



Research



Cite this article: Wang G, Chen L, Cichy RM, Kaiser D. 2025 Enhanced and idiosyncratic neural representations of personally typical scenes. *Proc. R. Soc. B* **292**: 20250272.

<https://doi.org/10.1098/rspb.2025.0272>

Received: 31 January 2025

Accepted: 26 February 2025

Subject Category:

Neuroscience and cognition

Subject Areas:

behaviour, cognition, neuroscience

Keywords:

scene perception, typicality, individual differences, drawing, electroencephalography, deep neural networks

Author for correspondence:

Gongting Wang

e-mail: generalwgt@gmail.com

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.7702797>.

Enhanced and idiosyncratic neural representations of personally typical scenes

Gongting Wang^{1,2}, Lixiang Chen^{1,2}, Radoslaw Martin Cichy¹ and Daniel Kaiser^{2,3}

¹Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

²Department of Mathematics and Computer Science, Physics, Geography, Justus-Liebig-Universität Gießen, Gießen, Germany

³Center for Mind, Brain and Behavior (CMBB), Justus-Liebig-Universität Gießen, Philipps-Universität Marburg and Technische Universität Darmstadt, Marburg, Germany

GW, 0000-0002-2364-0639; **DK**, 0000-0002-9007-3160

Previous research shows that the typicality of visual scenes (i.e. if they are good examples of a category) determines how easily they can be perceived and represented in the brain. However, the unique visual diets individuals are exposed to across their lifetimes should sculpt very personal notions of typicality. Here, we thus investigated whether scenes that are more typical to individual observers are more accurately perceived and represented in the brain. We used drawings to enable participants to describe typical scenes (e.g. a kitchen) and converted these drawings into three-dimensional renders. These renders were used as stimuli in a scene categorization task, during which we recorded electroencephalography (EEG). In line with previous findings, categorization was most accurate for renders resembling the typical scene drawings of individual participants. Our EEG analyses reveal two critical insights on how these individual differences emerge on the neural level. First, personally typical scenes yielded enhanced neural representations from around 200 ms after onset. Second, personally typical scenes were represented in idiosyncratic ways, with reduced dependence on high-level visual features. We interpret these findings in a predictive processing framework, where individual differences in internal models of scene categories formed through experience shape visual analysis in idiosyncratic ways.

1. Introduction

The ways in which humans perceive their environment are, by and large, studied across groups of participants, harnessing the coarse inter-individual stability of the mechanisms that guide visual perception. The focus on the group level is based on one of the core insights in vision research, which is that the visual system is remarkably similar across individuals and even across species [1–3]. When taking a closer look, however, perceptual mechanisms differ across people: the cortical architecture of the visual system varies inter-individually, resulting in idiosyncratic biases in cortical processing [4–6]. On the behavioural level, individual differences are observed in the perception of faces, objects and expertise-related stimuli [7–10], as well as in the attentional allocation in faces, objects and scenes [11–14].

One explanation for these idiosyncrasies lies in individually specific visual diets. The visual inputs we receive over our lifetimes [15,16] sculpt the response properties on our visual systems in ways that give rise to individual perception and neural representation. Indeed, visual diets can be a powerful predictor for behaviour and brain architecture [17,18]. If we take this argument seriously, and our visual representations are indeed formed by

our unique visual diets, then what is a typical instance of a visual stimulus category, say a scene, is not necessarily typical for another individual—what is typical should vary as a function of what we learned about the scene category over our lifetimes.

Here, we test whether individual notions of typicality are linked to idiosyncrasies in visual processing, focusing on the perception of complex and naturalistic visual scenes. To quantify what constitutes a typical scene for each individual participant, we first asked our participants to draw typical instances of real-world scene categories (e.g. a kitchen or bathroom), a paradigm we developed in recent behavioural work [19]. We then conducted an electroencephalography (EEG) experiment, in which participants categorized scene renders that we constructed to be similar to their own typical scene drawings or scene renders that were similar to other participants' drawings. In line with previously reported data [19], scene renders that were typical for individual participants were categorized more accurately by these participants. Crucially, multivariate analyses on the EEG data allowed us to reveal how such effects emerge on the neural level. These analyses yielded two key results. First, perceptual representations of personally typical scenes emerging at 200 ms of visual analysis are enhanced, compared with more atypical scenes. Second, personally typical scenes are represented in idiosyncratic ways, rendering the representations less faithful to the visual attributes of the scene.

2. Results

(a) Personally typical scenes are better categorized

We used drawing as a behavioural readout of individual participants' conceptions of typical everyday scenes [19]. In a first drawing session, participants were asked to draw typical scenes from six categories (bathroom, bedroom, café, kitchen, living room, office). To control the potential effect of memory on the following task, participants also copied photographs from the six categories. To reduce the influence of individual differences in drawing ability and style, we transformed these drawings into standardized three-dimensional renders (figure 1A), which were used as experimental stimuli in the second experimental session. Here, we tested if participants are more accurate in categorizing scenes that are similar to their typical drawings (i.e. personally typical scenes). To this end, participants completed a difficult six-way categorization task (figure 1A). Participants viewed personally typical renders, based on their own typical drawings ('own' condition), renders typical for other participants, based on other participants' typical drawings ('other' condition) and renders based on copied photographs ('control' condition). Behavioural results showed categorization accuracies varied significantly across condition, $F(2, 100) = 11.53$, $p < 0.001$, partial $\eta^2 = 0.19$. Specifically, higher categorization accuracy for the own condition (figure 1B), compared with both the other ($t(50) = 3.70$, $p < 0.001$; figure 1B) and control ($t(50) = 4.25$, $p < 0.001$; figure 1B) conditions. Response times were not significantly different between conditions, $F(2, 100) = 0.73$, $p = 0.49$, partial $\eta^2 = 0.01$. These results indicate that scene perception is more accurate for scenes that are typical for individual participants.

As these behavioural data were partly reported before—our previous behavioural report features 30 out of the 46 datasets [19]—we further tested whether the effect was independently replicated in the 16 participants not included in the previous study. The results from this new sample indeed fully replicated the effect, with significant differences between conditions in categorization accuracies, $F(2,30) = 3.80$, $p = 0.04$, partial $\eta^2 = 0.20$. Specifically, significantly higher categorization accuracy in the own condition, compared with both the other ($t(15) = 2.33$, $p = 0.034$) and control ($t(15) = 2.26$, $p = 0.039$) conditions (figure 1C). Response times were not significantly different between conditions, $F(2,30) = 0.07$, $p = 0.94$, partial $\eta^2 = 0.004$.

Only results from this experiment were reported before [19]. None of the following EEG and modelling results were reported elsewhere.

(b) Personally typical scenes evoke enhanced neural representations

To reveal how the personal typicality of a scene affects its representation in the brain, we recorded participants' EEG signals while they performed the categorization task. We then conducted a time-resolved decoding analysis [20] on the EEG data to discriminate between the six scene categories (figure 2A), from -200 to 800 ms relative to scene onset, separately for the own, other and control conditions. The resulting decoding performance over time yielded an estimate of the representational quality across conditions. Decoding accuracy rapidly increased for all conditions, starting from around 70 ms and peaked around 200 ms (figure 2B). Critically, we found stronger decoding for the own condition, compared with both the other condition (from 175 to 225 ms) and the control condition (from 175 to 225 and at 275 ms). This result demonstrates an enhanced neural representation of personally typical scenes. The timing of this effect further suggests that scene typicality on the individual level facilitates neural processing during perceptual analysis [21–24].

(c) Personally typical scenes yield idiosyncratic neural representations

The EEG decoding results show that scenes that are typical for individual participants evoke enhanced cortical representations. These enhanced representations, however, could originate from two different underlying processes. On the one hand, these enhanced representations could stem from an enhanced representation of visual features, caused by shaper neural tuning to features that are prevalent in personally typical scenes. If this were the case, deep neural network (DNN) activations should show a stronger correspondence with EEG representations. In this scenario, a higher signal-to-noise ratio in the neural data would yield a better DNN–brain correspondence. On the other hand, the enhanced neural representations for personally typical scenes could stem from these scenes more readily evoking higher level representations that abstract away from visual

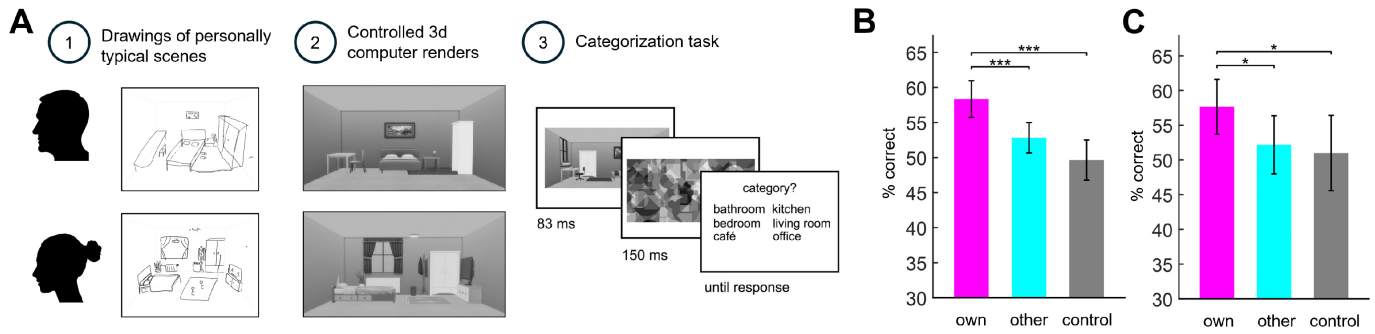


Figure 1. (A) Participants drew typical versions of six scene categories and copied scene photos from the same categories. To control for drawing ability differences across individuals, we converted drawings to three-dimensional renders based on each participant's own typical drawings, other participants' typical drawings and the copied scenes. During the following categorization task, participants categorized briefly presented renders into six scene categories. (B) Behavioural categorization accuracy was higher for renders based on participants' own drawings (own condition) than those based on other participants' drawings (other condition) or copies (control condition). (C) We also separately analysed the data from the 16 participants not yet reported before [19]; the results were independently replicated. Error bars represent s.e.m. * $p < 0.05$, *** $p < 0.001$.

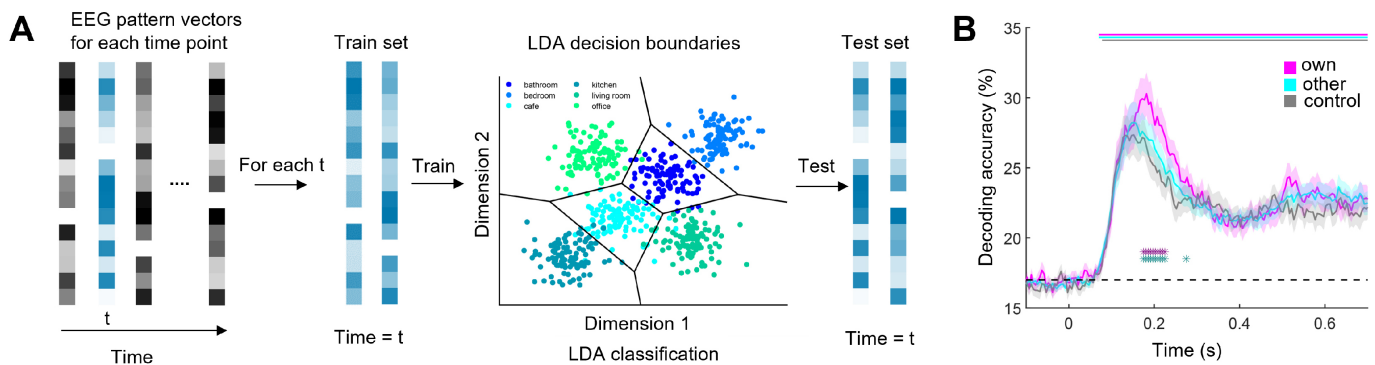


Figure 2. (A) To temporally track scene representations, we trained linear classifiers on the electroencephalography (EEG) activity pattern vectors at each time point. Classifiers were trained and tested on discriminating the six scene categories, using a 10-fold cross validation framework. (B) Between 175 and 225 ms after scene onset, decoding accuracy was higher in the own condition than in the other condition (indicated by dark purple significance markers) and the control condition (indicated by dark green significance markers). Error margins represent s.e.m. * $p < 0.05$ (corrected for multiple comparisons).

features, for instance by more readily activating representations of categorical prototypes. In this scenario, the DNN would not capture neural representations as well because it solely operates on the visual features of the images. If personally typical images indeed activated individual prototypes in the visual system, the DNN could not capture the transition towards these representations, thereby decreasing the DNN–brain correspondence [25]. Such higher level visual representations could yield highly category-specific response patterns that boost decoder performance.

To arbitrate between these possibilities, we employed a DNN model trained on scene categorization, which we used to quantify visual features extracted at different levels of the visual hierarchy. By quantifying how well the features extracted from the DNN predict brain activations in the EEG, we could test whether the enhanced neural representations of personally typical scenes are accompanied by a more or less pronounced representation of visual features. We first extracted activations from 12 layers along a googlenet [26] DNN trained on the Places365 dataset [27], for all scenes in the own, other and control conditions, separately for each participant. We then computed the pairwise similarity between these activations by correlating the activation vectors for each pair of scenes. This yielded a representational dissimilarity matrix (RDM) for each layer and condition. For the EEG, we first constructed RDMs based on the pairwise decoding analysis and then averaged RDMs across the time window in which we found a difference between the own and other conditions (175–225 ms). To quantify how well the representation in the time window of interest is predicted by the visual feature organization in the DNN, we correlated the DNN RDM at each layer with the EEG RDM, separately for the three conditions (figure 3A). The results showed that representations of visual features in early and intermediate DNN layers similarly predict neural representations across three conditions (figure 3B). Critically, in the last two DNN layers, neural representations in the own condition were predicted less well by the DNN features, compared with both the other (inception5b layer: $t(45) = 2.71$, $p_{\text{FDR}} = 0.041$; fully connected layer: $t(45) = 2.74$, $p_{\text{FDR}} = 0.018$) and control (inception5b layer: $t(45) = 3.01$, $p_{\text{FDR}} = 0.041$; fully connected layer: $t(45) = 3.50$, $p_{\text{FDR}} = 0.012$) conditions (figure 3B). This shows that scenes that are typical for individual participants evoke neural representations that are less faithful to higher level visual features extracted by the DNN. This supports the view that personally typical scenes more readily evoke categorical representations that are less dependent on high-level visual properties. Interestingly, for the own condition, the correspondence between the DNN features and the neural representations gradually declined along the DNN's feature hierarchy: when fitting a linear model on each participant's data, we found a significantly negative slope across participants ($t(45) = 3.05$, $p = 0.0038$), which was absent in the other ($t(45) = 0.63$, $p = 0.53$) and control ($t(45) = 0.47$, $p = 0.64$) conditions. This suggests that representations of personally typical scenes get progressively decoupled from visual features, compared with less typical scenes.

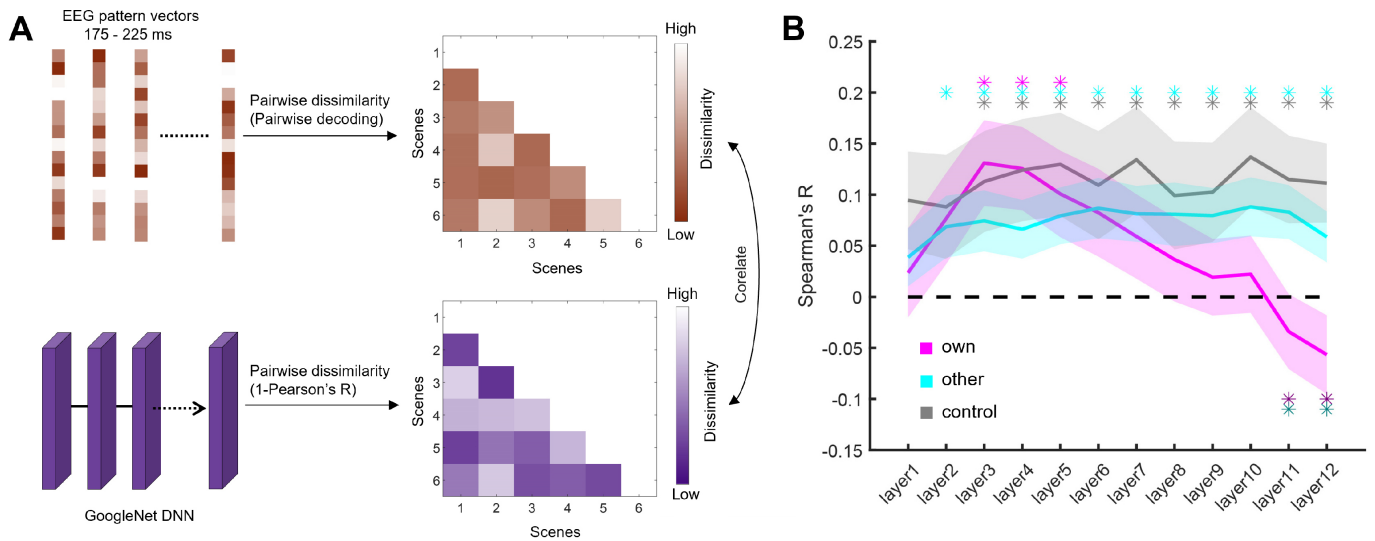


Figure 3. (A) To compare visual feature representations between a deep neural network (DNN) and the brain, we first constructed electroencephalography (EEG) representational dissimilarity matrices (RDMs) by computing pairwise decoding accuracies for all combinations of scenes, separately for the own, other and control conditions. We then averaged RDMs for the time period showing a difference between the own and other/control conditions (175–225 ms). Next, we constructed DNN RDMs by quantifying the pairwise dissimilarity of response vectors across the 12 layers of a googlenet trained on the Places365 dataset. Finally, we correlated the EEG and DNN RDMs for each condition, separately for each layer. (B) In the last two DNN layers, neural representations in the own condition were predicted less well by the DNN features, compared with both the other condition (indicated by dark purple significance markers) and the control condition (indicated by dark green significance markers). Error margins represent s.e.m. $*p < 0.05$ (corrected for multiple comparisons).

3. Discussion

Our findings demonstrate that the perception and neural representation of a scene vary as a function of whether the scene is typical for an individual participant. First, scenes that are more typical for individual participants are categorized more accurately, in line with previous results [19]. Second, this enhanced categorization of personally typical scenes is accompanied by an enhanced cortical representation, emerging during visual analysis at around 200 ms post-stimulus. Finally, these enhanced representations are more idiosyncratic, that is, more detached from high-level visual features of the images.

These results are parsimoniously explained by predictive processing theories [28–30], which posit that perception is mediated by a convergence of feedforward sensory analysis and feedback in the form of predictions derived from internal world models. On this view, individual people—based on their individual lifetime experience with the world—form idiosyncratic internal models of what the world should look like. These individually specific internal models in turn yield individual differences in perception and neural representations. The enhanced neural representations for typical—and hence more predictable—scenes observed here are in line with a sharpening of cortical representations for predictable visual inputs [31,32]. Personally typical scenes may yield sparser neural codes that facilitate the readout of information.

Typicality arises from visual diets and thus extensive experience with personally familiar scenes. This makes it challenging to fully disentangle typicality from long-term familiarity. We mitigated this issue by instructing participants not to draw their own rooms or scenes they are most familiar with, yet their drawings may still feature properties of scenes from their everyday life that they consider typical—or which, for their own scenes, they themselves arranged to be typical in the first place. Future studies could ask participants to provide photographs of their own everyday environments to delineate how much typical scene drawings are inspired by familiar real-life scenes.

Contrasting long-term familiarity, short-term familiarity acquired during the experiment needs to be controlled for. Here, we accounted for short-term familiarity using our copy-based control condition. Our data suggest that the effects are not driven by short-term familiarity acquired in the experiment. First, previous research has demonstrated that familiarity enhances neural representations already around 200 ms, but most prominently during late, presumably post-perceptual processing stages (>400 ms [33,34]). In contrast, the enhanced neural representations for personally typical scenes in our study emerged only around 200 ms, while no late effects were observed. Second, our DNN analysis shows that representations of typical scenes are less well explained by visual image properties. If familiarity drove high-quality representations through memory recall, we should see the opposite pattern. Finally, our previous study [19] suggests that enhanced behavioural categorization is not just observed for scenes constructed to be highly similar to a person's drawing, but that categorization parametrically varies with the similarity to the drawing. This further refutes the idea that the effects are solely driven by participants acquiring familiarity with a single scene during drawing. However, while our control condition successfully controlled for the drawing process and the visual exposure to the drawing, a tighter control for the mental composition of the scenes in the typical drawing condition could be desirable in future studies. For example, participants could be asked to draw typical scenes from a different context (e.g. 'Draw the King's bedroom!'), which requires comparable mental scene composition but does not yield scenes typical for the individual.

Our results also reveal how representational idiosyncrasies emerge across the visual feature processing hierarchy. The DNN analysis shows that neural representations of personally typical and atypical scenes are similarly predicted by representations of low- and mid-level features in the DNN, suggesting that basic levels of visual processing do not reflect whether or not a scene

taps into the internal model of individual observers (though one caveat here is that we purposefully matched the scenes in their low-level features). In contrast, representations are differently predicted by high-level features for scenes that are personally typical or not (which is consistent with previous modelling of behavioural responses [19]). High-level visual features are coded less faithfully when the scenes are typical for individual participants. It furthermore suggests that personally typical scenes are represented in more idiosyncratic ways that are not captured by a DNN's feature hierarchy. Neural processes that elude the modelling capacity of currently used DNN models include the activation of individually specific category prototypes [25,35] or the rapid recruitment of semantic representations that bridge perception and memory [36]. Such processes are likely to differ from individual to individual and are thus not captured by DNN models trained on a single dataset. Alternatively, personally typical scenes may yield stronger recurrent connectivity that shapes more complex representations that are not captured by feedforward-only DNNs [37].

This finding aligns with predictive theories [28–30], which posit that the interpretation of visual inputs depends on top-down predictions generated from internal models. The way these internal models are engaged may differ based on the personal typicality of the input: when a scene is more typical for an individual, top-down predictions may align better with the input and more strongly reshape processing dynamics. The impact of predictions may dominate neural representations at later stages, after the feedforward analysis indicated that the input likely stems from a specific category. During these late processing stages, consequently, the reliance on feedforward processes driven by visual features is reduced. For less typical scenes, top-down processes may reshape processing less fundamentally, given the divergence between the input and predictions.

While our results highlight the enhancement of representations for personally typical scenes, they cannot directly elucidate the mechanisms through which these enhanced representations emerge from the interplay between feedforward and feedback dynamics in the cortex. To make progress, future research could combine EEG and functional magnetic resonance imaging recordings to chart the spatiotemporal dynamics evoked by personally typical and atypical scenes. Such studies could arbitrate between a purely feedforward emergence of sharpened representations for typical scenes and a refinement of these representations through cortical feedback from higher level cortex or memory systems [38,39]. Future studies could also use DNN architectures that explicitly mimic feedback connectivity in cortex to disentangle feedforward and feedback information flows during the emergence of idiosyncratic representations in the visual system [37]. For capturing the idiosyncratic nature of visual processing more comprehensively, these models could further be enriched with personalized training regimes or explicit participant-specific priors, which would push the limits of predictive power on the individual-participant level.

In sum, our research reveals idiosyncrasies in the way scenes are perceived and represented as a function of how typical they are to individual participants. Critically, we demonstrate that representations of personally typical scenes are enhanced and more idiosyncratic. These findings highlight that a comprehensive understanding of visual representations in human cortex requires researchers to take individual differences into account. With our drawing-based approach, we provide a straightforward method for predicting such differences.

4. Material and methods

(a) Participants

Fifty-one participants (23.3 ± 2.9 years, 15/36 male/female) completed the experiment. One additional participant completed the drawing session but did not return for the EEG session. Thirty-five participants were tested at Freie Universität Berlin, including EEG recordings for 30 participants (23.0 ± 2.2 years, 8/22 male/female). Behavioural data for these participants have been reported before [19]. Another 16 participants (23.9 ± 4.3 years, 7/9 male/female) were tested at Justus-Liebig-Universität Gießen, using an identical set-up. In total, 51 participants completed the categorization task, including EEG recording for 46 participants (23.3 ± 3.1 years, 15/31 male/female). The sample size for the EEG study is comparable to recent multivariate EEG studies conducted in the laboratory [36,38,39]. Procedures were approved by the ethics committee of the Department of Education and Psychology, Freie Universität Berlin and the ethics committee of the Justus-Liebig-Universität Gießen, respectively, and adhered to the Declaration of Helsinki.

(b) Drawing sessions

Participants provided their drawings on an Apple iPad Pro using an Apple Pencil. Drawings were created using the Sketchbook App. Here, participants were asked to draw typical versions of six scene categories (bathroom, bedroom, café, kitchen, living room and office). For each drawing, they were given 30 s to plan the drawings and then 4 min to execute it. A perspective grid was provided to guide the arrangement of objects across the scene. We instructed participants to draw the most typical but not their own or most liked, rooms. We additionally asked them to copy a photograph for each of the scene categories, under the same time constraints. Full details on the drawing session can be found in our previous study [19].

(c) Scene renders

Given the variation in participants' drawing ability and style, we created experimental stimuli that removed these differences. Specifically, we created three-dimensional renders similar to participants' drawings by placing objects into an empty room, using the SIMS4 builder toolkit (see figure 1A). In total, 318 renders were created: 312 renders were based on the typical

drawings of 52 individual participants and 6 renders were based on the 6 control photographs (these control renders were identical for all participants).

(d) Categorization task

We used the Psychophysics Toolbox [40] for Matlab to set up the categorization task. Here, participants were asked to categorize the briefly presented scene renders into six categories (see figure 1A). On each trial, a render (7° horizontal visual angle) was presented for 83 ms, followed by a geometric pattern mask, presented for 150 ms. Participants viewed renders based on their own drawing of a typical scene ('own' condition), other participants' drawings of typical scenes ('other' condition) and their copied scenes ('control' condition). We grouped participants into groups of four, and each participant saw renders based on their own drawings, renders based on other three participants' drawings and the control renders. Each of these 30 stimuli was repeated 40 times, for a total of 1200 trials, presented in random order.

(e) Electroencephalography recording and preprocessing

We recorded EEG signals using an EASYCAP 64-electrode system and Brainvision actiCHamp amplifier. Electrodes were arranged according to the 10–10 system. The data were recorded with 1000 Hz sample rate and filtered online between 0.03 and 100 Hz. All electrodes were referenced online to the Fz electrode and re-referenced offline to the averaged signal from all channels. We used FieldTrip [41] to process offline EEG data. The continuous EEG data were segmented to the epoch into trials ranging from 200 ms before stimulus onset to 800 ms after stimulus onset, and baseline corrected by subtracting the mean of the pre-stimulus interval for each trial and channel separately. Channels containing excessive noise were removed by visual inspection. The removed channels (2.12 ± 0.76 channels) were interpolated by their mean signals of neighbouring channels. No trials were removed. Blinks and eye movement artefacts were removed using independent component analysis and visual inspection of the resulting components. The epoch data were downsampled to 200 Hz.

(f) Time-resolved decoding analysis

To trace the temporal representation of scenes across time, we performed a time-resolved multivariate decoding analysis using CoSMoMVPA[42]. In this analysis, we decoded between the six scene categories, separately for the three conditions (i.e. own, other and control). Specifically, we performed classification analysis from 200 ms before stimulus onset to 800 ms after onset, using a sliding time window (50 ms width, 5 ms resolution). For each time window separately, we used linear discriminant analysis (LDA) classifiers with 10-fold cross validation. We first allocated the EEG data to 10 folds randomly. LDA classifiers were then trained on data from 9 folds and then tested on data from the left-out fold. Prior to classification, principal component analysis (PCA) was performed on all data from the training set, and the PCA solution (retaining 99% of the variance) was projected onto the testing set [43]. The amount of data in the training set was always balanced across scene categories. Classification was done repeatedly until every fold was left out once, and accuracies were averaged folds. For the other condition, we performed separate classification analyses for the renders from each of the other three participants, and accuracy was averaged across these three analyses.

(g) Representational similarity analysis

To better understand how neural representations in the own, other and control conditions are predicted by visual image features, we related neural representations to a DNN model using representational similarity analysis (RSA). We first extracted neural RDMs separately for the EEG data and DNN model, and then tested how well the DNN RDMs predict the EEG RDMs. For the EEG data, RDMs were constructed using the same classification routine as in the original decoding analysis (see above), but now computing pairwise decoding accuracies for all possible combinations of six categories, separately for each of the three conditions. As we were particularly interested in representations during the time points that showed a difference between the own and other (and own and control) conditions, we performed this analysis on all time points between 175 and 225 ms, and then averaged RDMs across these time points. For the DNN, we extracted activation vectors from 12 layers (cov1, cov2, inception3a, inception3b, inception4a, inception4b, inception4c, inception4d, inception4e, inception5a, inception5b and full-connected) along the hierarchy of googlenet [26] trained on scene categorization using the Places365 dataset [27], as used in our previous study [19]. We extracted layer-wise features activations for each of the six scenes used in each condition. We then constructed layer-specific RDMs by quantifying the pairwise dissimilarity ($1 - \text{Pearson's } R$) of response patterns in each layer. Finally, we correlated (Spearman's R) the EEG RDMs with the DNN RDMs, separately for each layer and condition. We Fisher-transformed the correlation values prior to statistical analysis.

We additionally computed an empirical noise ceiling for the RSA. Due to the stimuli differing between participants in the 'own' and 'other' conditions, noise ceilings were estimated from the RDMs constructed for the 'control' condition where the stimuli were identical across participants. First, RDMs were averaged within the 175–225 ms time window to produce a single RDM for each participant. Following the method described [44], the lower bound was calculated as the mean correlation between each participant's RDM and the group mean RDM, excluding that participant. The upper bound was computed as the mean correlation between each participant's RDM and the group mean RDM, including that participant. The resulting noise ceiling estimates were 0.2236 (upper bound) and 0.1352 (lower bound).

(h) Statistical analysis

For the behavioural data analysis, responses slower than 5 s were discarded. Accuracies and response times were compared using repeated-measures analysis of variances and paired *t*-tests. Only trials with correct responses were analysed for the response times. For the decoding analysis, we used *t*-tests and threshold-free cluster enhancement [45] in CoSMoMVA to test decoding accuracies against chance level and assess differences between conditions. Multiple comparison correction across time was based on sign-permutation tests with null distribution created from 10 000 bootstrapping iterations. For the RSA, we used *t*-tests to compare correlations against chance level and compare the difference between conditions. Multiple comparison correction across layers was performed using false-discovery-rate correction.

Ethics. Procedures were approved by the ethics committee of the Department of Education and Psychology, Freie Universität Berlin (approval number 043.2021) and the ethics committee of the Justus-Liebig-Universität Gießen (approval number AZ111/22), and adhered to the Declaration of Helsinki.

Data accessibility. Data are available at: <https://osf.io/ctsxv/>.

Supplementary material is available online [46].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. G.W.: conceptualization, data curation, formal analysis, investigation, methodology, resources, software, validation, visualization, writing—original draft, writing—review and editing; L.C.: methodology, writing—review and editing; R.M.C.: funding acquisition, project administration, supervision, writing—review and editing; D.K.: conceptualization, funding acquisition, methodology, project administration, supervision, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. G.W. is supported by a PhD stipend from the China Scholarship Council (CSC). R.M.C. is supported by the Deutsche Forschungsgemeinschaft (DFG; CI241/1-1, CI241/3-1, CI241/7-1) and by a European Research Council (ERC) starting grant (ERC-2018-STG 803370). D.K. is supported by the DFG (SFB/TRR135, project number 222641018; KA4683/5-1, project number 518483074), 'The Adaptive Mind', funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art, and an ERC Starting Grant (PEP, ERC-2022-STG 101076057). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

References

- Haxby JV, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, Gobbini MI, Hanke M, Ramadge PJ. 2011 A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**, 404–416. (doi:10.1016/j.neuron.2011.08.026)
- Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M. 2013 The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* **17**, 26–49. (doi:10.1016/j.tics.2012.10.011)
- Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA. 2008 Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* **60**, 1126–1141. (doi:10.1016/j.neuron.2008.10.043)
- Himmelberg MM, Winawer J, Carrasco M. 2022 Linking individual differences in human primary visual cortex to contrast sensitivity around the visual field. *Nat. Commun.* **13**, 3309. (doi:10.1038/s41467-022-31041-9)
- Kanai R, Rees G. 2011 The structural basis of inter-individual differences in human behaviour and cognition. *Nat. Rev. Neurosci.* **12**, 231–242. (doi:10.1038/nrn3000)
- Moutsiana C, de Haas B, Papageorgiou A, van Dijk JA, Balraj A, Greenwood JA, Schwarzkopf DS. 2016 Cortical idiosyncrasies predict the perception of object size. *Nat. Commun.* **7**, 12110. (doi:10.1038/ncomms12110)
- Gauthier I, Tarr MJ. 2016 Visual object recognition: do we (finally) know more now than we did? *Annu. Rev. Vis. Sci.* **2**, 377–396. (doi:10.1146/annurev-vision-111815-114621)
- Ramon M. 2023 Super-recognizers—a novel diagnostic framework, 70 cases, and guidelines for future work. *Neuropsychologia* **158**, 107809. (doi:10.1016/j.neuropsychologia.2023.108724)
- Richler JJ, Tomarken AJ, Sunday MA, Vickery TJ, Ryan KF, Floyd RJ, Sheinberg D, Wong ACN, Gauthier I. 2019 Individual differences in object recognition. *Psychol. Rev.* **126**, 226–251. (doi:10.1037/rev0000129)
- White D, Burton AM. 2022 Individual differences and the multidimensional nature of face perception. *Nat. Rev. Psychol.* **1**, 287–300. (doi:10.1038/s44159-022-00041-3)
- Broda MD, de Haas B. 2024 Individual differences in human gaze behavior generalize from faces to objects. *Proc. Natl Acad. Sci. USA* **121**, e2322149121. (doi:10.1073/pnas.2322149121)
- De Haas B, Iakovidis AL, Schwarzkopf DS, Gegenfurtner KR. 2019 Individual differences in visual salience vary along semantic dimensions. *Proc. Natl Acad. Sci. USA* **116**, 11687–11692. (doi:10.1073/pnas.1820553116)
- Henderson JM, Luke SG. 2014 Stable individual differences in saccadic eye movements during reading, pseudoreading, scene viewing, and scene search. *J. Exp. Psychol.* **40**, 1390–1400. (doi:10.1037/a0036330)
- Risko EF, Anderson NC, Lanthier S, Kingstone A. 2012 Curious eyes: individual differences in personality predict eye movement behavior in scene-viewing. *Cognition* **122**, 86–90. (doi:10.1016/j.cognition.2011.08.014)
- Barrett HC. 2020 Towards a cognitive science of the human: cross-cultural approaches and their urgency. *Trends Cogn. Sci.* **24**, 620–638. (doi:10.1016/j.tics.2020.05.007)
- Hartley CA. 2022 How do natural environments shape adaptive cognition across the lifespan? *Trends Cogn. Sci.* **26**, 1029–1030. (doi:10.1016/j.tics.2022.10.002)
- Llera A, Wolfers T, Mulders P, Beckmann CF. 2019 Inter-individual differences in human brain structure and morphology link to variation in demographics and behavior. *eLife* **8**, e44443. (doi:10.7554/eLife.44443)
- Coutrot A *et al.* 2022 Entropy of city street networks linked to future spatial navigation ability. *Nature* **604**, 104–110. (doi:10.1038/s41586-022-04486-7)

19. Wang G, Foxwell MJ, Cichy RM, Pitcher D, Kaiser D. 2024 Individual differences in internal models explain idiosyncrasies in scene perception. *Cognition* **245**, 105723. (doi:10.1016/j.cognition.2024.105723)
20. Grootswagers T, Wardle SG, Carlson TA. 2017 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. (doi:10.1162/jocn_a_01068)
21. Martin Cichy R, Khosla A, Pantazis D, Oliva A. 2017 Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. *NeuroImage* **153**, 346–358. (doi:10.1016/j.neuroimage.2016.03.063)
22. Harel A, Groen IIA, Kravitz DJ, Deouell LY, Baker CI. 2016 The temporal dynamics of scene processing: a multifaceted EEG investigation. *eNeuro* **3**, 0139–16.2016. (doi:10.1523/eneuro.0139-16.2016)
23. Kaiser D, Turini J, Cichy RM. 2019 A neural mechanism for contextualizing fragmented inputs during naturalistic vision. *eLife* **8**, e48182. (doi:10.7554/elife.48182)
24. Kaiser D, Häberle G, Cichy RM. 2020 Real-world structure facilitates the rapid emergence of scene category information in visual brain signals. *J. Neurophysiol.* **124**, 145–151. (doi:10.1152/jn.00164.2020)
25. Blank H, Bayer J. 2022 Functional imaging analyses reveal prototype and exemplar representations in a perceptual single-category task. *Commun. Biol.* **5**, 896. (doi:10.1038/s42003-022-03858-z)
26. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D. 2015 Going deeper with convolutions. In *Proc. 2015 IEEE Conf. on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015*, pp. 1–9. (doi:10.1109/CVPR.2015.7298594)
27. Zhou B, Lapedriza A, Khosla A, Oliva A, Torralba A. 2018 Places: a 10 million image database for scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 2017. (doi:10.1109/tpami.2017.723009)
28. Clark A. 2013 Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204. (doi:10.1017/s0140525x12000477)
29. Friston K. 2005 A theory of cortical responses. *Phil. Trans. R. Soc. B* **360**, 815–836. (doi:10.1098/rstb.2005.1622)
30. Keller GB, Mrsic-Flogel TD. 2018 Predictive processing: a canonical cortical computation. *Neuron* **100**, 424–435. (doi:10.1016/j.neuron.2018.10.003)
31. Kok P, Jehee JFM, de Lange FP. 2012 Less is more: expectation sharpens representations in the primary visual cortex. *Neuron* **75**, 265–270. (doi:10.1016/j.neuron.2012.04.034)
32. Peelen MV, Berlot E, de Lange FP. 2024 Predictive processing of scenes and objects. *Nat. Rev. Psychol.* **3**, 13–26. (doi:10.1038/s44159-023-00254-0)
33. Ambrus GG, Eick CM, Kaiser D, Kovács G. 2021 Getting to know you: emerging neural representations during face familiarization. *J. Neurosci.* **41**, 5687–5698. (doi:10.1523/jneurosci.2466-20.2021)
34. Klink H, Kaiser D, Stecher R, Ambrus GG, Kovács G. 2023 Your place or mine? The neural dynamics of personally familiar scene recognition suggests category independent familiarity encoding. *Cereb. Cortex* **33**, 11634–11645. (doi:10.1093/cercor/bhad397)
35. Bowman CR, Iwashita T, Zeithamova D. 2020 Tracking prototype and exemplar representations in the brain across learning. *eLife* **9**, e59360. (doi:10.7554/elife.59360)
36. Steel A, Billings MM, Silson EH, Robertson CE. 2021 A network linking scene perception and spatial memory systems in posterior cerebral cortex. *Nat. Commun.* **12**, 2632. (doi:10.1038/s41467-021-22848-z)
37. Kietzmann TC, Spoerer CJ, Sörensen LKA, Cichy RM, Hauk O, Kriegeskorte N. 2019 Recurrence is required to capture the representational dynamics of the human visual system. *Proc. Natl Acad. Sci. USA* **116**, 21854–21863. (doi:10.1073/pnas.1905544116)
38. Chen L, Cichy RM, Kaiser D. 2023 Alpha-frequency feedback to early visual cortex orchestrates coherent naturalistic vision. *Sci. Adv.* **9**, eadi2321. (doi:10.1126/sciadv.adi2321)
39. Cichy RM, Oliva A. 2020 A M/EEG-fMRI fusion primer: resolving human brain responses in space and time. *Neuron* **107**, 772–781. (doi:10.1016/j.neuron.2020.07.001)
40. Brainard DH. 1997 The psychophysics toolbox. *Spat. Vis.* **10**, 433–436. (doi:10.1163/156856897x00357)
41. Oostenveld R, Fries P, Maris E, Schoffelen JM. 2011 FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869. (doi:10.1155/2011/156869)
42. Oosterhof NN, Connolly AC, Haxby JV. 2016 CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Front. Neuroinform.* **10**, 27. (doi:10.3389/fninf.2016.00027)
43. Chen L, Cichy RM, Kaiser D. 2022 Semantic scene-object consistency modulates N300/400 EEG components, but does not automatically facilitate object representations. *Cereb. Cortex* **32**, 3553–3567. (doi:10.1093/cercor/bhab433)
44. Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N. 2014 A toolbox for representational similarity analysis. *PLoS Comput. Biol.* **10**, e1003553. (doi:10.1371/journal.pcbi.1003553)
45. Smith S, Nichols T. 2009 Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage* **44**, 83–98. (doi:10.1016/j.neuroimage.2008.03.061)
46. Wang G, Chen L, Cichy RM, Kaiser D. 2025 Supplementary material from: Enhanced and idiosyncratic neural representations of personally typical scenes. Figshare. (doi:10.6084/m9.figshare.c.7702797)