experience mapping labels to instances (Labov, 1973; Schyns & Rodet, 1997; Tanaka & Taylor, 1991). Even within a single language community, individuals show stable yet idiosyncratic boundary placements along perceptual continua (Nosofsky, 1986; Goldstone, 1994; Malt et al., 1999). These differences reflect persistent internal category representation that guides naming, similarity judgments, and decision times.

Despite this, neural models of object categorization have largely treated inter-individual variability as either measurement noise or as variability in low-level feature sensitivity, rather than as differences in the representational space itself. Even in Gwilliams & King (2020), individual differences were modeled primarily in terms of perceptual ambiguity, while the category partitions themselves were implicitly assumed to be shared. This leaves open a more specific question: Do neural dynamics operate primarily over stimulus-driven structure that is shared across observers, or whether representations align with the observer's own learned category boundaries when the stimulus is ambiguous.

From a computational perspective, recurrent dynamics are not only well suited for noise reduction or evidence

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

accumulation, but also for reshaping representational manifolds to align with learned decision axes. In attractor networks, predictive coding frameworks, and modern recurrent deep vision models, recurrence reshapes sensory spaces such that category boundaries become stable low-dimensional separatrices in population activity (Kriegeskorte & Kievit, 2013; Yamins & DiCarlo, 2016; Kietzmann et al., 2019). If the human cortex operates under similar principles, recurrence should not merely sharpen categorical decisions but should reorganize population codes according to the structure of the category space itself. Under this view, early neural responses should primarily reflect shared perceptual similarity for all exemplars, whereas later recurrent stages should encode participant-specific category representations.

We test this hypothesis using MEG recordings obtained while participants performed a two-alternative forced-choice (2AFC) labeling task on objects sampled from continua spanning basic-level category pairs. We combine mass-univariate with time-resolved multivariate decoding and response-locked analyses to examine how these different forms of structure are expressed over time in the MEG signal. Specifically, we computed temporal generalization matrices (King & Dehaene, 2014), in which classifiers trained at time t were tested at all time points t'. This approach allows us to assess whether the neural representations supporting decoding are transient, sustained, or dynamically evolving over time. To establish the behavioral foundation for these neural inquiries, we first characterize the structure and stability of category representations at the individual level. Further, if these representations are behaviorally relevant, their structure should predict decision dynamics. We test this using a distance-to-bound approach linking neural activity to reaction times (Ritchie & Carlson, 2016). We predict that (a) early neural activity will primarily reflect shared perceptual structure for both continuum types; (b) later, temporally generalized representations will encode participant-specific category boundaries; and (c) the distance of neural population states to these individual boundaries will predict response times, linking recurrent dynamics to behaviorally meaningful category representation.

Identifying the Boundaries of Basic Categories at the Individual Level

To characterize how category boundaries are structured across individuals, participants performed a 2AFC task on constructed visual continua between pairs of related everyday objects by parametrically manipulating diagnostic features.

Stimuli. Each continuum consisted of seven steps that spanned a gradual transition from one category prototype to the other, thereby allowing category boundaries to be estimated from labeling behavior. We created two types of continua that differed in the nature of the features defining category membership. In the *Experimental Continua*, category membership was determined by continuously varying diagnostic features, yielding fuzzy boundaries and graded transitions between categories (e.g., depth along a Plate–Bowl continuum). These continuous diagnostic dimensions allow boundary placement to vary across languages as a function of how features are weighted. In the *Control Continua*, category membership was determined by the presence or absence of a single diagnostic feature, producing sharp, essentially binary boundaries that are expected to align across languages (e.g., the presence of a spout distinguishing a teapot from a mug).

We created the stimuli images as realistic 3D models using Blender, a free and open-source 3D graphics software (Community, B.O., 2018). We displayed them as 400x400px images. We selected six object continua with clear two-category structure and no intermediate category (e.g., no "loveseat" between "couch" and "chair"), based on a norming procedure in which a separate set of participants labeled the endpoints and midpoints. Continua were retained only if the most frequent label for each endpoint accounted for at least 75% of responses, and if the midpoint did not elicit a third category. The final set included three continua with graded boundaries (i.e., *Experimental Continua*: Plate–Bowl, Bucket–Pot, Spatula–Ladle) and three with sharp, discrete boundaries (i.e., *Control Continua:* Watercan–Bucket, Teapot–Mug, Hammer–Ax; Fig. 1a).

Participants. We recruited 24 monolingual English speakers (13 females, 11 males; M_{age} = 24.50 years, SD = 7.72). All participants reported no significant proficiency in any second language (above "3" on a 7-point scale), normal or corrected-to-normal vision and no history of neurological or language disorders. They provided informed consent, received compensation for their time, and the study was approved by the Institutional Review Board at New York University.

Procedure. Before starting the experimental task participants underwent a calibration phase where they saw the objects corresponding to the first and last steps of each continuum accompanied by their category labels. They then completed a short practice of 12 trials on the 2AFC task using the endpoint objects, during which trial-by-trial feedback was provided. No feedback was given during the main experiment. The experimental task included 84 unique images (6 continua x 7 steps x 2 colors), each shown 8 times, totaling 672 trials. Color was task-irrelevant and served only to increase low-level visual variability across exemplars. We excluded participants who did not assign the correct label to the unambiguous exemplars (steps 1 and 7) at least 75% of the time or did not respond to at least 80% of the trials. Additionally, we removed all trials where participants didn't respond, resulting in a total rejection of 3.36% trials.

Each trial started with a fixation cross (300ms), followed by presentation of the target object. After 800ms, two possible labels appeared under the object (e.g., "Plate" and "Bowl" for the Plate–Bowl continuum), and participants were instructed to pick the best label for the object using a button box before a

2000ms timeout (Fig. 1a). These labels were presented on either the left or right side in a counterbalanced manner across trials. Inter-stimulus intervals were randomly sampled from a 400-600ms uniform distribution. Trial order was semi-randomized such that no consecutive trials showed objects from the same continuum. The experiment was presented using (PsychoPy, Version 2023.1.3; Peirce et al., 2019).

Boundaries of Basic Categories are Stable but can be Idiosyncratic

Analysis. To assess the stability of individual-specific category representations and quantify how idiosyncratic these representations were across participants, we modeled behavioral responses using subject-level psychometric functions. For each participant and continuum, we computed the proportion of category-consistent responses at each of step and fit a three-parameter logistic function estimating the category boundary (μ), slope (σ), and lapse rate. We estimated psychometric parameters separately for each subject and averaged them to obtain group-level estimates. We compared variability in category boundaries (μ) between experimental and control continua and assessed equality of variance using Levene's test. To directly test whether category boundaries are idiosyncratic across individuals yet stable within individuals, we compared within-subject variability to between-subject variability. Specifically, for each participant and continuum we computed (i) within-subject boundary shifts ($|\mu_{odd} - \mu_{even}|$) and (ii) between-subject deviation, defined as the absolute difference between an individual's boundary and a leave-one-subject-out (LOO) group boundary estimate ($|\mu_{\_subject} - \mu_{\_group\_LOO}|$). These measures were analyzed using a linear mixed-effects model with fixed effects of Continuum Type (Experimental vs Control) and Variability Source (within vs between), and random intercepts for participants and continua. To further characterize this relationship, we conducted paired t-tests separately for each continuum type comparing within-subject and between-subject variability. Greater between-subject than within-subject variability indicates that category boundaries differ systematically across individuals while remaining stable within individuals.

Results. Our results indicate that category boundaries in the *Experimental Continua* vary substantially across individuals, showing significantly greater between-subject variance than *Control Continua* ($F = 4.54$, $p < 0.05$). Across both continuum types, between-subject variability exceeded within-subject variability, but this effect differed markedly by condition. For Experimental continua, between-subject deviations were substantially larger than within-subject shifts ($M_{between} = 0.74$ vs. $M_{within} = 0.34$), $t(107) = 12.07$, $p < .001$. For Control continua, this difference was smaller but still significant (M_between = 0.37 vs. M_within = 0.30), $t(107) = 2.50$, $p = .014$. A linear mixed-effects model confirmed this pattern, revealing a significant interaction between Continuum Type and Variability Source ($\beta = -0.321$, $p < .001$), indicating that the divergence between within- and between-subject variability was significantly larger for Experimental than Control continua. Together, these results demonstrate that category boundaries are stable within individuals yet systematically differ across individuals, particularly for continua with graded category structure.

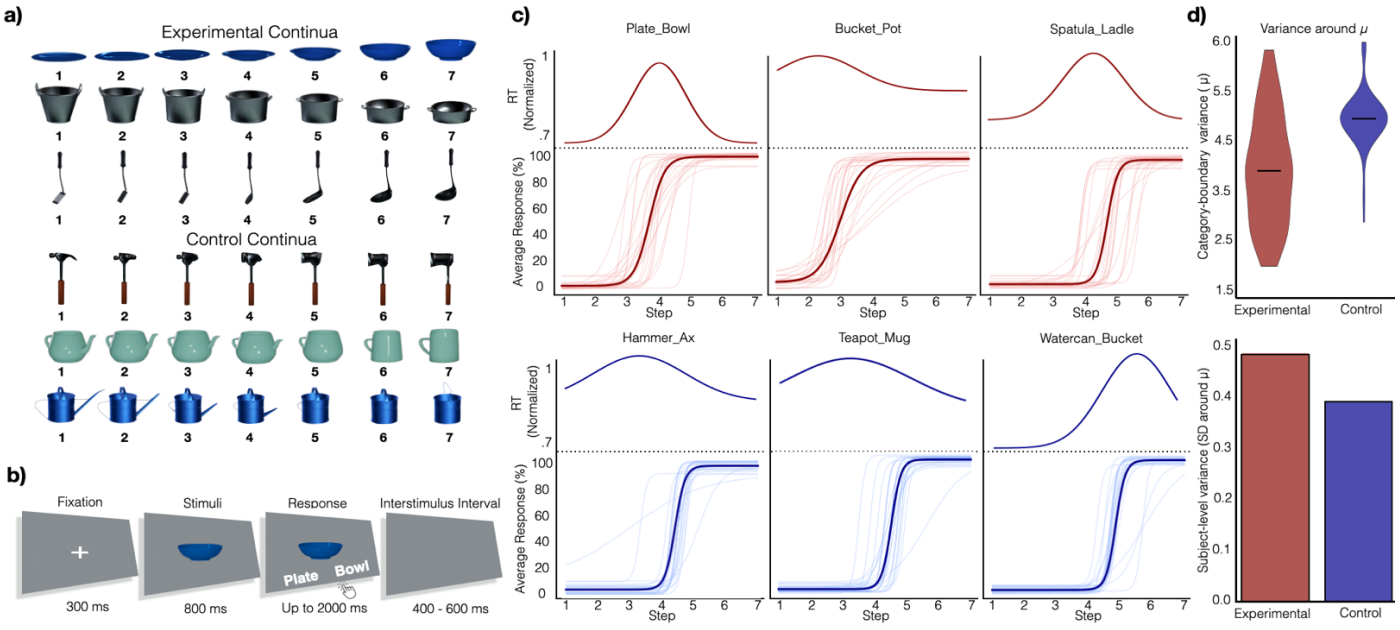

Figure 1: Behavioral Results. a) 2AFC trial design. b) *Experimental* Continua (top) and *Control* Continua (bottom). c) Reaction time data (higher plots) and psychometric curves (lower plots) from 2AFC task. Lighter lines show individual curves and darker lines show group averages. d) Top violin plot depicts the full distribution of variance around μ across continua. Bottom bar plot summarizes this variability by showing the mean standard deviation (*SD*) around μ for each continua type (*Experimental* in red, *Control* in blue).

Neural Representation of Basic Categories is Individual Specific

If recurrent visual dynamics instantiate learned category structure, then neural representations should evolve from reflecting shared perceptual similarity to reflecting observer-specific category representation. We tested this prediction by examining how perceptual, group-defined, and individual-defined category boundaries were expressed over time in the MEG signal.

MEG Acquisition. MEG data were collected in the KIT/NYU MEG Lab in New York using a whole-head 157 channel axial gradiometer system (Kanazawa Institute of Technology, Kanazawa, Japan) as participants lay in a dimly lit, magnetically shielded room. MEG data were sampled at 1000 Hz with a 200 Hz low-pass filter. Data were preprocessed and analyzed with MNE-Python (Gramfort et al., 2013, 2014) and Eelbrain package (https://pythonhosted.org/eelbrain). We 1) rejected all epochs that contained amplitudes >2500 fT/cm for any sensor after noise reduction, 2) rejected any epoch that contained sudden increases in the magnitude of the signal caused by artifacts (e.g., muscular movements), and 3) applied a 40 Hz low-pass filter that eliminated any remaining movement artifacts from our data, given that the gamma-frequency range (>40 Hz) is reportedly the one affected by muscle artifact contamination such as phasic contractions (Yuval-Greenberg & Deouell, 2009). After epoch rejection and behavioral trial rejection process, 10.11% of trials were excluded from further analysis. For multivariate decoding analyses, we did not equalize trial counts to maximize number of trials used for training.

Analysis. We segmented data into initial epochs spanning -300 to 2800 ms relative to stimulus-only onset, with baseline correction applied using the 300 ms preceding target onset. From these initial segments, we further derived two distinct sets of epochs for analysis. Stimulus-locked epochs were generated by cropping the data to a window of -200 to 800 ms relative to stimulus onset. Response-locked epochs were constructed by re-centering each trial's time axis such that the zero-point corresponded to the participant's specific reaction time for that trial. We then cropped these response-aligned segments to a window of -400 to 400 ms around the response. We utilized single-trial data to preserve the trial-wise variance necessary for classification. For each time reference frame, we quantified three types of category information using time-resolved multivariate decoding: 1) Perceptual decoding, contrasting unambiguous steps from opposite ends of the continua (steps 1–2 versus 6–7); 2) Group-category decoding, using category boundaries defined by the group-average boundary for each continuum; and 3) Individual-category decoding, using category boundaries defined by each participant's own psychometric curve. Decoding was performed separately per continua and results were averaged across continua of the same type.

To capture the temporal evolution of neural category representation, we implemented time-resolved decoding using MNE-Python. Decoding was performed in sensor space to preserve the full spatial structure of the measured magnetic fields. At each time point, classifiers were trained using the full pattern of activity across all MEG sensors (the predictors being the amplitude values from each sensor within a specific time window). We averaged non-overlapping 10ms windows to improve signal-to-noise while preserving temporal precision. The target output for the classifiers varied by analysis: stimulus identity for perceptual decoding, the category label predicted by the group-average boundary for group-category decoding, and the label derived from the individual's psychometric curve for individual-category decoding. We trained classifiers using a pipeline of z-score normalization followed by L2-regularized logistic regression, selecting regularization strength via five-fold cross-validation. We evaluated performance via stratified k-fold cross-validation and quantified it as area under the ROC curve (AUC; chance = 0.5). We assessed statistical reliability at each time point using nonparametric cluster-based permutation tests. We generated null distributions by randomly permuting the sign of subject-level accuracy values 1,000 times and computed permutation-corrected p-values for each cluster.

Results. In Stimuli-locked epochs, all three decoding schemes were significantly above chance for both *Experimental* and *Control Continua*. However, perceptual decoding showed the earliest onset of reliable information (~150ms for *Experimental Continua* and ~175ms for *Control Contina*), consistent with the canonical feedforward cascade from perceptual to categorical processing. Importantly, decoding peaks were nearly identical for Group- and Individual-defined boundaries (~200ms for *Experimental* and ~225ms for *Control Continua*), indicating that early neural activity primarily reflected shared perceptual representation rather than idiosyncratic category structure.

A different pattern emerged in Response-locked epochs. Here, perceptual decoding was weakest across both continuum types. Critically, for *Experimental continua,* decoding based on individual-category boundaries was both more accurate and more sustained than decoding based on group boundaries. This advantage was absent for *Control continua*, where Individual- and Group-based decoding performance was nearly identical (Fig. 2). Taken together, these results demonstrate that neural activity surrounding the categorization response encodes category information in a manner that reflects each observer's internal category representation. These results demonstrate a temporal transition in representational geometry: while early activity is dominated by stimulus-driven features shared across observers, late-stage activity reorganizes to reflect each observer's unique internal decision space.

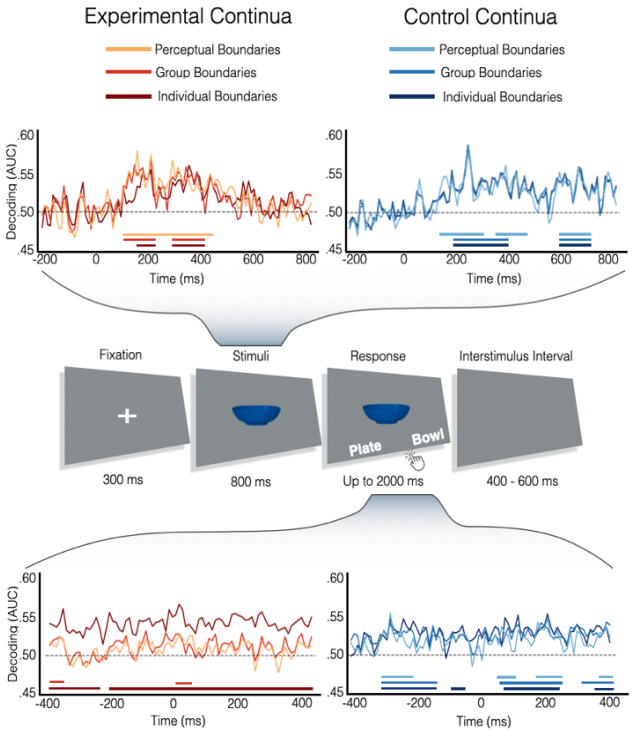

Figure 2: Temporal Decoding Results. Top plots indicate results from Stimuli epochs (200ms before image onset to 800ms after). Bottom plots indicate results from Response epochs (400ms before each participant's own response and 400ms after). Left plots show results from *Experimental Continua* and right from *Control Continua.* Lines below indicate points where decoding is significant.

Recurrent Dynamics Stabilize Individual Specific Categories

Analysis. We trained decoders at each time point and evaluated their performance across all other time points, producing temporal generalization matrices that capture whether category information is transient and time-specific (on-diagonal) or stable and maintained across time (off-diagonal generalization). As in previous analyses, decoding was performed separately using perceptual, group-defined, and individual-defined category boundaries for *Experimental* and *Control Continua.*

Results. For *Experimental Continua*, individual-category decoding yielded pronounced off-diagonal generalization, forming sustained square-shaped clusters across training and testing times. This pattern indicates that once category information emerges, it is maintained over time in a stable representational format, consistent with recurrent stabilization of category structure. In contrast, group-category decoding for these continua showed more limited temporal generalization, suggesting that the group-defined boundary did not align with the stable representation expressed in the neural signal. For *Control Continua*, temporal generalization patterns were highly similar for individual- and group-category decoding. Both exhibited comparable on-diagonal and off-diagonal structure. These results suggest that recurrent dynamics do not simply stabilize category information in general, but preferentially stabilize representations aligned with each observer's internal category representation (Fig. 3).

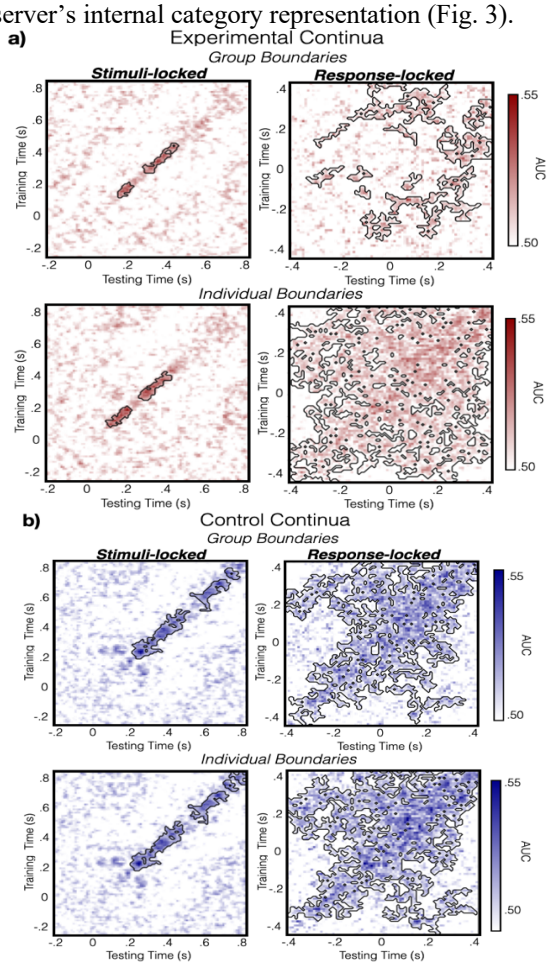

Figure 3: Temporal generalization results. a) *Experimental Continua.* Results from group average category boundaries for decoding on top and bottom show results from using each participant's own category boundaries for decoding. b) *Control Continua.* Results from group average category boundaries for decoding on top and bottom show results from using each participant's own category boundaries for decoding. For both continua types, left plots show generalized decoding during Stimuli epochs while right show results during Response epochs.

Neural Representation is Behaviorally Meaningful

While decoding analyses establish whether category information is present in neural activity, they do not directly specify how such representations relate to behavior. To bridge this gap, we adopted a distance-to-bound approach, which links multivariate neural representations to behavioral performance by leveraging the geometry of the decision boundary learned by a classifier. In this framework, the distance of a neural

pattern from the classifier's decision boundary provides a graded measure of decision evidence: stimuli represented farther from the boundary are expected to elicit faster and more reliable responses, whereas stimuli closer to the boundary reflect greater perceptual uncertainty and longer decision times. Accordingly, as category representations unfold over time, stimuli should become increasingly separable in neural space, resulting in a progressive increase in their distance from the decision boundary such that at peak decoding performance the stimuli should be ordered by their distance from the decision boundary. For our design, that would result in steps to the end of a given continuum (more prototypical category members) lying farther from the boundary and middle steps (more ambiguous items) lying closer to it (schematized in Fig. 4a). Consistent with this view, prior work has shown that distance-to-bound measures predict reaction times in visual categorization tasks and provide an "inner psychophysics" linking neural representational structure to behavioral variability (Ritchie & Carlson, 2016; Ritchie et al., 2015). By applying this approach to our time-resolved decoding analyses, we test whether the neural category representations, particularly those reflecting individual-specific category boundaries, systematically predicts participants' reaction times, thereby directly relating neural representational dynamics to categorization behavior.

Analysis. For each continuum, we first computed group-average RTs as a function of stimulus step by averaging RTs across trials at each step. We fit a Gaussian function to the stepwise RT profile in stimulus space and then evaluated the fitted function on a dense grid of stimulus steps and normalized the resulting curve to its maximum, yielding a smooth, scaled estimate of expected RT as a function of stimulus position. We next applied this behavioral model to single-trial neural data. For each participant and continuum in Response-locked epochs, we temporally smoothed the data using non-overlapping 10 ms windows and assigned category labels based on the individual category boundary. At each time point, we trained a z-scored, L2-regularized logistic regression classifier on the neural data and computed the signed distance of each trial's neural pattern to the classifier's decision boundary. We took the absolute value of this distance as a measure of neural decision evidence, with larger distances indicating stronger category evidence. We then quantified the relationship between neural evidence and behavior by computing the Pearson correlation between trial-wise distance-to-bound values and fitted RTs at each time point. This procedure yielded a time-resolved measure of how strongly neural category representations predicted behavioral response times.

Results. This analysis allows us to ask whether the neural category boundary we decoded corresponds to the decision variable participants use during categorization. For *Experimental Continua,* we observed significant negative correlations between neural distance-to-bound and reaction times, whereby objects in a continuum located further from the decision boundary provided stronger category evidence and supported faster categorization. These correlations emerged well before the overt response, spanning approximately −400 to −300ms relative to response onset, and peaked around 150ms after the response, consistent with the accumulation and stabilization of category evidence over time. In contrast, *Control Continua* showed no significant relationship between neural distance-to-bound and reaction times at any time point (Fig 4). Notably, reaction times for Control Continua were relatively flat across stimulus steps, consistent with the presence of salient diagnostic features that reduce perceptual ambiguity and minimize trial-by-trial variability in decision evidence. Together, these results demonstrate that late neural representations are not merely descriptive of category structure but corresponds to the decision axis participants use to resolve ambiguity during categorization. The distance of neural states to this boundary predicts how quickly participants decide, revealing that recurrently stabilized category representations directly support behavior.

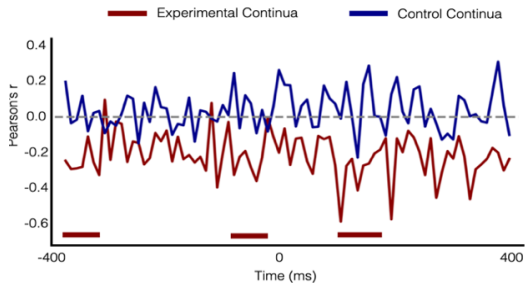

Figure 4. Distance-to-bound approach. Time-resolved correlations during the response epoch for individual boundary decoding. Red lines show Experimental Continua and blue lines Control Continua; bars indicate significant correlations.

Discussion

Recurrent visual dynamics do not merely sharpen stimulus representations; they actively reflect an observer's learned category structure. While early activity encoded shared perceptual features, later response-locked activity selectively expressed individual-specific boundaries, specifically when categories were learned through experience rather than physically diagnostic. Temporal generalization confirmed these representations were stable, consistent with recurrent processing maintaining decision-relevant information. Critically, this neural representation predicted behavior: distance-to-bound relationships with reaction time emerged for idiosyncratic continua, indicating that representational uncertainty constrains decision dynamics when evidence is probabilistic. These results demonstrate that individual differences in category learning are instantiated in neural population codes, supporting models where recurrent processing aligns neural representation with an observer's internal decision space

References

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: the first 1000 ms. Journal of vision, 13(10), 1-1.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature neuroscience*, *17*(3), 455-462.

Community, B. O. (2018). *Blender - a 3D modelling and rendering package*. Stichting Blender Foundation, Amsterdam. Retrieved from http://www.blender.org

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

Grootswagers, T., Robinson, A. K., Shatek, S. M., & Carlson, T. A. (2019). Untangling featural and conceptual object representations. *NeuroImage*, *202*, 116083.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroinformatics*, *7*, 267.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *neuroimage*, *86*, 446-460.

Gwilliams, L., & King, J. R. (2020). Recurrent processes support a cascade of hierarchical decisions. *Elife*, *9*, e56603.

Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. *Journal of neurophysiology*, *111*(1), 91-102.

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*, *116*(43), 21854-21863.

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in cognitive sciences*, *18*(4), 203-210.

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in cognitive sciences*, *17*(8), 401-412.

Labov, W. (1973). The boundaries of words and their meanings. New ways of analyzing variation in English, ed. by C. *J. Bailey and R. Shuy*.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language,40*(2), 230–262

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

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51,* 195–203.

Ritchie, J. B., & Carlson, T. A. (2016). Neural decoding and "inner" psychophysics: A distance-to-bound approach for linking mind, brain, and behavior. *Frontiers in neuroscience*, *10*, 190.

Ritchie, J. B., Tovar, D. A., & Carlson, T. A. (2015). Emerging object representations in the visual system predict reaction times for categorization. *PLoS computational biology*, *11*(6), e1004316.

Rosch, E. (1978). Principles of categorization. *Cognition and categorization/Erlbaum*.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, *8*(3), 382-439.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*(3), 457-482.

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(3), 681.

Singh, A., Taliaferro, M., Lindsay, G., & Blanco-Elorrieta, E. (2025). Blending Boundaries: A Computational Approach to How Bilinguals Reconcile Cross-Linguistic Categorization. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 47).

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, *19*(3), 356-365.

Yuval-Greenberg, S., & Deouell, L. Y. (2009). The broadband-transient induced gamma-band response in scalp EEG reflects the execution of saccades. *Brain topography*, *22*(1), 3-6.