Title

THINGS-EEG: Human electroencephalography recordings for 1,854 concepts presented in rapid serial visual presentation streams

Authors

Tijl Grootswagers[1,2], Ivy Zhou[2], Amanda K. Robinson[2], Martin N. Hebart[3], Thomas A. Carlson[2]

Affiliations

1. The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia
2. School of Psychology, The University of Sydney, Sydney, Australia
3. Vision and Computational Cognition Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

corresponding author: Tijl Grootswagers (t.grootswagers@westernsydney.edu.au)

Abstract

The neural basis of object recognition and semantic knowledge have been the focus of a large body of research but given the high dimensionality of object space, it is challenging to develop an overarching theory on how brain organises object knowledge. To help understand how the brain allows us to recognise, categorise, and represent objects and object categories, there is a growing interest in using large-scale image databases for neuroimaging experiments. Traditional image databases are based on manually selected object concepts and often single images per concept. In contrast, 'big data' stimulus sets typically consist of images that can vary significantly in quality and may be biased in content. To address this issue, recent work developed THINGS: a large stimulus set of 1,854 object concepts and 26,107 associated images. In the current paper, we present THINGS-EEG, a dataset containing human electroencephalography responses from 50 subjects to all concepts and 22,248 images in the THINGS stimulus set. The THINGS-EEG dataset provides neuroimaging recordings to a systematic collection of objects and concepts and can therefore support a wide array of research to understand visual object processing in the human brain.

Background & Summary

Humans are able to visually recognise and meaningfully interact with a large number of different objects, despite drastic changes in retinal projection, lighting or viewing angle, and the objects being positioned in cluttered visual environments. Object recognition and semantic knowledge, our ability to make sense of the objects around us, have been the subject of a large amount of cognitive neuroscience research[1–3]. However, previous neuroimaging research in this field has often relied on a manual selection of a small set of images[1,4,5]. In contrast, recent developments in computer vision have produced very large image sets for training artificial intelligence, but the individual images in these sets are minimally curated and therefore make them often unsuitable for research in psychology and neuroscience. To overcome these issues, recent work has created large, curated image sets that are designed for studying the cognitive and neural basis of human vision[5,6]. One of these is THINGS[5], which is an image set containing 1,854 object concepts representing the most frequent concepts used in the English language, accompanied by 26,107 associated manually-curated high-quality image exemplars and human behavioural annotations. This rich collection of stimuli and behavioural data has already been used to study the core representational dimensions underlying human similarity judgements[7]. The next phase is to collate corresponding neural responses to stimuli in THINGS. This would contribute to an emerging landscape of large datasets of neural responses to curated image sets that accelerate research in visual, computational, and cognitive neuroscience[8].

Collecting neurophysiological data for the THINGS dataset, with over 26,000 images, is unachievable in a traditional neuroimaging experiment: Typically, classic object vision experiments present around one image per second [e.g., 9–11]. Collecting one trial for each image in THINGS would take more than seven hours, which is infeasible to achieve with a single session and would require a complex design involving multiple scanning sessions. However, we have recently shown that it is possible to uncover detailed information about visual stimuli presented in rapid serial visual presentation (RSVP) streams using electroencephalography (EEG)[12–15]. In one of these studies[12], participants viewed over 16,000 visual object presentations at 5 and 20 images per second, in a single 40-minute EEG session. Results from multivariate pattern classification and representational similarity analysis revealed detailed temporal dynamics of object processing that were similar to work that used slower presentation speeds (around one image per second). Therefore, fast presentation paradigms are highly suitable for collecting neural responses to the large number of visual object stimuli in the THINGS database.

Here, we recorded human brain responses of 50 participants to all 1,854 object concepts in THINGS using 22,248 of the THINGS stimuli (12 images per concept). Participants were also presented with a separate validation set of 200 images that were repeated 20 times. Here, we describe the THINGS-EEG dataset and recording procedure, and present initial technical validation analyses. The THINGS-EEG dataset presents a rich resource of time-varying neural recordings that we hope will be of great value for studying the temporal dynamics of human object processing.

Methods

50 individuals volunteered to take part in the experiment in return for course credit. This comprised 36 females and 14 males, mean age 20.44 (sd 2.72), age range 17 – 30. Participants had different language profiles, with 26 native English speakers, 24 non-native speakers, 24 monolinguals, and 25 bilinguals. All participants reported normal or corrected-to-normal vision. There are 4 participants marked for potential exclusion due to notably poor signal quality or equipment failure (marked in the *participants.tsv* file). These participants are included in the release for completeness. The study was approved by the University of Sydney ethics committee. Informed consent was obtained from all participants at the start of the experiment.

Stimuli were obtained from the THINGS database[5] (Figure 1A). THINGS contains 1,854 objects concepts, with 12 or more images per concept. The first 12 images for each concept were used for this experiment, resulting in 22,248 different visual images, which we divided into 72 sequences of 309 stimuli. Individual concepts were never repeated within one sequence. To be able to assess within- and between-subject variance on the same images, we presented another 200 validation images from the THINGS database 12 times to every subject after the main part of the experiment. For every subject we used the same 200 validation images (listed in the *test_images.csv* file). We repeated these images in 12 sequences of 200, where we presented them in random order. The experiment thus contained 84 sequences in total. The subjects were not explicitly made aware of the two different parts.

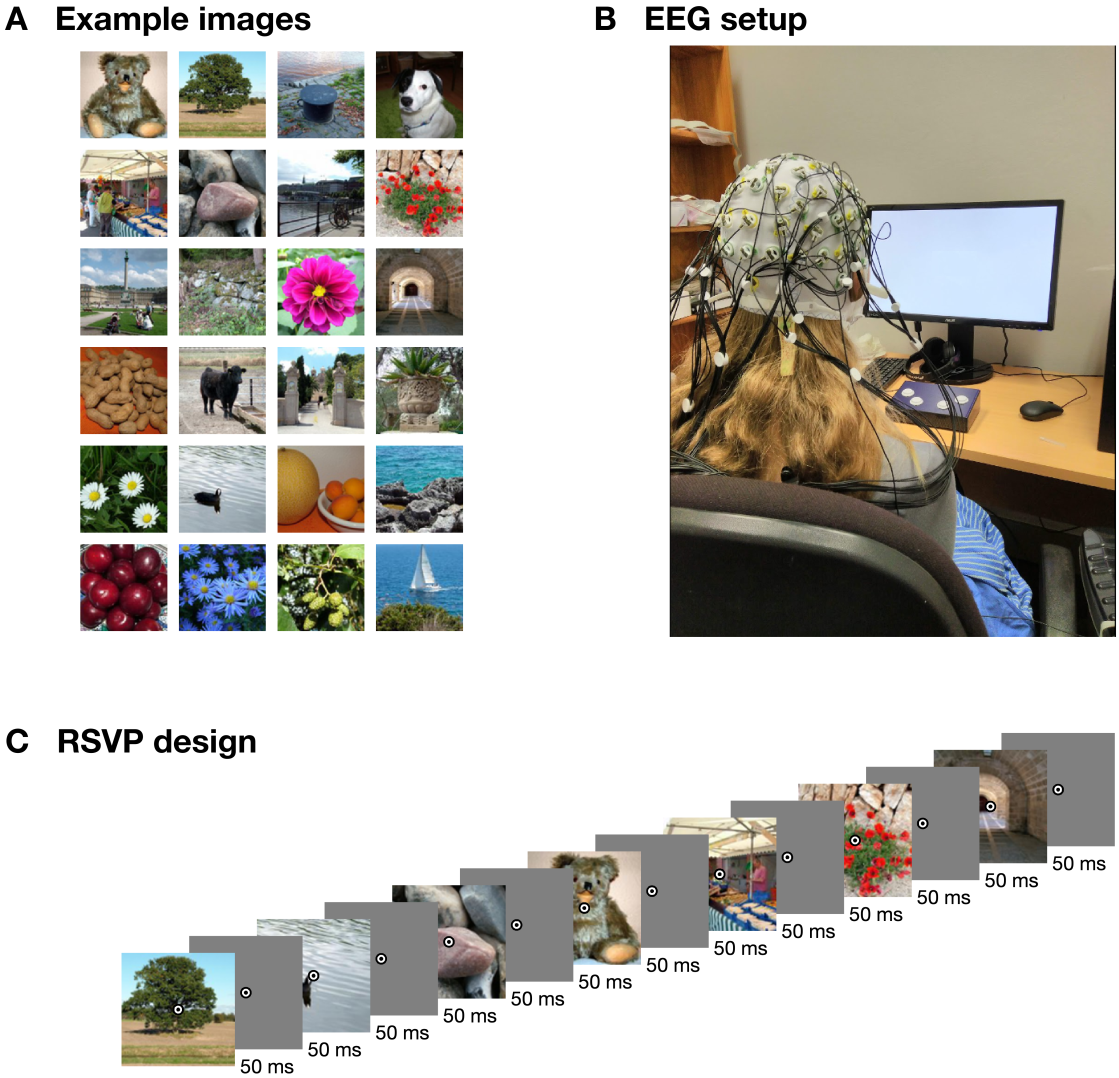

Figure 1: Example stimuli, design, and EEG setup image. A) Example images similar to the stimuli used in the experiment. B) EEG experimental setup (photo credit AKR). C) Rapid serial visual presentation design. For illustration purposes, only part of the sequence is shown. For this figure, all images were replaced by public domain images with similar appearance (obtained from PublicDomainPictures.net: Brunhilde Reinig).

The experiment was programmed in Python, using the Psychopy[16] library. The sequences were presented at 10Hz, with a 50% duty cycle (Figure 1C). That is, each image was presented for 50ms, followed by a 50ms blank screen. Participants were seated about 57cm from the screen, and the stimuli subtended approximately 10 degrees visual angle. Overlaid at the centre of each image was a bullseye (0.5 degrees visual angle) to help participants maintain fixation. To increase attention and engagement, each sequence contained 2 to 5 random target events, where the bullseye turned red for 100ms, and the participants were instructed to press a button on a button box. At the end of each sequence, the display showed the progress through the experiment, and participants were able to start the next sequence with a button press. Participants were asked to sit still and minimise eye movements during the sequences and to use the time between sequences as breaks and relax, and start the next sequence using a button press when they were ready. The experiment lasted around one hour.

We used a BrainVision ActiChamp system to record continuous data while participants viewed the sequences (Figure 1B). Conductive gel was used to reduce impedance at each electrode site below 10 kOhm where possible. The median electrode impedance was under 18 kOhm in

40/50 participants and under 60 kOhm in all participants. We used 64 electrodes, arranged according to the international standard 10–10 system for electrode placement[17,18]. The signal was digitised at a 1000-Hz sample rate with a resolution of 0.0488281µV. Electrodes were referenced online to Cz. An event trigger was sent over the parallel port at the start of each sequence, and at every stimulus onset and offset event.

To perform basic quality checks and technical validation, for each subject, we ran a standard decoding analysis. We decoded pairwise images for the 200 validation images, and we created the full time-varying 1,854×1,854 Representational Dissimilarity Matrix[19,20] reflecting the pairwise decoding accuracies between all 1,854 object concepts. We used a minimal preprocessing pipeline derived from our previous RSVP-MVPA studies[12–15]. Using Matlab and the EEGlab toolbox[21], data were filtered using a Hamming windowed FIR filter with 0.1Hz highpass and 100Hz lowpass filters, re-referenced to the average reference, and downsampled to 250Hz. Epochs were created for each individual stimulus presentation ranging from [-100 to 1000ms] relative to stimulus onset. No further preprocessing steps were applied (e.g., baseline correction or epoch rejection), as in our previous work using similar presentation paradigms[12–15]. The channel voltages at each time point served as input to the decoding analysis.

Decoding analyses were performed in Matlab using the CoSMoMVPA toolbox[22]. We first decoded between the 200 validation images. For a given pairs of images, we used a leave-one-sequence out (total: 12 sequences) cross-validation procedure and trained a regularised (λ=0.01) linear discriminant classifier to distinguish between the images. The mean classification accuracy on the image pair in the left-out sequences were stored in a 200×200×275×50 (image×image×time point×subject) RDM, which is symmetrical across its first diagonal. A similar procedure was performed for the main experiment, using the 1,854 different image concepts. This resulted in an 1,854×1,854×275×50 (concept×concept×time point×subject) RDM. For the 200 validation images, we also computed noise ceilings by comparing between subject RDMs, as described in previous work[23]. The noise ceilings estimate the lower and upper bound of the highest achievable performance of a model that attempts to explain variance in the data.

Data & Code Availability

All data and code are publicly available from the Open Science Framework (https://osf.io/hd6zk/). The raw EEG recordings are hosted on Figshare in BIDS[24,25] format (https://doi.org/10.6084/m9.figshare.14721282). The preprocessed (Matlab/EEGLAB[21] format), and epoched (Matlab/CoSMoMVPA[22] format) data are included for convenience. The code to reproduce the technical validation analyses and figures presented in this manuscript, as well as the RDMs for the full set and the RDMs for the validation images (Matlab format) are also available from the Open Science Framework (https://osf.io/hd6zk/), which also contains links to the above repositories.

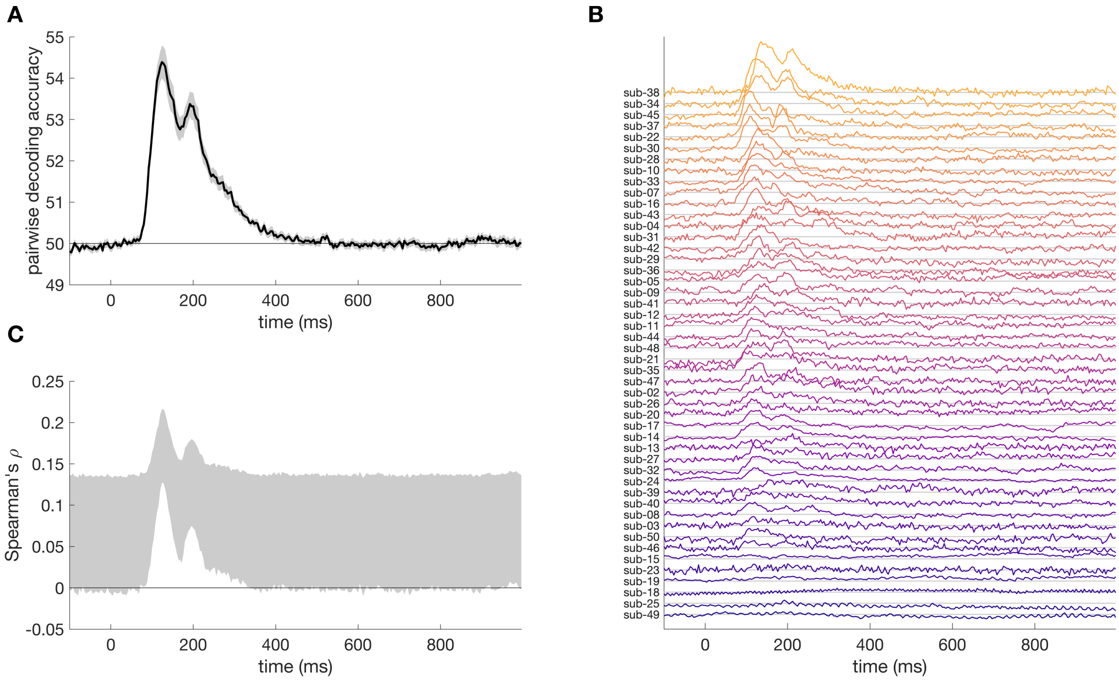

Figure 2: Results for the 200 validation images that were repeated 12 times at the end of the session. A) mean pairwise classification accuracy over time. B) Mean pairwise decoding over time, per subject, sorted by peak classification accuracy. Subjects 1 and 6 are not shown as they did not have data on the validation images. C) Noise ceiling over time shows the expected correlation of the 'true' model with the RDMs of the validation images and reflects the between-subject variance in the RDMs.

Technical Validation Analyses

We computed representational dissimilarity matrices for the 200 validation images, by calculating time-varying decoding accuracy between all pairs of images. Mean, subject-wise decoding accuracy (Figure 2A) showed an initial peak around 100ms, and a second, lower peak around 200ms after stimulus onset. The shape of the time-varying decoding was similar to previous object decoding studies[e.g., 9,10], and was also similar to previous results on images presented in fast succession[12–14], indicating data quality was similar to these studies. For most subjects, this shape was apparent from their individual data (Figure 2B). The noise ceiling (Figure 2C) indicates an average similarity (correlation) of up to 0.2 between the subject-specific dissimilarity matrices.

Next, we computed the full RDM for all 1,854 images concepts (1,717,731 pairs). The average accuracy within this RDM (Figure 3A) was lower than for the validation images, which is likely due to the fact that accuracy reflects generalised concept-similarity across images. Figure 3B shows the full RDM at one time point (200ms). To test if the values in the RDM contain meaningful information, we computed the correlation between the full RDM and four example categorical models (Figure 3C). The models coded for the presence of a certain category (e.g., animal). Figure 3C shows each model reaches an above-zero correlation, with the 'natural' model reaching the highest correlation, around 200ms.

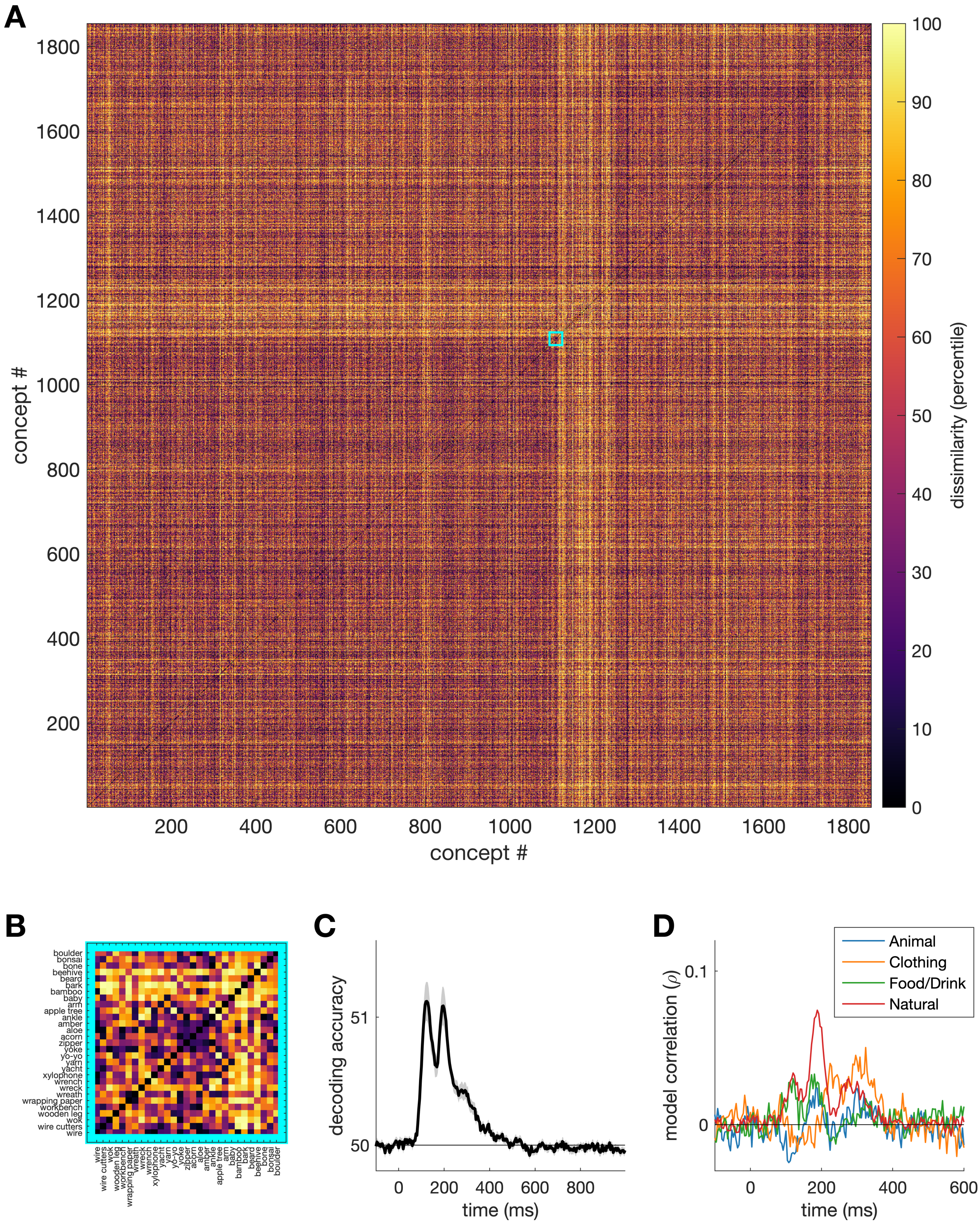

Figure 3: Results for the 1,854 image concepts that were repeated 12 times (using a different image each time). A) Full 1,854×1,854 RDM at 200ms, arranged by high-level category. B) Zoomed in section of the full RDM. C) Mean pairwise classification accuracy between concepts over time. D) Correlation over time between the neural RDM and four high-level categorical models.

Discussion

Here, we presented THINGS-EEG, a dataset of human electro-encephalography responses to 22248 images from all 1,854 concepts in the THINGS object database. In the main session, each image was repeated once, and in the second part, 200 validation images were repeated 12 times each to be able to assess data quality and compare the data to future datasets acquired with other modalities. In total, 26,248 visual images were presented in a 1-hour EEG session. This was achieved using a rapid serial visual processing paradigm. Technical validation results indicated the dataset contains detailed neural responses to images, which shows that the dataset is a high-quality resource for future investigations into the neurobiology of visual object recognition.

In this study, we presented over 25,000 trials in a 1-hour EEG experiment. While this is an exceptionally large number in a visual object perception study, the paradigm has several limitations. Firstly, by presenting the images in rapid succession at 10Hz, new information is being presented while previous trials are still being processed. While our previous work has shown that a great amount of detail about objects can be extracted from brain responses to RSVP streams[12–15], the images are being forward and backward masked, and therefore the data does not capture the full brain response to each image[13]. For example, cognitive functions such as memories or emotions may not have enough time to be instantiated at such rapid presentation rates. Our design also involved one presentation per image, which makes image-specific analyses challenging, placing the focus of this work at the level of the 1,854 object concepts. The benefit of this is that the data has a built-in control for image-level confounds. For example, visual regularities that are specific to an image will not affect the analysis at the concept level.

The THINGS-EEG dataset has a lot of potential for investigating the neurobiology of visual object recognition and semantic knowledge. The dataset presents a very large set of non-invasive neural responses to visual stimuli in human participants. We foresee many possible uses of this dataset. For example, the dataset could be used to test models of visual object representation, such as different semantic models, or deep neural networks. The rapid presentation paradigm allows to examine sequential effects in the data, such as how a specific object concept influences the encoding of the subsequent presentations. Finally, the data could be used to test the generalisability of previous studies that were limited by small stimulus sets[4]. In sum, as THINGS is a high-quality stimulus set of record size, THINGS-EEG accompanies this resource with a comprehensive set of human neuroimaging recordings.

Acknowledgements

This research was supported by ARC DP160101300 (TAC), ARC DP200101787 (TAC), and ARC DE200101159 (AKR). The authors acknowledge the University of Sydney HPC service for providing High Performance Computing resources.

Author contributions

TG: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Data Curation, Writing – Original Draft, Writing – Review & Editing, Supervision, Project administration.
IZ: Conceptualization, Investigation, Data Curation, Writing – Review & Editing, Project administration.
AKR: Conceptualization, Methodology, Writing – Review & Editing.
MNH: Conceptualization, Methodology, Writing – Review & Editing.
TAC: Conceptualization, Methodology, Writing – Review & Editing, Supervision, Funding acquisition, Project administration.

Competing interests

The authors declare no competing interests.

References

1. Wardle, S. G. & Baker, C. Recent advances in understanding object recognition in the human brain: deep neural networks, temporal dynamics, and context. *F1000Research* **9**, 590 (2020).
2. DiCarlo, J. J. & Cox, D. D. Untangling invariant object recognition. *Trends Cogn. Sci.* **11**, 333–341 (2007).
3. Gauthier, I. & Tarr, M. J. Visual Object Recognition: Do We (Finally) Know More Now Than We Did? *http://dx.doi.org/10.1146/annurev-vision-111815-114621* http://www.annualreviews.org/doi/10.1146/annurev-vision-111815-114621 (2016).
4. Grootswagers, T. & Robinson, A. K. Overfitting the literature to one set of stimuli and data. *ArXiv210209729 Q-Bio* (2021).
5. Hebart, M. N. *et al.* THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLOS ONE* **14**, e0223792 (2019).
6. Mehrer, J., Spoerer, C. J., Jones, E. C., Kriegeskorte, N. & Kietzmann, T. C. An ecologically motivated image dataset for deep learning yields better models of human vision. *Proc. Natl. Acad. Sci.* **118**, (2021).
7. Hebart, M. N., Zheng, C. Y., Pereira, F. & Baker, C. I. Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nat. Hum. Behav.* **4**, 1173–1185 (2020).
8. Naselaris, T., Allen, E. & Kay, K. Extensive sampling for complete models of individual brains. *Curr. Opin. Behav. Sci.* **40**, 45–51 (2021).
9. Carlson, T. A., Tovar, D. A., Alink, A. & Kriegeskorte, N. Representational dynamics of object vision: The first 1000 ms. *J. Vis.* **13**, 1 (2013).
10. Cichy, R. M., Pantazis, D. & Oliva, A. Resolving human object recognition in space and time. *Nat. Neurosci.* **17**, 455–462 (2014).
11. Kaneshiro, B., Guimaraes, M. P., Kim, H.-S., Norcia, A. M. & Suppes, P. A Representational Similarity Analysis of the Dynamics of Object Processing Using Single-Trial EEG Classification. *PLOS ONE* **10**, e0135697 (2015).
12. Grootswagers, T., Robinson, A. K. & Carlson, T. A. The representational dynamics of visual objects in rapid serial visual processing streams. *NeuroImage* **188**, 668–679 (2019).
13. Robinson, A. K., Grootswagers, T. & Carlson, T. A. The influence of image masking on object representations during rapid serial visual presentation. *NeuroImage* **197**, 224–231 (2019).
14. Grootswagers, T., Robinson, A. K., Shatek, S. M. & Carlson, T. A. Untangling featural and conceptual object representations. *NeuroImage* **202**, 116083 (2019).
15. Grootswagers, T., Robinson, A. K., Shatek, S. M. & Carlson, T. A. The neural dynamics underlying prioritisation of task-relevant information. *Neurons Behav. Data Anal. Theory* **5**, 1–17 (2021).
16. Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods* **51**, 195–203 (2019).
17. Jasper, H. H. The ten-twenty electrode system of the International Federation. *Electroencephalogr Clin Neurophysiol* **10**, 371–375 (1958).
18. Oostenveld, R. & Praamstra, P. The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* **112**, 713–719 (2001).
19. Grootswagers, T., Wardle, S. G. & Carlson, T. A. Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *J. Cogn. Neurosci.* **29**, 677–697 (2017).

20. Kriegeskorte, N., Mur, M. & Bandettini, P. A. Representational Similarity Analysis - Connecting the Branches of Systems Neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).

21. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).

22. Oosterhof, N. N., Connolly, A. C. & Haxby, J. V. CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Front. Neuroinformatics* **10**, (2016).

23. Nili, H. *et al.* A Toolbox for Representational Similarity Analysis. *PLoS Comput Biol* **10**, e1003553 (2014).

24. Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* **3**, 160044 (2016).

25. Pernet, C. R. *et al.* EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Sci. Data* **6**, 103 (2019).