Behavioral/Cognitive

Dynamic Representation of Multidimensional Object Properties in the Human Brain

Lina Teichmann,[1,2] Martin N. Hebart,[1,3,4,5] and Chris I. Baker[1]

[1]Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892, [2]Department of Basic Neuroscience, Faculty of Medicine, University of Geneva, Geneva 1206, Switzerland, [3]Vision and Computational Cognition Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig 04103, Germany, [4]Department of Medicine, Justus Liebig University Giessen, Giessen 35392, Germany, and [5]Center for Mind, Brain and Behavior (CMBB), Universities of Marburg, Giessen, and Darmstadt, Justus Liebig University Giessen, Giessen 35392, Germany

Our visual world consists of an immense number of unique objects and yet, we are easily able to identify, distinguish, and reason about the things we see within a few hundred milliseconds. Here, we used a large-scale and comprehensively sampled stimulus set and developed an analysis approach to capture how rich, multidimensional object representations unfold over time in the human brain. We modeled time-resolved MEG signals of four humans (two females and two males) viewing single presentations of tens of thousands of object images based on millions of behavioral judgments. Extracting behavior-derived object dimensions from similarity judgments, we developed a data-driven approach to guide our understanding of the neural representation of the object space and found that every dimension is reflected in the neural signal. Studying the temporal profiles for different object dimensions, we found that the time courses fell into two broad types, with either a distinct and early peak (~125 ms) or a slow rise to a late peak (~300 ms). Further, early effects were stable across participants, in contrast to later effects which showed more variability, suggesting that early peaks may carry stimulus-specific and later peaks more participant-specific information. Dimensions with early peaks appeared to be primarily visual dimensions and those with later peaks more conceptual, suggesting that conceptual representations are more variable across people. Together, these data provide a comprehensive account of how behavior-derived object properties unfold in the human brain and form the basis for the rich nature of object vision.

Key words: MEG; objects; timeseries; visual perception

Significance Statement

Humans are excellent at identifying, distinguishing, and reasoning about a huge number of objects—all of which requires comparing visual information to internal representations and assign what we see to object categories. Simultaneously, we also process properties relevant to behavior. Seeing a cat, for instance, involves recognizing both its physical properties (fur, size, ears, claws) and many other types of properties (living, moving, playful) that add up to our idea of "cat." In our study, we investigated the time course of the neural response using MEG neuroimaging. We found that a diverse array of object properties relevant to behavior contributes to the neural signal and reveal how such rich object representations unfold over time in the human brain.

Introduction

A central aspect of vision is our ability interact with a huge variety of different objects that are associated with a wide range of perceptual, functional, and conceptual properties. What neural representations support our ability to make sense of our visual world in the span of just a few hundred milliseconds? To capture the multidimensional nature of object representations and how they unfold over time, here

Received May 29, 2025; revised Nov. 21, 2025; accepted Jan. 9, 2026.

Author contributions: L.T., M.N.H., and C.I.B. designed research; M.N.H. contributed unpublished reagents/analytic tools; L.T. analyzed data; L.T. and C.I.B. wrote the paper.

We thank Anna Corriveau, Alexis Kidder, Adam Rockter, and Maryam Vaziri-Pashkam for their help collecting the data and Philipp Kaniuth for help generating the predicted similarity embedding. Additional thanks to Tom Holroyd and Jeff Stout for technical support and discussions. We thank Grace Edwards and Susan Wardle for helpful comments on earlier versions of this manuscript. We utilized the computational resources of the NIH HPC Biowulf cluster to run the MEG analyses (http://hpc.nih.gov). The work presented here was supported by the Intramural Research Program of the National Institutes of Health (ZIA-MH-002909), under National Institute of Mental Health Clinical Study Protocol 93M-1070 (NCT00001360). Additional support was received by an Ambizione Grant awarded to L.T. by the Swiss National Science Foundation. Furthermore, M.N.H. received support from a research group grant by the Max Planck Society, the ERC Starting Grant project COREDIM (ERC-StG-2021-101039712), and the Hessian Ministry of Higher Education, Science, Research, and the Arts (LOEWE Start Professorship to M.N.H. and Excellence Program "The Adaptive Mind").

The authors declare no competing financial interests.

Correspondence should be addressed to Lina Teichmann at lina.teichmann@unige.ch.

This paper contains supplemental material available at: https://doi.org/10.1523/JNEUROSCI.1057-25.2026

https://doi.org/10.1523/JNEUROSCI.1057-25.2026

we moved beyond the traditional, stimulus-focused approach of studying object processing by directly examining how continuous, behavior-derived object dimensions are reflected in the neural signal.

Prior studies that have typically focused on specific types of stimuli or properties have demonstrated both univariate (e.g., N170 for faces) and multivariate differences in the brain response over time. While these studies have revealed some general features of the object response, they are often based on relatively small, hand-selected sets of stimuli (Carlson et al., 2013; Cichy et al., 2014; Bankson et al., 2018) which do not sample the object space in a representative way and cannot adequately capture the richness of the object response (Grootswagers and Robinson, 2021). Hand-selecting sets of stimuli may lead to a sampling bias as some types of objects are often overrepresented (e.g., animals) while others may be absent or underrepresented (e.g., furniture).

To provide a more comprehensive understanding of object vision, we focused on addressing two key challenges: (1) sampling an expansive set of object stimuli across the thousands of object types we can identify and interact with and (2) accounting for the rich meaning and behavioral relevance associated with individual objects beyond discrete labels (Contier et al., 2024). To do this, we turned to THINGS-data (Hebart et al., 2023), which contains MEG data for 1,854 systematically sampled object concepts (Hebart et al., 2019) as well as rich behavioral data comprising 4.7 million similarity judgments that have been used to derive 66 underlying dimensions of objects (Hebart et al., 2020, 2023). Using these data, we developed a data-driven approach to uncover the temporal dynamics of object processing by directly examining how behavior-derived dimensions are reflected in the evolving object representations in the human brain.

In contrast to previous work requiring category-based stimulus selection and labelling, here we use behavioral embeddings that characterize each image across multiple dimensions capturing similarity relationships directly. In addition, the dimension values for each object are continuous (e.g., jellybeans are more colorful than sunflowers, but sunflowers are more colorful than sugar), allowing for fine-grained modeling of similarity in the neural data and thus capturing the richness of object vision. In contrast to common approaches such as representational similarity analysis (RSA), our method allows us to directly study evoked neural representations at the global level (i.e., across dimensions) as well as at the local level (i.e., each dimension separately). Critically, our approach goes beyond studying object identification and categorization and allows us to determine the time course of response specific to each of the behavior-derived dimensions.

Our results show that every dimension is reflected in the neural signal. The temporal profiles evoked by the dimensions tended to group according to the relative strength of two phases of processing (~125 and ~300 ms) as well as the presence or absence of an offset related response (~500–600 ms). Early responses as well as stimulus offset effects were spatially tied to posterior sensors while the second phase of processing was spatially more distributed. Early effects were more generalizable across participants than later effects suggesting that stimulus-specific information is reflected in the early parts of the signal while subject-specific information unfolds later in time. An exception to this were dimensions that are primarily associated with physical properties of the object, which generalized well across participants throughout the timeseries. By focusing on behavioral relevance of object dimensions, our results collectively provide a comprehensive characterization of the temporal unfolding of visual object responses in the human brain.

Materials and Methods

Dataset

We used the publicly available THINGS dataset (Hebart et al., 2023) which contains densely sampled MEG data of four human participants (two males, two females) as well as crowdsourced behavioral data. The MEG portion of the data contained neural recordings from four participants who each viewed a total of 27,048 unique images over the course of 12 sessions. Every image was shown centrally for 500 ms with an inter-stimulus interval of 800–1,200 ms. Participants completed a target-detection task, looking for artificially generated images of objects that do not exist. Of the 27,048 trials, 22,248 trials were experimental trials showing unique images from the THINGS database (Hebart et al., 2019), which were used for the analysis.

Each image belonged to one of 1,854 object concepts (e.g., aardvark, clock, chicken wire, among many others). Unique image exemplars for each of the concepts were repeated 12 times over the course of the MEG experiment (one image per concept per session). In addition to the image concepts, we used a behavior-derived embeddings to model MEG sensor responses. The embeddings contained weights on 66 dimensions which capture trial-by-trial responses for 4.7 million odd-one-out judgments on triplets of the 1,854 object concepts (Hebart et al., 2023). Each dimension describes a certain object property (e.g., circular/round, colorful, food-related); however, these dimensions were derived in a data-driven way based on the behavioral data.

The original embedding was trained at the concept-level (one image per concept) and hence could miss visual variability across exemplars. For example, the model may have been trained on the concept "egg" with a photo of a brown egg, but participants saw images of both white and brown eggs. This means that any image-specific dimensions cannot generalize to other examples of the same object if the visual appearance (e.g., color) is different. In order to obtain image-level embeddings, we used a neural network model (CLIP-ViT; Radford et al., 2021) that can predict image–text pairs and has also been shown to be able to predict similarity judgments with high accuracy (Hebart et al., 2022; Muttenthaler et al., 2022). We used a separate ridge regression model for each dimension and fit it for the behavioral data corresponding to 1,854 images. Then we examined the activity patterns in the final layer of the image encoder. We then used ridge regression to predict dimension weights for each of the 66 dimensions for all ~27,000 images in the THINGS database. To model the evoked neural response measured with MEG, we used the image-level predicted weights along the 66 dimensions. Please note that, while this analysis relies on features derived from a neural network model, the human similarity embedding showed good fits at the level of individual dimensions, demonstrating that these effects were not merely driven by projecting the CLIP image embedding to 66 arbitrary unidimensional spaces. Using the image-based embedding did result in an overall increase in model performance but the overall pattern of results remained similar even without the use of CLIP-ViT (see Supplementary Fig. S1).

Preprocessing

Our preprocessing pipeline was built using *mne-python* (Gramfort et al., 2013) and described in detail in the dataset release (Hebart et al., 2023). The preprocessing steps included filtering (1–40 Hz), baseline correction using *z*-scoring, and epoching the data from −100 to 1,300 ms relative to stimulus onset. No artifacts or eyeblinks were removed. Raw and preprocessed data can be directly downloaded from OpenNeuro (https://doi.org/10.18112/openneuro.ds004212.v3.0.0).

Analyses

Modeling MEG data based on multidimensional similarity judgments: within-participant regression. To model how the behavior-derived dimensions unfold over time in the human brain, we fitted multiple ridge regression models at every timepoint to learn the association between the multivariate MEG sensor response and the scores along each dimension. We trained the model on data from 11 out of the 12 sessions (20,394

trials) and tested on the remaining one (1,854 trials). This process was repeated so that every session was used as testing data once. The model was trained and tested for each participant separately. A separate model was trained and tested at every timepoint. Models were fit in Python using ridge regression models with a tuned alpha regularization term at each timepoint (possible alpha regularization terms included 100 linearly spaced values between 0.000001 and 30,000). The input data was scaled using the *sci-kit learn StandardScaler* function and the regularization was tuned in an internal cross-validated way using the *sci-kit learn RidgeCV* function (Pedregosa et al., 2011).

We assessed the model's performance by correlating the predicted dimension score of all left-out trials with behavioral embeddings for each of the images. These correlations were interpreted as amount of information in the neural signal associated with a given dimension. To assess statistical significance, we generated subject-level null distributions. We reran the identical analysis using a session-wise cross-validation but shuffled the testing labels 1,000 times. The shuffling index was kept consistent across cross-validation splits and the resulting correlations were averaged. Thus, for each dimension we had a null distribution at every timepoint. We defined the 99th percentile of each distribution as a threshold and then looked for the maximum value across timepoints and dimensions to test the model performance against to correct for multiple comparisons.

To gain insights into which sensors primarily drove the effects, we ran the regression analysis on smaller subsets of sensors. We ran this sensor-searchlight analysis in an identical way as the full regression analysis outlined above, with the only difference being the dimensionality of the data: In the original analysis we used the data from 271 sensors, whereas the sensor-searchlight was run on clusters of 5–10 sensors. The channel neighborhoods were defined with standardized adjacency maps included in sci-kit learn which are based on automatic, symmetric triangulation of the 2D CTF MEG layout. The sensor-searchlight analysis allows us to project the model's performance (correlation between predicted dimension score of all left-out trials with behavioral embeddings for each of the images) to the sensor layout to see which sensor groups can predict the behavioral embeddings at a given timepoint the best.

Lastly, we also trained a linear model to predict the activation of each sensor using the multidimensional similarity judgments. We used a session-wise cross-validation approach and ran this analysis for each participant separately. The model's performance was assessed by correlating the predicted sensor activations for the test set and the true sensor activations at every timepoint. In contrast to the sensor-searchlight analysis, this analysis does not allow to readily interpret the spatial distribution of each dimension but rather leverages the breadth of all behavioral dimensions to make predictions about neural data measured at each sensor location.

Examining time course similarities across people: across-participant regression. To examine whether time course profiles are consistent across participants, we trained the a model to learn the association between the multivariate MEG sensor activation pattern at every timepoint and the behavioral dimension profiles using data from one of the participants and testing it on the others. We did a session-wise cross-validation to match the within-participant analysis in terms of training and testing set size (see (a)). We repeated this process until all pairwise comparisons of participants were used as training and testing set. A separate model was trained and tested at every timepoint. Models were fit in Python using *sci-kit learn* ridge regression models following the same hyperparameter tuning method as in the within analysis.

We plotted the difference between the within- and the across-participant model for all 66 dimensions at select timepoints (100, 200, 300, 400, 500, 600 ms). We used Bonferroni-corrected pairwise t tests to test whether the differences between the models were significant (defined as $p < 0.01$) at these timepoints. The results of the tests are reported in the supplementary materials (Fig. S4).

Examining time course similarities across dimensions: dynamic time warping. The results of the regression models were timeseries of correlations for each dimension. To compare the shapes of these timeseries and assess overall similarities and differences, we used dynamic time warping implemented in the *dtaisdistance* toolbox (Meert et al., 2020). In contrast to correlation measures which are compression based, DTW is shape based and is well suited to investigate timeseries similarities that may have a temporal drift (Aghabozorgi et al., 2015). The goal of DTW is to find matches between patterns of two timeseries by assessing how much one timeseries has to be warped to look like the other one. This is achieved by generating distance matrices filled with pairwise Euclidean distances between timepoints and finding the shortest path through this matrix while adhering to several rules: The start and end of the timeseries have to align, the path cannot go back in time, and it has to be continuous.

The DTW similarity measure represents the sum of the Euclidean distances along the shortest path. We extracted this measure for smoothed time courses and generated a similarity matrix. Given that we were interested in the relative shape of the timeseries and not the differences in signal-to-noise ratio, we normalized the timeseries before running the dynamic time warping by calculating z-scores for each dimension timeseries and each participant. Then we averaged across participants and calculated the DTW similarity measure for all dimension comparisons. Because z-scoring and timeseries with only a few timepoints that are above zero can amplify the effect of noisy timeseries, we sorted the time courses by number of significant timepoints across all participants and excluded the dimension time courses with the lowest 5% of significant timepoints. This resulted in three dimension time courses to be excluded [fire-related, yellow, masculine (stereotypical)]. As this cutoff is arbitrary, we also tested to exclude more dimensions or basing the exclusion on a different metric (e.g., peak amplitude before z-scoring). The bottom dimensions are always the same and the number of excluded dimensions does not influence the conclusions from the DTW analysis in any substantial way.

Running hierarchical clustering on the resulting DTW distance matrix allowed us to establish a qualitative measure of different prototypical timeseries characteristics. We set the threshold for the hierarchical clusters to be at 0.5× the maximum distance observed. This threshold is arbitrary, and the number of clusters extracted can be influenced by changing this threshold. However, the overall conclusion that the two main characteristics differentiating the clusters are an early and sharp peak versus a slow rise to a late, sustained peak are robust.

Open science practices

All data is publicly available under a Creative Commons license and can be downloaded from OpenNeuro: https://doi.org/10.18112/openneuro.ds004212.v3.0.0. The analysis code for all analyses in this paper is available on GitHub: https://github.com/Section-on-Learning-and-Plasticity/THINGS-MEG.

Results

The overarching goal of the current study was to characterize how multidimensional representations unfold over time by combining large-scale MEG data with behavior-derived similarity embeddings. Our primary aims were to (1) extract time courses from the MEG response that are associated with each behavior-derived multidimensional profile, (2) reveal how these time courses vary across dimensions and participants, and (3) identify prototypical temporal characteristics shared between response profiles of individual dimensions.

THINGS-MEG measured evoked neural responses in four participants viewing >27,000 unique natural images associated with 1,854 object concepts. To associate object dimensions with these natural images, we used behavioral embeddings derived from similarity judgments, based on 4.7 million judgments on 1,854 object concepts in a triplet odd-one out task (Hebart et al., 2023). Thus, the stimuli used in the MEG study are associated with both object concept labels (e.g., nail polish) and weights on behavior-derived dimensions (e.g., colorful; Hebart et al., 2020). The dimensions cover a broad range of object properties, with some being strongly linked to visual properties

(e.g., colorfulness) and others linked more to functional or contextual properties (e.g., childhood-related). The behavioral similarity embeddings were based on concept-level judgments (i.e., one image per object concept), potentially missing some of the visual variability in the MEG stimuli (12 images per object concept). To overcome this issue, we used an artificial neural network (CLIP-ViT; Radford et al., 2021) to augment the behavioral dataset and generate image-level embeddings for later predictions (Hebart et al., 2022). This model has been successfully used to provide image-level estimates of object dimensions (Muttenthaler and Hebart, 2021; Wang et al., 2023; Contier et al., 2024; Conwell et al., 2024). Post hoc analyses of our data showed a consistent improvement in prediction scores for all 66 dimensions when using image-level versus concept-level embeddings (Supplementary Fig. 1). While the results were consistently stronger, the overall pattern of results remained similar even without the use of CLIP-ViT.

We used the scores on the 66 behavior-derived dimensions to model the evoked neural response to >26,000 images recorded with MEG. In particular, we used both decoding and encoding models to associate the multivariate MEG sensor response with the 66-dimensional behavioral embedding. In contrast to previous work, our method allows us to effectively examine multidimensional object profiles that are obtained from behavioral data in a data-driven way. We can capture and examine the relationships between objects across many dimensions, as two images that are similar along one dimension may be very different along another dimension (e.g., beads and nail polish are both colorful but not necessarily childhood related). Thus, this approach does not rely on selecting and contrasting object classes but instead uses the same images and experimental trials with a relabeling according to behavior-derived dimensional profiles.

Distinct time courses can be derived for each behavior-derived dimension

We fitted multiple linear regression models to learn the association between the MEG sensor activation patterns and the behavioral embeddings at every timepoint (Fig. 1). The linear models were fit on MEG data from 11 sessions (20,394 trials). Using the left-out, independent session (1,854 trials) as a test set, we predicted the continuous value along each dimension from the MEG sensor activation patterns. Correlating these predicted scores with the true behavioral dimensional profiles resulted in timeseries of dimension information in the neural response for all four participants. The analysis revealed behavior-derived multidimensional information from ∼80 ms onward (Fig. 2*A*). The time course showed an early peak at ∼100 ms which was maintained over time up to 1,000 ms after stimulus onset, with an overall peak at ∼300 ms. To estimate the spatial distribution of this effect, we ran two additional analyses focusing on the localized response. First, we ran a sensor-searchlight regression (Fig. 2*A*). Specifically, we trained models to predict the behavioral embedding of each dimension based on a group of adjacent MEG sensors at every timepoint. Second, we leveraged the breadth of the behavioral dimensions to fit linear models to predict each sensor activity pattern separately (Fig. S2). In particular, we trained models to predict the univariate response in each MEG sensor at every timepoint using the multidimensional behavioral values associated with each stimulus. Both analyses show that, while posterior sensors had the strongest effects particularly early in time, the effects became more distributed later in time (>150 ms; Fig. 2*A*) with an increased relative contribution of temporal and frontal sensors.

Extracting the correlation time courses for each dimension separately, we found that there was significant information about most dimensions in the signal from 100 ms onward (for at least 3 participants, 56/66 dimensions carried significant information by 100 ms and 65/66 dimensions by 150 ms). Furthermore, our results revealed a different unfolding of neural responses across time for different dimensions. For example, dimensions such as "plant-related," "colorful/playful," and "white" showed distinct, early peaks (∼125 ms) and strong contributions of posterior sensors. In contrast, other dimensions such as "body-/people-related," "food-related," and "transportation-related" yielded a slower rise to a later peak (∼300 ms) and displayed a more distributed pattern across MEG sensors. In addition, several of the dimensions yielding distinct early peaks exhibited a stimulus offset effect at ∼500 ms. In contrast, several other dimensions did not show a distinct early peak or offset response but unfolded slowly over time and rose to a late peak (>300 ms). While signal-to-noise ratio differed across dimensions, all dimension time courses exceeded zero at some point over time (see Fig. 2*B* for representative example time courses selected based on peak amplitude and Supplementary Fig. 3 for all time courses). Strikingly, the behavior-derived dimensional profiles were evident in the neural response even though MEG participants completed an orthogonal detection task. This demonstrates that the dimensions are automatically reflected in the neural data without a task that requires participants to engage with the object properties directly. Overall, these results highlight that a wide range of behavior-derived dimensions are reflected in distinct temporal profiles and that their information is distributed across MEG sensors.

Time courses vary across dimensions and are consistent across participants

Building on the finding that a range of behavior-derived dimensions are reflected in distinct neural profiles measured with MEG, we next tested to what degree these time courses are consistent across participants. As above, we used multiple linear regression models, but this time with a subject-based cross-validation scheme, to examine whether the temporal characteristics we found for each dimension are idiosyncratic or consistent across participants. Specifically, we trained the model on MEG data from each participant and tested its performance on the data from the remaining ones. We used a session-wise cross-validation scheme, so that the model was trained and tested on the same number of trials as the within-participant analysis. This is a very stringent test for generalizability, as the model trains and tests on completely different datasets.

Our results show that for most dimensions, the across-participant models revealed similar timeseries characteristics as the within-participant model (Fig. 3, Supplementary Fig. S5), highlighting that the temporal profiles we uncovered for each dimension were robust and not idiosyncratic to specific individuals of our study. As Figure 3*A* shows, the early peaks (∼125 ms) in particular were consistent in amplitude and timing when comparing the within- and the across-participant model. In contrast, later effects (e.g., 200, 400, 600 ms) did not generalize as well across participants for most dimensions. For many dimensions, we observed a substantial drop in performance for the across-participant model performance relative to the within-participant model at ∼200 ms before the performance improved again (Fig. 3*B*,*C*). This indicates that the differences we observed between the within- and across-participant model were not solely driven by a time-dependent decrease in signal-to-noise ratio. The

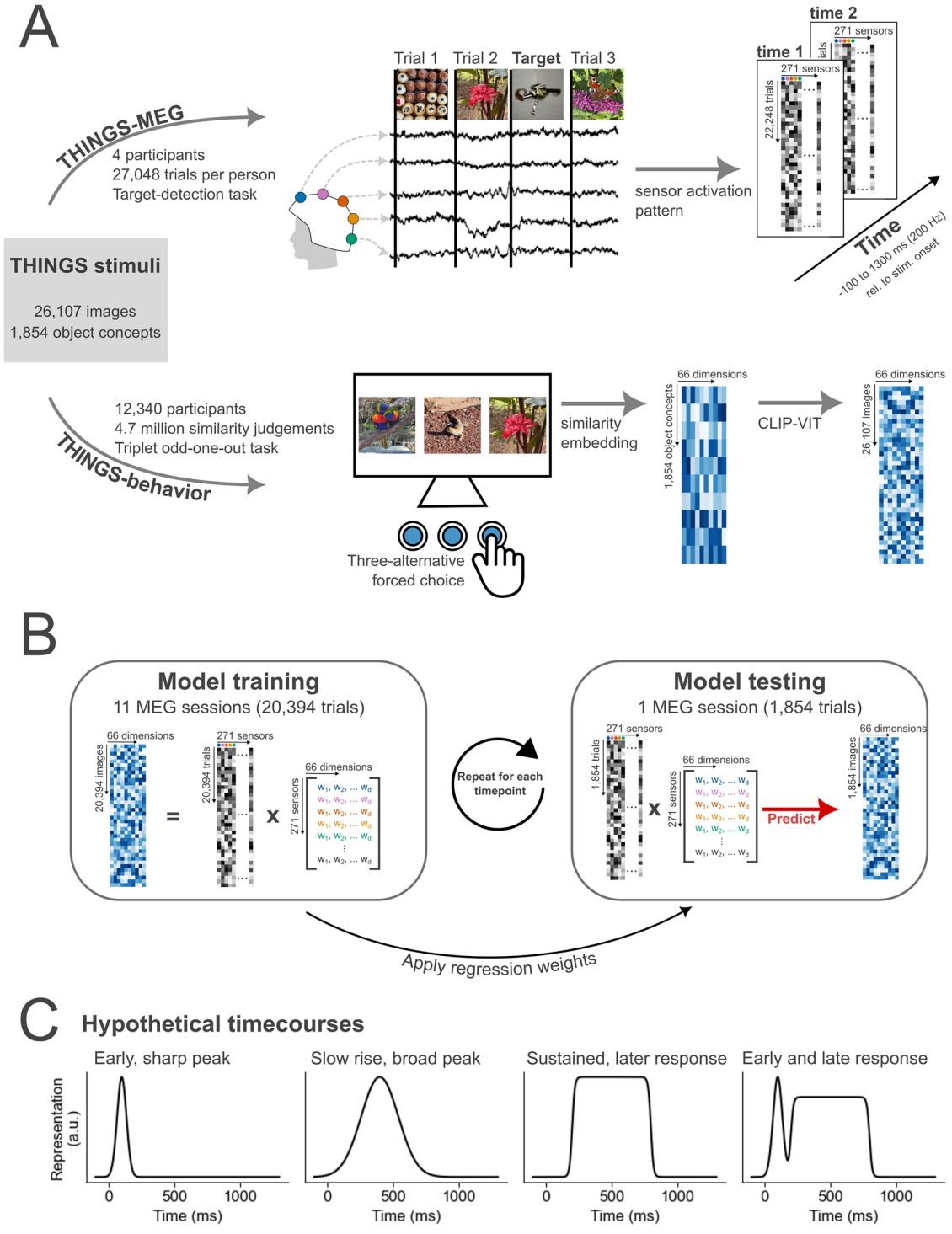

Figure 1. Summary of THINGS-MEG and THINGS-Behavior datasets and the methodological approach to combine them. ***A***, Summary of the datasets used. Evoked responses to images from the THINGS image-database were recorded over time using MEG. In total, four participants completed 12 sessions, resulting in >100,000 trials in total. During the MEG session, participants were asked to detect computer-generated images of non-nameable objects. In the behavioral task, a separate set of participants viewed three objects from the THINGS image-database at the time and were asked to pick the odd-one-out. A computational model was then trained to extract similarities along 66 dimensions for all 1,854 object concepts. The data for behavioral data was crowdsourced via Amazon Mechanical Turk. In total >12,000 participants completed a total of 4.7 million similarity judgements. Using CLIP-ViT, we extended the embedding to capture similarities for all images used in the MEG experiment. Please note that example stimuli shown in this figure are photos that are part of the public domain (Flickr, Author: Lina Teichmann). The original THINGS stimuli were released as part of Hebart et al. (2019) and can be downloaded here https://doi.org/10.17605/OSF.IO/JUM2F. ***B***, Overview of the methodological approach of combining these two datasets with the goal of understanding how multidimensional object properties unfold in the human brain. To train the model, we extract the sensor activation pattern at each timepoint across the MEG sensors and use the behavioral embeddings to learn an association between the two datasets. The linear regression weights are then applied to sensor activation patterns of independent data to predict the behavioral embedding. To evaluate the model's performance, we correlated the predicted and true embedding scores. ***C***, Hypothetical time courses that could be observed for different dimensions.

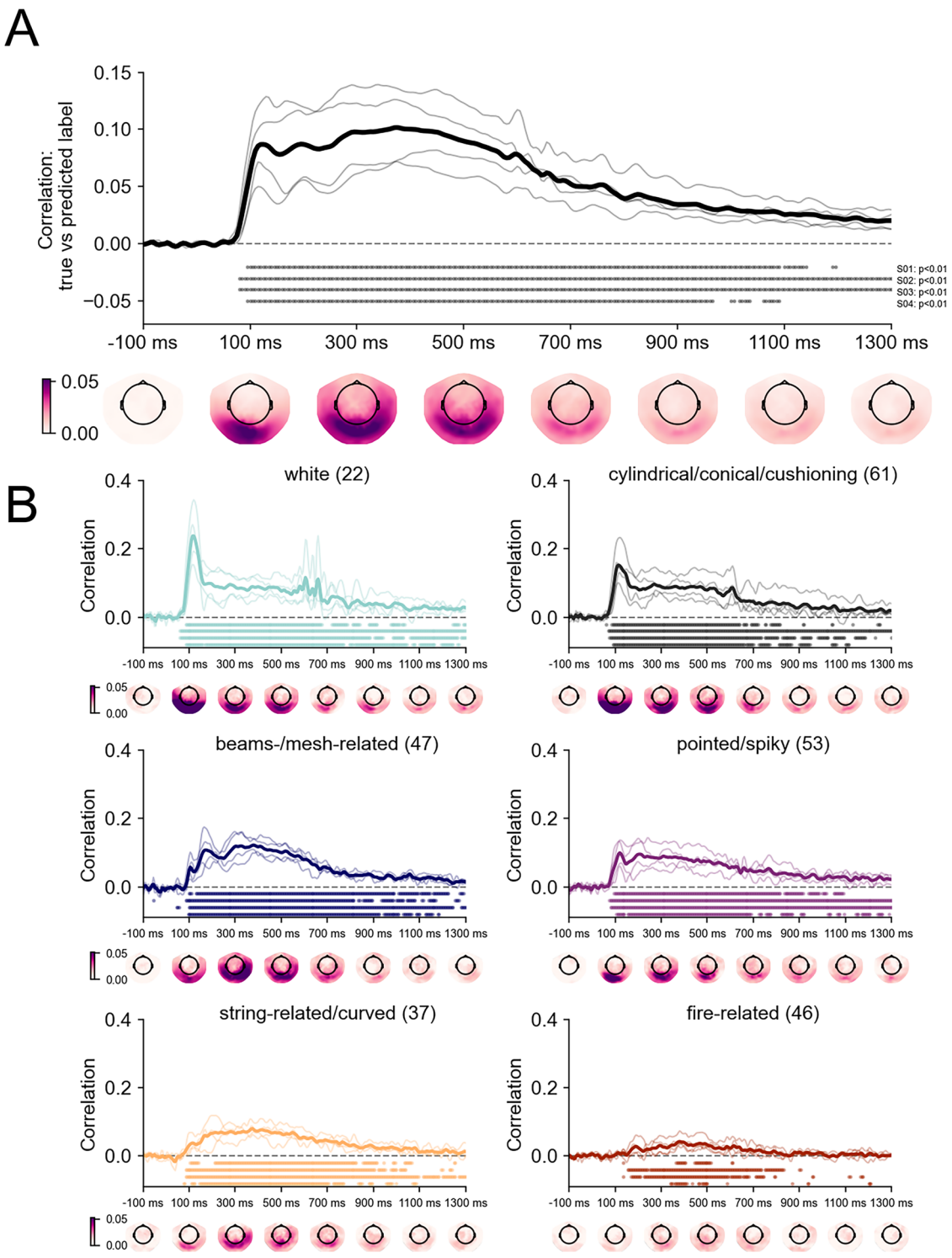

Figure 2. Modeling results for within-participant models of MEG data and multidimensional similarity judgments. ***A***, Correlation between the predicted and true behavioral embeddings across all dimensions over time. The thick, black line shows the average across all participants, and the thin gray lines show individual participants. The dots below the time courses show significant timepoints for each of the four participants at a threshold of $p < 0.01$ established based on individual-level null distributions generated using label permutations (see Materials and Methods). The topographical maps below the time courses show results from the sensor-searchlight regression at single timepoints averaged across all dimensions. Darker colors show a higher correlation between the predicted and true behavioral embedding when the model is fit on a small group of sensors (see Materials and Methods). ***B***, Example time courses for six dimensions. Time courses were first sorted by peak amplitude, and then we picked every eleventh time course to show a representative sample of time courses with different signal-to-noise ratios. The images within each subplot show the six stimuli with the highest weight on that dimension (derived from behavior). Topographical maps below the time courses show results from the sensor-searchlight regression at single timepoints for each of the six dimensions. All individual time courses can be found in the Supplementary Materials (Fig. S3). A dynamic version of the multidimensional unfolding across all 66 dimensions can be accessed here.

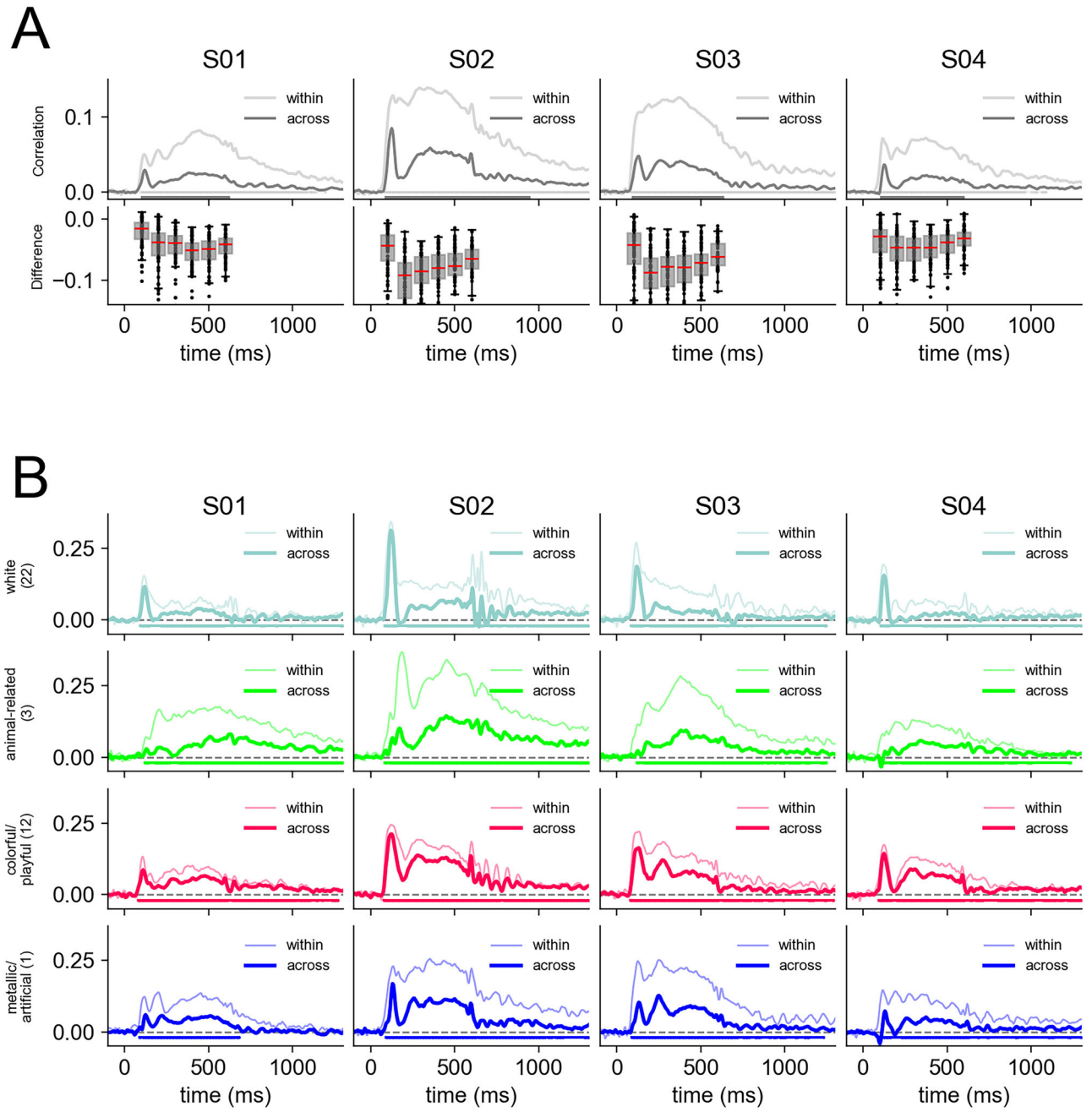

Figure 3. Differences between within- and across-participant model. ***A***, Average performance of the model across all dimensions when fitted as a within-participant model (session-wise cross-validation) and an across-participant model (participant-wise cross-validation). The dots below the time courses show significant timepoints for each of the four participants for the across-participant model at a threshold of $p < 0.01$ established based on null distributions generated using label permutations (see Materials and Methods). The difference between the two models is plotted at select time windows of interest at 100, 200, 300, 400, 500, and 600 ms. Statistics of the pairwise comparisons for each time window are reported in supplementary material (Fig. S4). ***B***, Each subplot shows an example dimension timeseries when the model is fit within each participant (light color) and across different participants (dark color). All other dimension time courses can be found in the supplementary materials (Fig. S5).

magnitude of the differences between the within- and across-participant models was stable after the initial drop from ~250 ms onward (Fig. 3*C*). In addition, the results show that strong stimulus-offset effect at 500–600 ms observed in some dimensions for the within-participant models (e.g., the color dimensions) were also present when the model was trained and tested across participants. Together, these findings suggest that early effects (~125 ms) may carry largely stimulus-specific information that generalizes well across participants, while slightly later effects (~200 ms) are more subject-specific. However, for dimensions that are visually more homogenous (e.g., white, colorful), we found that the within- and across-participant models perform similarly throughout time, including the response associated with stimulus offset.

Peak timing and relative amplitude are prototypical temporal characteristics of different dimension time courses

Comparing the dimension time courses visually suggests some commonalities across dimensions. For example, some dimensions shared a strong early peak and others showed a slower,

gradual rise. To quantify the similarity of time course shapes across dimensions, we used dynamic-time warping (DTW). DTW captures the similarity between a pair of timeseries by assessing how much one of the timeseries has to be warped to resemble another one. The result of this analysis is a time-time-matrix with cost values indicating the amount of warping that has to be done at every timepoint. To measure the similarity of a given timeseries pair, we extracted the sum of the Euclidean distances along the path of lowest cost. If the path falls on the diagonal of the time-time-matrix, it means that the timeseries are identical. If it veers off the diagonal, the timeseries are more dissimilar (Fig. 4*A*,*B*).

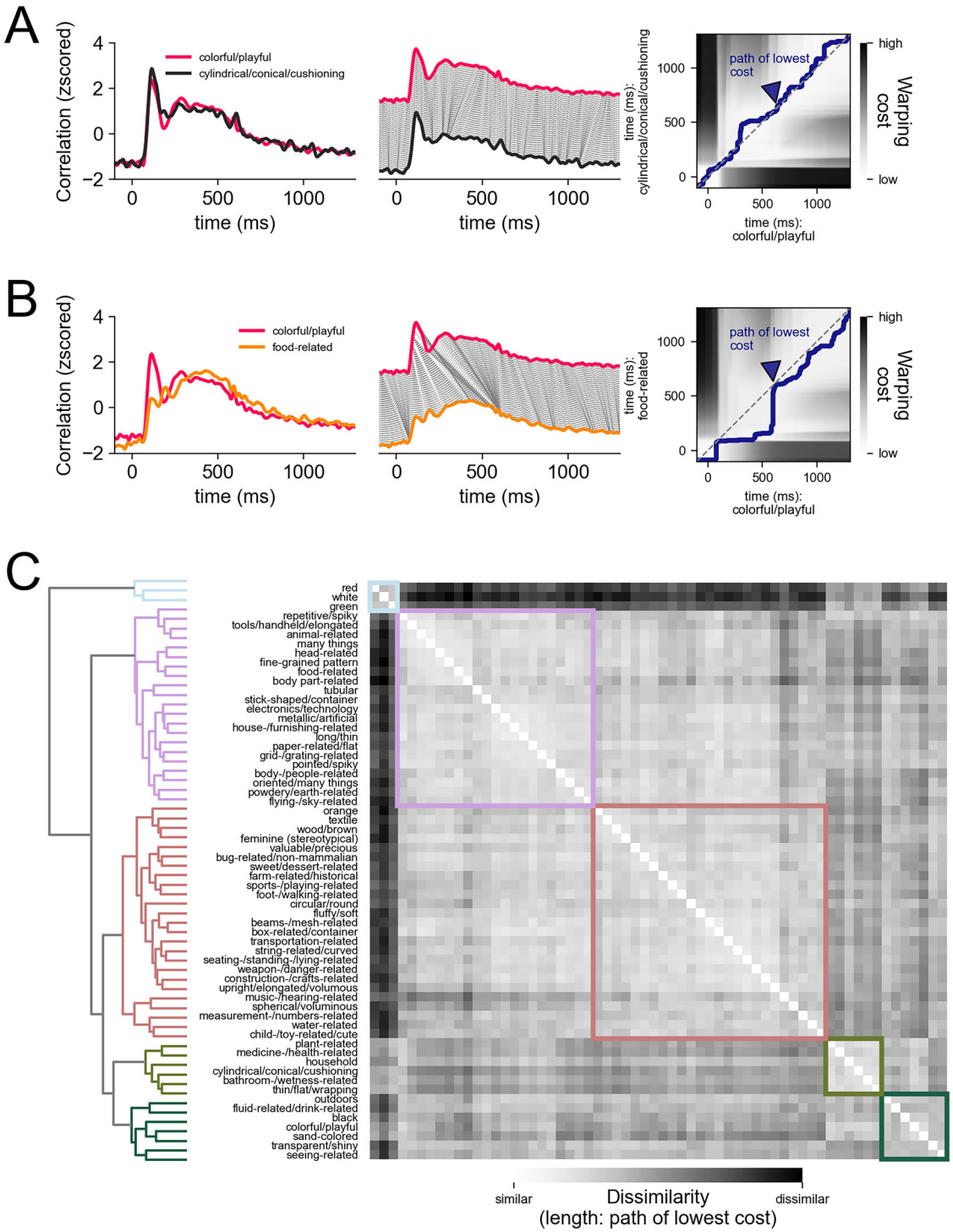

Figure 4. Dynamic time warping (DTW) as a method to compare timeseries similarities. ***A***, The DTW approach for two similar time courses ("colorful" vs "cylindrical"). The correlations were scaled and plotted over time (left panel). DTW assesses how much one timeseries needs to be warped to resemble the other one (middle panel). This warping cost can be established by calculating the Euclidean distance between all timepoints (right panel) to assess dissimilarity. The path of lowest cost describes the path through the matrix that minimizes the warping costs while adhering to some rules (see Materials and Methods). ***B***, The same as ***A*** but for two timeseries that are more dissimilar ("colorful" vs "food-related"). ***C***, Dissimilarity matrix based on the DTW dissimilarity measure (Euclidean distance along the path of lowest cost) for all timeseries pairs. Timeseries with low signal-to-noise ratio were excluded (see Materials and Methods). Hierarchical clustering on this matrix was applied to sort the dimension timeseries accordingly (left).

Applying DTW to our data, we generated a distance matrix for dimension timeseries pairs and ran hierarchical clustering on that matrix to determine which dimensions evoked similar time courses (Fig. 4*C*). Cluster A first separated from all other dimensions. This cluster contained dimensions describing colors ("red," "white," "green"). Next, Cluster B separated from Cluster C, which then separated from D and E. Similar to Cluster A, Clusters D and E contained dimensions describing properties related to color (e.g., "sand-colored," "black," "plant-related," "bathroom-related") as well as other properties such as shape (e.g., "cylindrical/conical/cushioning," "thin/flat"). In contrast, Clusters B and C contained dimensions that are visually more varied (e.g., "animal related," "many things").

After running the clustering, we sorted and averaged the cluster correlations and also plotted the averaged sensor-searchlight results for the dimensions in each cluster to examine prototypical timeseries characteristics (Fig. 5). The primary feature that appeared to distinguish the different timeseries was the relative

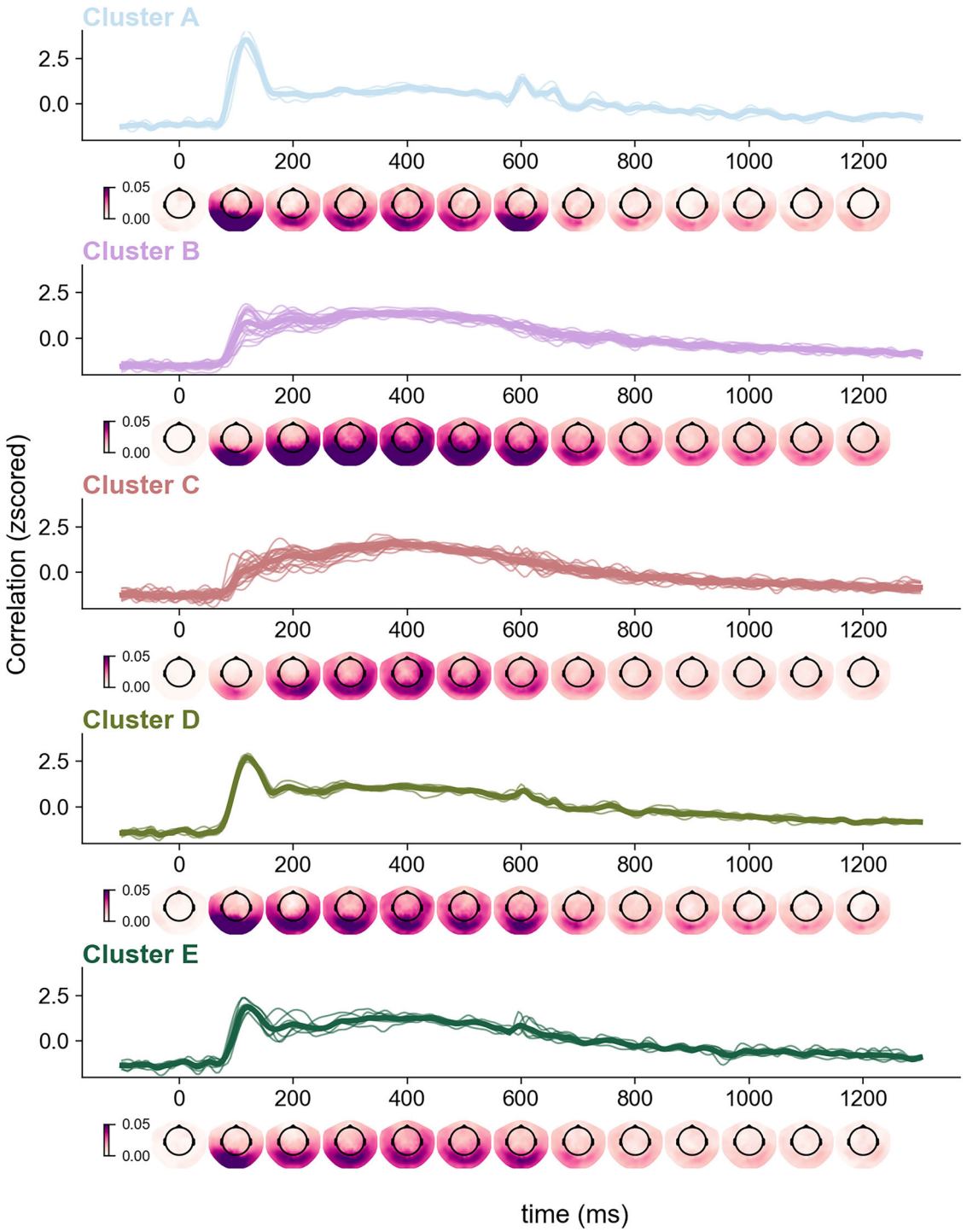

Figure 5. Examining prototypical timeseries characteristics. Every subplot shows the timeseries as well as the averaged sensor-searchlight results for dimensions within each cluster. Thin lines show individual dimension timeseries, and the thicker lines show the averages per cluster.

strength of the early (~125 ms) and late (>200 ms) correlations. The presence of an early peak (clusters A, D, and E) was also often accompanied by a second local peak around the time of stimulus offset. At both early and offset timepoints, posterior sensor clusters had the highest correlations pointing to a spatially localized effect. Dimensions within these three clusters included "red," "green," "thin/flat," "transparent/shiny," and "colorful/playful." These were all dimensions that appear to reflect a specific visual properties (e.g., color, shape) contained within the images. Thus, the early peak and strong offset response might be driven by underlying visual consistencies in objects with high scores on these dimensions.

In contrast, the other clusters (B and C) showed the strongest correlation after 200 ms, with a slow rise to a late and prolonged maximum, but differing in the relative size of the early and late correlations. The effects for both of these clusters were distributed across the sensors. Cluster B had a distinct early peak with posterior sensors being the main driver of this effect while cluster C had lacked an early peak and showed only minimal involvement of the posterior sensors. Dimensions within these clusters included "farm-related," "flying-related," and "body-/people-related." Notably, these clusters accounted for the majority of the dimensions, highlighting the importance of the later response for behaviorally relevant properties.

Overall, the clustering suggests that there are two broad classes of temporal profiles, those with a distinct, early peak and a stimulus offset effect and those with a late peak, often without any clear early peak. The clusters with strong early effects, which showed better generalization across participants, tended to reflect more stimulus-specific information (e.g., green, colorful). In contrast, the clusters with strong late peaks, which showed weaker generalization, appeared to correspond to more conceptual properties, possibly reflecting a greater contribution of subject-specific information.

Discussion

Resolving incoming visual information to make sense of our environment is a challenging task that our brain solves within just a few hundred milliseconds. Here, we used a similarity embedding based on millions of behavioral judgements to model temporally resolved neural responses and showed that behavior-derived dimensions of objects are represented in the brain over time with distinct temporal profiles. Neural dimension time courses tended to group according to the relative strength of two phases of processing (~125 and ~300 ms) as well as the presence or absence of an offset related response (~500–600 ms). Early responses as well as stimulus offset effects were spatially tied to posterior sensors while the second phase of processing at ~300 ms was more distributed spatially. Notably, early effects were more generalizable across participants while later effects were more variable across people. Overall, our work highlights that representations of diverse behavior-derived object dimensions emerge and evolve at different timepoints in the neural signal and contribute to the rich nature of object vision.

Our results build on findings from earlier studies on object vision which often examined how differences between broad object categories (e.g., faces, animals, body parts) arise in the human brain (Liu et al., 2002; Carlson et al., 2013; Clarke et al., 2013; van de Nieuwenhuijzen et al., 2013; Cichy et al., 2014; Rossion, 2014; Goddard et al., 2016; Hebart et al., 2018; Grootswagers et al., 2019). To disentangle what might be driving differences in the neural response, some studies have used stimulus sets with perceptually similar stimuli (e.g., glove and hand) or stimuli that straddle category bounds of object properties such as animacy (e.g., robots; Kaiser et al., 2016; Proklova et al., 2019; Contini et al., 2021). Others have tried to separate the contribution of visual and semantic object properties to the neural signal by using cross-exemplar generalization to determine when we can distinguish objects across different exemplars (Carlson et al., 2013; Bankson et al., 2018; Grootswagers et al., 2019; Teichmann et al., 2020), across object position and size (Isik et al., 2014), or modeling the data using visual and semantic models (Clarke et al., 2015). Here, we substantially extended these studies by using the THINGS-MEG dataset that uses >22,000 stimuli which are systematically and densely sampled from object space. The images were not repeated throughout the experiment, but the breadth of sampling allows us to model the object response in the four MEG participants directly using the crowd-sourced behavioral data. While the sample size is small, systematic and dense sampling of the stimulus space allowed us to avoid common issues such as oversampling individual categories leading to an overestimation of generalization performance (Shirakawa et al., 2024) and selection and category assignment biases (Grootswagers and Robinson, 2021). Thus, the methodological approach we developed here allows us to probe time courses of object vision in a representative and broad manner.

We uncovered different temporal dynamics for different object dimensions of visual object processing and our data-driven approach broadly revealed distinct temporal features. We found that some dimensions had a transient and distinct, early peak <150 ms and an offset response at ~100 ms after stimulus-offset. In contrast, other dimensions lacked the early peak completely or showed it more subtly. These dimensions tended to slowly rise to a later peak (~300 ms) that was more sustained over time. We found that early effects were more consistent across participants than later ones, suggesting that early peaks reflect stimulus-specific and later peaks subject-specific information. Indeed, when looking at which dimensions had a distinct early peak, we found that stimulus-specific visual properties such as dimensions capturing color differences seemed to drive early distinctions. In contrast, later effects seemed to be more associated with concept-related properties, and critically our results suggest that the impact of such dimensions was variable across participants. The variability at later timepoints was accompanied by increased information distribution across MEG sensors in contrast to early timepoints when information was spatially focused on the posterior sensors. This further supports the idea that early timepoints carry primarily stimulus-specific information while later timepoints carry more subject-specific information. It is important to note that the stimulus-specific effects we observed here are not tied to specific exemplars, as every unique image was shown only once, and all analyses were based on cross-exemplar generalizations. More research is needed to examine the origin of the variability of subject-specific information at later timepoints. It is possible that these differences are related to individual differences in the sources or temporal structure of top-down feedback signals or in differences of information content relayed up from visual areas, perhaps reflecting differences in individual experience or knowledge. Importantly, it is unlikely that the differences between early and late timepoints can be explained solely by a time-dependent signal-to-noise difference. In particular, we observed a large difference between the across- and within-participant models at ~200 ms but that the difference decreased right after. In addition, the difference between the two models was stable after 250 ms. Lastly, the stimulus offset effect in

some dimensions at 500–600 ms can be seen in the within- and the across-participant models. All of this speaks against the variability at later timepoints being driven by a time-dependent signal-to-noise difference. Overall, our results highlight that specific temporal profiles are associated with different behavior-derived dimensions but that some broad characteristics can distinguish between stimulus- and subject-specific information.

Our method of using behavior-based embeddings to model the MEG response stands in contrast to previous work that has focused on category or property labels assigned to individual stimuli, which may not capture our broader understanding of objects and the specific properties that may be shared between them (Ritchie et al., 2024). Other groups have addressed this issue with different methods such as modeling the neural data using feature norms (McRae et al., 2005). In one particular MEG study (Clarke et al., 2015), this approach was used to model semantic content of stimuli from several categories which was then contrasted with output from a computational model of object vision. A drawback of this method is that feature norms rely on verbally naming properties which means key visual or conceptual properties may be missed while other properties may be overemphasized. In contrast, our approach extracts object dimensions that are behavior-derived using a visual odd-one-out task which does not require properties to be explicitly nameable. These behavioral embeddings are continuous similarity scores along visual, conceptual, and functional dimensions. One of the strengths of our method is that the MEG data going into each regression model remained the same, but the label value capturing how strongly each image is associated with the dimension at hand differed. This approach is powerful as it makes use of all the data while allowing to study many dimensions simultaneously. Furthermore, it captures the complexities of object vision where objects are associated with many dimensions. The data here show that information related to all 66 behavior-derived dimensions can be read-out from the MEG signal and have specific temporal profiles. Together, our results highlight that modeling neural responses using continuous behavior-derived embeddings offers a more comprehensive understanding of the visual object space than a focus on broader object category or specific dimensions.

Our results here show that using a data-driven approach to define continuous dimension labels successfully reveals different temporal dynamics of the neural object response. However, it is important to note that while this advances our ability to interpret the neural dynamics with more fine-grained distinctions it also comes with constrains. Specifically, the embedding is based on successfully predicting choice behavior in an odd-one-out task of the stimuli used. While the embedding dimensions are diverse, interpretable, and robust (Hebart et al., 2019, 2023), it is important to acknowledge that there are aspects of the object response that may not be captured by this task or by the embedding. Furthermore, the behavioral embedding was derived from a separate set of participants than those from which neural responses were collected. While our data are consistent enough to be generalizable across participants, we find that generalization performance is better for earlier peaks and dimensions that capture perceptually homogenous objects (e.g., red, green, colorful). This may partially reflect the fact that our behavioral embeddings are derived from crowdsourced data and thus may prioritize dimensions that tend to be shared across individuals. Future work should quantify the effect of behavioral demands on the embedding and focus on the individual differences more closely to understand how the object space may be skewed given personal experiences. Although the behavioral embedding may not be sufficient to fully explain the complexities of object vision, it allows us to interpret the neural dynamics associated with object vision beyond category-wise decoding done previously on the M/EEG-THINGS datasets (Gifford et al., 2022; Grootswagers et al., 2022; Hebart et al., 2023).

In conclusion, by using behavioral judgments of similarity to guide our understanding of the neural representation of the object space, we find that different aspects of the object response emerge at different timepoints and together create the experience of meaningful visual object processing.

References

Aghabozorgi S, Shirkhorshidi AS, Wah TY (2015) Time-series clustering–a decade review. Inf Syst 53:16–38.

Bankson BB, Hebart MN, Groen II, Baker CI (2018) The temporal evolution of conceptual object representations revealed through models of behavior, semantics and deep neural networks. Neuroimage 178:172–182.

Carlson TA, Tovar DA, Alink A, Kriegeskorte N (2013) Representational dynamics of object vision: the first 1000ms. J Vis 13:1.

Cichy RM, Pantazis D, Oliva A (2014) Resolving human object recognition in space and time. Nat Neurosci 17:455–462.

Clarke A, Taylor KI, Devereux B, Randall B, Tyler LK (2013) From perception to conception: how meaningful objects are processed over time. Cereb Cortex 23:187–197.

Clarke A, Devereux BJ, Randall B, Tyler LK (2015) Predicting the time course of individual objects with MEG. Cereb Cortex 25:3602–3612.

Contier O, Baker CI, Hebart MN (2024) Distributed representations of behaviour-derived object dimensions in the human visual system. Nat Hum Behav 8:2179–2193.

Contini EW, Goddard E, Wardle SG (2021) Reaction times predict dynamic brain representations measured with MEG for only some object categorisation tasks. Neuropsychologia 151:107687.

Conwell C, Prince JS, Kay KN, Alvarez GA, Konkle T (2024) A large-scale examination of inductive biases shaping high-level visual representation in brains and machines. Nat Commun 15:9383.

Gifford AT, Dwivedi K, Roig G, Cichy RM (2022) A large and rich EEG dataset for modeling human visual object recognition. Neuroimage 264: 119754.

Goddard E, Carlson TA, Dermody N, Woolgar A (2016) Representational dynamics of object recognition: feedforward and feedback information flows. Neuroimage 128:385–397.

Gramfort A, et al. (2013) MEG and EEG data analysis with MNE-Python. Front Neurosci 7:767.

Grootswagers T, Robinson AK, Shatek SM, Carlson TA (2019) Untangling featural and conceptual object representations. Neuroimage 202:116083.

Grootswagers T, Zhou I, Robinson AK, Hebart MN, Carlson TA (2022) Human EEG recordings for 1,854 concepts presented in rapid serial visual presentation streams. Sci Data 9:3.

Grootswagers T, Robinson AK (2021) Overfitting the literature to one set of stimuli and data. Front Hum Neurosci 15:682661.

Hebart MN, Bankson BB, Harel A, Baker CI, Cichy RM (2018) The representational dynamics of task and object processing in humans. Elife 7: e32816.

Hebart MN, Dickter AH, Kidder A, Kwok WY, Corriveau A, Van Wicklin C, Baker CI (2019) THINGS: a database of 1,854 object concepts and more than 26,000 naturalistic object images. PLoS One 14:e0223792.

Hebart MN, Zheng CY, Pereira F, Baker CI (2020) Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. Nat Hum Behav 4:1173–1185.

Hebart MN, Kaniuth P, Perkuhn J (2022) Efficiently-generated object similarity scores predicted from human feature ratings and deep neural network activations. J Vis 22:4057.

Hebart MN, Contier O, Teichmann L, Rockter AH, Zheng CY, Kidder A, Corriveau A, Vaziri-Pashkam M, Baker CI (2023) THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. Elife 12:e82580.

Isik L, Meyers EM, Leibo JZ, Poggio T (2014) The dynamics of invariant object recognition in the human visual system. J Neurophysiol 111:91–102.

Kaiser D, Azzalini DC, Peelen MV (2016) Shape-independent object category responses revealed by MEG and fMRI decoding. J Neurophysiol 115: 2246–2250.

Liu J, Harris A, Kanwisher N (2002) Stages of processing in face perception: an MEG study. Nat Neurosci 5:910–916.

McRae K, Cree GS, Seidenberg MS, McNorgan C (2005) Semantic feature production norms for a large set of living and nonliving things. Behav Res Methods 37:547–559.

Meert W, Hendrickx K, Van Craenendonck T, Robberechts P, Blockeel H, Davis J (2020) DTAIDistance (v2.3.10). Zenodo.

Muttenthaler L, Dippel J, Linhardt L, Vandermeulen RA, Kornblith S (2022) Human alignment of neural network representations. arXiv Preprint arXiv:2211.01201.

Muttenthaler L, Hebart MN (2021) THINGSvision: a Python toolbox for streamlining the extraction of activations from deep neural networks. Front Neuroinform 15:679838.

Pedregosa F, et al. (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830.

Proklova D, Kaiser D, Peelen MV (2019) MEG sensor patterns reflect perceptual but not categorical similarity of animate and inanimate objects. Neuroimage 193:167–177.

Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J (2021) Learning transferable visual models from natural language supervision. International Conference on Machine Learning, pp 8748–8763.

Ritchie JB, Wardle SG, Vaziri-Pashkam M, Kravitz DJ, Baker CI (2024) Rethinking category-selectivity in human visual cortex (No. arXiv:2411.08251). arXiv.

Rossion B (2014) Understanding face perception by means of human electrophysiology. Trends Cogn Sci 18:310–318.

Shirakawa K, Nagano Y, Tanaka M, Aoki SC, Majima K, Muraki Y, Kamitani Y (2024) Spurious reconstruction from brain activity. arXiv Preprint arXiv:2405.10078. https://arxiv.org/abs/2405.10078

Teichmann L, Quek GL, Robinson AK, Grootswagers T, Carlson TA, Rich AN (2020) The influence of object-color knowledge on emerging object representations in the brain. J Neurosci 40:6779–6789.

van de Nieuwenhuijzen ME, Backus AR, Bahramisharif A, Doeller CF, Jensen O, van Gerven MA (2013) MEG-based decoding of the spatiotemporal dynamics of visual category perception. Neuroimage 83:1063–1073.

Wang AY, Kay K, Naselaris T, Tarr MJ, Wehbe L (2023) Better models of human high-level visual cortex emerge from natural language supervision with a large and diverse dataset. Nat Mach Intell 5:1415–1426.