Multidimensional object properties are dynamically represented in the human brain

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

[1]Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda MD, USA
[2]Vision and Computational Cognition Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[3] Department of Medicine, Justus Liebig University, Giessen, Germany

Corresponding author: lina.teichmann@nih.gov

Acknowledgements:
We thank Anna Corriveau, Alexis Kidder, Adam Rockter, and Maryam Vaziri-Pashkam for their help collecting the data. Additional thanks to Tom Holroyd and Jeff Stout for technical support and discussions. We thank Grace Edwards 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, ZIC-MH002968), under National Institute of Mental Health Clinical Study Protocol 93 M-1070 (NCT00001360).

Abstract

Our visual world consists of an immense number of unique objects and yet, we are easily able to identify, distinguish, interact, and reason about the things we see within several hundred milliseconds. This requires that we flexibly integrate and focus on different object properties to support specific behavioral goals. In the current study, we examined how these rich object representations unfold in the human brain by modelling time-resolved MEG signals evoked by viewing thousands of objects. Using millions of behavioral judgments to guide our understanding of the neural representation of the object space, we find distinct temporal profiles across the object dimensions. These profiles fell into two broad types with either a distinct and early peak (~150 ms) or a slow rise to a late peak (~300 ms). Further, the early effects are stable across participants in contrast to later effects which show more variability across people. This highlights that early peaks may carry stimulus-specific and later peaks subject-specific information. Given that the dimensions with early peaks seem to be primarily visual dimensions and those with later peaks more conceptual, our results suggest that conceptual processing is more variable across people. Together, these data provide a comprehensive account of how a variety of object properties unfold in the human brain and contribute to the rich nature of object vision.

Introduction

A core feature of vision is our ability to identify, distinguish, interact, and reason about a huge variety of different objects. What neural mechanisms support such a challenging task in the span of just a few hundred milliseconds? While many studies have focused on broad distinctions between objects (e.g., categories), it is clear that our understanding of objects extends well beyond our ability to label or discriminate them. Here, we used the publicly available THINGS dataset (Hebart, Contier, Teichmann et al., 2023) to determine the temporal dynamics of object vision in the human brain. We combined the large-scale THINGS-MEG and a multidimensional model of object processing derived from millions of behavioral responses to develop a novel analytical approach and gain a comprehensive understanding of the temporal unfolding of object vision.

To understand the temporal dynamics of object vision, many prior studies have compared responses to different types of object stimuli (e.g., faces, scenes). For example, Electroencephalography (EEG) and Magnetoencephalography (MEG) studies have revealed differences in responses to faces and general objects that peak around 170 ms (Liu et al., 2002; Rossion, 2014). More recently, using multivariate and machine learning analyses, studies have further shown that M/EEG signals evoked by broad object categories (e.g., animals, plants, body parts) can be distinguished within the first 200 ms and have focused on what might be driving differences in the neural response (Carlson et al., 2013; Cichy et al., 2014; Clarke et al., 2013; Goddard et al., 2016; Grootswagers et al., 2019; Hebart et al., 2018; van de Nieuwenhuijzen et al., 2013). For example, to disentangle which object features contribute to differences in the neural signal, some studies have used stimulus sets with perceptually similar stimuli (e.g., glove and hand) or stimuli that straddle category bounds (e.g., robots) of object categories such as animacy (Contini et al., 2021; Kaiser et al., 2016; Proklova et al., 2019). Others have tried to separate the contribution of visual and semantic object properties to the neural signal by using cross-exemplar generalization showing when we can distinguish objects using different exemplars (Bankson, Hebart et al., 2018; Carlson et al., 2013), testing for tolerance across object position and size (Isik et al., 2014), or modelling the data using visual and semantic models (Clarke et al., 2015). Overall, the results of these studies suggest that early responses (<150 ms) reflect primarily visual information and later responses reflect more conceptual or semantic information.

While these studies have revealed some general features of the dynamic object response, they are typically based on relatively small, hand-selected sets of stimuli which do not sample the object space in a representative way and cannot adequately capture the richness of the object response (Grootswagers & Robinson, 2021). Hand-selecting sets of stimuli may lead to a sampling bias because some object categories

are considered to be important (e.g., faces, animals) and may be overrepresented while other categories may be completely absent. Our goal in the current study was to provide a much more comprehensive view of object processing by: (1) analyzing responses to tens of thousands of diverse object images, and (2) focusing on the specific behavioral relevance of properties associated with these objects. To do this, we turned to THINGS-data (Hebart, Contier, Teichmann et al., 2023), which contains MEG data for 1,854 objectively-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 in a data-driven manner (e.g., colorful, plant-related, transportation-related) (Hebart et al., 2020, Hebart, Contier, Teichmann et al., 2023). Previous time-resolved analyses of response to the THINGS stimuli (THINGS-MEG: Hebart, Contier, Teichmann et al., 2023) and THINGS-EEG (Grootswagers et al., 2022) have revealed differential neural signals evident within the first 200 ms (Grootswagers et al., 2022; Hebart, Contier, Teichmann et al., 2023) that enable object decoding. As with previous work, however, it is unclear what object properties drive these effects and whether the relative contribution of different properties varies over time.

Here, we developed a novel data-driven approach to uncover the temporal dynamics of object processing, focusing on similarities and differences across the 66 object dimensions extracted from the behavioral data. This approach allows us to study which object dimensions drive the neural object response at every given timepoint. We use the behavioral data to label each of the >22,000 images shown during the MEG data collection. All 66 dimensions can be studied simultaneously as every image has a ranking in all dimensions, making stimulus-selection unnecessary and allowing us to model the relationship between objects directly. In addition, the rankings are continuous along each dimension (e.g., sugar is less colorful than sunflowers, but sunflowers are less colorful than jellybeans), capturing the richness of object vision. Our approach thus goes beyond object categorization, giving us a deeper understanding of visual object processing.

We find distinct temporal profiles for the different object dimensions even though the MEG participants completed a target-detection task and thus did not focus on object properties directly. The temporal profiles tend to group according to the relative strength of two phases of processing (~150 ms and ~300 ms) as well as the presence or absence of an offset related response. Critically, we find that early effects are generalizable across participants while later effects are more variable across people highlighting that stimulus-specific information is reflected in the early parts of the signal while subject-specific information unfolds later. Collectively, our approach focusing on behavioral relevance of object properties provides a comprehensive insight into how the visual object response unfolds in the human brain.

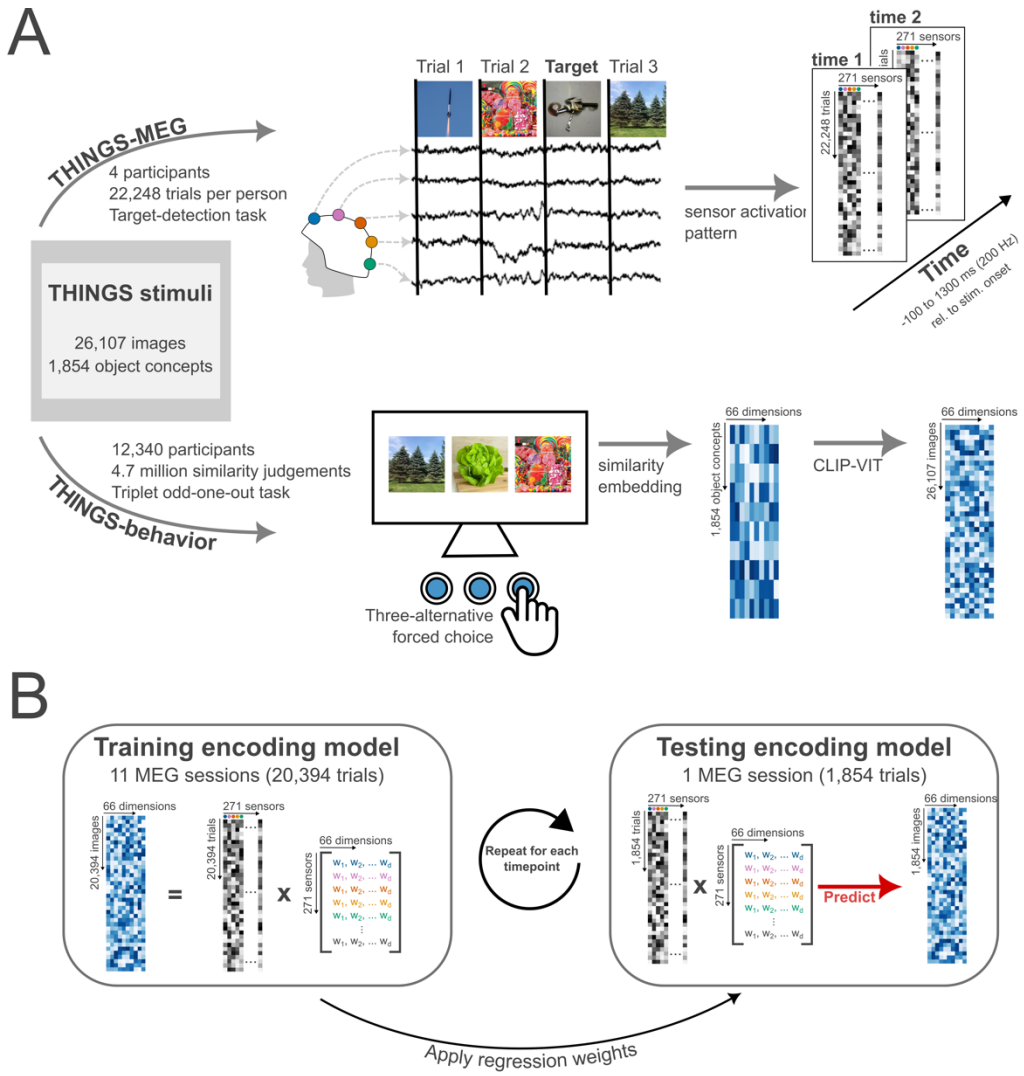


Figure 1. Description of THINGS-MEG and THINGS-Behavior and the methodological approach to combine them. (A) shows a summary of the datasets used. The MEG data was collected using a CTF MEG system with 271 intact sensors. Evoked responses to images from the THINGS image-database were recorded over time. In addition, there were 200 catch images which were computer-generated images of non-nameable object which participants had to detect. In total, four participants completed 12 sessions, resulting in >88,000 experimental trials total. In the behavioral task, Amazon Mechanical Turk 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 object concepts. Using CLIP-VIT we extended the embedding to capture similarities for every image. In total >12,000 participants completed a total of 4.7 million similarity judgements. (B) shows an 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 271 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.

Results

The overarching goal of the current study is to characterize how multidimensional representations unfold over time by combining large-scale MEG data with behaviorally-relevant similarity embeddings. Our primary aims were to determine (a) how behaviorally-relevant multidimensional profiles are reflected in the timecourse of neural responses, (b) how the temporal dynamics vary across people and across dimensions, and (c) what the prototypical temporal characteristics of the responses are.

THINGS-MEG contains evoked neural responses in four participants viewing >22,000 unique natural images associated with 1,854 object concepts. To associate object dimensions with these natural images, we used behavioral embeddings that are based on similarity judgments (Hebart, Contier, Teichmann et al., 2023). The similarity embeddings link each of the 1,854 object concepts with weights along 66 dimensions derived from crowdsourced data of 4.7 million odd-one-out judgments of image triplets. Thus, the stimuli used in the MEG study are associated with concept labels (e.g., nail polish) as well as scores on the behaviorally-relevant dimensions (e.g., colorful) (Hebart et al., 2020). The dimensions cover a broad range of object features with some being strongly linked to visual features (e.g., colorfulness) and others that are linked to more functional or contextual features (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. To overcome this issue, we used an artificial neural network (CLIP-VIT, Radford et al., 2021) to generate image-level embeddings (Hebart et al., 2022).

We used the scores on the 66 behaviorally-relevant dimensions to model the evoked neural response recorded with MEG. In particular, we used a linear encoding model 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 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 re-ordering according to behaviorally-relevant dimensional profiles.

(a) How are behaviorally-relevant multidimensional profiles reflected in the timecourse of neural responses?

We fitted multiple linear regression models to learn the association between the MEG sensor activation patterns and the behavioral embeddings at every timepoint (Figure 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 testing set, we predicted the score along each dimension from the MEG sensor activation patterns. Correlating these predicted scores with the true behavioral dimensional profiles results in timeseries for all four participants showing that there is behaviorally-relevant multidimensional information in the neural signal from ~100 ms onwards (Figure 2A). To examine which MEG sensors tend to drive the effects, we also fitted linear models to predict the activation of each sensor separately. We find that posterior sensors have the strongest effect across time, however the relative contribution of frontal sensors increases later in time (>150 ms). Together, these results indicate that the neural signal can be modelled using continuous, behaviorally-relevant dimensions and that the information is distributed across sensors.

Extracting the correlation timecourses for each dimension (Figure 2A rose plots, Supplementary Figure S1), we can see that different dimensions unfold in the neural signal in a different way. For example, dimensions such as “plant-related”, “colorful/playful”, “white” have distinct, early peaks (~150 ms) whereas other dimensions such as “body-/people-related”, “food-related”, and “transportation-related” show a slower rise to a later peak (~300 ms). In addition, we find that several of the dimensions driving distinct early peaks have a stimulus offset effect at ~500 ms. In contrast, there are several other dimensions that do not show a distinct early peak or offset response but unfold slowly over time and rise to a late peak (>300 ms). While signal to noise ratio differs, all dimension timecourses exceed zero at some point over time (see Figure 2B for representative example timecourses selected based on peak amplitude and Supplementary Figure 1 for all timecourses). Remarkably, the behaviorally-relevant dimensional profiles are evident in the neural signal although the MEG participants completed an orthogonal task. This demonstrates that the dimensions exist 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 behaviorally-relevant dimensions are reflected in distinct temporal profiles.

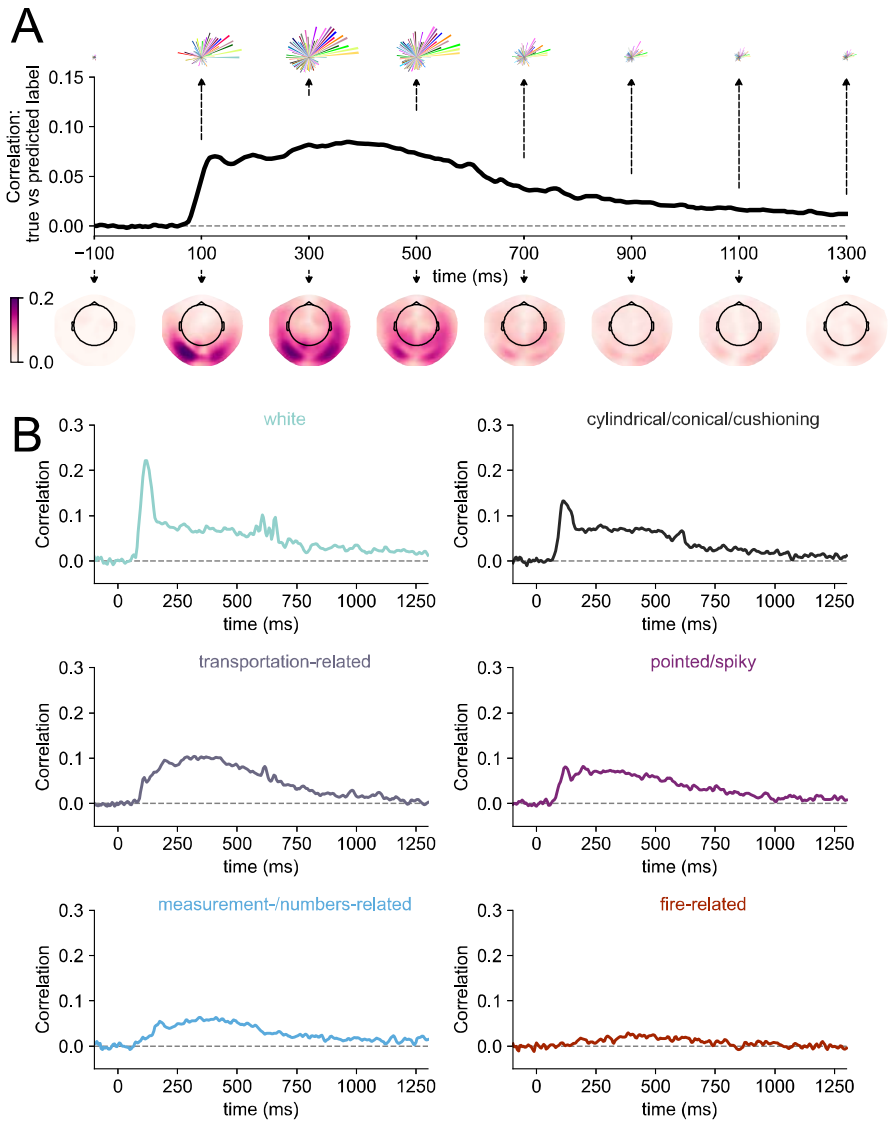
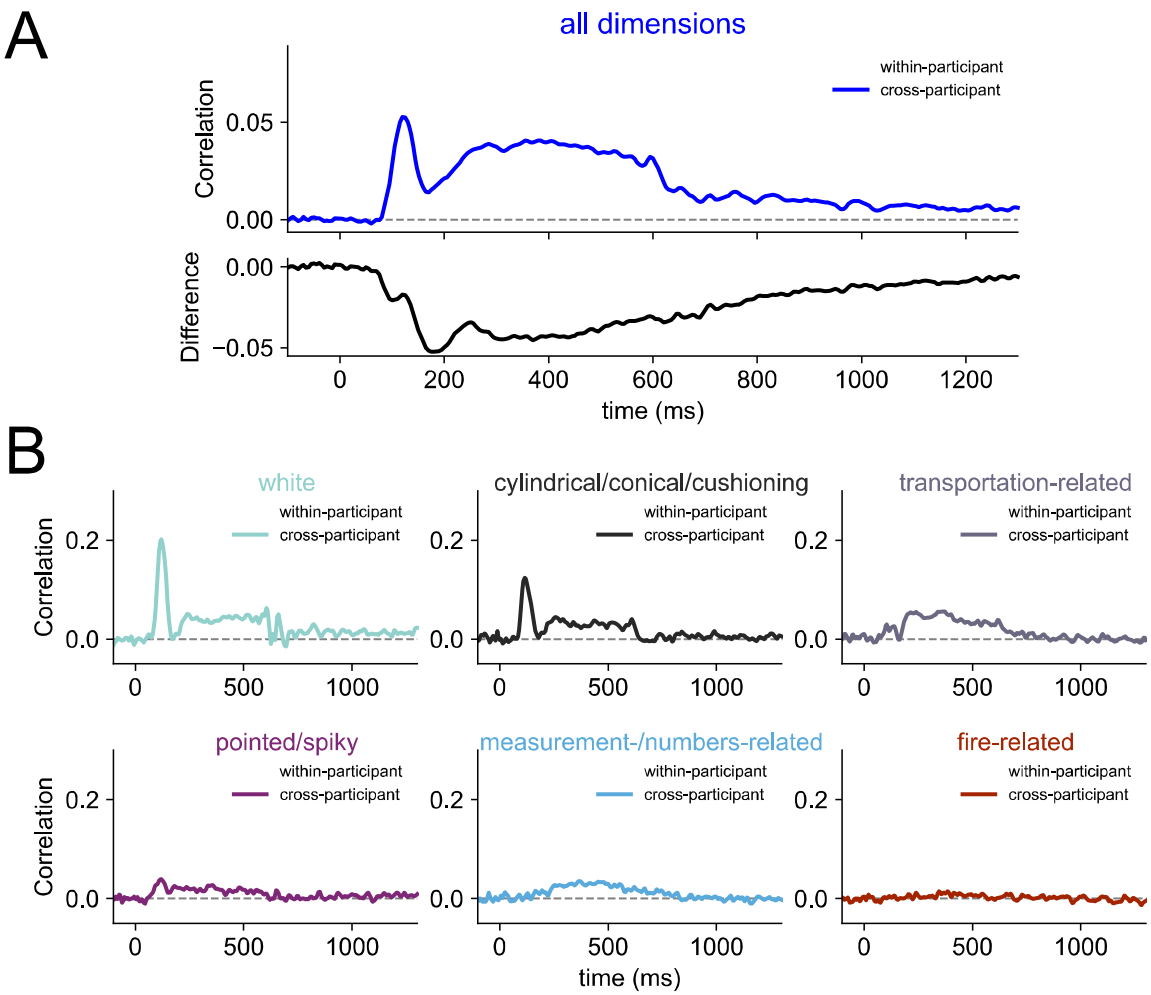


Figure 2. Modelling results for within-participant models of MEG data and multidimensional similarity judgments. (A) shows the correlation between the predicted and true behavioral embeddings across all dimensions over time. The thick, black line shows the average across all participants, the thin grey lines show individual participants. The rose plots above the timeseries show snapshots of the model performance across all 66 dimensions. The order of the petals is based on maximum peak amplitude. Longer petals indicate that there is more information associated with a given dimension in the signal at this timepoint. A dynamic version of these plots can be accessed here. The topographical maps below the timecourses show snapshots of the model performance when it is fitted to individual sensors. Darker colors show a higher correlation between the predicted and true weights at each sensor location. A dynamic version of the topographical plots can be accessed here. (B) shows example timecourses for six timeseries. Timecourses were first sorted by peak amplitude and then we picked every 11th timecourse to show a representative sample of timecourses with different signal-to-noise ratios. All individual timecourses can be found in the Supplementary Materials (Figure S1).

(b) How do the temporal dynamics vary across people and dimensions?

To test whether the different temporal profiles for different object properties can be consistently observed across participants, we used again multiple linear regression models (see (a)), but this time with a subject-based cross-validation scheme. Specifically, we trained the model on MEG data from three participants and tested its performance on the data from the remaining one. We find that the cross-participant models perform similarly to the within-participant model (Figure 3), highlighting that the temporal profiles we uncovered for each dimension are robust. In particular, early peaks (<150 ms) were consistent in amplitude and timing when comparing the within- and the cross-participant model. In contrast, later effects did not generalize as well across participants. This indicates that early effects carry stimulus-specific information while later ones carry subject-specific information. In general, being able to extract temporal profiles for different object properties across participants highlights that the effects are distributed and not sensor-specific, as head location in the MEG differs across participants.


A
all dimensions
within-participant
cross-participant
Correlation
0.05
0.00
Difference
0.00
−0.05
0
200
400
600
800
1000
1200
time (ms)
B
white
cylindrical/conical/cushioning
transportation-related
pointed/spiky
measurement-/numbers-related
fire-related
Correlation
0.2
0.0
500
1000
time (ms)

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 (Figure 4A).

Applying DTW to our data, we generated a distance matrix for all dimension timeseries pairs and ran k-means clustering on that matrix to determine which dimensions evoke similar brain activation patterns (Figure 4B). To determine the optimal number of k-clusters, we calculated the gap statistic (see supplementary) and used k=5 for our data. After running the clustering, we sorted and averaged the cluster correlations to examine prototypical timeseries characteristics. We found that the presence or absence of a strong, early peak is a key feature of timeseries dimensions such as “red”, “green”, “thin/flat”, and “colorful/playful”. For these dimensions we also observed a stronger offset effect after stimulus disappearance at ~500 ms. The early peak and strong offset response presumably reflect the underlying visual consistencies in objects with high scores on these dimensions. Other dimensions such as “farm-related”, “flying-related”, and “body-/people-related” show a more gradual, later rise to a prolonged maximum before tapering off. While similar timeseries shapes have been found in previous studies focusing on comparing stimulus difference along a single dimension (e.g., comparing colors), it is important to note the different temporal profiles here are evoked by a re-ordering of the same stimuli according to behavioral similarity embeddings. This highlights that our approach allows us to extract meaningful information across dimensions simultaneously and offers novel opportunities to study several stimulus properties using the same data. 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 slow, gradual rise to a late peak. The early effects tend to carry more stimulus-specific information while the later effects seem to correspond closer to subject-specific information.

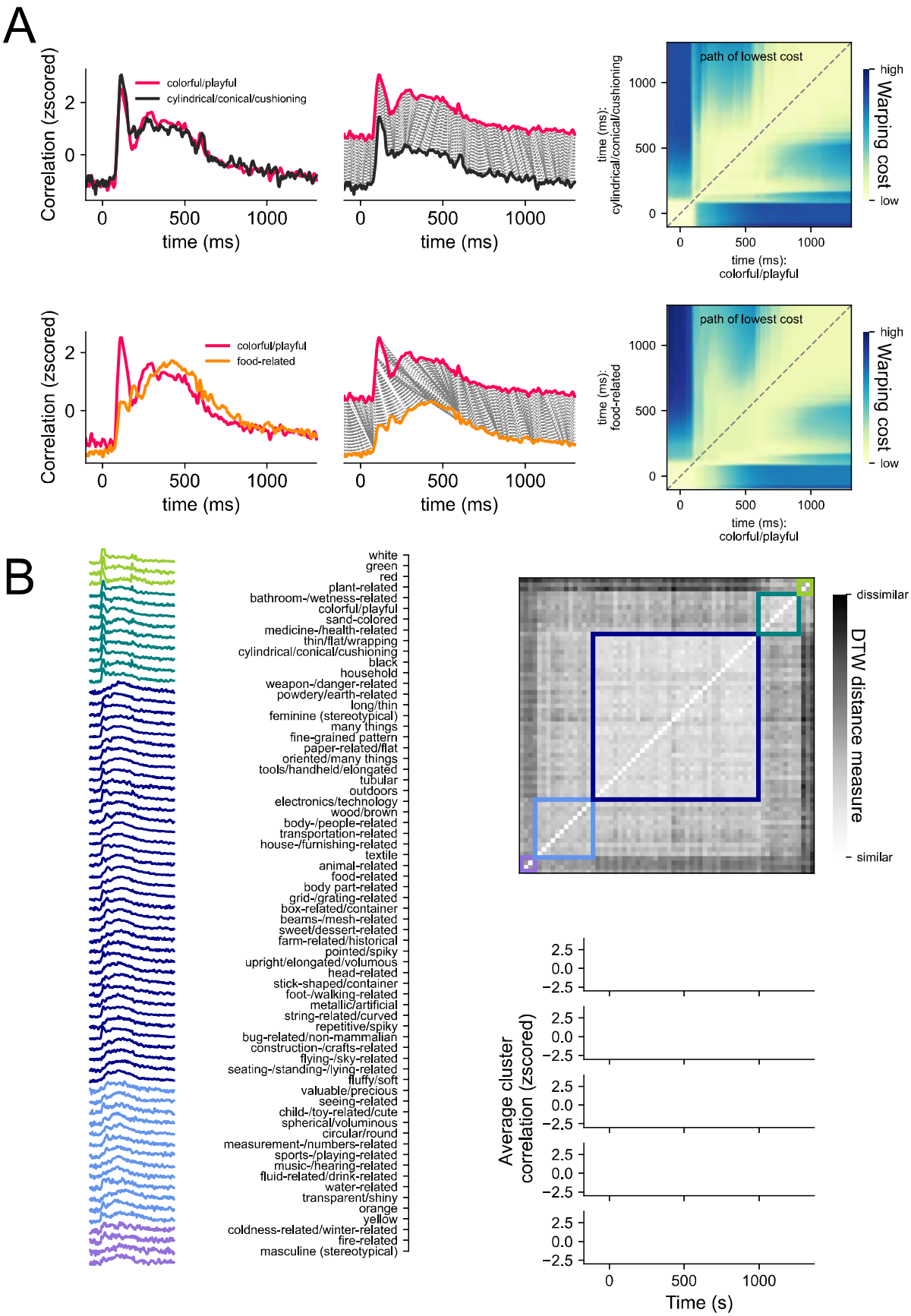

A
Correlation (zscored)
colorful/playful
cylindrical/conical/cushioning
time (ms)
path of lowest cost
time (ms): cylindrical/conical/cushioning
time (ms): colorful/playful
high
Warping cost
low
food-related
time (ms): food-related
B
white
green
red
plant-related
bathroom-/wetness-related
colorful/playful
sand-colored
medicine-/health-related
thin/flat/wrapping
cylindrical/conical/cushioning
black
household
weapon-/danger-related
powdery/earth-related
long/thin
feminine (stereotypical)
many things
fine-grained pattern
paper-related/flat
oriented/many things
tools/handheld/elongated
tubular
outdoors
electronics/technology
wood/brown
body-/people-related
transportation-related
house-/furnishing-related
textile
animal-related
food-related
body part-related
grid-/grating-related
box-related/container
beams-/mesh-related
sweet/dessert-related
farm-related/historical
pointed/spiky
upright/elongated/voluminous
head-related
stick-shaped/container
foot-/walking-related
metallic/artificial
string-related/curved
repetitive/spiky
bug-related/non-mammalian
construction-/crafts-related
flying-/sky-related
seating-/standing-/lying-related
fluffy/soft
valuable/precious
seeing-related
child-/toy-related/cute
spherical/voluminous
circular/round
measurement-/numbers-related
sports-/playing-related
music-/hearing-related
fluid-related/drink-related
water-related
transparent/shiny
orange
yellow
coldness-related/winter-related
fire-related
masculine (stereotypical)
dissimilar
DTW distance measure
similar
Average cluster correlation (zscored)
Time (s)

Figure 4. Dynamic time warping (DTW) as a method to compare timeseries similarities and extract prototypical timeseries characteristics. (A) shows the DTW approach for two pairs of timeseries. The correlations were scaled and plotted over time (left panel). Using DTW means we are assessing 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 (top right panel). The path of lowest cost describes the path through the matrix that minimizes the warping costs while adhering to some rules (see methods). To summarize the warping cost in a single number, we summed the Euclidean distances along the path as a similarity measure. (B) shows the similarity matrix containing the DTW similarity measures for all possible timeseries pairs (right panel). Running k-means clustering on this matrix allows us to sort the dimension timeseries (left panel) and dimensions labels (middle panel). Averaging the timeseries for each cluster (bottom right panel) allows to examine prototypical timeseries characteristics.

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 the neural data in order to understand how behaviorally-relevant object dimensions are represented in the brain over time. We find that individual object dimensions are directly reflected in the neural response with distinct temporal profiles for different dimensions. In particular, while some dimensions show strong representation shortly after stimulus onset resulting in a pronounced early peak in the timeseries, other dimensions show a slower rise to a later peak. Critically, we find that the early effects are highly similar across individuals allowing for cross-participant generalization. In contrast, later effects tend to be inconsistent across people with more limited generalization. Collectively, these results suggest that the early peak may correspond to representation of physical stimulus properties while the later response may reflect subject-specific conceptual information. Overall, the current work highlights that behaviorally-relevant object properties emerge and evolve at different timepoints in the neural signal and contribute to the rich nature of object vision.

Our data-driven approach focused on behaviorally-relevant object dimensions contrasts with more typical category or feature-driven approaches. First, prior work has often used small, hand-selected stimulus sets (e.g., Bankson, Hebart et al., 2018; Carlson et al., 2013; Cichy et al., 2014; Grootswagers et al., 2018; Kriegeskorte et al., 2008; Teichmann et al., 2020) with findings that may be tied closely to the specific stimuli chosen (Grootswagers & Robinson, 2021). While larger stimulus sets potentially avoid this problem, it is often unclear how the stimuli were sampled and biases within the stimulus set may still constrain the results. For example, the COCO stimuli used in the

large-scale MRI Natural Scenes Dataset (Allen et al., 2022) contain a disproportionately large number of giraffes. Second, the analyses in prior work are often focused on category or feature labels assigned to individual stimuli, which ignore our broader understanding of objects and the specific properties that may be shared between different objects. One alternative approach used in prior work is to model the neural data using feature norms (e.g., McRae et al., 2005). In one particular MEG study (Clarke et al., 2015), this approach was used to model semantic content of stimuli from only 11 categories which was then contrasted with output from a computational model of object vision. In contrast to our approach which rely on properties that are behaviorally-relevant to distinguish between objects, feature norms rely on verbally naming properties potentially missing key features. Similarly, visual models such as HMax are based on very basic visual features such as orientation or frequency ignoring complex visual features such as material properties (e.g., Schmid et al., 2023; Schmidt et al., 2022) or object color (e.g., Rosenthal et al., 2018) which are all relevant in visual object processing.

Our approach overcomes these issues by using the >22,000 images of the THINGS database which come with a behavioral embedding derived from 4.7 million similarity judgements. Using this dataset means reduced bias in stimulus selection and category assignment as the THINGS concepts are selected systematically (Hebart et al., 2019) and the behavioral embeddings are data-driven, relying on crowdsourced similarity judgements (Hebart et al., 2020, Hebart, Contier, Teichmann et al., 2023). Instead of analyzing image-specific effects that reflect one category or another, we model the neural data for all object images in terms of continuous similarity scores along 66 dimensions. That means the MEG data going into the analysis always remain the same, but the ranking differs depending on dimension. 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 properties. The data here show that information related to all 66 behaviorally-relevant dimensions can be read-out from the MEG signal and have distinct temporal profiles. Together, our results highlight that modelling neural data using continuous behaviorally-relevant embeddings offers a more comprehensive understanding of the visual object space.

We found different temporal dynamics for different object dimensions of visual object processing with a data-driven approach that broadly show two distinct temporal characteristics. We observe that some dimensions have a transient and distinct, early peak <150 ms and a stimulus-offset response at ~500 ms. In contrast, other dimensions lack the early peak completely or show it more subtly. These dimensions tend to slowly rise to a later peak (~300 ms) which is more sustained over time. We find that early

effects are 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 have a distinct early peak, we find that stimulus-specific visual properties such as color drive early distinctions. In contrast, later effects seem to be more associated with concept-related properties and critically our results suggest that the impact of such properties is variable across subjects. It is important to note that the stimulus-specific effects we observe here are not tied to specific exemplars as every unique image was shown only once and all analyses are based on cross-exemplar generalizations. Previous work has used cross-exemplar generalization as a method to disentangle visual and conceptual object properties (e.g., Bankson, Hebart et al., 2018; Carlson et al., 2013), however, this approach does not allow us to distinguish which object properties drive the effects at different timepoints. Our approach uses multidimensional behavioral embeddings and can therefore tease these differences apart by showing which behaviorally-relevant property contributes to the neural signal at a given timepoint. Overall, the results highlight that distinct temporal profiles are associated with different behaviorally-relevant dimensions but that some broad characteristics can distinguish between stimulus- and subject-specific information.

One limitation of our work is the lack of subject-specific behavioral embeddings to model the neural data. While our data is consistent enough to generalizable across participants, we find that generalization performance is better for earlier peaks. This may partially be the case because our behavioral embeddings are derived from crowdsourced data and thus may prioritize dimensions that tend to be shared across individuals. Another drawback of the crowdsourced data is that we do not have any insights into what drove the decision making in the triplet task. That means the dimension labels may not accurately reflect the similarity dimension for every participant. Future work should investigate individual differences more closely to understand how the object space may be skewed given personal experience and the task at hand.

In conclusion, our work provides novel insights into how multidimensional object representations unfold in the human brain. Using behavior 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.

Methods

Dataset

We used the publicly available THINGS dataset (Hebart, Contier, Teichmann, et al., 2022) which contains densely sampled MEG data as well as crowdsourced behavioral data. The MEG portion of the data contains neural recordings from four participants who each viewed a total of 27,048 unique images over the course of a 12-sessions. Every image was shown centrally for 500 ms with an inter-stimulus interval of 800-1200 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). Unique image exemplars for each of the concepts 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 behaviorally-relevant 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, Contier, Teichmann 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. 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 started by examining the activity patterns in the final layer of the image encoder for each of the 1,854 objects. We then used a ridge regression to predict dimension weights for each of the 66 dimensions for all images in the THINGS database. To model the evoked neural response measured with MEG, we then used the image-level predicted weights along the 66 dimensions.

Preprocessing

Our preprocessing pipeline was built using *mne-python* (Gramfort et al., 2013) and described in detail in the dataset release (Hebart, Contier, Teichmann, et al., 2022). The preprocessing steps included filtering (1Hz – 40Hz), baseline correction using z-scoring and epoching the data from -100 to 1300 ms relative to stimulus onset. Preprocessed data can be directly downloaded from OpenNeuro (https://openneuro.org/datasets/ds004212/versions/2.0.0).

Analyses

Modelling MEG data based on multidimensional similarity judgments: within-participant regression

To model how the behaviorally-relevant dimensions unfold over time in the human brain, we fitted a multiple linear regression model 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 sci-kit learn linear regression models with default parameters (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. We ran 10,000 iterations of the model fit with permuted weights in each dimension, to establish a 95% confidence interval representing chance model performance.

To gain insights into which sensors primarily drive the effects, 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.

(b) Examining timecourse similarities across people: Cross-participant regression

To examine whether timecourse profiles are consistent across participants, we also trained a model with a participant-wise cross validation scheme. We trained the model to learn the association between the multivariate MEG sensor activation pattern at every timepoint and the behavioral dimension profiles using data from three of the four participants. We then tested its performance using the data from the left-out participant. We repeated this process until every participant's data were used as testing data once. A separate model was trained and tested at every timepoint. Models were fit in Python using *sci-kit learn* linear regression models with default parameters (Pedregosa et al., 2011).

(c) Examining timecourse 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 all of our pairwise timecourse comparisons to generate 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. Running clustering on the resulting DTW distance matrix allowed us to establish a qualitative measure of different prototypical timeseries characteristics. We used k-means clustering and determined the optimal number of k-clusters by calculating the gap statistic (Tibshirani et al., 2001) for n clusters ranging from n=2 to n=30 (see Supplementary Figure S3). As there was no clear peak in the gap statistic across cluster sizes, we settled on n=5 clusters, as this is when the change in gap statistic decreased substantially.

Open Science Practices

All data is publicly available under a Creative Commons license and can be downloaded from OpenNeuro: https://openneuro.org/datasets/ds004212/versions/2.0.0 *[this repository does not contain all aggregate behavioral data to run the analyses yet. It will be populated and re-uploaded as version 3.0.0 as soon as the review process is finalized].* The analysis code for all analyses in this paper are available on GitHub: https://github.com/Section-on-Learning-and-Plasticity/THINGS-MEG.

References

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., & Charest, I. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, *25*(1), 116–126.

Bankson, B. B., Hebart, M. N., Groen, I. I., & Baker, C. I. (2018). The temporal evolution of conceptual object representations revealed through models of behavior, semantics and deep neural networks. *NeuroImage*, *178*, 172–182.

Carlson, T. A., 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.

Clarke, A., Devereux, B. J., Randall, B., & Tyler, L. K. (2015). Predicting the Time Course of Individual Objects with MEG. *Cerebral Cortex*, *25*(10), 3602–3612. https://doi.org/10.1093/cercor/bhu203

Clarke, A., Taylor, K. I., Devereux, B., Randall, B., & Tyler, L. K. (2013). From perception to conception: How meaningful objects are processed over time. *Cerebral Cortex*, *23*(1), 187–197.

Contini, E. W., Goddard, E., & Wardle, S. G. (2021). Reaction times predict dynamic brain representations measured with MEG for only some object categorisation tasks. *Neuropsychologia*, *151*, 107687.

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

Grootswagers, T., Cichy, R. M., & Carlson, T. A. (2018). Finding decodable information that can be read out in behaviour. *NeuroImage*, *179*, 252–262.

Grootswagers, T., & Robinson, A. K. (2021). Overfitting the literature to one set of stimuli and data. *Frontiers in Human Neuroscience*, *15*, 682661.

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

Grootswagers, T., Zhou, I., Robinson, A. K., Hebart, M. N., & Carlson, T. A. (2022). Human EEG recordings for 1,854 concepts presented in rapid serial visual presentation streams. *Scientific Data*, *9*(1), 3.

Hebart, M. N., Bankson, B. B., Harel, A., Baker, C. I., & Cichy, R. M. (2018). The representational dynamics of task and object processing in humans. *Elife*, *7*, e32816.

Hebart, M. N., Contier, O., Teichmann, L., Rockter, A. H., Zheng, C. Y., Kidder, A., Corriveau, A., Vaziri-Pashkam, M., & Baker, C. I. (2023). THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. *Elife*, *12*, e82580.

Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Van Wicklin, C., & Baker, C. I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PloS One*, *14*(10), e0223792.

Hebart, M. N., Kaniuth, P., & Perkuhn, J. (2022). Efficiently-generated object similarity scores predicted from human feature ratings and deep neural network activations. *Journal of Vision*, *22*(14), 4057–4057.

Hebart, M. N., Zheng, C. Y., Pereira, F., & Baker, C. I. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nature Human Behaviour*, *4*(11), 1173–1185.

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.

Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016). Shape-independent object category responses revealed by MEG and fMRI decoding. *Journal of Neurophysiology*, *115*(4), 2246–2250.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*(9), 910–916.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559.

Muttenthaler, L., Dippel, J., Linhardt, L., Vandermeulen, R. A., & Kornblith, S. (2022). Human alignment of neural network representations. *ArXiv Preprint ArXiv:2211.01201*.

Proklova, D., Kaiser, D., & Peelen, M. V. (2019). MEG sensor patterns reflect perceptual but not categorical similarity of animate and inanimate objects. *NeuroImage*, *193*, 167–177.

Radford, A., Kim, J. W., 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*, 8748–8763.

Rosenthal, I., Ratnasingam, S., Haile, T., Eastman, S., Fuller-Deets, J., & Conway, B. R. (2018). Color statistics of objects, and color tuning of object cortex in macaque monkey. *Journal of Vision*, *18*(11), 1–1.

Rossion, B. (2014). Understanding face perception by means of human electrophysiology. *Trends in Cognitive Sciences*, *18*(6), 310–318.

Schmid, A. C., Barla, P., & Doerschner, K. (2023). Material category determined by specular reflection structure mediates the processing of image features for perceived gloss. *Nature Human Behaviour*.

Schmidt, F., Hebart, M. N., & Fleming, R. W. (2022). *Core dimensions of human material perception*.

Teichmann, L., Quek, G. L., Robinson, A. K., Grootswagers, T., Carlson, T. A., & Rich, A. N. (2020). The influence of object-color knowledge on emerging object representations in the brain. *Journal of Neuroscience*, *40*(35), 6779–6789.

van de Nieuwenhuijzen, M. E., Backus, A. R., Bahramisharif, A., Doeller, C. F., Jensen, O., & van Gerven, M. A. (2013). MEG-based decoding of the spatiotemporal dynamics of visual category perception. *Neuroimage*, *83*, 1063–1073.